Statistical Models of Overdispersed Spatial Defects for Predicting the Yield of Integrated Circuits

Tao Yuan, Suk Joo Bae, and Yue Kuo, *Fellow, IEEE*

***Abstract*—Defects generated in semiconductor manufacturing processes have serious effects on the yield of integrated circuits (ICs). Accurate modeling of the defect counts on IC chips is crucial for predicting the yield. The conventional Poisson yield model tends to underestimate the true yield by ignoring overdispersed patterns of defects on the wafer. This article uses various models based on the generalized Poisson (GP) distribution and/or HZ distributions to explore the overdispersed defect counts on semiconductor wafers. Real wafer map data are used to compare the performance of both nonregression and regression modeling approaches in terms of the log-likelihood, AIC, and relative bias for yield estimation. Analytical results indicate that the GP distribution is a competitive alternative to the negative binomial (NB) distribution for modeling defect counts on IC chips because the GP distribution can model overdispersion, underdispersion, or no dispersion. In particular, HZ models based on the NB and GP distributions show good potential for predicting the yield of IC chips on wafers.**

***Index Terms*—Clustered defects, generalized Poisson (GP) distribution, hurdle-at-zero (HZ) models, maximum likelihood estimation, prediction of IC yield, zero-inflated models.**

Nomenclature

Acronyms

AC	Adjacency clustering.
AIC	Akaike information criterion.
GP	Generalized Poisson.
HZGP	Hurdle-at-zero generalized Poisson.
HZNB	Hurdle-at-zero negative binomial.
HZP	Hurdle-at-zero Poisson.
IC	Integrated circuit.
MLE	Maximum likelihood estimate(or).
MRF	Markov random field.
NB	Negative binomial.
ZIGP	Zero-inflated generalized Poisson.
ZINB	Zero-inflated negative binomial.
ZIP	Zero-inflated Poisson.

Notation

a	NB dispersion parameter.
$\mathrm{f}(\mathrm{z})$	Spatial covariates for μ.
$\mathrm{g}(\mathrm{z})$	Spatial covariates for π.
$\mathrm{h}(\mathrm{z})$	Spatial covariates for ω.
$\mathcal{L}$	Likelihood function.
ℓ	Log-likelihood function.
n	Number of chips on a wafer.
n_0	Number of defect-free chips on a wafer.
n_p	Number of chips with defects on a wafer.
(r, φ)	Polar coordinates of a chip's center.
S_0	Zero count set.
S_p	Positive count set.
X	Random defect count.
x	Observed defect count data on a wafer.
y	Vector of covariates.
Y	Yield.
z	Location of a chip center on a wafer.
$\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}$	Vectors of regression coefficients.
$\boldsymbol{\theta}$	Parameter vector of a model.
μ	Mean parameter.
π	Probability of extra zeros.
$\boldsymbol{\Sigma}$	Variance–covariance matrix.
ϕ	GP dispersion parameter.
ω	Probability of positive counts.

Manuscript received September 19, 2018; revised December 11, 2018, April 14, 2019, and July 8, 2019; accepted September 15, 2019. Date of publication October 15, 2019; date of current version June 1, 2020. This work was supported by the National Science Foundation through Project 1633500 and Project 1633580. The work of S. J. Bae was supported by the Basic Science Research Program through the National Research Foundation of Korea under Grant 2018R1D1A1A09083149 funded by the Ministry of Education. Associate Editor: W. Chu. *(Corresponding author: Suk Joo Bae.)*

T. Yuan is with the Department of Industrial and Systems Engineering, Ohio University, Athens, OH 45701 USA (e-mail: yuan@ohio.edu).

S. J. Bae is with the Department of Industrial Engineering, Hanyang University, Seoul 04763, South Korea (e-mail: sjbae@hanyang.ac.kr).

Y. Kuo is with the Department of Chemical Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: yuekuo@tamu.edu).

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

Digital Object Identifier 10.1109/TR.2019.2943925

I. Introduction

SEMICONDUCTOR manufacturing is a complex multistage process that requires monitoring various interrelated critical parameters. In today's rapidly changing technological environment, semiconductor manufacturers strive to secure defect-free products in a fairly short time period by adopting sophisticated manufacturing technologies. However, design geometries are also shrinking continuously as manufacturing technologies evolve, resulting in a reactive approach of eliminating defect sources that is prohibitively slow. Defects generated in IC manufacturing processes seriously lower the manufacturing yield and can sometimes lead to reliability issues. The yield has been widely used as a performance metric in semiconductor manufacturing. To survive intense competition among leading semiconductor manufacturers, it has become a challenging task to increase the yield for new products via quick response to yield excursions at early stages of IC production. Accurate

0018-9529 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.

modeling of the defect distributions on IC chips is crucial for yield prediction and improvement, production cost estimation, and reliability enhancement [1].

To briefly review existing yield models in semiconductor manufacturing, the classical yield model adopts a Poisson distribution, assuming that the defects are uniformly distributed on a wafer. The random number of defects (i.e., the defect count) on a chip, X, follows the Poisson distribution with the following probability mass function (pmf):

$$P(X = x|\mu) = \frac{e^{-\mu}\mu^x}{x!}, \quad x = 0, 1, 2, \ldots \tag{1}$$

where $\mu(>0)$ is the mean parameter. Note that the Poisson variance is also equal to μ. However, it has been widely reported that the defects generated during IC fabrication tend to occur close to one another on a semiconductor wafer; thus, the observed defect count data typically exhibit overdispersion with extra zeros. The Poisson distribution tends to underestimate the probability of zero. To describe the overdispersed defect counts, various compound Poisson yield models have been introduced, e.g., the NB yield model, Murphy's yield model, and Seed's yield model, which are mixtures of Poisson distributions [2]. For example, the widely used NB distribution is a Poisson–Gamma mixture that computes the probability that an IC chip contains x defects as

$$P(X = x|\mu, a) = \frac{\Gamma(a+x)}{\Gamma(a)\Gamma(x+1)}\left(\frac{\mu}{\mu+a}\right)^x\left(\frac{a}{\mu+a}\right)^a$$

for $x = 0, 1, 2, \ldots$, where $\mu(>0)$ and $a(>0)$ are the mean parameter and dispersion parameter (or the so-called clustering coefficient), respectively. The variance of the NB distribution is $\mathrm{Var}(X) = \mu + \mu^2/a$. A larger a value implies less dispersion in the defect count data and less severe clustering of defects. When a approaches infinity, the NB distribution converges to the Poisson distribution, indicating homogeneously distributed defects with no clustering. Tyagi and Bayoumi [3] took a different approach toward the clustering effect and introduced a generalized double-Poisson distribution to model the count of clustered defects. The basic assumptions of this generalized double-Poisson distribution yield model are that cluster centers are homogeneously distributed, and defects within a cluster are randomly distributed. Therefore, the number of clusters and the number of defects within a cluster are assumed to be two independent Poisson random variables. The authors demonstrated the capabilities of the generalized double-Poisson distribution as a yield model in IC redundancy designs. Under no parametric assumption for the yield model, Tong and Chao [4] employed a general regression neural network to predict wafer yield for ICs with clustered defects. Recently, Hochbaum and Liu [5] proposed a yield prediction model called "AC model," which is a form of the MRF minimum energy model. The AC model adopted the MRF to identify defect patterns that enable the diagnosis of failure causes in the semiconductor manufacturing process.

When the overdispersion in a count dataset is mainly caused by excess zeros, zero-inflated distributions (e.g., ZIP, ZINB, and ZIGP) have been frequently adopted to model the extra dispersion due to excess zeros. A zero-inflated distribution is a mixture of the Bernoulli distribution and a count distribution (e.g., the Poisson or NB distribution). For example, the ZIP distribution has the following pmf:

$$P(X = 0|\pi, \mu) = \pi + (1-\pi)e^{-\mu}$$

$$P(X = x|\pi, \mu) = (1-\pi)\frac{e^{-\mu}\mu^x}{x!}, \quad x = 1, 2, \ldots$$

where $0 \le \pi \le 1$ is the probability of extra zeros. Including the spatial locations of the chips on a semiconductor wafer as covariates, Bae *et al.* [1] proposed Poisson regression, NB regression, and ZIP regression yield models to reflect spatially nonhomogeneous defect patterns. Later, Yuan *et al.* [6] extended those regression modeling approaches under a hierarchical Bayesian framework. Additionally, they proposed a hierarchical Bayesian ZINB regression yield model, showing that the Bayesian ZINB regression yield model provides the most accurate yield estimation among the four regression yield models.

In this article, we attempt to apply several models based on the GP distribution and/or hurdle-at-zero (HZ) distributions to analyze the overdispersed defect counts on semiconductor wafers. The GP distribution is also a mixture of Poisson distributions, like the NB distribution. The GP/ZIGP distributions and various HZ distributions (e.g., the HZP, and HZNB) have been widely used in various fields to analyze discrete count data with overdispersion and/or extra zeros. For example, Joe and Zhu [7] fitted the NB, GP, ZINB, and ZIGP distributions to overdispersed spinal tumor count data. Their results indicate that NB fits the dataset better than GP, but ZIGP is the best among the four distributions. Yip and Yau [8] applied the ZIP, ZINB, ZIGP, and zero-inflated double-Poisson distributions to model automobile insurance claim data, which include a lot of zeros. Srivastava and Chen [9] applied the GP distribution to deep sequencing of ribonucleic acids and showed that the GP distribution outperforms the Poisson distribution in terms of identifying differentially expressed genes. Gupta *et al.* [10] applied the ZIGP distribution to analyze overdispersed fetal movement data and death count data. Angers and Biswas [11] used the ZIGP distribution to fit fetal movement data in a Bayesian framework. Gurmu and Trivedi [12] observed that the HZNB distribution has potential in modeling the number of recreational boating trips with overdispersion and excessive zeros. However, the GP-based models and HZ models have not been applied to predict the manufacturing yield of IC chips in semiconductor manufacturing.

Covariate information can be added to the aforementioned count distributions, resulting in regression models. For example, a Poisson regression model may be constructed by assuming that the Poisson mean μ in (1) depends on a vector of covariates y as $\ln \mu = \mathrm{y}^T\beta$, where β is a vector of regression coefficients that includes an intercept term. Similarly, the logit transformation of the probability of extra zeros in a zero-inflated distribution or the probability of positive counts in an HZ distribution can be assumed to be a linear combination of the covariates. There have been numerous applications of these regression models in count data analysis. Lewsey and Thomson [13] applied the ZIP regression and ZINB regression models to count data related to dental epidemiology. Lambert [14] applied the ZIP regression

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.

model to defects in a manufacturing process. Hall [15] employed random-effect ZIP regression and zero-inflated binomial regression to model horticultural count data. Boucher and Denuit [16] compared the performance of fixed-effect and random-effect Poisson regression models using motor insurance claim counts. The Poisson, NB, ZIP, and ZINB regression models have been applied to predict the yield in semiconductor manufacturing [1], [2], [6]. This article also explores the application of regression models based on the GP distribution and/or HZ distributions for yield estimation by using the spatial location of each IC chip on a wafer as a covariate for the corresponding defect count listed in a wafer map.

This article only focuses on statistical modeling of the defect counts on semiconductor wafers to predict the yield of IC chips. Other modeling approaches using distributed defect patterns, such as spatial nonhomogeneous Poisson processes [17], are not discussed here. We mainly consider fixed-effect models and adopt the maximum likelihood (ML) method to estimate the parameters of the yield models. The remainder of this paper is organized as follows. Section II presents the GP, ZIGP, and HZGP distributions, as well as their regression extensions, to model defect counts on semiconductor wafers. In Section III, the parameters of interest of these models are estimated by the ML method, and the inference on yield estimates is provided based on the asymptotic variance–covariance of the estimators. Section IV uses three real wafer map datasets to compare the different models. Finally, Section V concludes this article.

II. Defect Count Models

This section illustrates several modeling approaches to analyze defect counts listed on a wafer map to predict the yield of IC chips in semiconductor manufacturing. The GP, ZIGP, and HZGP distributions, as well as their spatial regression extensions, are discussed. Whereas the NB model has a more general variance specification than the Poisson model, the GP, ZIGP, and HZGP models differ from the NB model in terms of the first moment specification. The phenomenon of excess zeros (inflated) can be caused by clustering. HZ or zero-inflated specification can account for these excess zeros, although those models may cause overfitting problems. HZ and zero-inflated models can be thought of as refined models of truncation and censoring [12].

Suppose that a semiconductor wafer consists of n mutually exclusive chips, C_i, for $i = 1, 2, \ldots, n$. Additionally, let $\mathbf{z}_i$ denote the spatial location of the ith chip center and X_i denote the random defect count in C_i. Furthermore, let $\mathbf{x} \equiv \{x_1, x_2, \ldots, x_n\}$ represent the observed defect counts on a wafer.

A. GP Distribution

First, we ignore the spatial information and assume that the random defect counts X_is are independent and identically distributed GP random variables. The GP distribution for a nonnegative integer-valued random variable X has the following pmf:

$$P(X = x|\mu, \phi) = \frac{\mu(\mu + (\phi - 1)x)^{x-1}}{x!}\phi^{-x} \exp\left(-\frac{\mu + (\phi - 1)x}{\phi}\right) \quad (2)$$

$x = 0, 1, 2, \ldots$, where $\mu(>0)$ is the mean parameter and $\phi(>0)$ is the dispersion parameter. The mean and variance of the GP distribution are $E(X) = \mu$ and $\mathrm{Var}(X) = \mu\phi^2$, respectively. The probability of zero is $P(X = 0|\mu, \phi) = e^{-\mu/\phi}$. Clearly, the Poisson distribution is a special case of the GP distribution when the dispersion parameter $\phi = 1$. The cases with $\phi > 1$ and $0 < \phi < 1$ correspond to the overdispersion and underdispersion relative to the Poisson distribution, respectively. Unlike the NB distribution, which can only handle overdispersion, the GP distribution can model overdispersion, underdispersion, or no dispersion in the count data. Joe and Zhu [7] proved that the GP distribution is a Poisson mixture, like the NB distribution. The GP distribution can be used to control the thickness of the tail of the Poisson distribution by introducing the dispersion parameter. When the first two moments are fixed, the GP distribution has a heavier tail, while the NB distribution has a higher probability mass at zero. However, in many situations, the two distributions have pmfs with very similar shapes; thus, they may fit equally well with real count data. The GP distribution is preferable, especially when the occurrence of events is not likely to be independent (e.g., clustered defects in semiconductor wafers).

Various other parameterizations of the GP distribution have been adopted in the literature. For example, Joe and Zhu [7] used the following pmf:

$$P(X = x|\theta, \eta) = \frac{\theta(\theta + \eta x)^{x-1}}{x!} \exp(-\theta - \eta x)$$

for $\theta > 0$ and $\eta < 1$, where the mean and variance are given by $E(X) = \theta/(1 - \eta)$ and $\mathrm{Var}(X) = \theta/(1 - \eta)^3$, respectively. The two parameters η and θ are related to the parameters in model (2) as $\eta = (\phi - 1)/\mu$ and $\theta = \mu/\phi$. Wang and Famoye [18] adopted the following pmf:

$$P(X = x|\mu, \alpha) = \left(\frac{\mu}{1 + \alpha\mu}\right)^x \frac{(1 + \alpha x)^{x-1}}{x!} \exp\left(-\frac{\mu(1 + \alpha x)}{1 + \alpha\mu}\right)$$

for $\theta > 0$ and $-\infty < \alpha < \infty$. The mean and variance under this parameterization are $E(X) = \mu$ and $\mathrm{Var}(X) = \mu(1 + \alpha\mu)^2$, respectively. Gupter *et al.* [10] considered a ZIGP distribution, where the GP pmf has the form

$$P(X = x|\alpha, \theta) = \frac{(1 + \alpha x)^{x-1}}{x!} \cdot \frac{(\theta e^{-\alpha\theta})^x}{e^\theta}$$

for $\theta > 0$, $0 \le \alpha \le 1$, and $|\alpha\theta| < 1$. Here, the mean and variance are $E(X) = \theta/(1 - \alpha\theta)$ and $\mathrm{Var}(X) = \theta/(1 - \alpha\theta)^3$, respectively.

In this article, the parameterization given by (2) will be used; this is the case because the two parameters μ and ϕ have clear and intuitive meanings. In addition to having a simple mean and variance structure, this parameterization is also more intuitive

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.

when developing regression models by assuming that the mean parameter μ depends on spatial covariates.

B. ZIGP Distribution

As a useful model for handling count data with excess zeros, the pmf of the ZIGP distribution is given by

$$P(X=0|\pi,\mu,\phi) = \pi + (1-\pi)e^{-\mu/\phi}$$
$$P(X=x|\pi,\mu,\phi) = (1-\pi)\frac{\mu(\mu+(\phi-1)x)^{x-1}}{x!}\phi^{-x} \times \exp\left(-\frac{\mu+(\phi-1)x}{\phi}\right), x=1,2,\ldots \quad (3)$$

where $0 \le \pi \le 1$ is the probability of extra zeros. The ZIGP distribution is a mixture of a point mass at zero (a Bernoulli distribution) and a GP distribution. The setup in (3) is introduced to accommodate excess zeros (i.e., excess ICs with no defects for yield prediction). It is also possible to allow for a decreasing proportion of zero if $[-e^{-\mu/\phi}/((1-e^{-\mu/\phi})] < \pi < 1$, where the extreme case of $\pi = -e^{-\mu/\phi}/((1-e^{-\mu/\phi})$ yields the truncated-at-zero distribution [12]. The parameter π may be further parameterized, e.g., a logit function of some covariates, as suggested by Lambert [14].

C. HZGP Distribution

Another type of count distribution for count data with excess zeros is the so-called HZ distributions [19], e.g., the HZP, HZNB, and HZGP distributions. The basic idea underlying HZ formulations is that a binomial probability model governs the binary outcome of whether a count variate has a zero or a positive realization. If the realization is positive, the *hurdle* is crossed, and the conditional distribution of the positives is governed by a truncated-at-zero count data model. The HZ distributions assume that the zeros and positive counts are from two separate data-generating processes, e.g., Bernoulli distribution and a truncated-at-zero count distribution, respectively. For example, the pmf of the HZGP distribution is

$$P(X=0|\omega) = 1-\omega$$
$$P(X=x|\omega,\mu,\phi) = \left(\frac{\omega}{1-e^{-\mu/\phi}}\right)\frac{\mu(\mu+(\phi-1)x)^{x-1}}{x!}\phi^{-x} \times \exp\left(-\frac{\mu+(\phi-1)x}{\phi}\right), x=1,2,\ldots$$

where $0 \le \omega \le 1$ is the probability of positive counts. Note that the probability of zero is independent of both μ and ϕ. The zero-inflated and HZ distributions are mathematically equivalent when $\omega = (1-\pi)(1-e^{-\mu/\phi})$. However, the two distributions are based on different assumptions for data-generating processes. The zero-inflated distribution assumes that the zeros and nonzeros are from the same data-generating process, which is described as a mixture of a point mass at zero and a GP distribution. Alternatively, the zeros and nonzeros in the HZ distribution are from two different data-generating processes.

D. Spatial Regression Models Based on the GP Distribution

This section presents the GP regression, ZIGP regression, and HZGP regression models, considering the spatial covariates. The HZP regression is not covered here because the HZP distribution is a special case of HZGP. The GP regression model assumes that a random defect count at the ith chip, X_i, follows the GP distribution with a location dependent mean parameter, μ_i, i.e.,

$$P(X_i=x_i|\boldsymbol{\theta}) = \frac{\mu_i(\mu_i+(\phi-1)x)^{x_i-1}}{x_i!}\phi^{-x_i} \times \exp\left(-\frac{\mu_i+(\phi-1)x_i}{\phi}\right), x_i=0,1,\ldots$$

for $i=1,2,\ldots,n$. The mean number of defects at the ith chip with a center location $\mathbf{z}_i$, $E[X_i] \equiv \mu_i$, is assumed to have the form

$$\zeta(\mu_i) = \mathrm{f}(\mathbf{z}_i)^T\boldsymbol{\beta}$$

where $\mathrm{f}(\mathbf{z}_i)$ denotes the $(q\times 1)$ vector of spatial covariates evaluated at the ith chip location $\mathbf{z}_i$, and $\boldsymbol{\beta}$ is the $(q\times 1)$ unknown vector of regression coefficients including an intercept term. The number of parameters (q) should be less than the number of chips (n) to be estimable. The spatial covariates may be Cartesian coordinates of the circuit centers or polar coordinates measured by using the center of the wafer as the origin. The dominant effects of defect clustering are expressed via spatial variables $\mathrm{f}(\mathbf{z}_i)$. $\zeta(\cdot)$ represents a link function to variables that are functions of the chip location. Here, the functional form of $\zeta(\cdot)$ is chosen as $\zeta(\mu_i) = \ln(\mu_i) = \mathrm{f}(\mathbf{z}_i)^T\boldsymbol{\beta}$. The parameter vector of the GP regression model to be estimated is $\boldsymbol{\theta} \equiv (\boldsymbol{\beta},\phi)$.

The ZIGP regression model is given by

$$P(X_i=0|\boldsymbol{\theta}) = \pi_i + (1-\pi_i)e^{-\mu_i/\phi}$$
$$P(X_i=x_i|\boldsymbol{\theta}) = (1-\pi_i)\frac{\mu_i(\mu_i+(\phi-1)x_i)^{x_i-1}}{x_i!}\phi^{-x_i} \times \exp\left(-\frac{\mu_i+(\phi-1)x_i}{\phi}\right), x_i=1,2,\ldots$$

for $\boldsymbol{\theta} \equiv (\boldsymbol{\beta},\boldsymbol{\gamma},\phi)$. We employ the log and logit link functions for μ and π, respectively: $\ln(\mu_i) = \mathrm{f}(\mathbf{z}_i)^T\boldsymbol{\beta}$ and logit $(\pi_i) = \mathrm{g}(\mathbf{z}_i)^T\boldsymbol{\gamma}$, for $i=1,2,\ldots,n$. Similarly, the HZGP regression model is given by

$$P(X_i=0|\boldsymbol{\theta}) = 1-\omega_i,$$
$$P(X_i=x_i|\boldsymbol{\theta}) = \left(\frac{\omega_i}{1-e^{-\mu_i/\phi}}\right)\frac{\mu_i(\mu_i+(\phi-1)x_i)^{x_i-1}}{x_i!}\phi^{-x_i} \times \exp\left(-\frac{\mu_i+(\phi-1)x_i}{\phi}\right), x_i=1,2,\ldots.$$

We choose the same functional forms of $\ln(\mu_i) = \mathrm{f}(\mathbf{z}_i)^T\boldsymbol{\beta}$ and $\mathrm{logit}(\omega_i) = \mathrm{h}(\mathbf{z}_i)^T\boldsymbol{\eta}$. The parameter set is $\boldsymbol{\theta} \equiv (\boldsymbol{\beta},\boldsymbol{\eta},\phi)$.

Following the approach proposed by Bae *et al.* [1], we choose $\mathbf{z}_i \equiv (r_i,\varphi_i)$, i.e., the polar coordinates of the ith chip center relative to the wafer center and $\mathrm{f}(\mathbf{z}_i) = \mathrm{g}(\mathbf{z}_i) = \mathrm{h}(\mathbf{z}_i) = (1, r_i, \cos\varphi_i, \sin\varphi_i, r_i\cos\varphi_i, r_i\sin\varphi_i)^T$. This is done because

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.

the distributed defects tend to present radial and angular variation on the wafers [1]. Herein, we set $q = 6$.

III. Inference and Yield Estimation

This section outlines the ML method used to estimate the parameters of the models presented in Section II and the yield of IC chips on a wafer. When estimating the yield, we simply assume that every defect causes a yield loss; thus, the yield is equal to the probability of zero estimated by a model. We partition the observed defect counts into two sets: the zero count set $S_0 \equiv \{i|x_i = 0\}$ and the positive count set $S_p \equiv \{i|x_i > 0\}$ for $i = 1, 2, \ldots, n$. Let n_0 represent the number of zeros in a dataset, that is, the cardinality of set S_0 is defined as $n_0 = |S_0| = \sum_{i=1}^{n} \mathcal{I}_{\{x_i=0\}}$, where $\mathcal{I}_{\{\cdot\}}$ is an indicator function. In addition, we denote the number of nonzeros in a dataset by $n_p \equiv n - n_0$. Numerical methods (e.g., the $fsolve$ function in MATLAB or GNU Octave) are needed to solve the systems of nonlinear equations required by the ML method. Alternatively, the *ZIGP* package in R provides a convenient function that can be used to find the MLEs of the GP/ZIGP distributions and the GP/ZIGP regression models.

A. GP Distribution

Given the observed defect counts x on a wafer, the model parameters μ and ϕ of the GP distribution can be estimated by the ML method, which finds the parameter values that maximize the likelihood function

$$\mathcal{L}(\mu, \phi; \mathrm{x}) = \prod_{i=1}^{n} P(X = x_i|\mu, \phi)$$
$$= \prod_{i=1}^{n} \frac{\mu(\mu + (\phi - 1)x_i)^{x_i - 1}}{x_i!} \phi^{-x_i} \exp\left(-\frac{\mu + (\phi - 1)x_i}{\phi}\right).$$

It is practically more convenient to maximize the log-likelihood function $\ell(\mu, \phi; \mathrm{x}) \equiv \ln \mathcal{L}(\mu, \phi; \mathrm{x})$. MLEs of μ and ϕ, which are denoted by $\hat{\mu}$ and $\hat{\phi}$, respectively, are solutions of the following two equations:

$$\frac{\partial \ell(\mu, \phi; \mathrm{x})}{\partial \mu} = \frac{n}{\mu} - \frac{n}{\phi} + \sum_{i=1}^{n} \frac{x_i - 1}{\mu + (\phi - 1)x_i} = 0$$

$$\frac{\partial \ell(\mu, \phi; \mathrm{x})}{\partial \phi} = \sum_{i=1}^{n} \frac{x_i(x_i - 1)}{\mu + (\phi - 1)x_i} - \frac{2\sum_{i=1}^{n} x_i}{\phi} + \frac{n\mu + (\phi - 1)\sum_{i=1}^{n} x_i}{\phi^2} = 0.$$

Once MLEs of the GP parameters are obtained, MLEs of any functions of the model parameters can be computed. For example, the MLE of the probability of zero, which is equivalent to the IC chip yield, is

$$\hat{Y} \equiv \widehat{P(X = 0)} \equiv P(X = 0|\hat{\mu}, \hat{\phi}) = e^{-\hat{\mu}/\hat{\phi}}.$$

We can construct confidence intervals for the model parameters and yield based on the standard errors derived from the (observed) Fisher information matrix. A large-sample approximation of estimated standard errors of the MLEs of μ and ϕ is given by the inverse of the (observed) Fisher information matrix [20]

$$\hat{\boldsymbol{\Sigma}}_{\hat{\mu},\hat{\phi}} \equiv \begin{bmatrix} \widehat{\mathrm{Var}}(\hat{\mu}) & \widehat{\mathrm{Cov}}(\hat{\mu}, \hat{\phi}) \\ & \widehat{\mathrm{Var}}(\hat{\phi}) \end{bmatrix}$$
$$= \begin{bmatrix} -\frac{\partial^2 \ell}{\partial \mu^2}\Big|_{\hat{\mu},\hat{\phi}} & -\frac{\partial^2 \ell}{\partial \mu \partial \phi}\Big|_{\hat{\mu},\hat{\phi}} \\ & -\frac{\partial^2 \ell}{\partial \phi^2}\Big|_{\hat{\mu},\hat{\phi}} \end{bmatrix}^{-1}$$

where the second partial derivatives of the log-likelihood function are given by

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\frac{n}{\mu^2} - \sum_{i=1}^{n} \frac{x_i - 1}{(\mu + (\phi - 1)x_i)^2}$$

$$\frac{\partial^2 \ell}{\partial \mu \partial \phi} = \frac{n}{\phi^2} - \sum_{i=1}^{n} \frac{x_i(x_i - 1)}{(\mu + (\phi - 1)x_i)^2}$$

$$\frac{\partial^2 \ell}{\partial \phi^2} = -\frac{2(n\mu - \sum_{i=1}^{n} x_i)}{\phi^3} + \frac{\sum_{i=1}^{n} x_i}{\phi^2} - \sum_{i=1}^{n} \frac{x_i^2(x_i - 1)}{(\mu + (\phi - 1)x_i)^2}.$$

These second partial derivatives are evaluated at $\mu = \hat{\mu}$ and $\phi = \hat{\phi}$. The standard errors of $\hat{\mu}$ and $\hat{\phi}$ are estimated by $\widehat{\mathrm{s.e.}}(\hat{\mu}) = \sqrt{\widehat{\mathrm{Var}}(\hat{\mu})}$ and $\widehat{\mathrm{s.e.}}(\phi) = \sqrt{\widehat{\mathrm{Var}}(\hat{\phi})}$, respectively. One can then compute an approximate $(1 - \alpha)100\%$ confidence interval for a model parameter (e.g., for μ, $\hat{\mu} \pm z_{\alpha/2} \cdot \widehat{\mathrm{s.e.}}(\hat{\mu})$). Because μ is positive, we instead construct an approximate $(1 - \alpha)100\%$ confidence interval for μ based on the log transformation as [20]

$$\left[\hat{\mu}/\exp(z_{\alpha/2} \cdot \widehat{\mathrm{s.e.}}(\hat{\mu})/\hat{\mu}), \quad \hat{\mu} \times \exp(z_{\alpha/2} \cdot \widehat{\mathrm{s.e.}}(\hat{\mu})/\hat{\mu})\right].$$

The standard error of the yield estimate $\hat{Y}$ can be estimated by the delta method as

$$\widehat{\mathrm{s.e.}}(\hat{Y}) \equiv \sqrt{\widehat{\mathrm{Var}}(\hat{Y})} = \sqrt{\nabla\hat{Y}^T \hat{\boldsymbol{\Sigma}}_{\hat{\mu},\hat{\phi}} \nabla\hat{Y}}$$

where $\nabla\hat{Y} \equiv (\frac{\partial Y}{\partial \mu}|_{\hat{\mu},\hat{\phi}}, \frac{\partial Y}{\partial \phi}|_{\hat{\mu},\hat{\phi}})^T$. Thus, for the GP distribution $Y = e^{-\mu/\phi}$, we have

$$\frac{\partial Y}{\partial \mu} = -\frac{e^{-\mu/\phi}}{\phi} \quad \text{and} \quad \frac{\partial Y}{\partial \phi} = \frac{\mu e^{-\mu/\phi}}{\phi^2}.$$

An approximate $(1 - \alpha)100\%$ confidence interval can be constructed for Y based on the logit transformation [20]

$$\left[\frac{\hat{Y}}{\hat{Y} + (1 - \hat{Y}) \times w}, \frac{\hat{Y}}{\hat{Y} + (1 - \hat{Y})/w}\right]$$

where $w = \exp(z_{\alpha/2} \cdot \widehat{\mathrm{s.e.}}(\hat{Y})/\hat{Y}(1 - \hat{Y}))$.

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.

B. ZIGP Distribution

The likelihood function of the ZIGP distribution is given by

$$\mathcal{L}(\pi,\mu,\phi;\mathbf{x}) = \left[\pi + (1-\pi)e^{-\mu/\phi}\right]^{n_0} \times \prod_{i\in S_p}\left\{(1-\pi)\frac{\mu(\mu+(\phi-1)x_i)^{x_i-1}}{x_i!}\phi^{-x_i} \exp\left(-\frac{\mu+(\phi-1)x_i}{\phi}\right)\right\}.$$

MLEs of the three parameters are obtained by setting the first partial derivatives of the log-likelihood function with respect to the three model parameters. As a result, MLEs of π, μ, and ϕ are solutions of the following three equations:

$$\begin{aligned}
&\pi + (1-\pi)e^{-\mu/\phi} = \frac{n_0}{n}\\
&-\frac{n(1-\pi)e^{-\mu/\phi}}{\phi} + \frac{n_p}{\mu} - \frac{n_p}{\phi} + \sum_{i\in S_p}\frac{x_i-1}{\mu+(\phi-1)x_i} = 0\\
&\frac{n\mu(1-\pi)e^{-\mu/\phi}}{\phi^2} + \sum_{i\in S_p}\frac{x_i(x_i-1)}{\mu+(\phi-1)x_i} - \frac{2\sum_{i\in S_p}x_i}{\phi}\\
&+\frac{n_p\mu + (\phi-1)\sum_{i\in S_p}x_i}{\phi^2} = 0.
\end{aligned} \tag{4}$$

According to (4), the yield estimated by the ZIGP distribution is

$$\hat{Y} = \hat{\pi} + (1-\hat{\pi})e^{-\hat{\mu}/\hat{\phi}} = \frac{n_0}{n}.$$

Therefore, the estimated probability of zero (i.e., the yield) of the ZIGP distribution is always equal to the observed fraction of zeros in the dataset. This is true for any other zero-inflated distributions, including the ZIP and ZINB distributions.

Similar to the GP distribution, the variance–covariance matrix for the MLEs $\hat{\boldsymbol{\theta}} \equiv (\hat{\pi},\hat{\mu},\hat{\phi})$ is estimated via the (observed) Fisher information matrix

$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}} \equiv \begin{bmatrix} \widehat{\mathrm{Var}}(\hat{\pi}) & \widehat{\mathrm{Cov}}(\hat{\pi},\hat{\mu}) & \widehat{\mathrm{Cov}}(\hat{\pi},\hat{\phi}) \\ & \widehat{\mathrm{Var}}(\hat{\mu}) & \widehat{\mathrm{Cov}}(\hat{\mu},\hat{\phi}) \\ & & \widehat{\mathrm{Var}}(\hat{\phi}) \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2\ell}{\partial\pi^2}\Big|_{\hat{\boldsymbol{\theta}}} & -\frac{\partial^2\ell}{\partial\pi\partial\mu}\Big|_{\hat{\boldsymbol{\theta}}} & -\frac{\partial^2\ell}{\partial\pi\partial\phi}\Big|_{\hat{\boldsymbol{\theta}}} \\ & -\frac{\partial^2\ell}{\partial\mu^2}\Big|_{\hat{\boldsymbol{\theta}}} & -\frac{\partial^2\ell}{\partial\mu\partial\phi}\Big|_{\hat{\boldsymbol{\theta}}} \\ & & -\frac{\partial^2\ell}{\partial\phi^2}\Big|_{\hat{\boldsymbol{\theta}}} \end{bmatrix}^{-1}$$

where

$$\begin{aligned}
\frac{\partial^2\ell}{\partial\pi^2} &= -\frac{n_0(1-p_0)^2}{Y^2} - \frac{n_p}{(1-\pi)^2},\\
\frac{\partial^2\ell}{\partial\pi\partial\mu} &= \frac{n_0}{\phi}\cdot\frac{p_0}{Y^2}\\
\frac{\partial^2\ell}{\partial\pi\partial\phi} &= -\frac{n_0\mu}{\phi^2}\cdot\frac{p_0}{Y^2}\\
\frac{\partial^2\ell}{\partial\mu^2} &= \frac{n_0}{\phi^2}\cdot\frac{\pi(1-\pi)p_0}{Y^2} - \frac{n_p}{\mu^2} - \sum_{i\in S_p}\frac{x_i-1}{[\mu+(\phi-1)x_i]^2}\\
\frac{\partial^2\ell}{\partial\mu\partial\phi} &= \frac{n_0}{\phi^2}\cdot\frac{(1-\pi)p_0}{Y} - \frac{n_0\mu}{\phi^3}\cdot\frac{\pi(1-\pi)p_0}{Y^2} + \frac{n_p}{\phi^2}\\
&\quad - \sum_{i\in S_p}\frac{x_i(x_i-1)}{[\mu+(\phi-1)x_i]^2}\\
\frac{\partial^2\ell}{\partial\phi^2} &= -\frac{2n_0\mu}{\phi^3}\cdot\frac{(1-\pi)p_0}{Y} + \frac{n_0\mu^2}{\phi^4}\cdot\frac{\pi(1-\pi)p_0}{Y^2}\\
&\quad - \frac{2(n_p\mu - \sum_{i\in S_p}x_i)}{\phi^3} - \sum_{i\in S_p}\frac{x_i^2(x_i-1)}{[\mu+(\phi-1)x_i]^2}\\
&\quad + \frac{\sum_{i\in S_p}x_i}{\phi^2}.
\end{aligned}$$

Note that $Y = \pi + (1-\pi)e^{-\mu/\phi}$ is the probability of zero (i.e., the yield) of the ZIGP distribution, while $p_0 = e^{-\mu/\phi}$ is the probability of zero of the GP distribution. In addition, $\nabla\hat{Y} \equiv (\frac{\partial Y}{\partial\pi}|_{\hat{\boldsymbol{\theta}}}, \frac{\partial Y}{\partial\mu}|_{\hat{\boldsymbol{\theta}}}, \frac{\partial Y}{\partial\phi}|_{\hat{\boldsymbol{\theta}}})^T$ with

$$\frac{\partial Y}{\partial\pi} = 1-p_0, \quad \frac{\partial Y}{\partial\mu} = -\frac{(1-\pi)p_0}{\phi}, \quad \frac{\partial Y}{\partial\phi} = \frac{(1-\pi)\mu p_0}{\phi^2}.$$

C. HZGP Distribution

The likelihood function of the HZGP distribution is given by

$$\mathcal{L}(\omega,\mu,\phi;\mathbf{x}) = (1-\omega)^{n_0}\left(\frac{\omega}{1-e^{-\mu/\phi}}\right)^{n_p} \times \prod_{i\in S_p}\frac{\mu(\mu+(\phi-1)x_i)^{x_i-1}}{x_i!}\phi^{-x_i}\exp\left(-\frac{\mu+(\phi-1)x_i}{\phi}\right)$$

and MLEs of the three model parameters are solutions of the following three equations:

$$\begin{aligned}
&1-\omega = \frac{n_0}{n}\\
&-\frac{n_x e^{-\mu/\phi}}{\phi(1-e^{-\mu/\phi})} + \frac{n_p}{\mu} - \frac{n_p}{\phi} + \sum_{i\in S_p}\frac{x_i-1}{\mu+(\phi-1)x_i} = 0\\
&\frac{n_x\mu e^{-\mu/\phi}}{\phi^2(1-e^{-\mu/\phi})} + \sum_{i\in S_p}\frac{x_i(x_i-1)}{\mu+(\phi-1)x_i}\\
&-\frac{2\sum_{i\in S_p}x_i}{\phi} + \frac{n_p\mu+(\phi-1)\sum_{i\in S_p}x_i}{\phi^2} = 0.
\end{aligned} \tag{5}$$

Again, the yield estimated by the HZGP distribution, or any other HZ distributions, is equal to the observed fraction of zeros in the dataset, which is $\hat{Y} = 1-\hat{\omega} = n_0/n$ according to (5). The variance–covariance matrix for the MLEs $\hat{\boldsymbol{\theta}} \equiv (\hat{\omega},\hat{\mu},\hat{\phi})$ is

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.

estimated by

$$\hat{\Sigma}_{\hat{\theta}} \equiv \begin{bmatrix} \widehat{\text{Var}}(\hat{\omega}) & 0 & 0 \\ & \widehat{\text{Var}}(\hat{\mu}) & \widehat{\text{Cov}}(\hat{\mu},\hat{\phi}) \\ & & \widehat{\text{Var}}(\hat{\phi}) \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 \ell}{\partial \omega^2}\Big|_{\hat{\theta}} & 0 & 0 \\ & -\frac{\partial^2 \ell}{\partial \mu^2}\Big|_{\hat{\theta}} & -\frac{\partial^2 \ell}{\partial \mu \partial \phi}\Big|_{\hat{\theta}} \\ & & -\frac{\partial^2 \ell}{\partial \phi^2}\Big|_{\hat{\theta}} \end{bmatrix}^{-1}.$$

Note that $\hat{\omega}$ is independent of $\hat{\mu}$ and $\hat{\phi}$. Herein, the second partial derivatives of the log-likelihood function are given by

$$\frac{\partial^2 \ell}{\partial \omega^2} = -\frac{n_0}{(1-\omega)^2} - \frac{n_p}{\omega^2}$$

$$\frac{\partial^2 \ell}{\partial \mu^2} = \frac{n_p p_0}{\phi^2(1-p_0)^2}$$

$$\frac{\partial^2 \ell}{\partial \phi^2} = -\frac{n_p \mu p_0}{\phi^3(1-p_0)^2} + \frac{n_p}{\phi^2(1-p_0)} - \sum_{i\in S_p} \frac{x_i(x_i-1)}{[\mu+(\phi-1)x_i]^2}$$

$$\frac{\partial^2 \ell}{\partial \mu \partial \phi} = \frac{n_p \mu^2 p_0}{\phi^4(1-p_0)^2} - \frac{2 n_p \mu}{\phi^3(1-p_0)} + \frac{(\phi+2)\sum_{i\in S_p} x_i}{\phi^3} - \sum_{i\in S_p} \frac{x_i^2(x_i-1)}{[\mu+(\phi-1)x_i]^2}.$$

Because the estimated yield is only dependent on ω, $\nabla \hat{Y}$ is simply equal to $(-1,0,0)^T$.

D. Spatial Regression Models Based on the GP Distribution

The likelihood function of the GP regression model is given by

$$\mathcal{L}(\boldsymbol{\theta};\mathrm{x}) = \prod_{i=1}^{n} P(X = x_i|\mu_i,\phi) = \prod_{i=1}^{n} \frac{\mu_i(\mu_i+(\phi-1)x_i)^{x_i-1}}{x_i!}\phi^{-x_i}\exp\left(-\frac{\mu_i+(\phi-1)x_i}{\phi}\right)$$

for $i = 1,2,\ldots,n$, where $\boldsymbol{\theta} \equiv (\beta,\phi)$. Herein, $\ln(\mu_i) = \mathrm{y}_i^T\beta$ for $\mathrm{y}_i \equiv \mathrm{f}(\mathrm{z}_i) = (1, r_i, \cos\varphi_i, \sin\varphi_i\ r_i\cos\varphi_i, r_i\sin\varphi_i)^T$ and $\beta = (\beta_1,\beta_2,\ldots,\beta_6)^T$, which includes the intercept term β_1. MLEs of the model parameters $\beta_1,\beta_2,\ldots,\beta_6$ and ϕ are solutions of the following equations:

$$\sum_{i=1}^{n}\frac{x_i(x_i-1)}{\mu_i+(\phi-1)x_i} + \sum_{i=1}^{n}\frac{\mu_i+(\phi-1)x_i}{\phi^2} - \frac{2\sum_{i=1}^{n}x_i}{\phi} = 0$$

$$\sum_{i=1}^{n} y_{ij} + \sum_{i=1}^{n}\frac{(x_i-1)\mu_i y_{ij}}{\mu_i+(\phi-1)x_i} - \sum_{i=1}^{n}\frac{\mu_i y_{ij}}{\phi} = 0, j = 1,\ldots,6$$

where y_{ij} is the jth element of the covariate vector y_i, for $j = 1,2,\ldots,q$. The yield estimated by the GP regression model is

$$\hat{Y} = \frac{1}{n}\sum_{i=1}^{n} P(X_i = 0|\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n} e^{-\hat{\mu}_i/\hat{\phi}}$$

where $\hat{\mu}_i = \exp(\mathrm{y}_i^T\hat{\beta})$ for $i = 1,2,\ldots,n$. The variance–covariance matrix for $\hat{\boldsymbol{\theta}} \equiv (\hat{\beta},\hat{\phi})$ is estimated via the (observed) Fisher information matrix. Detailed derivations of the second partial derivatives for the GP, ZIGP, and HZGP regression models are not given in this article for conciseness.

The likelihood function of the ZIGP regression model is given by

$$\mathcal{L}(\boldsymbol{\theta};\mathrm{x}) = \prod_{i\in S_0}(\pi_i + (1-\pi_i)e^{-\mu_i/\phi}) \times \prod_{i\in S_p}\left\{(1-\pi_i)\frac{\mu_i(\mu_i+(\phi-1)x_i)^{x_i-1}}{x_i!}\phi^{-x_i} \times \exp\left(-\frac{\mu_i+(\phi-1)x_i}{\phi}\right)\right\}$$

where $\ln(\mu_i) = \mathrm{y}_i^T\beta$ and $\text{logit}(\pi_i) = \mathrm{y}_i^T\gamma$ for the set of parameters $\boldsymbol{\theta} \equiv (\beta,\gamma,\phi)$. MLEs of the model parameters are obtained by solving the following equations:

$$\sum_{i\in S_0}\frac{\mu_i(1-\pi_i)e^{-\mu_i/\phi}}{\phi^2(\pi_i+(1-\pi_i)e^{-\mu_i/\phi})} + \sum_{i\in S_p}\frac{x_i(x_i-1)}{\mu_i+(\phi-1)x_i} + \frac{1}{\phi^2}\sum_{i\in S_p}(\mu_i+(\phi-1)x_i) - \frac{2\sum_{i\in S_p}x_i}{\phi} = 0$$

$$-\sum_{i\in S_0}\frac{\mu_i y_{ij}(1-\pi_i)e^{-\mu_i/\phi}}{\phi(\pi_i+(1-\pi_i)e^{-\mu_i/\phi})} + \sum_{i\in S_p}\frac{(x_i-1)\mu_i y_{ij}}{\mu_i+(\phi-1)x_i} + \sum_{i\in S_p} y_{ij} - \frac{1}{\phi}\sum_{i\in S_p}\mu_i y_{ij} = 0, j = 1,\ldots,6$$

$$\sum_{i\in S_0}\frac{1-e^{-\mu_i/\phi}}{\pi_i+(1-\pi_i)e^{-\mu_i/\phi}}\frac{e^{\mathrm{y}_i^T\gamma}}{(e^{\mathrm{y}_i^T\gamma}+1)^2}y_{ij} - \sum_{i\in S_p}\frac{1}{1-\pi_i}\frac{e^{\mathrm{y}_i^T\gamma}}{(e^{\mathrm{y}_i^T\gamma}+1)^2}y_{ij} = 0, j = 1,\ldots,6.$$

The yield estimated by the ZIGP regression model is given by

$$\hat{Y} = \frac{1}{n}\sum_{i=1}^{n} P(X_i = 0|\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\pi}_i + (1-\hat{\pi}_i)e^{-\hat{\mu}_i/\hat{\phi}})$$

where $\hat{\mu}_i = \exp(\mathrm{y}_i^T\hat{\beta})$ and $\hat{\pi}_i = e^{\mathrm{y}_i^T\hat{\gamma}}/(e^{\mathrm{y}_i^T\hat{\gamma}}+1)^2$ for $i = 1,2,\ldots,n$. The variance–covariance matrix for $\hat{\boldsymbol{\theta}} \equiv (\hat{\gamma},\hat{\beta},\hat{\phi})$ is estimated via the (observed) Fisher information matrix.

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.

The likelihood function of the HZGP regression model is

$$\mathcal{L}(\boldsymbol{\theta};\mathrm{x}) = \prod_{i\in S_0}(1-\omega_i) \times \prod_{i\in S_p}\left\{\frac{\omega_i}{1-e^{-\mu_i/\phi}}\cdot\frac{\mu_i(\mu_i+(\phi-1)x_i)^{x_i-1}}{x_i!} \phi^{-x_i}\exp\left(-\frac{\mu_i+(\phi-1)x_i}{\phi}\right)\right\}$$

where $\ln(\mu_i)=\mathbf{y}_i^T\boldsymbol{\beta}$, and $\mathrm{logit}(\omega_i)=\mathbf{y}_i^T\boldsymbol{\eta}$ for the set of paramters $\boldsymbol{\theta}\equiv(\boldsymbol{\beta},\boldsymbol{\eta},\phi)$. MLEs of the model parameters are solutions of the following equations:

$$\sum_{i\in S_p}\left[\frac{\mu_i e^{-\mu_i/\phi}}{\phi^2(1-e^{-\mu_i/\phi})}\frac{x_i(x_i-1)}{\mu_i+(\phi-1)x_i}-\frac{2x_i}{\phi} +\frac{\mu_i+(\phi-1)x_i}{\phi^2}\right]=0$$

$$\sum_{i\in S_p}\left[y_{ij}+\frac{(x_i-1)\mu_i y_{ij}}{\mu_i+(\phi-1)x_i}-\frac{\mu_i y_{ij}}{\phi}\cdot\frac{1}{1-e^{\mu_i/\phi}}\right]=0$$

$$-\sum_{i\in S_0}\frac{1}{1-\omega_i}\cdot\frac{e^{\mathbf{y}_i^T\boldsymbol{\eta}}}{(e^{\mathbf{y}_i^T\boldsymbol{\eta}}+1)^2}y_{ij} +\sum_{i\in S_p}\frac{1}{\omega_i}\frac{e^{\mathbf{y}_i^T\boldsymbol{\eta}}}{(e^{\mathbf{y}_i^T\boldsymbol{\eta}}+1)^2}y_{ij}=0$$

for $j=1,2,\ldots,6$. The yield estimated by the HZGP regression model is given by

$$\hat{Y}=\frac{1}{n}\sum_{i=1}^{n}P(X_i=0|\hat{\theta})=\frac{1}{n}\sum_{i=1}^{n}(1-\hat{\omega}_i)$$

where $\hat{\omega}_i=e^{\mathbf{y}_i^T\hat{\boldsymbol{\eta}}}/(e^{\mathbf{y}_i^T\hat{\boldsymbol{\eta}}}+1)^2$ for $i=1,2,\ldots,n$.

The variance–covariance matrix for $\hat{\boldsymbol{\theta}}\equiv(\hat{\eta},\hat{\beta},\hat{\phi})$, $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$, is estimated by

$$\begin{bmatrix}\left[-\frac{\partial^2\ell}{\partial\eta_j\partial\eta_k}\Big|_{\hat{\boldsymbol{\theta}}}\right]_{6\times6} & \mathbf{0}_{6\times6} & \mathbf{0}_{6\times1}\\ & \left[-\frac{\partial^2\ell}{\partial\beta_j\partial\beta_k}\Big|_{\hat{\boldsymbol{\theta}}}\right]_{6\times6} & \left[-\frac{\partial^2\ell}{\partial\beta_j\partial\phi}\Big|_{\hat{\boldsymbol{\theta}}}\right]_{6\times1}\\ & & -\frac{\partial^2\ell}{\partial\phi^2}\Big|_{\hat{\boldsymbol{\theta}}}\end{bmatrix}^{-1}$$

for $j,k=1,2,\ldots,6$. For the HZGP regression model, $\frac{\partial^2\ell}{\partial\eta_j\partial\beta_k}=\frac{\partial^2\ell}{\partial\eta_j\partial\phi}=0$. This is due to the assumption of separated data-generating processes for zero and positive counts.

The ZIGP and HZGP regression models are no longer mathematically equivalent due to their different ways of covariates introduction. The estimation of parameters η, as well as the probability of zero for the HZGP regression model, is separated from the estimation of other parameters (β and ϕ). Thus, the probability of zero estimated by the HZGP regression model is the observed fraction of zero. As a result, the HZGP regression model may provide more accurate yield estimation than the ZIGP regression model.

This section outlined the ML method to estimate the unknown model parameters. There are other parameter estimation methods that can be applied to the proposed models in this article. For example, the expectation-maximization (EM) algorithm may be applied to fit the zero-inflated models [14]. The EM algorithm defines an unobserved indicator variable δ_i for each count observation x_i in order to indicate whether x_i is an inflated zero. Hall [15] also mentioned the Newton–Raphson and Newton–Raphson with Fisher scoring algorithms as alternatives for maximizing the log-likelihood functions of ZIP and binomial regression models.

IV. Data Example

This section presents the analytical results collected when the different count models were applied to the data of three real wafer maps. Fig. 1 shows the defect count data on the three wafers. The data were analyzed by Yuan *et al.* [6] using the Poisson regression, NB regression, ZIP regression, and ZINB regression models for yield prediction in a hierarchical Bayesian framework. Each wafer consists of $n=473$ IC chips (or dies). Table I summarizes the descriptive statistics of defect counts from the three wafer maps. There are an excessive number of zeros on each wafer, which result from defect clustering [21]. We modeled the three wafer map data by 18 models, including the nonregression Poisson, NB, GP, ZIP, ZINB, ZIGP, HZP, HZNB, and HZGP distributions, along with their corresponding regression models. The *fsolve* function in GUN Octave was employed to find MLEs of the applied models. Additionally, Table II lists the R packages that can be used to fit some of the models.

Table III lists the MLEs and 95% confidence intervals of the model parameters (shown in parentheses) of the nine nonregression distributions. The dispersion parameter ϕ of the GP distribution is greater than 1 for all three wafers, indicating the phenomenon of overdispersion in all three datasets. The ϕ value of the ZIGP and HZGP distributions is less than the ϕ value of the corresponding GP distribution; after modeling the excess zeros, the remainder of the data become less dispersive. For the same reason, the a values of the ZINB and HZNB distributions are higher than the a value of the NB distribution for each dataset. The MLEs for the regression models are not presented in this article for conciseness. Variable selections can be conducted to build best regression models that include only significant covariate terms. However, for comparison purpose, we always used the full regression models that include all of the covariate terms.

Table IV presents the comparison between the estimated yield values ($\hat{Y}$) and observed yield values (n_0/n) for the three wafers, along with their 95% confidence intervals. Table V compares the performance of various applied models using three criteria: the maximum log-likelihood ($\hat{\ell}$), AIC, and relative bias for yield estimation. The AIC value of a model is calculated according to [22] as $\mathrm{AIC}=2\,k-2\hat{\ell}$, where $\hat{\ell}$ is the maximum value of the log-likelihood function and k is the number of model parameters that measures the complexity of a model. The log-likelihood assesses the fit of a model, and the AIC includes a penalty to discourage overfitting. Among a set of candidate

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.

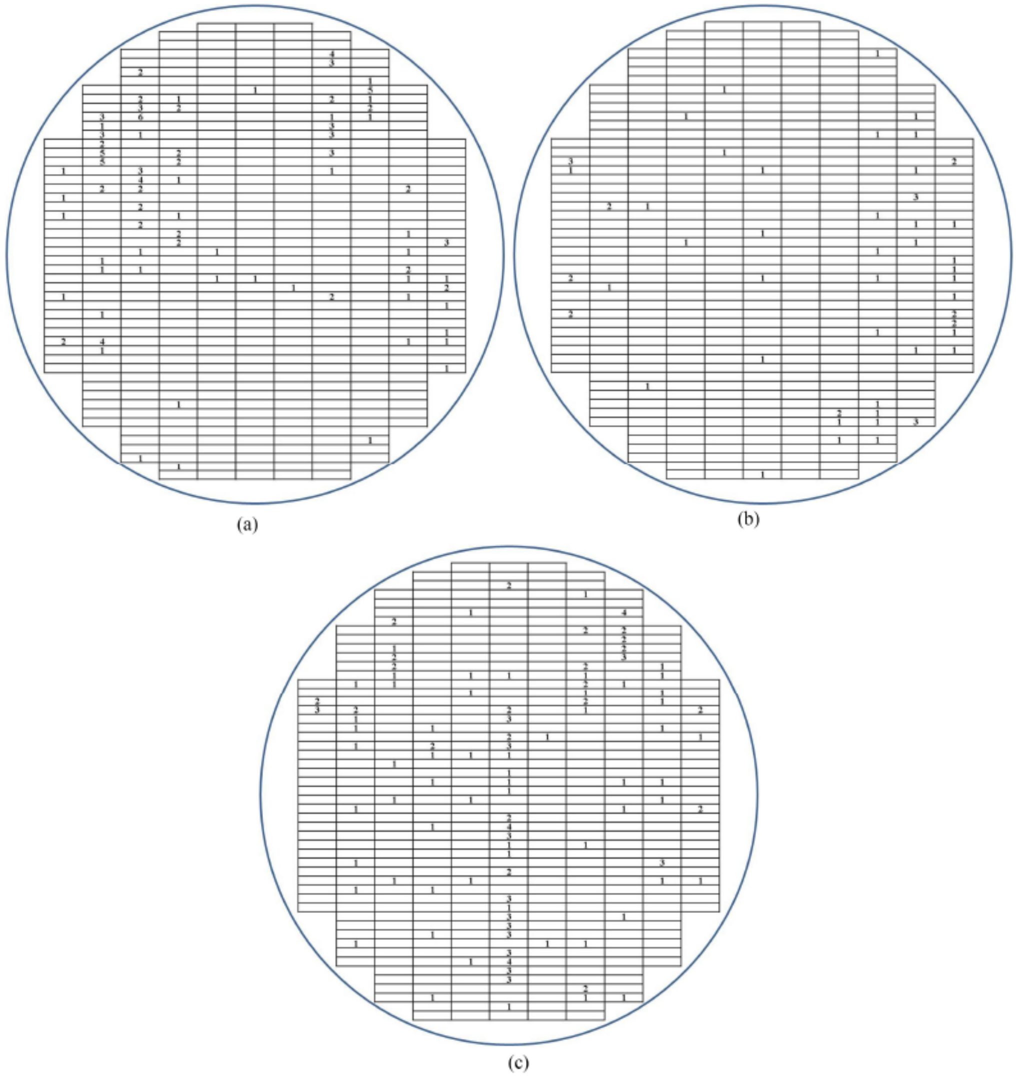

Fig. 1. (a)–(c) Examples of defect counts on three real wafers [6].

TABLE I
DESCRIPTIVE STATISTICS OF THE EXAMPLE DATASETS

Wafer	Relative Frequency				Mean	Variance
	0	1	2	≥ 3		
(a)	0.8436	0.0825	0.0402	0.0338	0.290	0.668
(b)	0.8985	0.0803	0.0148	0.0063	0.129	0.180
(c)	0.7928	0.1290	0.0444	0.0338	0.326	0.551

TABLE II
R PACKAGES USED FOR FITTING THE MODELS

Package	Function	Models
Stats	*glm*	Poisson, Poisson regression
MASS	*glm.nb*	NB, NB regression
pscl	*zeroinfl*	ZIP, ZINB, ZIP regression, ZINB regression
	hurdle	HZP, HZNB, HZP regression, HZNB regression
ZIGP	*est.zigp*	GP, ZIGP, GP regression, ZIGP regression

models, the model with the smallest AIC value is preferred. In addition, we assume that every defect is a fault; thus, the probability of zero is actually the yield of a wafer. We then define the relative bias for yield estimation of a model as

$$\frac{\text{estimated yield} - \text{observed yield}}{\text{estimated yield}} \times 100\%.$$

In Table V, there is no remarkable trend indicating that any one model outperforms the others (among all 18 models) in terms of the three performance criteria. However, we have observed some general trends, which will be discussed in the following.

A. Nonregression Distributions

We first compare the results from the nine nonregression distributions, which lead to the following observations.

1) The Poisson distribution does not fit the wafer data well. For every wafer, the Poisson distribution always has the lowest log-likelihood value, the highest AIC value, and the largest absolute value of the relative bias for yield estimation. In particular, the Poisson distribution significantly underestimates the true yield. The NB and GP

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.

TABLE III
MLEs OF THE NONREGRESSION DISTRIBUTIONS

Wafer	Poisson θ	Poisson $\hat{\theta}$	NB θ	NB $\hat{\theta}$	GP θ	GP $\hat{\theta}$	ZIP θ	ZIP $\hat{\theta}$	ZINB θ	ZINB $\hat{\theta}$	ZIGP θ	ZIGP $\hat{\theta}$	HZP θ	HZP $\hat{\theta}$	HZNB θ	HZNB $\hat{\theta}$	HZGP θ	HZGP $\hat{\theta}$
(a)	μ	0.290 (0.245, 0.342)	μ	0.290 (0.221, 0.380)	μ	0.290 (0.219, 0.383)	μ	1.390 (1.107, 1.747)	μ	1.014 (0.530, 1.940)	μ	1.056 (0.616, 1.810)	μ	1.390 (1.107, 1.747)	μ	1.014 (0.530, 1.940)	μ	1.056 (0.616, 1.810)
			α	0.177 (0.118, 0.266)	ϕ	1.676 (1.384, 2.029)			α	1.681 (0.215, 13.13)	ϕ	1.250 (0.961, 1.626)			α	1.681 (0.215, 13.13)	ϕ	1.250 (0.961, 1.626)
							π	0.792 (0.739, 0.836)	π	0.714 (0.509, 0.858)	π	0.726 (0.564, 0.844)	ω	0.156 (0.126, 0.192)	ω	0.156 (0.126, 0.192)	ω	0.156 (0.126, 0.192)
(b)	μ	0.129 (0.100, 0.166)	μ	0.129 (0.095, 0.174)	μ	0.129 (0.095, 0.174)	μ	0.500 (0.296, 0.845)	μ	0.305 (0.015, 6.200)	μ	0.332 (0.033, 3.327)	μ	0.500 (0.296, 0.845)	μ	0.305 (0.015, 6.200)	μ	0.332 (0.033, 3.327)
			α	0.298 (0.134, 0.663)	ϕ	1.202 (1.045, 1.383)			α	1.319 (0.001, 2462)	ϕ	1.097 (0.752, 1.599)			α	1.319 (0.001, 2462)	ϕ	1.097 (0.752, 1.599)
							π	0.742 (0.600, 0.847)	π	0.577 (0.007, 0.996)	π	0.611 (0.036, 0.985)	ω	0.101 (0.077, 0.132)	ω	0.101 (0.077, 0.132)	ω	0.101 (0.077, 0.132)
(c)	μ	0.326 (0.278, 0.381)	μ	0.326 (0.262, 0.404)	μ	0.326 (0.262, 0.404)	μ	0.984 (0.770, 1.258)	μ	0.888 (0.523, 1.507)	μ	0.897 (0.551, 1.461)	μ	0.984 (0.770, 1.258)	μ	0.888 (0.523, 1.507)	μ	0.897 (0.551, 1.461)
			α	0.377 (0.239, 0.594)	ϕ	1.379 (1.207, 1.575)			α	6.498 (0.063, 665.3)	ϕ	1.060 (0.838, 1.341)			α	6.498 (0.063, 665.3)	ϕ	1.060 (0.838, 1.341)
							π	0.669 (0.590, 0.740)	π	0.633 (0.433, 0.796)	π	0.637 (0.455, 0.787)	ω	0.207 (0.173, 0.246)	ω	0.207 (0.173, 0.246)	ω	0.207 (0.173, 0.246)

TABLE IV
YIELD ESTIMATION

Wafer	Observed Yield	Regression/ Nonregression	Poisson	NB	GP	ZIP	ZINB	ZIGP	HZP	HZNB	HZGP
(a)	0.8436	Nonregression	0.7485 (0.7105, 0.7831)	0.8423 (0.8065, 0.8726)	0.8413 (0.8053, 0.8716)	0.8436 (0.8080, 0.8736)	0.8436 (0.8080, 0.8736)	0.8436 (0.8080, 0.8736)	0.8436 (0.8080, 0.8736)	0.8436 (0.8080, 0.8736)	0.8436 (0.8080, 0.8736)
		Regression	0.7629 (0.7285, 0.7943)	0.8419 (0.8070, 0.8715)	0.8385 (0.8031, 0.8686)	0.8336 (0.7993, 0.8629)	0.8389 (0.8048, 0.8681)	0.8423 (0.8089, 0.8708)	0.8436 (0.8087, 0.8730)	0.8436 (0.8087, 0.8730)	0.8436 (0.8087, 0.8730)
b)	0.8985	Nonregression	0.8790 (0.8476, 0.9047)	0.8984 (0.8677, 0.9226)	0.8983 (0.8676, 0.9225)	0.8985 (0.8679, 0.9227)	0.8985 (0.8679, 0.9227)	0.8985 (0.8679, 0.9227)	0.8985 (0.8679, 0.9227)	0.8985 (0.8679, 0.9227)	0.8985 (0.8679, 0.9227)
		Regression	0.8831 (0.8535, 0.9073)	0.8975 (0.8672, 0.9214)	0.8991 (0.8701, 0.9223)	0.8894 (0.8608, 0.9127)	0.9007 (0.8712, 0.9241)	0.9000 (0.8708, 0.9232)	0.8985 (0.8690, 0.9220)	0.8985 (0.8690, 0.9220)	0.8985 (0.8690, 0.9220)
(c)	0.7928	Nonregression	0.7221 (0.6835, 0.7577)	0.7908 (0.7517, 0.8252)	0.7897 (0.7504, 0.8242)	0.7928 (0.7539, 0.8270)	0.7928 (0.7539, 0.8270)	0.7928 (0.7539, 0.8270)	0.7928 (0.7539, 0.8270)	0.7928 (0.7539, 0.8270)	0.7928 (0.7539, 0.8270)
		Regression	0.7251 (0.6871, 0.7601)	0.7905 (0.7516, 0.8248)	0.7904 (0.7515, 0.8245)	0.7848 (0.7462, 0.8189)	0.7898 (0.7519, 0.8232)	0.7906 (0.7532, 0.8236)	0.7928 (0.7543, 0.8266)	0.7928 (0.7543, 0.8266)	0.7928 (0.7543, 0.8266)

TABLE V
MODEL COMPARISON

Criterion	Wafer	Regression/ Nonregression	Poisson	NB	GP	ZIP	ZINB	ZIGP	HZP	HZNB	HZGP
$\hat{\ell}$	(a)	Nonregression	-366.5	-300.8	-301.9	-301.5	-299.6	-299.6	-301.5	-299.6	-299.6
		Regression	-334.7	-285.8	-290.3	-274.9	-272.5	-270.9	-280.8	-280.6	-280.7
	(b)	Nonregression	-196.2	-186.9	-187.0	-187.0	-186.9	-186.9	-187.0	-186.9	-186.9
		Regression	-181.6	-177.0	-174.0	-171.6	-171.6	-167.9	-170.0	-170.0	-169.8
	(c)	Nonregression	-374.2	-343.1	-343.8	-341.8	-341.6	-341.6	-341.8	-341.6	-341.6
		Regression	-368.3	-339.9	-339.6	-328.1	-325.8	-325.6	-333.1	-333.1	-333.0
AIC	(a)	Nonregression	735.1	605.6	607.7	607.0	605.1	605.2	607.0	605.1	605.2
		Regression	681.3	585.5	594.6	573.8	571.1	567.7	567.7	587.3	587.5
	(b)	Nonregression	394.3	377.9	378.0	378.1	379.8	379.8	378.1	379.8	379.8
		Regression	375.2	368.0	362.0	367.2	369.2	361.8	364.1	366.1	365.6
	(c)	Nonregression	750.4	690.2	691.6	687.5	689.2	689.2	687.5	689.2	689.2
		Regression	748.6	693.8	693.3	680.2	677.6	677.1	690.1	692.1	692.0
Relative bias for yield estimation	(a)	Nonregression	-11.26%	-0.14%	-0.27%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Regression	-9.56%	-0.19%	-0.60%	-1.19%	-0.55%	-0.14%	0.00%	0.00%	0.00%
	(b)	Nonregression	-2.17%	-0.01%	-0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Regression	-1.72%	-0.12%	0.07%	0.25%	0.25%	0.17%	0.00%	0.00%	0.00%
	(c)	Nonregression	-8.92%	-0.25%	-0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Regression	-8.54%	-0.29%	-0.31%	-1.01%	-0.38%	-0.28%	0.00%	0.00%	0.00%
Average of the absolute relative bias for yield est.		Nonregression	7.45%	0.13%	0.23%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Regression	6.61%	0.20%	0.33%	0.82%	0.39%	0.20%	0.00%	0.00%	0.00%

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.

distributions greatly improve the yield estimation results compared to the Poisson distribution. For all three wafers, the NB distribution produces slightly better results than the GP distribution. Joe and Zhu [7] pointed out that the NB distribution has a larger probability of zero than the GP distribution when the first two moments are fixed. Therefore, it may provide a better fit to datasets with excess zeros.

2) An HZ distribution and its corresponding zero-inflated distribution produce identical results because they are mathematically equivalent, as discussed in Section II-B. All six of the zero-inflated and HZ distributions correctly estimate the observed yield values, which has been discussed in Section III.
3) The ZINB, ZIGP, HZNB, and HZGP distributions give the best (and almost identical) results for every wafer according to the maximum log-likelihood and relative bias for yield estimation. However, they may not always be the preferred models according to the AIC values. The NB distribution and the ZIP distribution have the lowest AIC values for wafers (b) and (c), respectively.
4) Note that the approximate 95% confidence intervals for the dispersion parameter a in the ZINB/HZNB distributions are very wide, especially for wafers (b) and (c). After accounting for extra zeros in the dataset, the remainder of the dataset may not exhibit overdispersion. This is evidenced by the fact that the ϕ values of the ZIGP/HZGP distributions are almost equal to one.
5) The GP distribution might be a better distribution than the NB distribution to build zero-inflated and/or HZ models because the GP distribution can handle all of the cases (i.e., no dispersion, overdispersion, and underdispersion), while the NB distribution can model overdispersion only.

Overall, the ZINB, ZIGP, HZNB, and HZGP distributions provide the best fit to the data by modeling extra zeros in the wafers, providing more accurate yield estimates for the count data. The improvements of the ZIGP/HZGP distributions over the GP distribution are more substantial than the improvements of the ZINB/HZNB distributions over the NB distribution. This result is consistent with the findings of Joe and Zhu [7], who compared the NB, GP, ZINB, and ZIGP distributions. They observed that the fits to their example data by the four distributions were ranked in the order of "ZIGP $>$ ZINB $>$ NB $>$ GP," and they concluded that the ZIGP distribution may be more sensitive for detecting extra zeros than the ZINB distribution.

B. Regression Models

Next, we examine the results from the nine regression models, which lead to observation of the following general trends.

1) Regression models always provide better fits to defect counts on the wafers than the corresponding nonregression distributions according to the maximum log-likelihood values; this is the case because more parameters are included in the regression models. However, the regression models do not always produce lower AIC values.
2) The ZIGP regression model is preferred among all of the nonregression and regression models in terms of both the AIC and log-likelihood criteria for all three example wafers.
3) Zero-inflated regression models and their corresponding HZ regression models do not produce identical results. It is noted that the HZ regression models provide perfect yield estimates that are equal to the observed yields.
4) The HZNB and HZGP regression models tend to produce very similar results. There is no definite conclusion regarding which one is better in terms of the AIC and log-likelihood criteria. The HZGP regression model is better for wafers (b) and (c), while the HZNB regression model is preferred for wafer (a).

The HZNB and HZGP regression models are likely to be very promising for yield prediction based on defect counts on wafers. Due to its ability to model overdispersion or underdispersion, as well as no dispersion, the GP distribution is advantageous compared to the NB distribution to build zero-inflated or HZ regression models. It is possible that after accounting for the extra zeros in a dataset, the remainder of the data becomes homogeneous or even underdispersed. In such a situation, GP-based zero-inflated/HZ models may fit the data better than the corresponding NB-based models. For example, when the HZNB regression model is applied to fit the data of wafer (b), the dispersion parameter is estimated to be $\hat{a} = 5255$, indicating no overdispersion for the positive counts. As a result, the HZNB and HZP regression models produce identical estimation results for wafer (b). However, when the HZGP regression is applied to wafer (b), the estimated dispersion parameter is $\hat{\phi} = 0.885 < 1$, implying an underdispersion for the positive counts; thus, the HZGP regression model fits the data of wafer (b) better than the HZNB/HZP regression models.

This section compared the yield models using three real wafer map datasets. There have been the studies from a various perspective on defect patterns and yield estimation in semiconductor manufacturing using simulation studies (e.g., [6], [21]). There may exist a major limitation to generate the real defect patterns occurring in semiconductor manufacturing. Defects in semiconductor manufacturing are generated in the highly complex processes and caused by various causes. The existing studies simulated defect patterns in the wafer based on some simplified assumptions. For example, defects in a curvilinear or ring shaped pattern were simulated by assuming that the defects are uniformly distributed along and about the curve or circle [21]. Therefore, the simulated defect patterns may not fully resemble the real observed defect patterns. For this reason, this article only illustrated the performance of the various yield models using real wafer map data.

V. Conclusion

This article compared a variety of models (including regression models) for modeling defect counts to predict IC chip yields in semiconductor manufacturing. Our results indicated that the GP distribution was a competitive alternative to the NB distribution to build models that could predict defect counts

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.

on IC chips, while the HZ approach was promising for yield modeling. Particularly, HZ models based on the NB and GP distributions provided perfect yield prediction for our example datasets. If simpler models were preferred, the ZINB, ZIGP, HZNB, and HZGP distributions were recommended to predict the yield of IC chips on the wafers. GP-based models and their NB-based counterparts usually produce comparable (or identical) results. There is no definite preference between the two models in terms of the yield prediction.

This article assumed that every defect is a fault that causes a reduction in the yield. However, in practice, not all defects are fatal to the function of IC chips. When a defect is located in a defect-sensitive area (i.e., "critical area"), it becomes a fault. The fault probability derived from the critical area and defect-size distribution can be incorporated into the yield computation of IC chips on the wafers. This article only covered fixed-effect models and used the ML method to estimate model parameters. Random-effect models or Bayesian approaches can be further explored in the future. In addition, based on the yield models presented in this article, the yield–reliability relation could be further investigated to predict the extrinsic device reliability.

References

[1] S. J. Bae, J. Y. Hwang, and W. Kuo, "Yield prediction via spatial modeling of clustered defect counts across a wafer map," *IIE Trans.*, vol. 39, no. 12, pp. 1073–1083, 2007.

[2] W. Kuo, W.-T. K. Chien, and T. Kim, *Reliability, Yield, and Stress Burn-in*. Boston, MA, USA: Kluwer, 1998.

[3] A. Tyagi and M. A. Bayoumi, "Defect clustering viewed through generalized Poisson distribution," *IEEE Trans. Semicond. Manuf.*, vol. 5, no. 3, pp. 196–206, Aug. 1992.

[4] L.-I. Tong and L.-C. Chao, "Novel yield model for integrated circuits with clustered defects," *Expert Syst. Appl.*, vol. 34, pp. 2334–2341, 2008.

[5] D. S. Hochbaum and S. Liu, "Adjacency-clustering and its application for yield prediction in integrated circuit manufacturing," *Oper. Res.*, vol. 66, no. 6, pp. 1457–1759, 2018

[6] T. Yuan, S. Z. Ramadan, and S. J. Bae, "Yield prediction for integrated circuits manufacturing through hierarchical Bayesian modeling of spatial defects," *IEEE Trans. Rel.*, vol. 60, no. 4, pp. 729–741, Dec. 2011.

[7] H. Joe and R. Zhu, "Generalized Poisson distribution: The property of mixture of Poisson and comparison with negative binomial," *Biometrical J.*, vol. 47, no. 2, pp. 219–229, 2005.

[8] K. C. H. Yip and K. K. W. Yau, "On modeling claim frequency data in general insurance with extra zeros," *Insurance: Math. Econ.*, vol. 36, no. 2, pp. 153–163, 2005.

[9] S. Srivastava and L. Chen, "A two-parameter generalized Poisson model to improve the analysis of RNA-seq data," *Nucleic Acids Res.*, vol. 38, no. 17, 2010, Art. no. e170.

[10] P. L. Gupta, R. C. Gupta, and R. C, Tripathi,"Analysis of zero-inflated count data," *Comput. Statist. Data Anal.*, vol. 23, pp. 207–218, 1996.

[11] J. Angers and A. Biswas, "A Baysian analysis of zero-inflated generalized Poisson model," *Comput. Statist. Data Anal.*, vol. 42, pp. 37–46, 2003.

[12] S. Gurmu and P. K. Trivedi, "Excess zeros in count models for recreational trips," *J. Bus. Econ. Statist.*, vol. 14, no. 4, pp. 469–477, 1996.

[13] J. D. Lewsey and W. M. Thomson, "The utility of the zero-inflated Poisson and zero-inflated negative binomial models: A case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status," *Community Dentistry Oral Epidemiol.*, vol. 32, pp. 183–189, 2004.

[14] D. Lambert, "Zero-inflated Poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, pp. 1–14, 1992.

[15] D. B. Hall, "Zero-inflated Poisson and binomial regression with random effects: A case study," *Biometrics*, vol. 56, pp. 1030–1039, 2000.

[16] J. P. Boucher and M. Denuit, "Fixed versus random effects in Poisson regression models for claim counts: A case study with motor insurance," *Astin Bull.*, vol. 36, pp. 285–301, 2006.

[17] J. Y. Hwang, W. Kuo, and C. Ha, "Modeling of integrated circuit yield using a spatial nonhomogeneous Poisson process," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 377–384, Aug. 2011.

[18] W. Wang and F. Famoye, "Modeling household fertility decisions with generalized Poisson regression," *J. Population Econ.*, vol. 10, no. 3, pp. 273–283, 1997.

[19] J. Mullahy, "Specification and testing of some modified count data models," *J. Econometrics*, vol. 33, pp. 341–365, 1986.

[20] W. M. Meeker and L. A. Esconar, *Statistical Methods for Reliability Data*. New York, NY, USA: Wiley, 1998.

[21] T. Yuan, W. Kuo, and S. J. Bae, "Detection of spatial defect patterns generated in semiconductor fabrication processes," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 392–403, Aug. 2011.

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

Authorized licensed use limited to: OHIO UNIVERSITY. Downloaded on December 14,2020 at 00:38:45 UTC from IEEE Xplore. Restrictions apply.