Quality Engineering

ISSN: 0898-2112 (Print) 1532-4222 (Online) Journal homepage: http://www.tandfonline.com/loi/lqen20

A discrete spatial model for wafer yield prediction

Hao Wang, Bo Li, Seung Hoon Tong, In-Kap Chang & Kaibo Wang

To cite this article: Hao Wang, Bo Li, Seung Hoon Tong, In-Kap Chang & Kaibo Wang (2018) A discrete spatial model for wafer yield prediction, Quality Engineering, 30:2, 169-182, DOI: 10.1080/08982112.2017.1328063

To link to this article: https://doi.org/10.1080/08982112.2017.1328063

Accepted author version posted online: 16 May 2017.
Published online: 12 Jul 2017.

QUALITY ENGINEERING
2018, VOL. 30, NO. 2, 169–182
https://doi.org/10.1080/08982112.2017.1328063

Taylor & Francis
Taylor & Francis Group

A discrete spatial model for wafer yield prediction

Hao Wang[a], Bo Li[b], Seung Hoon Tong[c], In-Kap Chang[c], and Kaibo Wang[a]

[a]Department of Industrial Engineering, Tsinghua University, Beijing, China; [b]School of Economics and Management, Tsinghua University, Beijing, China; [c]Department of Quality Assurance, Memory Business, Samsung Electronics, Hwasung, Korea

ABSTRACT

Yield analysis is one of the key concerns in the fabrication of semiconductor wafers. An effective yield analysis model will contribute to production planning and control, cost reductions and the enhanced competitiveness of enterprises. In this article, we propose a novel discrete spatial model based on defect data on wafer maps for analyzing and predicting wafer yields at different chip locations. More specifically, based on a Bayesian framework, we propose a hierarchical generalized linear mixed model, which incorporates both global trends and spatially correlated effects to characterize wafer yields with clustered defects. Both real and simulated data are used to validate the performance of the proposed model. The experimental results show that the newly proposed model offers an improved fit to spatially correlated wafer map data.

KEYWORDS

Bayesian model; generalized linear model; spatial data; wafer; yield prediction

Introduction

Integrated circuits (ICs) are some of the most wildly used electronic components in industrial production and daily life. As realizations of ICs, chips are sets of electronic circuits that are interconnected on semiconductor material plates (wafers) to fulfill complex electronic functions. The relationship between a semiconductor wafer and the IC chips that are produced from it is illustrated in Figure 1. Through a series of complex production procedures, hundreds or even thousands of chips, such as memory, microcomputers, and sensors, can be simultaneously fabricated on the same silicon wafer. With the rapid development of fabrication techniques, the integration scale of modern ICs has become extremely large. In recent decades, the modern IC production scale has continued to increase and has reached the level of billions (Schaller 1997; Thompson and Parthasarathy 2006).

However, since the fabrication of the first generation of such chips in the 1960s, the difficulties encountered in chip production have never been entirely overcome and, indeed, have become more challenging. Improving yield is still critical for enhancing manufacturing performance. Weber (2004) proposed a model that aligns yield with profitability. Milor (2013) attributed a firm yield concern to the short life cycles and rapid price decline in the chip industry. The author also noted that, with the same cost of materials and time, a higher yield is equal to greater revenue. Fabricating a chip on a wafer requires hundreds of physical and chemical steps and weeks or months of time. As observed by Maly et al. (2006), the variability in a manufacturing process is usually the primary reason for the production of nonconforming chips. In a memory production process, nonconforming chips are identified in a functional testing step before packaging; defective chips can then be plotted on a colored wafer map. These wafer maps help technicians to diagnose abnormal failure patterns and explore potential root causes of failures.

The occurrence of noncomforming chips directly affects the yield of a process, which is critical for production productivity, material planning and quality control. Accurate yield prediction is always appealing for proper production planning and control (Joseph and Adya 2002). In wafer fabrication, yield is commonly defined as the proportion of the total number of chips that are successfully produced. Many researchers have proposed different yield models for process characterization and other purposes. In general, yield losses

CONTACT Kaibo Wang kbwang@tsinghua.edu.cn Department of Industrial Engineering, Tsinghua University, No. 1 Qinghua Yuan, Haidian District, Beijing 100084, China.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lqen.

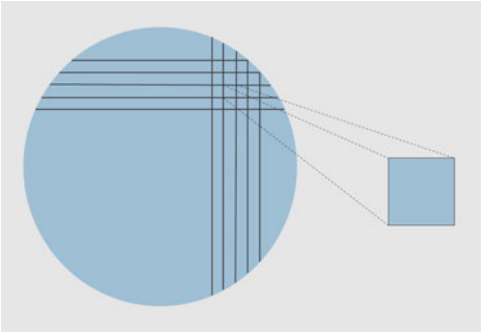

Figure 1. Relationship between semiconductor wafer (left) and IC chip (right).

can be regarded as the product of four components: the wafer process yield Y_{wp}, referring to the percentage of wafers arriving at the wafer probe step; the wafer probe yield Y_{cp}, referring to the percentage that pass the probe step; the assembly yield Y_{ap}, associated with the assembly process; and the final test yield Y_{ft}, referring to the percentage that survive through the final electrical test (Kuo and Kim 1999). Figure 2 summarizes these four yield losses in a flow chart.

Thus, the overall yield can be expressed as the product of the above components:

$$Y_{total} = Y_{wp}Y_{cp}Y_{ap}Y_{ft}. \quad [1]$$

Among these four components, the yield loss from the wafer probe step is the highest in the chip fabrication process (Milor 2013). In the wafer probe step, chips may fail to pass the necessary functional tests because of open circuits, short circuits or other causes (Kumar et al. 2006). Further studies of failure patterns reveal that losses identified in the probe step can be divided into two types: losses related to gross deformation in continuous regions and losses arising from random local deformations at specific sites (Mirza et al. 1995). In the literature, these failures have been treated differently.

Wafer probe yield models for random failures are usually constructed under the assumption of spatial independence among chips on a wafer; the distributions of conforming and nonconforming chips are assumed to be identical and independent, regardless of their spatial locations. Based on this assumption, several statistical yield models have been proposed and widely used in previous research focusing on the wafer probe yield to monitor the process and to predict the level of production capability. The Poisson model and the negative binomial model have been the most commonly used models for this purpose.

The Poisson yield model assumes that the defect number on every chip follows a Poisson distribution with a fixed parameter. If the parameter λ of this Poisson distribution is known, the probability that any specific chip contains x defects can be obtained as follows:

$$P(x) = \frac{e^{-x}\lambda^x}{x!}, x = 0, 1, 2, \ldots \quad [2]$$

Chips that contain zero defects are conforming and are included in the numerator for calculating the wafer probe yield. However, extensive engineering practice implies that the process conditions for adjacent chips on a wafer, such as the temperature and pressure in a local area, should usually be similar; thus, the defective chips on a wafer are likely to occur with a certain clustering tendency rather than entirely randomly and independently. Kuo and Kim (1999) showed that, when certain spatial patterns occur, a Poisson distribution underestimates the observed yield because it fails to consider any spatial information about a wafer. Figure 3 shows sixteen round wafer maps obtained after the wafer probe stage. Black dots in each map indicate defective chips. Many of the maps exhibit clustering in their defect patterns, whereas only a few do not exhibit such clustering.

To partially overcome this problem, the negative binomial (NB) model has been proposed to capture the clustering phenomenon (Kumar et al. 2006). Under the NB model, the occurrence probability of an event increases with other existing occurrences in the surrounding area. It is assumed that the density of defects D is a random variable that follows a gamma distribution with parameters α and β:

$$f(D) = \frac{D^{\alpha-1}e^{-D/\beta}}{\Gamma(\alpha)\beta^{\alpha}}, \quad [3]$$

where α is the shape parameter and β is the scale parameter. Then, the probability that a specific location

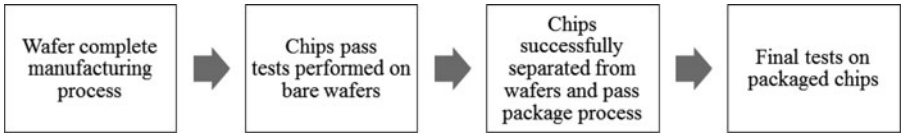

Figure 2. Four types of yield losses corresponding to different steps of the manufacturing process.

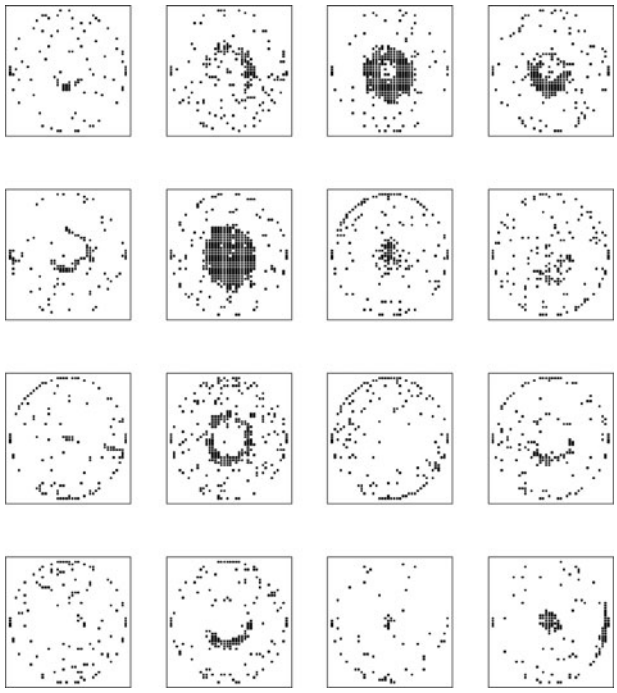

Figure 3. Real wafer maps indicating the yield on each chip, where defective chips are marked with black dots.

contains x defects is

$$P(x) = \int_0^\infty \frac{e^{-D}D^x}{x!}\frac{D^{\alpha-1}e^{-D/\beta}}{\Gamma(\alpha)\beta^\alpha}dD. \qquad [4]$$

Although the negative binomial model is significantly more flexible than the Poisson model, a univariate gamma distribution still fails to capture the spatial correlations among the defect densities at different location. A multivariate gamma distribution, however, will produce conceptual and computational difficulties (Banerjee et al. 2014).

As an alternative to the NB model, zero-inflated models are another commonly used type of model. These models are designed to address dominant zeros in datasets, which can be regarded as a coupled problem of clustered defects. Since the introduction of zero-inflated Poisson (ZIP) regression by Lambert (1992), various zero-inflated models have been proposed to fit data with many zero values such as the zero-inflated Poisson (ZIP) model, the zero-inflated binomial (ZIB) model and the zero-inflated negative binomial (ZINB) model (Yau and Lee 2001; Fatahi et al. 2012; He et al. 2012). Zero-inflated models are constructed based on the assumption that there is a random shock leading to a Poisson or NB process and that this random shock occurs independently with probability p.

However, in a wafer fabrication process, real data present a two-dimensional spatial structure. Specifically, both conforming and nonconforming chips are counted from die to die, and dies are spatially correlated with each other in a lattice on a wafer. One cannot ignore the fact that chips are fabricated simultaneously and geographically adjacent. Although zero-inflated models are useful for fitting certain types of data relevant to chip production, these models still ignore the spatial correlation between a chip and its neighbors. Considerable evidence shows that proximity implies correlation (Hansen et al. 1997; Hwang and Kuo 2007). It has also been shown that the defect density presents a significant radial dependency on the location (Ferris-Prabhu et al. 1987). In the study of Meyer and Pradhan (1989), a center-satellite model is discussed to describe the clustered distribution of defects. Stine et al. (1997) proposed a model that extracts die-level, wafer-level, and die-wafer interactions from the raw data. Chen and Liu (2000) applied an ART1 neural network to recognize the spatial defect pattern, and Di Palma et al. (2005) performed unsupervised spatial pattern classification. Kim et al. (2016) developed a step-down spatial randomness test based on multiple spatial fail bin test maps. Although pattern classification has become a major branch of research for discriminating wafer failures, it still cannot relate the yield information at a particular site to its location.

Despite the lack of spatial autocorrelation analysis in yield modeling, autocorrelation has been considered in the context of the continuous-spatial statistical quality control field. For the modeling of continuous spatial data, models have been proposed that incorporate both spatial and non-spatial components (Bao et al. 2014). However, such models cannot be applied to the discrete spatial data that are naturally collected in wafer yield studies. The phenomenon of the spatial clustering of data has also been studied and modeled from many perspectives such as the estimation of the profiles of major pollution. However, in semiconductor production and other quality management fields, this type of analysis is mainly used for defect pattern classification (Hwang and Kuo 2007).

In the literature, the mixed modeling of random spatial effects has gradually attracted the interest of researchers. Diggle et al. (1998) summarized the available geostatistical prediction models; many followed this framework and considered the need for an a priori correlated process. Such models are usually written

in the form of a generalized linear mixed model (GLMM), which is an extension of a generalized linear (GL) model. In recent years, mixed models have been widely used in geostatistics and biometric studies. Krueger and Montgomery (2014) proposed a GLMM model in which the random effects of sampling plans are considered. In their model, the random errors from observations within and between groups are described using a batch-specific model that produces estimates of the mean after the link function transformation.

In this article, we propose a new spatial model for analyzing and predicting the wafer probe yield based on discrete spatial data. The new model combines the consideration of the spatial coordinates and random spatial errors to capture both the macro-scale tendencies and micro-scale correlations of a wafer map. A Bayesian hierarchical structure is introduced in our study to endow the model with flexibility. Both simulated and real datasets are tested against the proposed model; the results show that the new model performs better than previous models.

The remainder of this article is organized as follows. In the second section, we introduce our proposed model for yield prediction. As an innovative feature of our yield model, the mechanism and implementation of the intrinsic conditional autoregressive component will be illustrated in detail. In the third section, we assess the performance of the proposed model using real datasets and compare our model with existing models. In the fourth section, the model is thoroughly evaluated using more extensive simulated datasets. Finally, the last section summarizes this study and presents interesting topics for further research.

Hierarchical GL model with spatial coordinates and iCAR components

Previous GL model with spatial coordinates: GLSp

Before presenting our model, we briefly introduce a model that considered spatial information. Bae et al. (2007) first used spatial polar coordinates to incorporate spatial information for a single wafer map. In their study, the authors constructed a generalized linear model with spatial coordinates to capture the macro-scale tendency of variations in the defect density; the model was constructed on a single wafer, and they adopted frequent models. Yuan et al. (2011) extended the study of Bae et al. (2007) by adopting a Bayesian framework. Another difference is that they considered variations not only within a single wafer but also across several wafers.

In the study of Bae et al. (2007), the model was constructed on a single wafer, and they did not adopt hierarchical Bayesian models. Instead, they chose the traditional frequent model and the EM algorithm to solve the problem.

The density of the Poisson yield model is assumed to be of the form

$$log(\lambda_i) = f(\boldsymbol{X}_i)\boldsymbol{\beta},\ i = 1, 2, \ldots n, \qquad [5]$$

where λ_i is the expected number (density) of defects and $f(\boldsymbol{X}_i)$ is a function of the chip's polar coordinates $X_i = [r, \theta]$ at the ith location. As shown above, because of the underlying physical mechanisms of their formation, defects often cluster in the center of the wafer and may exhibit a radially increasing pattern. Therefore, polar coordinates have a considerable advantage over Cartesian coordinates, although it is possible to convert from one to the other. Considering interaction effects,

$$f(\boldsymbol{X}_i) = [r_i, cos_i, sin_i, r_i cos_i, r_i sin_i]. \qquad [6]$$

In our research, we choose to reconstruct Bae et al. (2007)'s Poisson model in a Bayesian framework (hereinafter referred to as the GLsp model), which is a special case of Yuan et al. (2011)'s work, as a benchmark. The proposed model is also under Bayesian framework. The model evaluation and comparison are between these two hierarchical Bayesian models.

The reason why we choose Poisson distribution is that, first, zero-inflated models and our proposed model shares same property of allowing more zeros in the observed data and second, compared with zero-inflated models, our proposed model could capture more spatial information. Third, the implementation of zero-inflated models will encounter with some problems. In addition, in the motivation project of our research, the observed data do not always present significant "zero-inflated" pattern, but sometimes contains only a small percentage of "zeros."

Hierarchical GL mixed model with spatial coordinates and intrinsic conditional autoregressive components: GLSp-iCAR

Similar to the spatial model study proposed by Bae et al. (2007), this study focuses on a yield model for a single wafer. The new model has a Bayesian structure, which offers reasonable flexibility in combining information from different sources; a spatial Gaussian Markov random field (GMRF) is used to capture spatial autocorrelations among sites.

In our proposed model, a spatial random component introduces enough variations for the Poisson distributions. In "zero clustered regions," the spatial dependence will be relatively high, and $log(\lambda_i) = \boldsymbol{f(X)\beta} + s_i$ will be affected by s_i. Thus, $log(\lambda_i)$ in this region could be a sufficiently small negative value, which corresponds to a nearly zero Poisson rate at chip location i. By adding spatial components, the overdispersion phenomenon in observed data can be effectively captured.

In the following, we construct our hierarchical model in three stages. In the first stage, we use Poisson distributions to fit the observations $\boldsymbol{Y}$, where y_i depends on the latent variable λ_i. In the second stage, we assign a generalized linear mixed model to λ_i with a GMRF as the prior. In the third stage, unknown prior parameters are assigned.

At stage 1, we assume that the number of defects at the ith chip location on the wafer, Y_i, has an independent Poisson distribution given the density λ_i. Note that other distribution models are also applicable here. Now, for a total of n sites on a wafer map, the numbers of defect occurrence are dominated by n different Poisson processes:

$$\text{STAGE 1:} \quad Y_i \sim Poisson(\lambda_i), \ i = 1, 2, \ldots, n. \qquad [7]$$

In stage 2, we further assume that the density of this Poisson distribution is influenced by both its polar coordinates, exhibiting large-scale effects, and random spatial effects that can be interpreted in terms of interrelated micro-scale shifts. The expected values of the observed data are linearly related to some predictive variables via a link function:

$$\text{STAGE 2.}A\text{:} \quad log(\lambda_i) = f(\boldsymbol{X}_i)\boldsymbol{\beta} + s_i, \ i = 1, 2, \ldots, n, \qquad [8]$$

where $\boldsymbol{X}_i = [r_i, \theta_i]$ represents the polar coordinates of the ith location on the wafer and $f(\boldsymbol{X}_i) = [1, x_{i1}, x_{i2}, \ldots, x_{im}]$, the new predictive variables, is a function of polar coordinates and has a linear relation with the logarithm of the density λ_i. Here, we follow the work of Bae et al. (2007) and set $f(\boldsymbol{X}_i) = [1, r_i, cos_i, sin_i, r_i cos_i, r_i sin_i]$. $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_m]^T$ is the regression parameter of interest, which reflects a large-scale spatial tendency across the wafer. s_i represents random effects on location i depending on the neighborhood of the site and is characterized by an autoregressive property.

Without random spatial effects, the density of the Poisson process at any location depends only on its linear predictive variables $f(\boldsymbol{X})$ and coefficients $\boldsymbol{\beta}$, as we assume that the spatial tendency parameters $\boldsymbol{\beta}$ are unknown but deterministic for a particular wafer. However, sometimes, the real spatial pattern characteristics are not fully consistent with the macro-scale tendency assumed by this coordinates structure, in which i is a site on a continuous and very smooth curved surface. To include the influence of local shifts in the model, we assume here that there exist not only a macro-scale tendency but also a local variation (usually manifesting as local clustering), which is reflected by the random spatial error s_i. Thus, the macro-scale tendency and local clustering phenomena can be captured simultaneously at stage 2.

In a continuous model, random spatial effects can be modeled using a Gaussian random field. In our discrete data model, we adopted a GMRF (Gaussian Markov Random Field) to model spatial dependence because the data are observed on a lattice structure, and the GMRF can simplify the estimation of parameters by defining a sparse precision matrix $\boldsymbol{Q}$.

Here, we use intrinsic GMRFs, also called conditional autoregression (CAR) (Besag 1974, 1975), $s_i \sim car(\kappa)$, to depict the local variation and model the spatially dependent correlation. Markov random fields have previously been used to depict the clustering of random defects (Hansen et al. 1997). An intrinsic GMRF differs with a GMRF in that it does not use full-rank precision matrices, although the form of its definition is otherwise quiet similar. The random spatial components $\boldsymbol{s} = [s_1, s_2, \ldots, s_n]^T$ are defined with respect to an undirected graph $\boldsymbol{\mathcal{G} = (\mathcal{V}, \mathcal{E})}$, whose density function can be expressed in the following form:

$$\pi(\boldsymbol{s}) = (2\pi)^{-n/2}(|\boldsymbol{Q}|^*)^{1/2} \times \exp\left(-\frac{1}{2}(\boldsymbol{s}-\boldsymbol{\mu})^T\boldsymbol{Q}(\boldsymbol{s}-\boldsymbol{\mu})\right), \qquad [9]$$

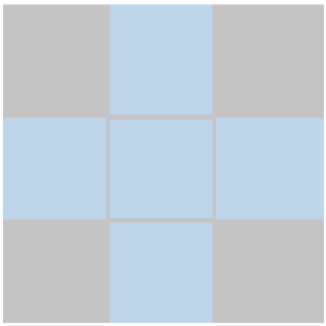

Figure 4. Neighborhood structure of the intrinsic GMRF on a wafer map; the light gray [blue] lattice sites are those designated as neighbors.

where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{Q}$ is the inverse of the symmetric semi-positive definite covariance matrix $\boldsymbol{\Sigma}$. The precision matrix $\boldsymbol{Q}$ plays a significant role in our study because it yields an intuitive interpretation of neighborhood structures and the conditional independence properties of $\boldsymbol{s}$. $Q_{ij} = 1$ if i and j are connected with each other. For a first-order intrinsic GMRF, we designate a structure for $\boldsymbol{Q}$ such that, for all i, $\sum_j Q_{ij} = 0$, which leads to a rank deficiency of the precision matrix. $|\boldsymbol{Q}|^*$ therefore becomes a generalized determinate. Here, "first order" refers to the following neighborhood structure selection strategy (Besag and Kooperberg 1995): let n_i denote the number of neighbors of chip i at location i, where only those sites that share an edge with the site of interest on the lattice wafer map are defined as neighbors, as shown in Figure 4. Then, the precision matrix $\boldsymbol{Q}$ of this first-order intrinsic GMRF is defined as

$$Q_{ij} = \kappa \begin{cases} n_i & i = j, \\ -1 & i \sim j, \\ 0 & \text{otherwise}, \end{cases} \qquad [10]$$

where κ is the random precision parameter of interest and the basic components of $\boldsymbol{Q}$ can be easily generated from the lattice wafer map.

From the above definition of the precision matrix, we can easily obtain

$$\text{STAGE 2.B}: \quad s_i|\boldsymbol{s}_{-i}, \kappa \sim N\left(\frac{1}{n_i}\sum_{j:j\sim i} s_j, \frac{1}{n_i\kappa}\right), \quad i = 1, 2, \ldots, n. \qquad [11]$$

The Markov property, $s_i \perp \boldsymbol{s}_{-\{i,nei(i)\}}|\boldsymbol{s}_{nei(i)}$, can be proven. Specifically, given all the neighbors' random effect values of s_i, s_i is conditionally independent of the non-neighbor nodes on the graph (Rue and Held 2005). Some of the properties of the GMRF are listed below:

$$E(s_i|\boldsymbol{s_{-i}}) = \mu_i - \frac{1}{Q_{ii}}\sum_{j:j\sim i} Q_{ij}(s_j - \mu_j),$$
$$Prec(s_i|\boldsymbol{s_{-i}}) = Q_{ii} \text{ and}$$
$$Corr(s_i, s_j|\boldsymbol{s_{-ij}}) = -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}, \ i \neq j. \qquad [12]$$

By incorporating the intrinsic GMRF, we introduce a spatial correlation $\boldsymbol{s}_i$ into our model after defining "neighbors" for every chip. Because our research focuses on a regular wafer map, which is a lattice system, a first-order neighborhood can be established without much effort. Thus, $\boldsymbol{s}$ follows a multivariate normal distribution with a specified covariance structure. The sparse matrix $\boldsymbol{Q}$ leads to the Markov property, which is highly convenient for the computation of parameters.

In stage 3, the prior distributions are prescribed by prior information. The unknown parameters include the regression parameter $\boldsymbol{\beta}$ and the random effect precision parameter κ. Because the reciprocal of κ, $\frac{1}{\kappa}$, reflects the scale of covariance, we define $\tau^2 = \frac{1}{\kappa}$ for defining the priors:

$$\text{STAGE 3.A}: \ \beta_j \sim N(m_j, v_j), \ j = 1, 2, \ldots, m$$
$$\text{STAGE 3.B}: \ \tau^2 \sim Inv - Gamma(a, b), \qquad [13]$$

where τ^2 is assigned an inverse-Gamma distribution. Thus, in prior, the precision parameter κ has a Gamma distribution. For generality, we set $a = 1$ and $b = 1$. An alternative is to use a uniform prior distribution $U(0, M_\tau)$ on an appropriate interval. $\boldsymbol{\beta}$ has independent Gaussian priors, where $m_j = 0$ and $v_j = 1000$, in our formulation.

Because $\boldsymbol{\beta}$ is also a GMRF with diagonal covariance and precision matrices, in the computation, we define an $(n + m)$-dimension GMRF $\boldsymbol{w} = (\boldsymbol{s}, \boldsymbol{\beta})^T$, which still has a sparse structure.

As the model is constructed under conditional framework, the common method used to estimate the parameters changes from the EM method to the Markov Chain Monte Carlo (MCMC) method under the above Bayesian structure. The MCMC method begins from an arbitrary value, generalizes proposed values until the Markov chain converges to the target distribution $\boldsymbol{\pi}()$ and then samples from the stable posterior distribution (Mengersen et al. 1999). Compared to other geostatistical models, the GMRF is very

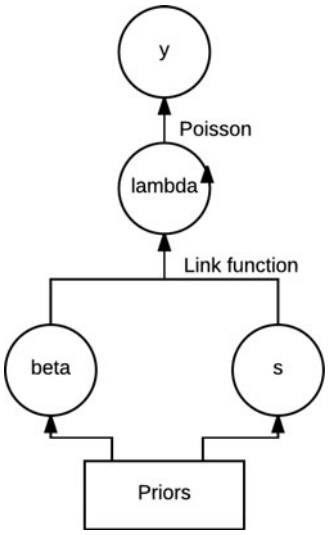

Figure 5. Hierarchical structure of the parameters in different stages.

computationally convenient. We can find the posterior distribution of $\boldsymbol{s}$ using a conditional sampling algorithm. The hierarchical structure of the parameters is shown in Figure 5.

As the problem size increases, the factorization of the precision matrix $\boldsymbol{Q}$ will become more time consuming. However, due to the sparse property of the precision matrix Q, computing the factorization of Q only requires $\mathcal{O}(n^{3/2})$ instead of $\mathcal{O}(n^3)$ flops. A computationally efficient alternative to MCMC is Integrated Nested Laplace Approximations (INLA), which provide a higher calculation speed $\mathcal{O}(nlog(n)^2)$ with reasonable accuracy (Held et al. 2010). Because the INLA approach produces a numerical approximation to the posteriors of interest and because the MCMC algorithm will guarantee the correct answer in the end, we did not adopt this method in our research. However, when the data size becomes extremely large, we could adopt the INLA method to obtain effective estimates. One real application example is the case where, if we use an ordinary PC to run the MCMC method for a 20 $\times$ 20 wafer map, the algorithm finishes within a minute. If the INLA method is applied to a wafer with 1,460 lattices, the algorithm finishes within only several seconds.

Implementation of the proposed model on a typical dataset

In this section, we employ a real wafer map (*RealData*) and three typical maps (*DATA-2, DATA-3, and DATA-4*) simulated in the study of Bae et al. (2007) to demonstrate the advantages of the novel model over previous models.

First, we present the regression results for *RealData*, which originates from the research of Tyagi and Bayoumi (1994). As Figure 6a shows, the defect clustering is significant near the corners and forms a continuous region from the center to the lower left.

As mentioned earlier, the proposed GLSp-iCAR model is distinguished from the GLSp model by its consideration of spatial random error. As an improvement over independent random error, the spatial correlation component acts as an excellent adjustment for local clustering. Using the CARBayes package in the R language, we can obtain the parameter estimates for both models. We burn in the first 5,000 samples of the 20,000 total samples. The point estimates and the confidence intervals are shown in Table 1.

The acceptance rate of the Metropolis-Hastings algorithm is approximately 40%, which implies a balance between distribution sampling and efficiency. The value of τ^2 in the GLSp-icar model reflects the level of spatial variance, whereas the value of τ^2 in the GLSp model reflects only the level of global heterogeneity accounting for clustering.

The results of applying Moran's I test to the residuals of the two models show that the GLSp-iCAR model explains more of the spatial dependency than does the GLSp model. The value of Moran's I statistic for the GLSp model is two times larger than that for the GLSp-iCAR model, confirming the weaker fitting performance achieved using only spatial coordinates. The results are shown in Table 2.

In addition, if we plot the linearly fitted predictors $f(\boldsymbol{X}_i)\hat{\boldsymbol{\beta}}$, the posterior mean of the spatial correlation component $\boldsymbol{s}$, and the total fitted $\log(\hat{\boldsymbol{\lambda}}) = f(\boldsymbol{X}_i)\hat{\boldsymbol{\beta}} + \boldsymbol{s}$ together, as shown in Figure 7, we find that, without $\boldsymbol{s}$, the fitted surface is rather smooth and therefore cannot effectively reflect the spatial clustering patterns.

In Table 3, the level of τ^2 when applying the GLSp-iCAR and GLSp models to the other three simulated datasets is shown. The defects are randomly scattered on wafer map *DATA-2* with clustering, show ringlike patterns on wafer map *DATA-3*, and are clustered in the bottom right on wafer map *DATA-4*. Note that, for *DATA-2*, the posterior median of τ^2 is relatively large compared with the others. This may be interpreted as the result of sudden local turbulence producing some

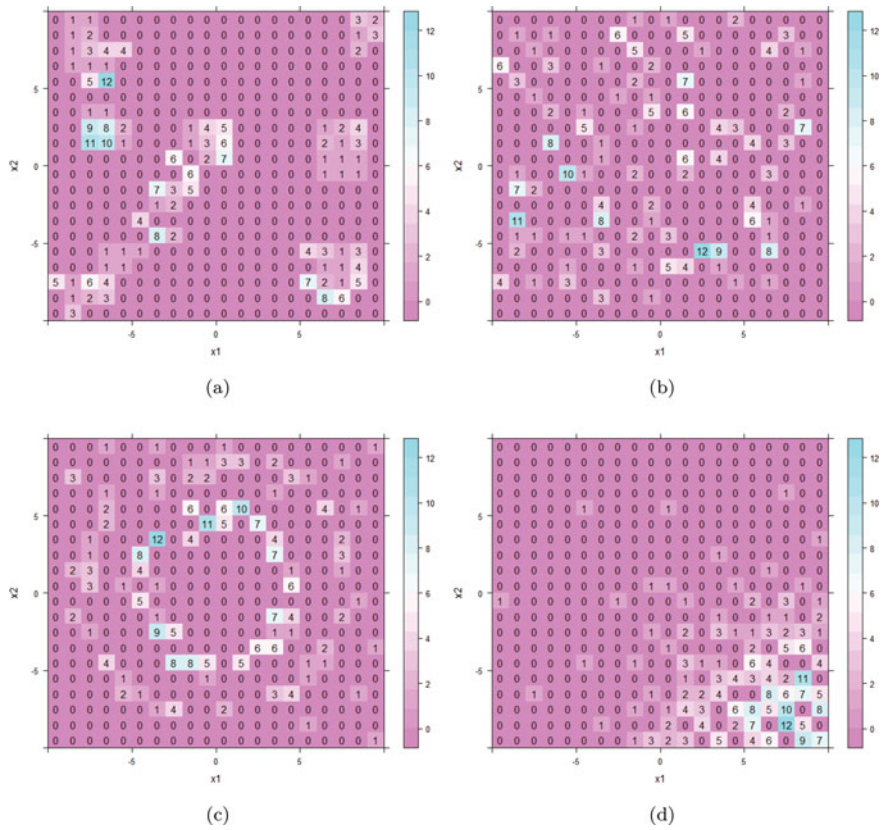

Figure 6. Real wafer map and typical simulated wafer map showing defect pattern and numbers for each chip.

extreme value, as well as the smaller clustering area. For *DATA-3*, the defect clustering area is also small, leading to a somewhat large τ^2. By contrast, in *DATA-4*, it is evident that the structure of the macro-scale tendency is sufficient to explain the defect pattern. As a result, both the GLSp-iCAR and GLSp models return small estimates of τ^2.

A comparison of the deviance information criteria (DIC) values for the four datasets is presented in Table 4. The DIC index is one of the most popular criteria used in Bayesian MCMC computations (Millar 2009; Linde 2005). Because the effective

Table 1. MCMC results for GLSp-iCAR and GLSp model on real data.

GLSp-iCAR	Median	2.5%	97.5%	accept
Intercept	−2.5735	−4.0105	−0.4463	43
r	−0.3812	−0.7183	−0.0754	43
sin	−1.5165	−3.6450	1.6065	43
cos	3.1287	1.3417	5.3303	43
rsin	0.0551	−0.2828	0.4145	43
rcos	−0.2036	−0.4556	0.0380	43
τ^2	11.9687	8.2584	18.5109	43
GLSp	**Median**	**2.5%**	**97.5%**	**accept**
Intercept	−3.5116	−4.8972	−2.4109	42.1
r	0.1046	−0.0372	0.2343	42.1
sin	−2.2524	−3.5729	−0.6700	42.1
cos	1.2310	−0.0272	2.5707	42.1
rsin	0.2241	0.0224	0.3837	42.1
rcos	−0.1570	−0.3114	−0.0262	42.1
τ^2	5.2966	3.8077	7.8430	100.0

Table 2. Moran I statistic under randomization for residuals in two models.

GLSp-iCAR Moran I	GLSp Moran I	Expectation	Variance
0.1401	0.4270	−0.0025	0.0013

Table 3. MCMC results for the GLSp-iCAR and GLSp models on Data-2, Data-3, and Data-4.

–	Median	2.5%	97.5%	accept
τ^2(GLSp-iCAR)	18.7175	13.4521	26.2994	43.7
τ^2(GLSp)	5.7086	4.1978	8.0634	100.0
τ^2(GLSp-iCAR)	15.0251	11.0328	22.1271	42.6
τ^2(GLSp)	5.1899	3.6929	7.5525	100.0
τ^2(GLSp-iCAR)	1.0009	0.4699	2.1753	41.6
τ^2(GLSp)	0.5225	0.2729	0.9911	100.0

Table 4. DIC values for the GLSp-iCAR models on real data and on Data-2, Data-3, and Data-4.

–	*RealData*	*DATA-2*	*DATA-3*	*DATA-4*	Mean
GLSp-iCAR	484.82	548.48	562.15	521.79	529.31
GLSp	561.86	562.88	571.12	508.46	552.08

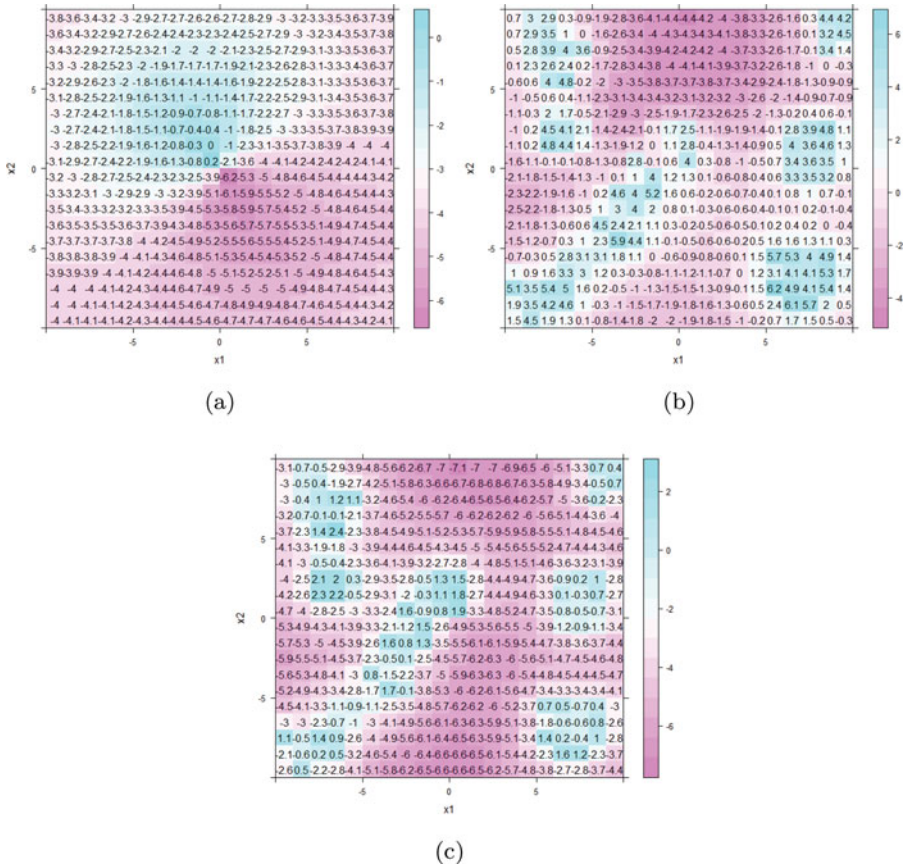

Figure 7. (a) linear fitted predictors $f(\boldsymbol{X}_i)\hat{\boldsymbol{\beta}}$, (b) posterior mean of spatial correlation component $\boldsymbol{s}$, and (c) total fitted $\log(\boldsymbol{\lambda}) = \mathrm{f}(\boldsymbol{X}_i)\hat{\boldsymbol{\beta}} + \boldsymbol{s}$.

number of parameters $p_D[y, \Theta, E(\theta|y)]$ is denoted by

$$p_D = \bar{D}_\theta - D(\bar{\theta}) = E_{\theta|y}\{-2\log L(y|\theta)\} + 2\log[P(y|E(\theta|y))] \quad [14]$$

we can obtain the DIC value by $\bar{D}_\theta = D(\bar{\theta}) + p_D$, which can be interpreted as a classical deviance measure plus a measure of complexity (Spiegelhalter et al. 2002).

The DIC provides an efficient means of quantitatively evaluating the fit performance of these two models. A smaller DIC value indicates a better model fit. Except for *DATA-4*, which presents a visibly continuous and smooth defect pattern, our GLSp-iCAR model achieves better performance than does the GLSp model for all datasets. Table 4 shows that its application to the real data results in a significant decrease in the DIC. To verify the superiority of the proposed model and explore the influence of the spatial clustering pattern on our model, in the next section, we present numerical experiments conducted using a large number of simulated datasets.

Simulation experiments comparing the performance of the GLSp-iCAR and GLSp models

To obtain a better understanding of the performance of the GLSp-iCAR and GLSp models, in this section, we use simulated data to conduct more extensive studies. The main reason of using a square wafer map in this article is for the convenience of performance comparison. However, it should be noted that our proposed model can be applied to round-shape wafer map directly without any modification. More specifically, in GMRF the model for stage 3, the shape of the map is uniquely characterized by a "neighbourhood structure." Therefore, the GMRF is in fact capable of handling any arbitrary map shapes.

We simulate spatially correlated defect data on a 20×20 wafer map. Each small square on the map represents a chip, and the number on each square indicates the corresponding generated "defect" count. Following the Bayesian framework, we generate the Poisson densities for 400 different lattices.

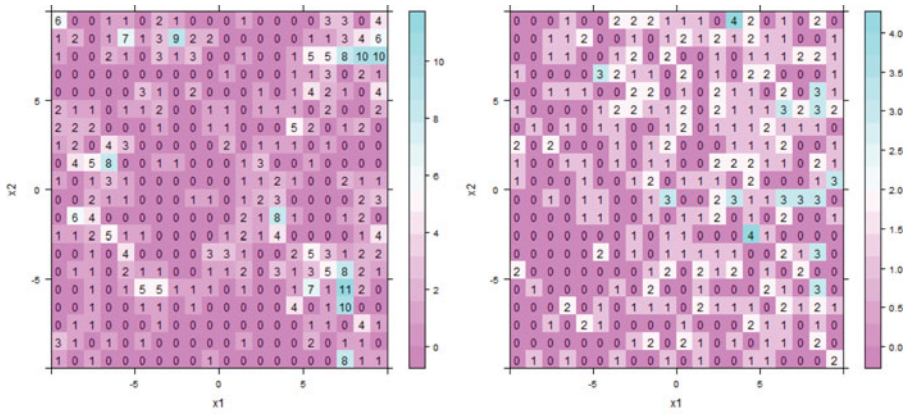

Figure 8. Different defect generation on the wafer map: (a) with and (b) without spatial correlation components s_i, where $a = 1.6$ and $b = 0.5$.

For the residuals in the stage 2 model, we simply assume i.i.d. normal distributions $N(\mu_i, \sigma_i^2)$ and fix the variance at $\sigma_i^2 = 0.5$ and the mean at $\mu_i = 0$. Because the residuals are not the focus of our study, we do not vary these parameter settings during the experiments.

For the random spatial effects, we adopt a multinormal distribution $\boldsymbol{s} \sim N(0, \boldsymbol{\Sigma})$, with the components of the correlation matrix defined as

$$\begin{aligned} \Sigma_{ij} &= a \cdot \exp\{-b \cdot \text{Distance}(i, j)\} \\ a &\in \{0.1, 0.4, 0.7, 1.0, 1.3, 1.6\} \\ b &\in \{0.1, 0.5, 0.9\}. \end{aligned} \quad [15]$$

Note that this formulation of the random spatial effects is different from the assumption adopted for the model in the second section because the covariance matrix is no longer sparse. According to our first-order GMRF model described in the second section, a component of the precision matrix is zero if the two corresponding chips are not adjacent, whereas the strength of the relationship between two chips is related to distance in the model used for data generation. Here, we can vary the values of a and b to generate different types of variations in the spatial correlation. As a increases, the correlation coefficient between any pair of "neighbors" becomes larger, whereas the correlation decreases more rapidly with distance when b increases.

Once the random effects have been produced, we set the coefficients for the coordinates

$$\begin{aligned} \boldsymbol{\beta} &= \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5,]^T \\ &= [0, -0.05, 0.1, -0.1, -0.05, -0.05]^T \end{aligned} \quad [16]$$

and generate the logarithm of the density at each location:

$$\begin{aligned} \eta &= log(\lambda_i) \\ &= f(\boldsymbol{X})\boldsymbol{\beta} + s_i + \mu_i \\ &= \beta_0 + r_i\beta_1 + cos_i\beta_2 + sin_i\beta_3 + r_icos_i\beta_4 \\ &\quad + r_isin_i\beta_5 + s_i + \mu_i. \end{aligned} \quad [17]$$

Figure 8 provides an example of how the random spatial effect component s_i influences the defect generation. It can first be observed that, because of the chosen global trend parameters $\boldsymbol{\beta}$, the defect rate increases toward the top and right of the wafer. Moreover, it is obvious that including random spatial errors in the density function leads to more highly clustered continuous regions of non-defective and defective chips, as seen in (a), whereas the defects generated without spatial correlation tend to exhibit a more sporadic distribution, as shown in (b). When $a = 1.6$ and $b = 0.5$, the number of defects on a chip can be greater than 10 with the inclusion of spatial correlation components, which indicates a higher defect rate in the defective region compared with the case without spatial correlation.

From Figure 9, the previous observation regarding the spatial components is further confirmed. Figure 9 shows various realizations of defect generation on the wafer map with different values of a and b. The results for $\{a = 0.1, b = 0.9\}$ present the smallest spatial effects, and those for $\{a = 1.6, b = 0.1\}$ present the largest spatial effects. We find that, if b is restricted to 0.1, the defects on the wafer map will occur in a broader continuous defective region as the parameter a is increased from 0.1 to 1.6. If b is restricted to 0.5,

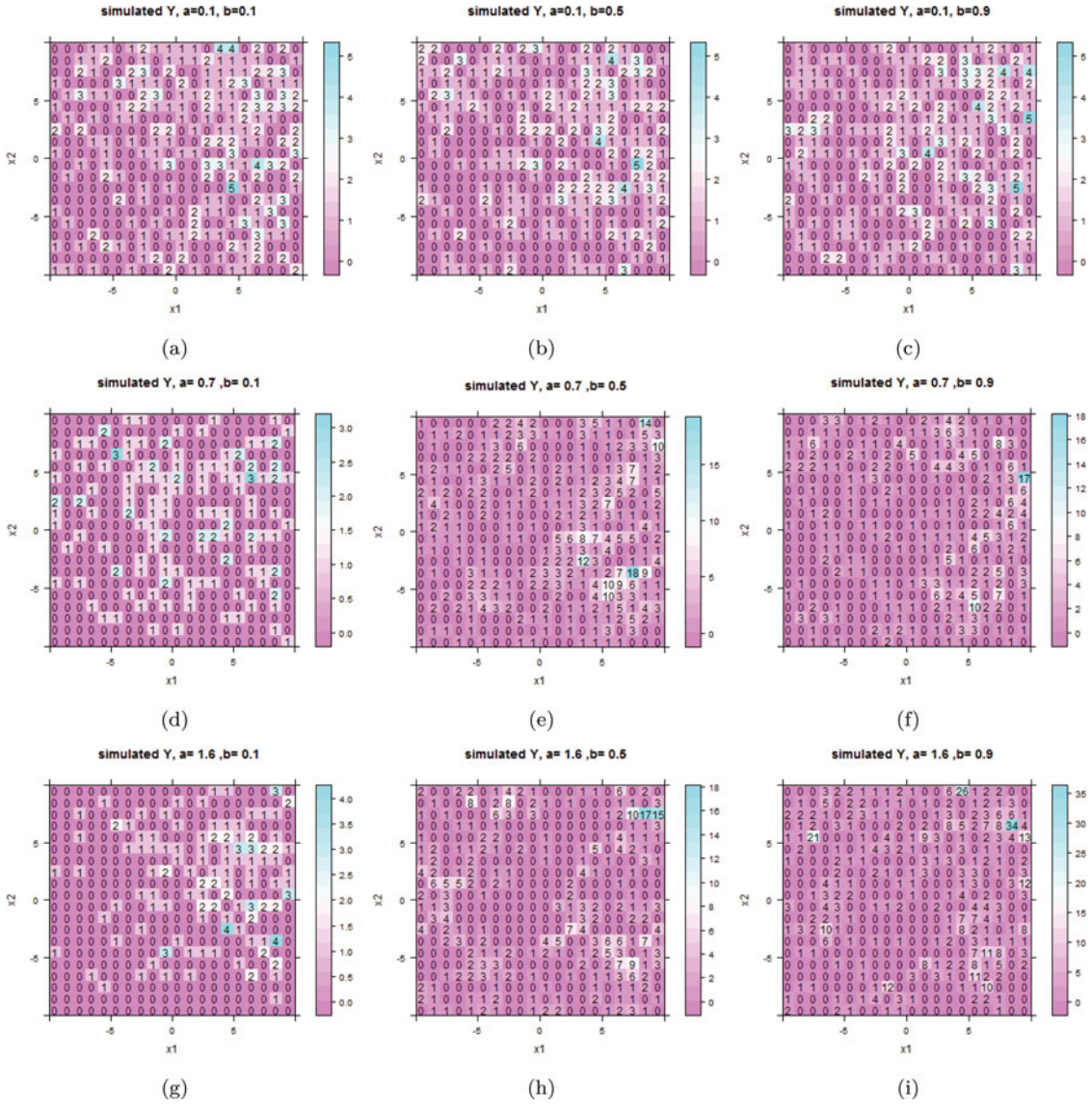

Figure 9. Various defect realizations on the wafer map for different values of *a* and *b*.

the defect pattern will transition to a curve or even a dot type as *a* increases. This phenomenon probably arises because of the rapid decrease in spatial correlation. Below, we demonstrate that excessively large values of *b* sometime cause poorer performance in simulation experiments.

By applying the different parameter settings for *a* and *b* defined in Eq. [15], we generate a total of 18 groups of defect-containing wafer maps, each consisting of 500 random samples. For each group, we apply the previous independent model (GLSp) and the proposed spatial model (GLSp-iCAR) to fit the data. The average DIC values are calculated to evaluate the fit performance of each model, and the results are presented in Table 5. According to Table 5, a deceasing trend in the DIC values from the independent model to the proposed spatial model can be observed. With the exception of the small shift in $\{a = 0.1\}$, all the experimental groups show an obvious decrease in the DIC value with the implementation of the proposed model. The differences in performance between the two models confirm the improvement achieved with the proposed model.

Figure 10 also illustrates the average performances under different parameter settings. It confirms the previous observation that, most of the time, the proposed model will at least not exhibit worse performance than the previous model, even if it does not achieve any improvement. With an increase in *a*, the average decrease in the DIC value may be as high as 60.

Table 5. Average DIC values in simulation experiments on different spatially correlated data.

b=0.1	GLSp-iCAR model	GLSp model (Baseline)	Change in DIC
a = 0.1	1035.5	1033.3	2.2
a = 0.4	942.6	958.7	− 16.1
a = 0.7	969.1	997.9	− 28.8
a = 1.0	996.6	1033.3	− 36.7
a = 1.3	1023.5	1072.9	− 49.4
a = 1.6	1042.2	1101.4	− 59.2
b = 0.5	GLSp-iCAR model	GLSp model (Baseline)	Change in DIC
a = 0.1	910.8	915.6	− 4.8
a = 0.4	962.6	989.5	− 26.9
a = 0.7	990.6	1028.1	− 37.5
a = 1.0	1017.9	1057.9	− 40.0
a = 1.3	1043.9	1087.0	− 43.1
a = 1.6	1088.2	1135.7	− 47.5
b = 0.9	GLSp-iCAR model	GLSp model (Baseline)	Change in DIC
a = 0.1	912.5	914.6	− 2.132
a = 0.4	978.6	998.2	− 19.6
a = 0.7	1023.2	1043.2	− 20.0
a = 1.0	1059.1	1083.3	− 24.2
a = 1.3	1086.3	1114.3	− 28.0
a = 1.6	1105.9	1139.9	− 34.0

In addition, the decreasing trend between $b = 0.1$ and $b = 0.5$ is more linear than that up to $b = 0.9$, which suggests that, without a large-scale neighborhood relation in the wafer map data, the new model will not provide a significant enhancement.

Taken together, the above results suggest that the fitting performance of the proposed spatial model tends to be better than that of the previous model in most situations in which spatial correlation is present in the wafer map. The model that considers random spatial correlation performs better than the model based on only coordinate predictors. Moreover, when the spatial dependence is larger, the performance of the proposed model will be improved.

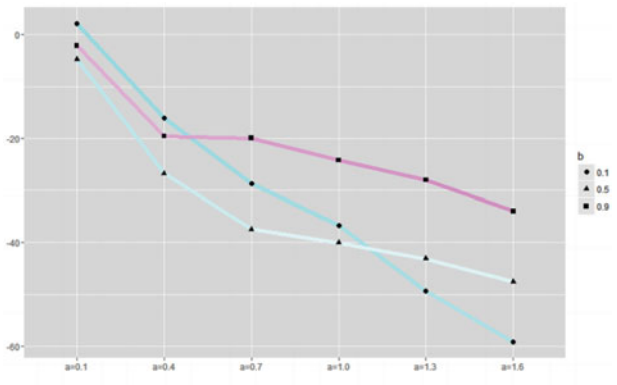

Figure 10. Average DIC values decrease under 500 simulation experiments, compared with GLSp models.

Conclusion

The spatial patterns characterizing wafer maps contain valuable information about potential manufacturing problems and will influence the final yield for a given wafer. In this study, a novel spatial model is proposed that can be viewed as an extension of the model proposed by Bae et al. (2007) and Yuan et al. (2011), which directly considers the spatial coordinate structure as prediction variables. In our model, the addition of random spatial effects enhances the model's flexibility in addressing local shifts or local clustering in a wafer map.

The advantages of including the spatially dependent errors were first confirmed through the application of the proposed model to real data, followed by analysis of several groups of simulated wafer defect maps. The numerical experiments indicate that, when the spatial dependency is high, the proposed model offers a greater improvement in performance. However, even when the spatial correlation is rather low, the model's fitting performance remains reasonable. Characteristics that are not captured by the macro-scale coordinate-based tendency predictors can be covered by the inclusion of random spatial errors.

In our proposed model, a spatial random component introduces sufficient variations for the defect rate distributions across a wafer. Simulated wafers in the forth section also show that the distribution of $log(\lambda_i) = \boldsymbol{f}(\boldsymbol{X_i})\boldsymbol{\beta} + s_i$ will be affected by the distribution of s_i. With higher spatial dependence, there tends to be more "zeros" occurring on the wafer map. Thus, by adding spatial components, the over-dispersion phenomenon in observed data can be effectively captured.

Using the Bayesian framework, it is possible to estimate parameters numerically and to construct a more suitable and complex hierarchical model. Benefiting from the MCMC method, the proposed hierarchical Bayesian model can be rapidly solved. When the wafer size increases, the MCMC method may encounter time consumption issues. We can adopt numerical approximations, such as INLA, to accelerate computation speed.

It is also worth noting that the method proposed in this article is not limited to a single distribution of data but rather can be extended as needed such as using a Bernoulli or negative binomial distribution. A straightforward method to adopt a zero-inflated

model is to model the random shock probability p_i as $\text{logit}(p_i) = \boldsymbol{f}(\boldsymbol{X_i})\boldsymbol{\beta} + s_i$ and to model the Poisson rate λ_i as $\log(\lambda_i) = \boldsymbol{g}(\boldsymbol{X_i})\boldsymbol{\beta}$. However, it is difficult to say that the spatial cluster phenomenon only exists in the distribution of p_i. Spatial dependency can be observed both in the zero's region and in the defect region. Thus, in this initial work, we do not consider zero-inflated models. The variations across sequential wafers could also be investigated.

About the authors

Hao Wang received her B.S. degree from Tianjin University in 2014. She is currently a Ph.D. candidate in the Department of Industrial Engineering at Tsinghua University. Her research focuses on statistical modelling, data analytics and quality engineering.

Bo Li is an Associate Professor in the School of Economic and Management at Tsinghua University, Beijing. He received his B.S. in B.S. in Mathematics from Peking University in 2002, and Ph.D. in Statistics form University of California, Berkeley in 2006. His research interests are large-scale complex data modeling and analysis, applied statistics, and econometrics in economics and business.

Seung Hoon Tong received a B.S. degree from Korea University in 1988, a M.S. degree from the Korea Advanced Institute of Science and Technology (KAIST) in 1991, and a Ph.D. degree from KAIST in 2006, under the sponsorship of Samsung Electronics Company Ltd., all in industrial engineering. In 1991, he joined Samsung Electronics Company Ltd., semiconductor business, where he has been engaged in the quality and reliability engineering area. His research interest is the factory integration based on manufacturing big data analytics and engineering statistics.

In-Kap Chang received B.S., M.S., and Ph.D. degrees from Seoul National University, Korea, in 1999, 2001, and 2008, respectively. He is currently working with Semiconductor Manufacturing Division of Samsung Electronics, Korea. He joined Samsung Electronics Company Ltd. in 2008. His research interests include statistical data mining, abnormal detection, sampling plans, quality assurance, and prediction for the customer failure rate in the semiconductor manufacturing process.

Kaibo Wang is a Professor in the Department of Industrial Engineering, Tsinghua University, Beijing, China. He received his B.S. and M.S. degrees in Mechatronics from Xi'an Jiaotong University, Xi'an, China, and his Ph.D. in Industrial Engineering and Engineering Management from the Hong Kong University of Science and Technology, Hong Kong. He is currently the Chair of the Quality, Statistics and Reliability (QSR) section of INFORMS, on the Editorial Review Board of Journal of Quality Technology, and also an Associate Editor for Quality Technology & Quantitative Management. His research focuses on statistical quality control and data-driven complex system modeling, monitoring, diagnosis, and control, with a special emphasis on the integration of engineering knowledge and statistical theories for solving problems from real industries. He is a senior member of ASQ, and a member of INFORMS and IIE.

Acknowledgments

The authors would like to thank the Editor-in-Chief and two anonymous referees for their valuable comments, which have helped us improve this work greatly.

Funding

The authors are grateful to Samsung Electronics for providing financial support and sample data to this research. Dr. Wang's research is partially supported by the National Natural Science Foundation of China under Grant No. 71471096 and the National Key Research and Development Plan (National Quality Infrastructure Project) under Grant No. 20l6YFF0204103.

References

Bae, S. J., J. Y. Hwang, and W. Kuo. 2007. Yield prediction via spatial modeling of clustered defect counts across a wafer map. *IIE Transactions* 39 (12): 1073–1083.

Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2014. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Crc Press.

Bao, L., K. Wang, and R. Jin. 2014. A hierarchical model for characterising spatial wafer variations. *International Journal of Production Research* 52 (6): 1827–1842.

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)* 192–236.

Besag, J. 1975. Statistical analysis of non-lattice data. *The Statistician* 179–195.

Besag, J., and C. Kooperberg. 1995. On conditional and intrinsic autoregressions. *Biometrika* 82 (4): 733–746.

Chen, F.-L., and S.-F. Liu. 2000. A neural-network approach to recognize defect spatial pattern in semiconductor fabrication. *Semiconductor Manufacturing, IEEE Transactions on* 13 (3): 366–373.

Di Palma, F., G. De Nicolao, G. Miraglia, E. Pasquinetti, and F. Piccinini. 2005. Unsupervised spatial pattern classification of electrical-wafer-sorting maps in semiconductor manufacturing. *Pattern Recognition Letters* 26 (12): 1857–1865.

Diggle, P. J., J. A. Tawn, and R. A. Moyeed. 1998. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 47 (3): 299–350.

Fatahi, A. A., R. Noorossana, P. Dokouhaki, and B. F. Moghaddam. 2012. Zero inflated poisson ewma control chart for monitoring rare health-related events. *Journal of Mechanics in Medicine and Biology* 12 (04): 1250065, .

Ferris-Prabhu, A. V., L. D. Smith, H. A. Bonges, and J. K. Paulsen. 1987. Radial yield variations in semiconductor wafers. *Circuits and Devices Magazine, IEEE* 3 (2): 42–47.

Hansen, M. H., V. N. Nair, and D. J. Friedman. 1997. Monitoring wafer map data from integrated circuit fabrication processes for spatially clustered defects. *Technometrics* 39 (3): 241–253.

He, S., W. Huang, and W. H. Woodall. 2012. Cusum charts for monitoring a zero-inflated poisson process. *Quality and Reliability Engineering International* 28 (2): 181–192.

Held, L., B. Schrödle, and H. Rue. 2010. Posterior and cross-validatory predictive checks: a comparison of mcmc and inla. In *Statistical Modelling and Regression Structures*. Springer. pp. 91–110.

Hwang, J. Y., and W. Kuo. 2007. Model-based clustering for integrated circuit yield enhancement. *European Journal of Operational Research* 178 (1): 143–153.

Joseph, V. R., and H. Adya. 2002. Prediction of yield in a multiproduct batch production environment. *Quality Engineering* 14 (1): 153–159.

Kim, B., Y.-S. Jeong, S. H. Tong, I.-K. Chang, and M.-K. Jeong. 2016. Step-down spatial randomness test for detecting abnormalities in dram wafers with multiple spatial maps. *IEEE Transactions on Semiconductor Manufacturing* 29 (1): 57–65.

Krueger, D. C., and D. C. Montgomery. 2014. Modeling and analyzing semiconductor yield with generalized linear mixed models. *Applied Stochastic Models in Business and Industry* 30 (6): 691–707.

Kumar, N., K. Kennedy, K. Gildersleeve, R. Abelson, C. M. Mastrangelo, and D. C. Montgomery. 2006. A review of yield modelling techniques for semiconductor manufacturing. *International Journal of Production Research* 44 (23): 5019–5036.

Kuo, W., and T. Kim. 1999. An overview of manufacturing yield and reliability modeling for semiconductor products. *Proceedings of the IEEE* 87 (8): 1329–1344.

Lambert, D. 1992. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics* 34 (1): 1–14.

Linde, A. 2005. Dic in variable selection. *Statistica Neerlandica* 59 (1): 45–56.

Maly, W., A. J. Strojwas, and S. W. Director. 2006. Vlsi yield prediction and estimation: A unified framework. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 5 (1): 114–130.

Mengersen, K. L., C. P. Robert, and C. Guihenneuc-Jouyaux. 1999. Mcmc convergence diagnostics: a review. *Bayesian Statistics* 6: 415–440.

Meyer, F. J., and D. K. Pradhan. 1989. Modeling defect spatial distribution. *Computers, IEEE Transactions on* 38 (4): 538–546.

Millar, R. B. 2009. Comparison of hierarchical bayesian models for overdispersed count data using dic and bayes' factors. *Biometrics* 65 (3): 962–969.

Milor, L. 2013. A survey of yield modeling and yield enhancement methods. *IEEE Transactions on Semiconductor Manufacturing* 26 (2): 196–213.

Mirza, A. I., G. O'Donoghue, A. W. Drake, and S. C. Graves. 1995. Spatial yield modeling for semiconductor wafers. In *Advanced Semiconductor Manufacturing Conference and Workshop, 1995. ASMC 95 Proceedings. IEEE/SEMI 1995* pp. 276–281. IEEE.

Rue, H., and L. Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL: CRC Press.

Schaller, R. R. 1997. Moore's law: past, present and future. *Spectrum, IEEE* 34 (6): 52–59.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (4): 583–639.

Stine, B. E., D. S. Boning, and J. E. Chung. 1997. Analysis and decomposition of spatial variation in integrated circuit processes and devices. *Semiconductor Manufacturing, IEEE Transactions on* 10 (1): 24–41.

Thompson, S. E., and S. Parthasarathy. 2006. Moore's law: the future of si microelectronics. *Materials Today* 9 (6): 20–25.

Tyagi, A., and M. A. Bayoumi. 1994. The nature of defect patterns on integrated-circuit wafer maps. *Reliability, IEEE Transactions on* 43 (1): 22–29.

Weber, C. M. 2004. Yield learning and the sources of profitability in semiconductor manufacturing and process development. *IEEE Transactions on Semiconductor Manufacturing* 17 (4): 590–596.

Yau, K. K. W., and A. H. Lee. 2001. Zero-inflated poisson regression with random effects to evaluate an occupational injury prevention programme. *Statistics in Medicine* 20 (19): 2907–2920.

Yuan, T., S. Z. Ramadan, and S. J. Bae. 2011. Yield prediction for integrated circuits manufacturing through hierarchical bayesian modeling of spatial defects. *Reliability, IEEE Transactions on* 60 (4): 729–741.