Methods for assessing empirical model parameters and calibration pattern measurements

Xin Zhou*, Eldar Khaliullin, Lan Luan
Luminescent Technologies, Palo Alto, CA, USA 94303

ABSTRACT

Assessing an empirical model for ILT or OPC on a full-chip scale is a non-trivial task because the model's fit to calibration input data must be balanced against its robust prediction on wafer prints. When a model does not fit the calibration measurements well, we face the difficult choice between readjusting model parameters and re-measuring wafer CDs of calibration patterns. On the other hand, when a model does fit very well, we will still likely have the nagging suspicion that an overfitting might have occurred. Here we define a few objective and quantitative methods for model assessment. Both theoretical foundation and practical use are presented.

Keywords: model calibration, model effectiveness, Fisher Information Matrix, Cramér-Rao Lower Bound

1. INTRODUCTION

Assessing an empirical model for ILT or OPC on a full-chip scale is a non-trivial task because the model's fit to calibration input data must be balanced against its robust prediction on wafer prints. When a model does not fit the calibration measurements well, we face the difficult choice between readjusting model parameters and re-measuring wafer CDs of calibration patterns. On the other hand, when a model does fit very well, we will still likely have the nagging suspicion that an overfitting might have occurred. Apparently we are in need of objective and quantitative methods for model assessment.

In practice, when we are afforded a cross-validation data set, a positive confirmation will make us feel comfortable with our model in hand. However, when a significantly larger error shows up in the cross-checking procedure than in the original calibration, we will be at a loss whether those data points in the verification set are suspect or the model itself ought to be made substantially different. The coupling between the model and the data is so strong that the cross-validation alone cannot answer all the questions when we are in doubt.

Some existing tools and concepts in the estimation theory of statistics provide a suitable theoretical framework for our purpose in model calibration and assessment. The next section summarizes the theoretical foundation detailed in a previously published paper [1].

2. THEORETICAL BACKGROUND

If a discrete data set represented by a vector $\boldsymbol{x} = \{x[0], x[1], \ldots, x[N-1]\}$ depends on a series of parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_p,\}$, we may design an estimator that finds the true value of $\boldsymbol{\theta}$ based on the data set

$$\hat{\boldsymbol{\theta}} = f(x[0], x[1], \ldots, x[N-1])$$

Here f is the parameter estimation function, which could be as simple as an algebraic function, or as complex as a lithographic process model calibration process.

If we use $P(\boldsymbol{x};\ \boldsymbol{\theta})$ to denote the probability density function (PDF) of obtaining measurements at $\boldsymbol{x}$ with the parameter at $\boldsymbol{\theta}$, the $p \times p$ Fisher information matrix $\boldsymbol{I}(\boldsymbol{\theta})$ is defined as

---

[*] xzhou@luminescent.com

$$[\boldsymbol{I}(\boldsymbol{\theta})]_{ij} = -E\left[\frac{\partial^2 \ln P(\boldsymbol{x};\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right] = -\int \frac{\partial^2 \ln P(\boldsymbol{x};\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} P(\boldsymbol{x};\boldsymbol{\theta}) d\boldsymbol{x}$$

The Cramér-Rao Lower Bound (CRLB) of parameter variance is

$$\mathrm{var}(\hat{\theta}_i) \geq [\boldsymbol{I}^{-1}(\boldsymbol{\theta})]_{ii} \tag{1}$$

where the true value of $\boldsymbol{\theta}$ is used when the derivatives are taken.

In an example of a signal sequence depending on a vector parameter $\boldsymbol{\theta}$

$$x[n] = s[n;\boldsymbol{\theta}] + w[n], \quad n = 0,1,\ldots,N-1 \tag{2}$$

The Fisher information matrix elements are

$$[\boldsymbol{I}(\boldsymbol{\theta})]_{ij} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial s[n;\boldsymbol{\theta}]}{\partial\theta_i}\frac{\partial s[n;\boldsymbol{\theta}]}{\partial\theta_j} \tag{3}$$

In the above formula, σ^2 is uniform variance of the white Gaussian noise *w*. The form of the deterministic signal $\boldsymbol{s(\theta)}$ is not yet restricted [2].

2.1 Uncertainty of parameters

We consider the entire calibration process of finding the true model parameters from measurements as a problem of computing a *p*-dimensional estimator $\boldsymbol{\theta}$. Furthermore, we assume that whatever systematic errors of metrology have been adequately absorbed in simulated CDs $s[n;\ \boldsymbol{\theta}]$, and that all aspects of lithographic process have been adequately modeled, so that σ^2 represents the true variance of noise $w[n]$ from measurement. Now we are ready to apply Eq.1 and Eq.3 to our calibrated empirical model to obtain parameter variance

$$\mathrm{var}(\theta_i) \geq \left[\sigma^2\left(\boldsymbol{J}^T \cdot \boldsymbol{J}\right)^{-1}\right]_{ii} \tag{4}$$

where $\boldsymbol{J}$ is the N×p Jacobian matrix of simulated CDs w.r.t. parameters

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial s[0;\boldsymbol{\theta}]}{\partial\theta_1} & \frac{\partial s[0;\boldsymbol{\theta}]}{\partial\theta_2} & \cdots & \frac{\partial s[0;\boldsymbol{\theta}]}{\partial\theta_p} \\ \frac{\partial s[1;\boldsymbol{\theta}]}{\partial\theta_1} & \frac{\partial s[1;\boldsymbol{\theta}]}{\partial\theta_2} & \cdots & \frac{\partial s[1;\boldsymbol{\theta}]}{\partial\theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial s[N-1;\boldsymbol{\theta}]}{\partial\theta_1} & \frac{\partial s[N-1;\boldsymbol{\theta}]}{\partial\theta_2} & \cdots & \frac{\partial s[N-1;\boldsymbol{\theta}]}{\partial\theta_p} \end{bmatrix} \tag{5}$$

The inequality Eq.4 specifies the minimal uncertainty of any calibrated model parameter.

2.2 Uncertainty of predictions

Now we present the result of the CRLB of N×N covariance matrix $\boldsymbol{C}_s$ whose diagonal elements correspond to the simulation variance at test patterns respectively. Assuming that we have performed Singular Value Decomposition (SVD) of $\boldsymbol{J=U\Sigma V}$

$$C_s \geq = \sigma^2 \boldsymbol{U} \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots & & \vdots \\ 0 & 0 & 1 & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix} \boldsymbol{U}^T \tag{6}$$

where the $N \times N$ diagonal matrix in the middle of two left-singular-vector unitary matrix $\boldsymbol{U}$ has p 1's and $(N\text{-}p)$ 0's; in other words, its upper left corner is a $p \times p$ identity matrix, and the rest of the elements are zeros.

The variance σ^2 is related to fitting error by

$$\sigma^2 \approx \frac{SSE}{N-p} = \frac{N}{N-p}(rms)^2 \tag{7}$$

where SSE is the sum-squared-error between simulation and measurement, N is the number of data points, p the number of parameters, and rms is the fitting error i.e. the usual metric for model accuracy.

The square root of CRLB of the right hand side of Eq.6 serves as a metric of prediction error of each individual test pattern. This will also be called the simulation uncertainty.

2.3 Effectiveness of a model

We define the trace of the right hand side of Eq.6 to be Ξ Model Effectiveness Index (***MEI***), which is the lower bound of total variability of a model:

$$\Xi \equiv \sigma^2 Tr \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots & & \vdots \\ 0 & 0 & 1 & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix} = \sigma^2 p = \frac{Np}{N-p}(rms)^2 \tag{8}$$

This figure of merit Ξ characterizes modeling uncertainty by balancing the fitting error (*rms*) with the number of parameters used (*p*), and the number of measurements taken (*N*). We can also use the square root of Ξ as another figure of merit, which has the same dimension as *rms*, Model Effectiveness Standard Index (***MESI***):

$$MESI \equiv \sqrt{\Xi} = \sqrt{\frac{Np}{N-p}} \cdot rms \approx \sqrt{p} \cdot rms \tag{9}$$

The right hand side of this definition holds when $N \gg p$. The smaller the value of Ξ is, the more effective the model is.

2.4 Identification of measurement outliers

We can use Eq.6 to compute the lower bound of expected simulation variance for each individual calibration test pattern. If we find that the actual measurement differs from the simulation by a certain factor, e.g. 6, of the square root of variance, we can be quite confident that that particular measurement point is an outlier and thus better to be re-measured or temporarily ignored.

3. EXPERIMENTAL

Using two sets of data points of wafer CD measurements from two different layers, we calibrated two models to demonstrate our method. For each model we have computed the model parameter variance and six times the square root of simulation variance at each data point, which we call "Uncertainty" in the plots.

3.1 Uncertainty of the model parameters

Fig.1 shows the variance of each model parameter, computed with Eq.4. Low variance of a parameter indicates that it is a stiffer parameter which is most sensitive to calibration measurement data.

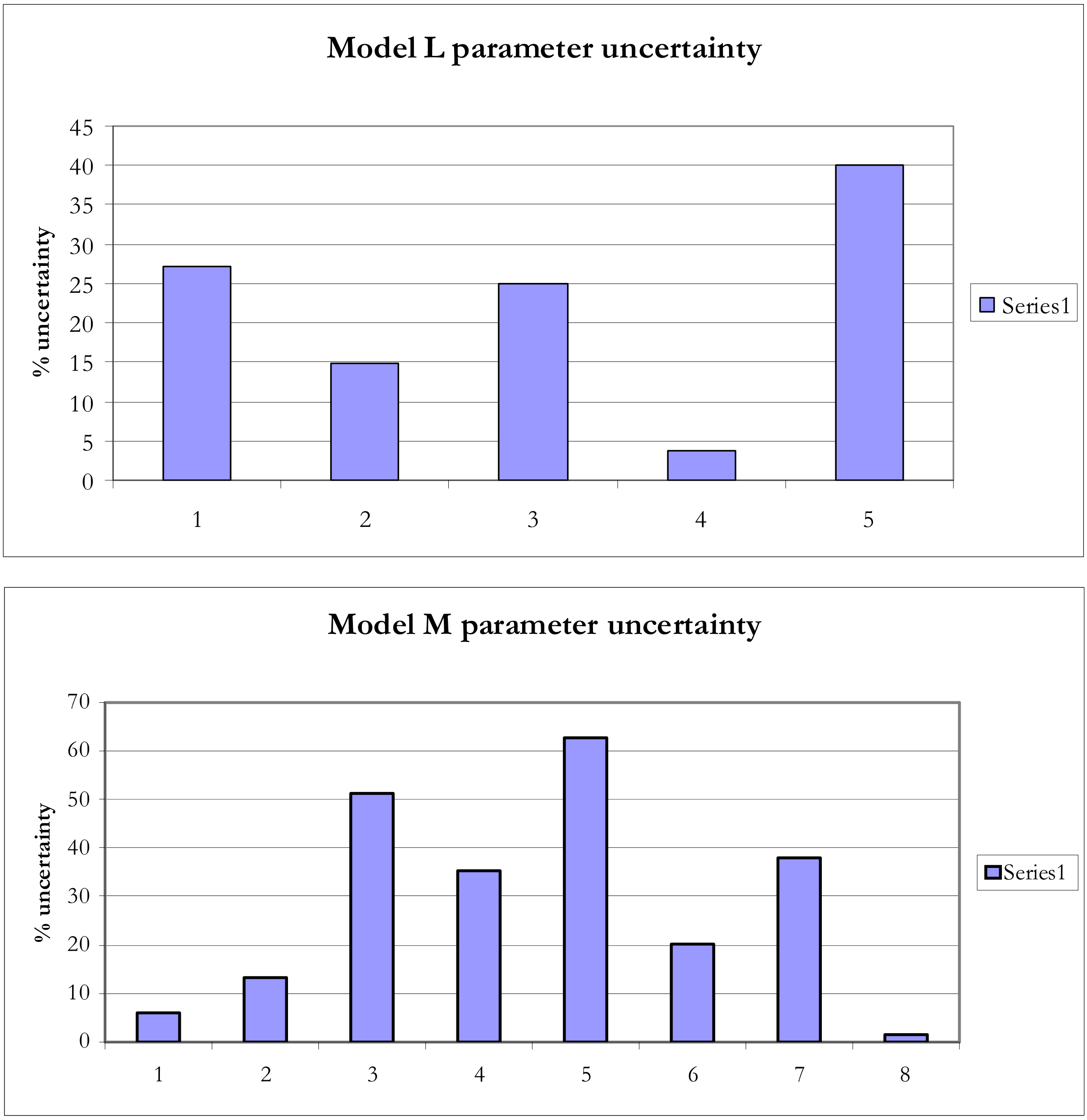

Figure 1. Expected uncertainty of calibrated model parameters

3.2 Comparing simulations with measurements

Judging by the model effective metric ***MESI***, Model L matches its data better than Model M : MESI of model L is 6.1, while that of Model M is 22.5.Using Eq.6 we compute model prediction error for each test pattern. Fig.2 Model L shows smaller (fitting error/simulation uncertainty) ratio than Fig.3 Model M.

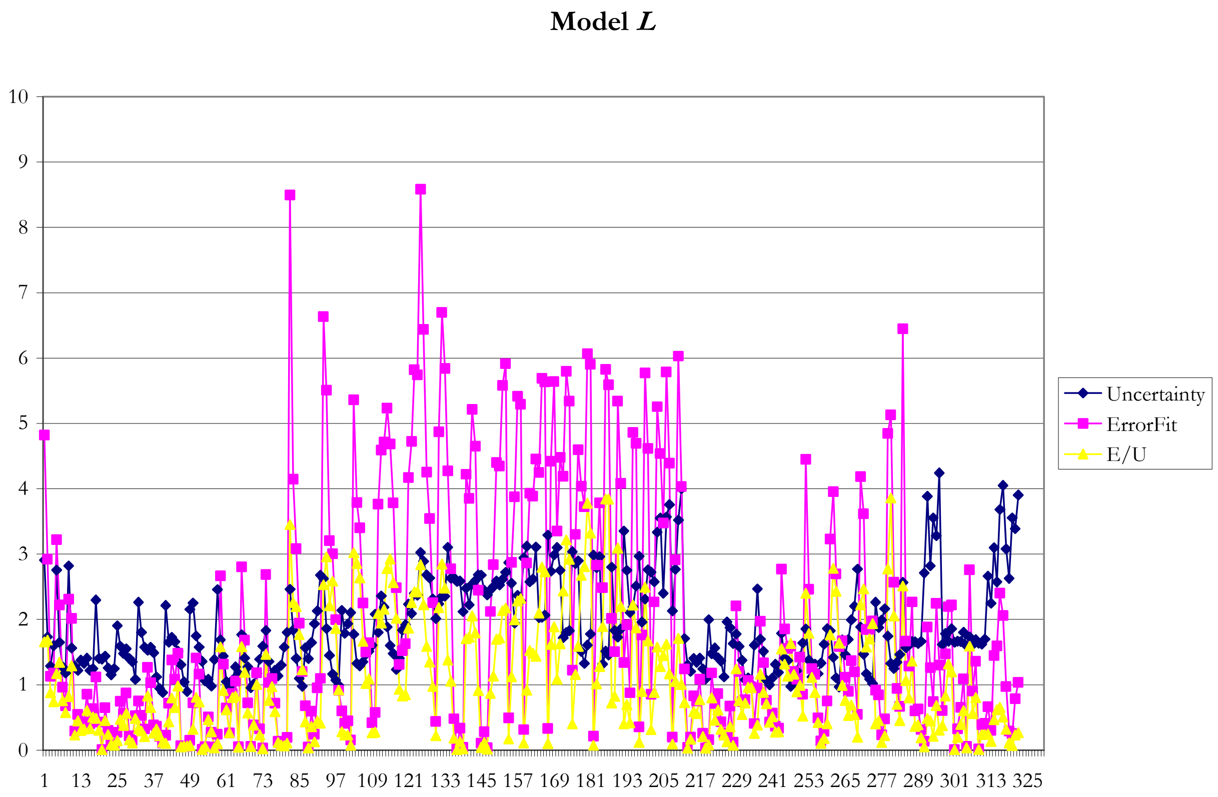

Figure 2.Model L has MESI = 6.1

The two examples show that when a model does not seem effective, as in the case of Model M, we should not yet jump to the easy conclusion that there is something wrong with the calibration process or the model formulation. The ErrorFit/Uncertainty ratio plot can tell us if examining raw data is the right direction to pursue. If there appear many outliers, we should take a step back and review the data collection process and see if printing or metrology plays the spoiler here.

Discussion with our customers revealed that Model M input data were not taken with repeated measurements of same patterns across wafer dies, but that they were measured just once per test pattern. The statistical variance of each data point in this case would be intrinsically infinite. However, the model calibration process takes into account all the input data simultaneously, so that an averaging process is taking place and a statistical significance can be reached even though any given individual data point is suspect. Because we don't have an independent and reliable measure of the metrology uncertainty in the single shot, single measurement data collection scheme, the resulting model fit of Model M is much less informative than it could be had its data been collected multiple times.

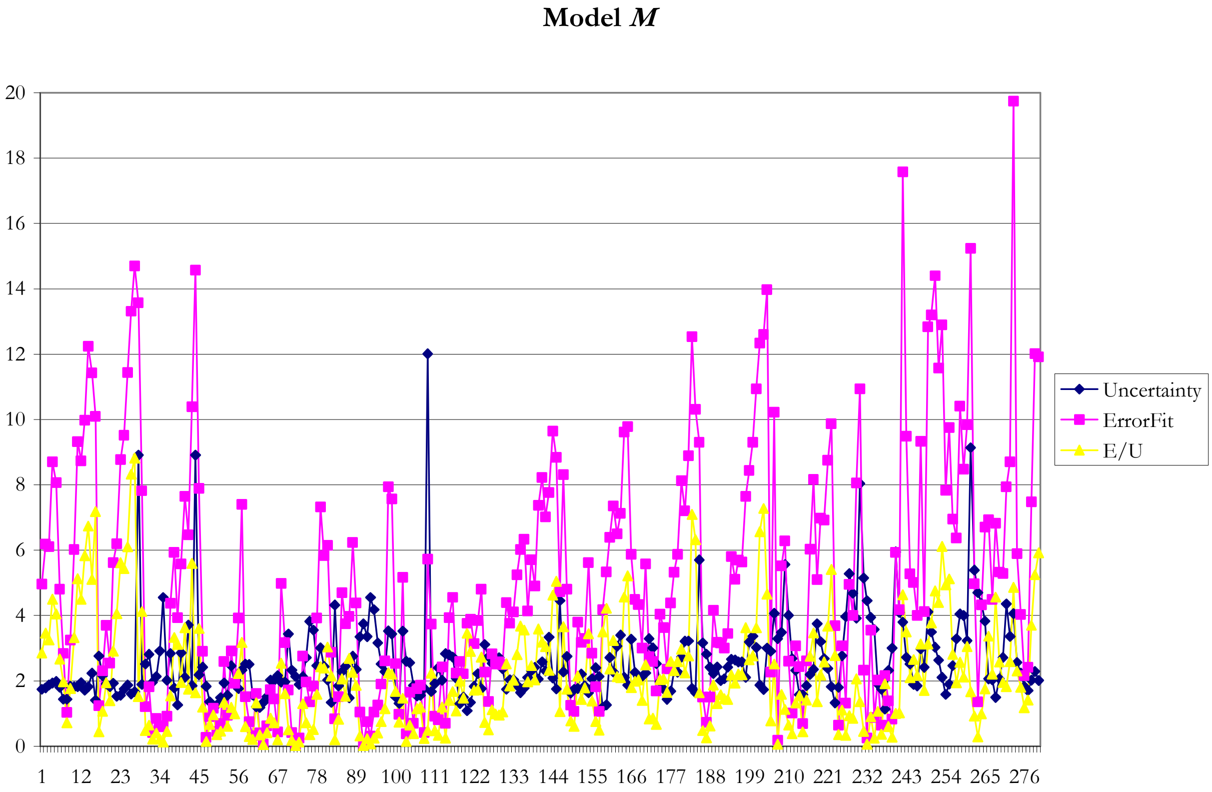

Figure 3.Model M shows larger measurement error, and MESI=22.5

4. CONCLUSIONS

We have set out to define some quantitative and objective methods to help evaluate empirical model parameters and the input measurement data. Model Effectiveness Standard Index (***MESI***) is defined as the square root of the trace of the CRLB of the simulation covariance matrix. The lower the ***MESI***, the better the model. We have also analyzed simulation uncertainty relative to measurement data, with the hope that the calibrated model itself can shed light on the question of whether some data points are true outliers.

Future research into this fascinating topic of model calibration will look at the coupling between model and data [3], and attempt to address the sufficiency question, i.e. whether our model formulation has adequately captured all underlying physical and chemical process.

REFERENCES

[1] Eldar Khaliullin, Yaogang Lian, Mark Davey, Xin Zhou, "What is a good empirical model?", Proc. SPIE 7488, 74883P-1 (2009).

[2] Samit Basu, Yoram Bresler, "The Stability of Nonlinear Least Squares Problems and the Cramér-Rao Bound", IEEE Transactions on Signal Processing, Vol.48, No.12, 3426-3436 (2000).

[3] Ralph E. Schlief, "Effect of data selection and noise on goodness of OPC model fit", Proc. SPIE 5754, 1147-1158 (2005).