Improved Methods for Lithography Model Calibration

Chris Mack

www.lithoguru.com, Austin, Texas

Abstract

Lithography models, including rigorous first principle models and fast approximate models used for OPC, require calibration using measured linewidth data. For models that predict process window behavior, the basic calibration data is linewidth versus focus and exposure over a range of feature sizes and types. The most common numerical method of finding the best fit model parameters is standard least-squares regression. While simple, this approach suffers from a number of well known problems. First, least-squares regression in not robust, meaning that even one bad data point can make the fit meaningless. Thus, outlier rejection becomes an important part of this approach. Both outlier rejection strategies and the use of robust fitting methods will be discussed. Second, standard least-squares may weight the data using the uncertainty in the measured linewidths, but uncertainty in the input variables, focus and exposure, is ignored. Often, at the extremes of focus and dose, errors in focus and dose actually dominate the resulting uncertainty in the measured linewidth. This can be accounted for using total least-squares regression. While often computationally difficult, in this paper an extremely fast and simple method for total least-squares regression will be presented for focus-exposure linewidth data. Finally, uncertainty in nominally fixed parameters, such as the linewidths of the features on the photomask used in the calibration, can lead to significant uncertainty in the resulting model parameters. The two standard approaches for dealing with this would be to leave these parameters fixed, or allow them to 'float' and be adjusted for best fit. Neither approach is satisfying. A better solution is to use Bayesian fitting, where *a priori* estimates of the mask feature widths and their uncertainties are used in the fitting merit function.

Keywords: Calibration, parameter estimation, regression, least-squares regression, total least-squares, Bayesian parameter estimation

1. Introduction to Parameter Estimation

Obviously, a primary goal of the use of a lithography model is for the output of that model to be predictive of some actual lithography process. Unfortunately, it is extremely unlikely that all of the input parameter settings for such a lithography model will be known *a priori* to an accuracy that allows the predictive capability of the model to be at its maximum. Thus, before using a lithography model, whether that model is based on first principles or is purely empirical, it will be necessary to estimate the input parameters of the model through the use of measured data. With respect to parameter estimation, all input parameters fall into two categories: parameters that can be directly measured and those that can't (or aren't).

Directly measured parameters can include such inputs to a lithography model as wavelength, numerical aperture, lens aberrations, source shape parameters (partial coherence), resist thickness, resist absorption coefficient, mask linewidth, etc. Other parameters, however, are more difficult to measure directly. As a result, their values are estimated by comparing model predictions to experimental lithography

Photomask and Next-Generation Lithography Mask Technology XIV, edited by Hidehiro Watanabe,
Proc. of SPIE Vol. 6607, 66071D, (2007) · 0277-786X/07/$18 · doi: 10.1117/12.728961

data in a process known as *model calibration*. Typically, model calibration data sets are made up of resist critical dimension (CD) as a function of mask CD and pitch, and often as a function of focus and exposure. Parameters that are often estimated through the process of model calibration include lens aberrations, source shape parameters, mask linewidth (or bias from designed linewidth), and many resist model parameters (acid diffusivity, development rate parameters, reaction rate constants, empirical resist model parameters, etc.).

[Note that there is significant overlap between the list of parameters that can be directly measured and those that are often estimated through indirect calibration. Not all parameters that can be measured are measured, often to the detriment of model predictability. Also, it is important to remember that even direct measurement of parameters leads to parameter uncertainty since all measurements include measurement error. This subject will be address below in the discussion of Bayesian parameter estimation.]

Model calibration involves adjusting the unknown input parameters of the model so that the predictions of the model are, in some way, a 'best fit' to the experimental calibration data. One of the keys to maximizing the predictability of the resulting model is in defining an appropriate metric of best fit for this calibration process. One of the most common approaches, which will be reviewed briefly here, is called *least-squares regression*. The least-squares regression method begins by defining a 'cost function' that compares model to data:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{\sigma_i} \right)^2 = \sum_{i=1}^{N} z_i^2 \qquad (1)$$

where χ^2 is the cost function (pronounced "chi squared"), N is the total number of experimental data points, y_i is the i^{th} experimental data point (with an associated set of input conditions, such as the dose, focus, mask CD, and pitch used in the experiment for that data point), $\hat{y}_i$ is the model prediction based on the same conditions as the i^{th} experimental data point, and σ_i is an *a priori* estimate of the measurement uncertainty for y_i. The term $\varepsilon_i = y_i - \hat{y}_i$ is called a model residual, and z_i is the reduced (normalized) residual.

The goal of least-squares regression is to find the set of model parameters that minimizes this cost function. There is an obvious question that arises: In what sense are the model parameters estimated in this way the 'best' parameters possible given this set of data? It can be readily shown that if the model residual is a normally distributed random variable with a mean of zero (and all residuals are independent of each other), the least-squares regression method produces the *maximum likelihood* parameters. This means that if different sets of experimental calibration data were to be collected under identical experimental conditions, the parameters generated by least-squares regression of the first calibration set would do the best possible job of predicting all of the other data sets in the limit of an infinite number of data sets.

The maximum likelihood interpretation of a least-squares regression makes this method of parameter estimation extremely appealing. However, some very restrictive conditions must be met before this interpretation is possible. In order for the model residuals to be normally distributed with a mean of zero, there must be no systematic experimental errors in the collection of the y_i data points, and the model must be a perfect representation of the experimental process (that is, with no systematic model errors). Further, the measurement uncertainty σ_i must be known *a priori*. In reality, none of the conditions are ever met – they can only be approached through very careful theoretical and experimental effort. Experimental CD data over a wide range of conditions will most certainly contain both random and systematic errors. And since lithographic models are continuing to evolve and improve over time there is little likelihood that they do not

contain systematic differences as compared to reality (this is especially true of the simplified models used for optical proximity correction).

Because of these problems, the least-squares regression technique does not in general produce the set of model parameters that maximizes the predictability of the model. In the sections below, a number of improvements to the least-squares technique will be described. These improved parameter estimation methods are well known and well established within the mathematical statistics community, and thus in one sense this paper represents a fairly straightforward review of the parameter estimation literature. However, it is this author's experience that none of these techniques are widely practiced (nor widely known) within the lithography community.

2. Regression Robustness and the Experimental Outlier

As mentioned above, one assumption of the least-squares regression is that all experimental errors are random. In fact, systematic errors always exist when making a measurement. Obviously, if these systematic errors were known they would be immediately corrected for. Thus, it is the unknown systematic errors in an experiment that are of concern. One instance of a systematic error that is extremely troublesome is the *data outlier* (or flyer). The classic "bad data point", an outlier is best defined as "an observation that deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism.[1]" As an example, consider the measurement of a focus exposure matrix using a top down CD SEM. If, at a certain exposure and focus condition, the pattern to be measured happens to experience pattern collapse (where the tall resist line mechanically fails and falls over on its side), the resulting CD value will be far off compared to what would otherwise be expected. If the model being calibrated with this data does not include pattern collapse mechanisms, it certainly would not be expected to do a good job of predicting this CD SEM measurement value.

Since data outliers are not at all uncommon in lithography model calibration, their influence on the resulting estimated parameters must be understood. Ideally, the parameter estimation method should not be overly sensitive to a few outliers. A lack of sensitivity to outliers is called model calibration *robustness*, which can be defined in several different ways. One simple description of robustness is the *breakdown point*[2], the minimum fraction of outliers in the calibration data set that can cause a change of arbitrary magnitude in the estimate of a parameter. One extremely problematic aspect of least-squares regression is that its breakdown point is zero: even one bad data point can completely invalidate the usefulness of the estimated parameter set. This situation arises from the squared nature of the cost function – the influence of a data point on the fit goes as the square of its difference from its expected value (i.e., from the model). Since data points with purely random errors typically deviate from the best-fit model by one standard deviation, one data point with a 5σ error will influence the fit as much 25 more typical data points.

The most common way to deal with the lack of robustness in least-squares regression is through *outlier rejection* – removing data points suspected of being outliers before using the data for calibration. As one might imagine, the challenge is in determining which data points are truly outliers, rather than valid data points that simply don't match our expectations. There are many outlier rejection methods[3], a few of which are listed below, in increasing order of statistical rigor (though some criterion are better under certain conditions, such as for small data sets):

- Manual removal
- Multiple of sigma
- Multiple of the interquartile range

- Q-test
- Chauvenet's criterion
- Pierce's criterion
- Grubb's test (extreme studentized deviate)

All of these methods, even the more rigorous Grubb's test[4], make the assumption that, other than the presence of a very small number of outliers, the residuals of the model fit are normally distributed random variables with a mean of zero. In other words, systematic errors in data or model are not considered when determining statistically which data points are outliers. Unfortunately, the application of this assumption for outlier removal tends to be self-reinforcing: data points which might indicate a significant problem in either data collection or the model used are systematically rejected from the analysis.

An alternative to outlier rejection is the use of a robust parameter estimator rather than the least-squares maximum likelihood estimator. The most robust estimator available is probably the least median of squares (LMS) regression[5], with a breakdown point that can approach the theoretical maximum of 0.5 (that is, nearly 50% of the data can be bad and a good fit can still be obtained). LMS regression, however, has very low efficiency (meaning that a large number of data points is required to get a good fit, as compared to a least-squares regression), and tends to be used only in extreme conditions of many expected outliers. By far the most popular robust estimator is the Huber's M-estimator[6]. The cost function uses the square of the reduced residual, z_i, when that reduced residual is small, then becomes linear with z_i for larger errors.

$$\text{Cost of data point } i = \begin{cases} {z_i}^2 & |z_i| \le c \\ c(2|z_i| - c) & |z_i| > c \end{cases}, \qquad z_i = \frac{y_i - \hat{y}_i}{\sigma_i} \tag{2}$$

The regression term c can be adjusted to optimize the trade-off between efficiency and robustness (higher values increase efficiency while decreasing robustness), with values near 1.3 being common. The Huber's M-estimator has a continuous first derivative (though a discontinuous second derivative), allowing for the use of reasonably effective minimization algorithms.

The use of robust parameter estimation methods is reasonably common in many different fields, but I am unaware of their use today for lithography model calibration.

3. Total Least Squares Regression

In the conventional least-squares regression, the residual between experiment and data is weighted by one over the estimated uncertainty in the experimental data. Generally, this uncertainty, expressed as the standard deviation of the data point, is estimated based on a basic knowledge of the measurement method. For example, a CD SEM might be characterized prior to data collection in order to determine its statistical precision as a function of the range of measurand types being used in the calibration (for example, precision as a function of feature size, or as a function of pitch). This precision is most often determined by repeated measurements of the same sample.

It is reasonably easy to provide a much more comprehensive estimate of data point uncertainty. Consider an experiment to measure one output y (for example, CD) as a function of K input variables v (for example, focus and exposure, so that $K = 2$). Uncertainty in the output is caused by measurement uncertainty, but it is also a function of the uncertainties in setting the input variables during the experiment. For example,

if for one data point the focus is set to -150 nm, the actual focus experienced by the resist feature to be measured will differ from this amount due to random focus errors built into the experimental process.

The impact of input variable uncertainty can be estimated using a standard *propagation of uncertainties* approach. Given $y = f(v_1, v_2, \ldots, v_K)$, and assuming the errors in the input variables are small enough so that the function f can be approximated as linear, the uncertainty in y can be estimated by

$$\sigma_y{}^2 = \sigma_{y-measurement}^2 + \sum_{j=1}^{K}\left(\frac{\partial y}{\partial v_j}\right)^2 \sigma_{v_j}{}^2 \tag{3}$$

Thus, an uncertainty in an input variable (σ_v) propagates to an uncertainty in the output based on the sensitivity of the output to the input (the partial derivative term evaluated at the nominal conditions of that data point). If the input variables are assumed to have no error, the uncertainty in the output is determined only by the uncertainty in the measurement of that output. If, however, the input settings have errors, then the output will have an added uncertainty that is independent of output measurement uncertainty. If the variance in output as calculated from equation (3) is used in the least-squares cost function, the regression is called a *total least-squares regression*, or an *error-in-variables regression*. The reduced residual becomes

$$z_i = \frac{y_i - \hat{y}_i}{\sqrt{\sigma_i^2 + \sum_{j=1}^{K}\left(\frac{\partial y_i}{\partial v_j}\right)^2 \sigma_{v_j}{}^2}} \tag{4}$$

Note that total least-squares regression is completely compatible with the use of robust estimators, such as the M-estimator described in the previous section.

Consider again the focus-exposure example. For a data set as shown in Figure 1, the sensitivity of the output to focus and exposure will be a minimum at best focus and exposure. Out of focus or underexposed, the partial derivatives of CD with respect to focus or dose will be higher, and often much higher. Thus, given estimates in the uncertainty in setting focus and exposure for each data point in the experiment, the out-of-focus and underexposed data points will have much larger uncertainty of the resulting CD value than the nominal dose/focus condition. In fact, errors in focus and exposure, propagated to the output CD, often dominate the resulting CD uncertainty at the extremes of focus and exposure. As such, those extreme dose/focus data points will be weighted less heavily in a model fit using total least-squares regression as compared to a conventional least-squares fit.

While conceptually simple, the total least-squares approach has one serious drawback – it is difficult to evaluate the partial derivatives of the output versus inputs function during the regression. There is a solution, however, that takes advantage of the fact that uncertainty estimates for each data point (the weights used in the regression) do not need to be extremely accurate in order to get good estimated parameters from the model fit. While we often want our model to predict CD to within a percent or two, calibration requires that the uncertainty estimates for each data point need only be on target to within 10 – 20% at best. Thus, estimating the data point uncertainties can use a much less rigorous approach than fitting the model to the data based on those uncertainties.

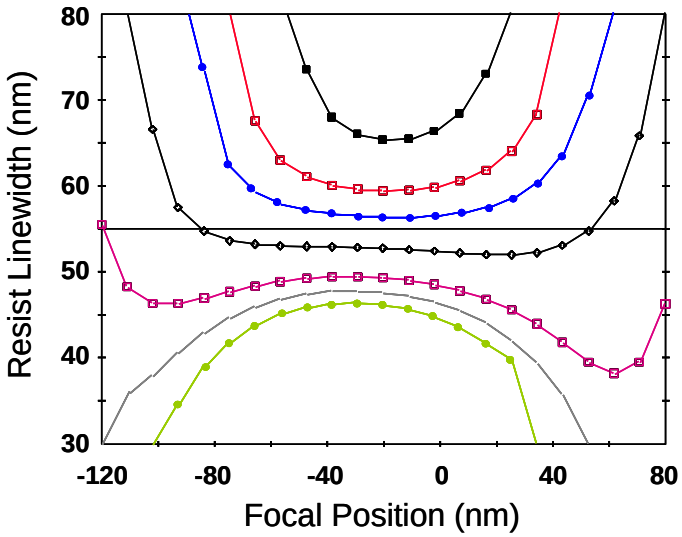

Figure 1. Typical focus-exposure matrix showing resist CD as a function of focus for different exposures (lines connecting data points of the same exposure dose are for visualization purposes only).

The following simple approach can be used to find the partial derivatives of the CD(focus, exposure) function and thus the weights to be used in the least-squares regression. First, fit the measured focus-exposure data set (using standard least-squares regression) to a simple but robust polynomial function (my favorite is based on a physical understanding of image formation and is commonly used for process-window calculations using experimental data[7]). Once a reasonable fit is obtained, the partial derivatives of the polynomial can be easily evaluated for each data point, allowing a regression weight to be calculated using equation (3). Given this new weight per data point, standard least-squares regression can be used to calibrate the lithography model. The result is a calibration that is not overly sensitive to highly uncertain data points at the extremes of the process window.

The application of total least-squares regression to focus-exposure data is reasonably straightforward, but I am unaware of its use today for lithography model calibration.

4. Bayesian Parameter Estimation

In the discussion above, better use of measurement error analysis results in better parameter estimation when fitting a model to calibration data. But as was mentioned in the Introduction, models also include parameters which are measured directly. And of course, all measurements include uncertainty, even the direct measurement of model parameters. How are estimates of the uncertainties in those directly measured parameters to be used when calibrating a model and estimating other parameters? In fact, there are some lithography model parameters which are frequently measured (or assumed) and fixed, such as mask CD and partial coherence, but are also frequently used as adjustable parameters in the fitting to calibration data. When is it most appropriate to fix a parameter, versus letting it 'float' as an adjustable parameter? Is there a way to avoid such an arbitrary distinction between 'known' and 'unknown' parameter values?

Bayesian statistics attempts to make the best use of all *a priori* information when fitting a model to a data set. It is named for the Reverend Thomas Bayes (1702 – 1761), a mathematician who first put the idea

of statistical induction on a firm mathematical footing. In terms of parameter estimation, the basic idea is this: If a model input parameter has an *a priori* best estimated value, with an estimated uncertainty in that best estimated value, this information should be used in the model calibration process. One very simple and straightforward way to use this information is through an additional cost-function term:

$$\tilde{\chi}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i - \hat{y}_i}{\sigma_i}\right)^2 + \frac{w_p}{M}\sum_{j=1}^{M}\left(\frac{p_j - \bar{p}_j}{\sigma_{p_j}}\right)^2 \qquad (5)$$

Here, there are M parameters of the model that are being adjusted to find the best fit, p_1, p_2, …, p_M. Each parameter has an *a priori* estimated value $\bar{p}_j$ and an uncertainty in that *a priori* value of σ_{p_j}. The value of the overall parameter weight w_p should be empirically determined based on experience with the problem domain, but can be set initially to one.

Looking at equation (5), one can see that the usual approaches to parameter fitting, where a parameter is either fixed or allowed to float freely, are simply extreme cases of the more general Bayesian fitting approach. If the uncertainty in the *a priori* parameter estimate, σ_p, goes to zero (meaning there is no uncertainty in our knowledge about the parameter), the cost function will always drive the parameter value to its mean – in other words, the parameter remains fixed at the best estimate of its value. If the uncertainty in the *a priori* parameter estimate, σ_p, is infinite (meaning we have no prior knowledge about the parameter), this term will not contribute to the cost and the parameter will be allowed to float freely during the fitting. Bayesian parameter estimation simply allows a continuum between these extreme cases, where our *a priori* knowledge about a parameter value is something between perfect and nonexistent.

As an example of where Bayesian parameter estimation can be used effectively for lithography model calibration, consider mask bias. Masks have a specification of mean-to-target CD, which when coupled with mask measurement uncertainty leads to a reasonably good estimate of the uncertainty of the mask CD at any given location on the mask. Since mask CD is an input to a lithography model, users are often faced with the choice of fixing or floating a bias between the actual mask CD and its design value when calibrating a lithography model. Bayesian parameter estimation, as used in equation (5), removes this false dichotomy and allows all of the information possessed by the user to be applied to the calibration problem.

I believe that Bayesian parameter estimation for lithography model has the potential to greatly improve the predictive accuracy of calibrated models, but further work should be done in assessing and optimizing its use for lithography model calibration.

5. A Note of Caution and Some Random Ideas

A simple output of a least-squares regression is the root-mean-square (RMS) difference between the best-fit model and the calibration data. Due to its convenience, many model users consider this RMS calibration error to be a simple measure of model goodness: given two competing models, the model that produces the lowest RMS calibration error is the better model. While such an interpretation is rife with problems for even standard least-squares regression, the final calibration RMS value is completely useless for any of the more advanced calibration techniques described here. By definition, standard least-squares will produce the lowest possible calibration RMS value. Thus, total least-squares, robust parameter estimation, and Bayesian parameter estimation techniques will always produce RMS calibration values that are higher than simple least-squares regression. Or course, this should never be thought of as a problem. The goal of

calibration is not to predict the calibration data set, but to predict *other* data sets with maximal accuracy. Maximal predictive accuracy is not achieved with standard least-squares regression, and so a model with minimum RMS difference to the calibration data will not be the best model.

This brings up the interesting subject of how to *validate* a model: given a model with a set of calibrated parameters, what is its predictive accuracy? The calibration RMS value offers very little information about predictive accuracy. And while the subject of model validation is an extremely broad one (and thus can not be adequately covered here), it is worth mentioning that model validation must rely on *out-of-sample* data, that is, data that was not used to calibrate the model. A very interesting experiment that could be performed is as follows. Consider a large data set collected for the purpose of model calibration. Randomly select half of the data and calibrate the model using the calibration process of record. Using this calibrated parameter set, find the RMS difference between the model and the second half of the data (the out-of-sample data that was not used for calibration). Now repeat these steps many times using different randomly selected halves of the data. The collection of RMS validation values will have a mean and a standard deviation that can be used to assess the true predictability of the model (for a given calibration data set size and a given calibration process).

The model validation experiment of the previous paragraph can also be used to compare different calibration approaches. If, for example, the above experiment was repeated for two different sample sizes, one could see how much the model predictability would increase with increasing calibration sample size. Additionally, different calibration processes (standard least-squares versus total regression, for example) can be compared using the ultimate metric of success, model predictability, as the judge.

Finally, as a random note, residual analysis is an underutilized tool for judging the quality of the data, the model, and the calibration process. Since a major assumption of standard least-squares regression is that the residuals are normally distributed, this assumption should be checked by examining the actual residuals of the calibration. For the large data sets that are commonly used for lithography model calibration, testing the normality of the residuals in not difficult and can be very revealing (I have never seen a set of residuals from a lithography model calibration that was even remotely close to normally distributed). Further, the properties of the chi-squared distribution under the assumption of normal residuals are well known (for example, the expectation value of the reduced chi-squared, $\tilde{\chi}^2 = \chi^2 / N$, is 1 and its variance is $2/N$). When the actual minimized chi-squared value differs significantly from its expected value, one or more of the assumptions of the fit are likely wrong. If the chi-squared is too large, there are likely to be systematic errors either in the data or in the model. If the chi-squared is too small, the model probably has too many adjustable parameters and has begun to fit the noise. Of course, there is always the chance that the *a priori* estimates of the data point uncertainties are significantly wrong (a not uncommon problem).

6. Conclusions

While least-squares regression is very commonly used for lithography model calibration (for both rigorous and simplified models), there are many improvements that could and should be made to this simple technique. Robust parameter estimators can significantly reduce the sensitivity of the fit to outliers, while also eliminating the dangers inherent in outlier rejection schemes. Total least-squares regression can improve the accuracy of estimated parameters by taking into account errors in the input variables of the experiment (a technique that is especially useful for focus-exposure matrix data sets). Finally, Bayesian approaches have the potential to further increase model predictability by using information about the uncertainty of nominally 'fixed' model parameters. The widespread use of one or more of these simple extensions to current calibration processes could significantly improve the predictability of the resulting calibrated models. And

predictability, after all, is the ultimate goal of the use of a lithography simulator. One of my guiding principles is this: Data is expensive, data analysis is cheap. It is incumbent on the modeler to extract as much information as possible from the data that is available.

Acknowledgements

Much of the research that underlies this paper was performed while I was the visiting Melchor chair professor at the University of Notre Dame in the fall of 2006. I am indebted to Notre Dame, and my sponsor Dr. Gary Bernstein, for providing me with such a wonderful opportunity and environment in which to learn.

7. References

[1] D. Hawkins, Identification of Outliers, Chapman & Hall (1980).
[2] P. J. Rousseeuw and A. M. Leroy, Robust Regression and Outlier Detection, John Wiley & Sons (New York: 1987).
[3] Boris Iglewicz and David C. Hoaglin , How to Detect and Handle Outliers, ASQ Quality Press (December, 1993).
[4] F. E. Grubbs, "Sample Criteria for Testing Outlying Observations", *Annals of Mathematical Statistics*, Vol. 21, No. 1 (Mar., 1950) pp. 27 – 58.
[5] P. J. Rousseeuw, "Least median of squares regression", *J. Amer. Statist. Assoc.*, Vol. 79 (1984), pp. 871-880.
[6] P. J. Huber, Robust Statistics, John Wiley& Sons (New York: 1981).
[7] C. A. Mack and J. D. Byers, "Improved Model for Focus-Exposure Data Analysis," *Metrology, Inspection and Process Control XVII, Proc.*, SPIE Vol. 5038 (2003) pp. 396-405.