Applied Surface Science xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Applied Surface Science

journal homepage: www.elsevier.com/locate/apsusc

Full Length Article

Model selection in spectroscopic ellipsometry data analysis: Combining an information criteria approach with screening sensitivity analysis

D.V. Likhachev*

GLOBALFOUNDRIES Dresden Module One LLC & Co. KG, Wilschdorfer Landstr. 101, D-01109 Dresden, Germany

ARTICLE INFO

Article history:
Received 31 July 2016
Received in revised form
26 September 2016
Accepted 27 September 2016
Available online xxx

Keywords:
Data analysis
Model selection
Information criteria
Sensitivity analysis
Parameterization
Spectroscopic ellipsometry

ABSTRACT

In the field of optical metrology, the selection of the best model to fit experimental data is absolutely nontrivial problem. In practice, this is a very subjective and formidable task which highly depends on metrology expert opinion. In this paper, we propose a systematic approach to model selection in ellipsometric data analysis. We apply two well-established statistical methods for model selection, namely, the Akaike (AIC) and Bayesian (BIC) Information Criteria, to compare different dispersion models with various complexities and objectively determine the “best” one from a set of candidate models. The information criteria suggest the most optimal way to quantify the balance between goodness of fit and model complexity. In combination with screening-type parametric sensitivity analysis based on so-called “elementary effects” (the Morris method) this approach allows to compare and rate various models, identify key model parameters and significantly enhance process of ellipsometric measurements evaluation.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

In spite of the well-known fact that the area of optical metrology in the semiconductor manufacturing industry is, typically, lagging behind the progression of the process tools (lithography, deposition, etch, etc.), spectroscopic ellipsometry as a method of choice for multiple in-line measurements has been evolving thereby providing the practitioners new capabilities for characterization of materials and structures as well as advanced process control [1–3]. One of the benefits of such progress is a full scope of material dispersion models available in modern ellipsometric data analysis software for production-grade spectroscopic ellipsometers previously accessible in research ellipsometers only. There are numerous dielectric function parameterizations which can be applied to various types of materials [4–7]. However, such abundance of available dispersion models would be the scholarly user's “a true paradise” but rather great perplexity and confusion for the “typical” industrial ellipsometry users. Here we come to a ubiquitous *model selection* problem which is the task of choosing a model of optimal complexity from a set of potential (or candidate) models based on a finite set of available experimental (training) data. The situation has been well reflected by Herzinger et al. [8]: “When performing a model dependent data analysis, simple models are preferable to complex ones if the fit quality is the same... The difficulty is in objectively evaluating the as each model complication is added to see if the fit really improved. Of course, if the fit does not get better with increasing complexity, that does not mean the complex model is necessarily wrong, but it does mean that one lacks sensitivity to allow a distinction and some other criteria must be employed.” Here, by the concept of rising complexity we mean both the number of variable parameters as well as the model structure as induced by a given functional form of used dispersion model and/or possible structural modifications such as interfaces between layers, roughness, grading, anisotropy, etc. Another important remark regarding model selection in ellipsometric data analysis has been made by Vedam in Ref. [9]: “The selection of the final model is based on the *simultaneous* fulfilment of the following five criteria: (i) a physically realistic model; (ii) a low value of σ the unbiased estimator of the mean square deviation; (iii) good agreement between the calculated and the observed [cos $\Delta(\lambda)$, tan $\Psi(\lambda)$] spectra over the entire spectral region studied; (iv) reasonably low values of the confidence limits of the variable parameters; and (v) acceptably low values of the cross-correlation coefficients between the evaluated parameters. Of course, too many fitting parameters or correlated parameters result in drastic increases in the confidence limits and

* Corresponding author.
E-mail address: dmitriy.likhachev@globalfoundries.com

http://dx.doi.org/10.1016/j.apsusc.2016.09.139
0169-4332/© 2016 Elsevier B.V. All rights reserved.

thus LRA [the least-squares regression analysis] ***must include a check against any tendency to add parameters indiscriminately just for the sake of reducing mean squares deviation***."

Model selection problem is closely related to two commonly used concepts such as overfitting and underfitting of data. Overfitting typically occurs if a very (too) flexible model, for instance, a model with excessive number of variables, is used to describe measured data. As John von Neumann famously pointed out that "...with four parameters I can fit an elephant, and with five I can make him wiggle his trunk." [10].[1] Specifically, too much flexible model will fit even the noise in the data, i.e., just the training dataset will be described perfectly. As a result of that, the estimations of model parameters will contain uncontrollable errors and the model will provide misleading predictions for new data [12]. In other words, despite the fact that we improve the fit by adding parameters to the model, meanwhile the estimations of parameters get worse since there are less data (or information) per parameter. On the contrary, an excessively simple model with only a few variables may produce an underfitting and, therefore, does not fit neither the training data nor new ones. Ideally, the best model should have an optimal balance between under- and overfitting, i.e., adequately fit measured ellipsometric data and provide reliable and unambiguous predictions while using the fewest parameters. But, as many practitioners know, this task is very difficult to do in practice and systematic approach is desperately needed to put the model selection in ellipsometric data analysis on a solid foundation.

This problem can be addressed by applying various statistical information criteria (IC) for model selection. The information criteria approach has been extensively applied in other fields. Over the last few decades, a couple of those criteria, namely, Akaike information criterion (AIC) [13,14] and Bayesian information criterion (BIC) [15] have become very popular and important tools in comparing and selecting models. AIC was established in the early 1970s and it is based on the information-theoretic framework [16–18]. BIC was introduced by Schwarz in the late seventies (sometimes also referred as Schwarz or SIC criterion) and it is not related to the information-theoretic approach but rather was named by analogy with AIC. AIC and BIC can be easily computed from most standard software outputs and are extensively used in data analysis. The information criteria quantify how experimental data support different models with various complexities and settle trade-off between data under- and overfitting. Usually, depending on parameter sensitivity, the introduction of additional parameters will improve model's fit to the data and the estimator of the goodness of fit will favour, among competing models, the model with the higher complexity (i.e., with the most parameters). However, the information criteria contain a penalty which controls overfitting, select the model that best balances under- and overfitting and, thus, determine an optimal number of parameters in the model. Of course, the use of IC is not a panacea and they do not identify the "true" or "final" model, – they select the best one for the given data from the set of candidate models. As well put by George Box et al. in Ref. [19,p. 440], "The most that can be expected from any model is that it can supply a useful approximation to reality: All models are wrong; some models are useful." Moreover, a lesson learnt from our previous study [20] shows that purely statistical model-selection scenario sometimes may not be persuasive enough since the information criteria do not take into account underlying physical meanings and technological aspects which are also highly important in ellipsometric measurement evaluations.

[1] As a matter of fact, it has been shown by Mayer et al. [11] that von Neumann's statement is actually correct and it is possible to reconstruct an elephantine shape with four complex parameters and even have a wiggling trunk by using the real part of the fifth parameter.

Incidentally, as has been noted by Yin and van Enk in Ref. [21], the usefulness of the information criteria approach in physical science and applications, where we typically know underlying model, lies mainly in ability to restrict number of model parameters and prevent overfitting. Yet, this is not a case for spectroscopic ellipsometry where we, typically, have a choice of physics-based dispersion models to describe a particular thin film.

In general, model selection often represents just a first step in spectroscopic ellipsometry data analysis. After that, we should continue with parametric sensitivity analysis (SA) [22–24] on the chosen model to evaluate relative importance of various model parameters. One of the most common and preferred by practitioners SA methods is based on "one-at-a-time" or "one-factor-at-a-time" (OAT/OFAT or "local") approach. OAT/OFAT estimates parameter sensitivity by evaluating model output changes while perturbing one parameter at a time by small amount and holding all other parameters fixed at their nominal values. After that, the parameter's value is restored to its original numerical value. This procedure, repeated for each parameter of interest, allows one to identify influential model parameters and rank them accordingly. However, it implies that the OAT/OFAT approach uses very restricted range of the input parameter variations around baseline and neglects possible parameters interactions. Therefore, in practice, good SA technique should estimate the effect one of the parameters of interest while all the others are allowed to float as well [25,26]. As it turns out, such an approach does exist and known as "Elementary Effect" (EE) method (or, the Morris method) and was proven to be an effective screening-type technique [27–32]. The EE method allows estimation of overall significance for each model parameter and describes its non-linear effect on model's output and/or interactions with other parameters.

The main purpose of this paper is to demonstrate how the information criteria approach and screening elementary effect method can be used together to select the most proper model and its optimal parameterization, reduce parameter uncertainties and classify model parameters according to their relative importance for the optical model output. Accurate and detailed consideration of all model-forming factors is an extremely complex undertaking. In this paper, we limit ourselves only to dispersion model selection assuming that an appropriate structural modeling has been already done based on a knowledge of involved processes and/or reference metrology. We show by example that the AIC and BIC model-selection procedures in combination with the EE method help to quantify the evidence for or against various dispersion models and determine relative importance of various model factors influencing an accuracy of ellipsometric data analysis.

2. A brief overview of the methodology

2.1. AIC and BIC

A typical non-linear least-squares data fitting problem can be mathematically abstracted as follows: we fit n experimental data points y_i, $i = 1,\ldots,n$, taken with some error at different values of independent variable x, by m different functional relations (candidate models) $f_j(x_i|\hat{\theta})$, $j = 1, \ldots, m$, where $\hat{\theta} = (p_1,\ldots,p_k)$ is the vector of model parameters:

$$y_i = f_j(x_i|\hat{\theta}) + \zeta_i, \tag{1}$$

where ζ_i are random variables assumed to be normally distributed with zero mean and uncertainty σ, $N(0,\sigma^2)$. Then an appropriate optimization algorithm applies to minimize the sum of the weighted squares of the residuals between the measured y_i and

Download English Version:

https://daneshyari.com/en/article/5347746

Download Persian Version:

https://daneshyari.com/article/5347746

Daneshyari.com