A surrogate model enables a Bayesian approach to the inverse problem of scatterometry

To cite this article: S Heidenreich et al 2014 J. Phys.: Conf. Ser. 490 012007

View the article online for updates and enhancements.
A surrogate model enables a Bayesian approach to the inverse problem of scatterometry

S. Heidenreich, H. Gross, M.-A. Henn, C. Elster and M. Bär
Physikalisch-Technische Bundesanstalt, Abbestr. 2-12, 10587 Berlin, Germany
E-mail: sebastian.heidenreich@ptb.de

Abstract. Scatterometry is an indirect optical method for the determination of photomask geometry parameters from scattered light intensities by solving an inverse problem. The Bayesian approach is a powerful method to solve the inverse problem. In the Bayesian framework estimates of parameters and associated uncertainties are obtained from posterior distributions. The determination the probability distribution is typically based on Markov chain Monte Carlo (MCMC) methods. However, in scatterometry the evaluation of MCMC steps require solutions of partial differential equations that are computationally expensive and application of MCMC methods is thus impractical. In this article we introduce a surrogate model for scatterometry based on polynomial chaos that can be treated by Bayesian inference. We compare the results of the surrogate model with rigorous finite element simulations and demonstrate its convergence. The accuracy reaches a value of lower than one percent for a sufficient fine mesh and the speed up amounts more than two order of magnitudes. Furthermore, we apply the surrogate model to MCMC calculations and we reconstruct geometry parameters of a photomask.

1. Introduction

Scatterometry provides a fast indirect optical method for the determination of photomask grating profile parameters that are obtained from scattered light intensities using inverse methods [1, 2]. In recent investigations profile parameters of rough absorber lines are reconstructed [5] from extreme ultraviolet light (EUV) measurements by employing a maximum likelihood method for parameter estimations [3, 4, 6]. This method provides accurate profile parameters with uncertainties of several nanometers. To further decrease uncertainties different measurements can be combined consistently by Bayesian interference.

In this paper, focus on an EUV photomask that consists of periodic absorber lines and a multilayer structure (see Fig. (1a) for details).

2. Inverse problem in scatterometry

In general the propagation of electromagnetic waves is described by Maxwell’s equations. For a grating geometry that consists of straight absorber lines, Maxwell’s equations reduce to the two dimensional Helmholtz equation: \( \Delta u(x, y) + k^2(x, y) u(x, y) = 0 \), where \( u \) is the transversal field component and \( k(x, y) \) is the wave number. The resulting boundary value problem is solved by a finite element method (FEM) which is computationally expensive. We have used the software package DIPOG 1 for our calculations.

1 developed at WIAS; http://www.wias-berlin.de/software /DIPOG
The mathematical model is defined by a map of geometry parameters onto light intensities. In particular, the set of parameters \( p_1, \ldots, p_N \) fixes a specific grating geometry and the solution of the boundary value problem with respect to the chosen geometry defines the map \( f_j(p_1) \ldots f_M(p): p \mapsto f_j(p) \). Here, the index \( j \) labels the diffraction orders of light diffracted intensities \( f_j \). Experimental data are modeled by the forward map \( f_j \) and additional noise, i.e., \( y_j = f_j + \epsilon_j \), where \( \epsilon_j \sim \mathcal{N}(0, \sigma_j) \) describes Gaussian noise with zero mean. The variance \( \sigma_j^2 = (a \cdot f_j(p))^2 + b^2 \), in which \( b \) models the background noise strength and \( a \) models noise of power fluctuations of the incident beam during the measurement [4]. The inverse problem in scatterometry is defined as the reconstruction of geometry parameter values \( p_i \) from \( y_j \).

3. Bayesian Approach

The Bayesian approach provides a statistical method to solve the inverse problem based on Bayes’ theorem,

\[
\pi(\theta; y) = \frac{\mathcal{L}(\theta; y)\pi_0(\theta)}{\int \mathcal{L}(\theta; y)\pi_0(\theta) d\theta} \tag{1}
\]

Here, the distribution \( \pi_0 \) includes prior knowledge about the parameter and \( \mathcal{L} \) determines the likelihood function. For scatterometry we choose prior knowledge \( \pi_0 = U[\theta_{\text{min}}, \theta_{\text{max}}] \), where \( U \) is a uniform distribution and the vector \( \theta \) consists of \( \{ p, a, b \} \). As in our previous studies we choose the likelihood function [3]

\[
\mathcal{L}(p, a, b; y) = \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi\sigma_j^2(p, a, b)}} \exp \left[ -\frac{(y_j - f_j(p))^2}{2\sigma_j^2(p, a, b)} \right], \sigma_j^2 = (a \cdot f_j(p))^2 + b^2. \tag{2}
\]

The expectation value of parameter \( \theta_i \) is given by the average with respect to the posterior measure, i.e., \( \bar{\theta}_i = E[\theta_i]_\pi \) and the variance by \( \text{var}(\theta_i) = E[\theta_i^2]_\pi - E[\theta_i]_\pi^2 \). In the Bayesian framework parameters are frequently estimated by MCMC sampling techniques which typically have slow convergence. For every sampling step the likelihood function and in turn the forward model has to be evaluated. Clearly, such an approach is not efficient for the inverse problem in scatterometry. In this respect we will apply a polynomial chaos method to construct a surrogate model that dramatically speeds up forward calculations.

4. Polynomial chaos

Polynomial chaos (PC) was introduced in 1938 by Wiener [7]. He used polynomial approximations for the Gaussian stochastic process. Later, Cameron and Martin [8] proved that every stochastic process with a finite second moment can be approximated by an infinite convergent sum of orthogonal polynomials. The approach became popular for uncertainty quantification after Xiu and Karniadakis [9] have generalized polynomial chaos by the introduction of the Wiener-Askey scheme for flow problems. The Wiener-Askey scheme relates stochastic processes to orthogonal polynomials which ensures good convergence results in the expansion. In this perspective we will focus on the non-intrusive variant of generalized polynomial chaos method in which proprietary software can be used as a black box for rigorous evaluations of the forward model necessary to determine expansion coefficients. In particular, the forward map for uncertain input geometry parameter \( p_i \) is approximated by a polynomial series

\[
f_j(p_1 \ldots n) = f_j(\mu_{1 \ldots n}) + \sum_{i=1}^{n} \sum_{k=1}^{R} f_{jk}(\mu_{1 \ldots n}) \Psi_k(\omega_i) + \ldots \tag{3}
\]

where \( \mu_i \) is the mean value of the related uniform prior distribution and \( \omega_i \) a random number. Specifically, a random parameter drawn from a prior distribution \( U[\theta_i^m, \theta_i^M] \) is given by \( \theta_i = \mu_i + \omega_i \). Therefore,
the range of random variables $\omega_i$ is set by its prior distribution, i.e., $\omega_i \in U[\theta_i^{\min} - \mu_i, \theta_i^{\max} - \mu_i]$. For simplifications we introduce the notation $U(\omega_i) = U[\theta_i^{\min} - \mu_i, \theta_i^{\max} - \mu_i]$.

According to the Wiener-Askey scheme [9] we choose Legendre polynomials as basis functions $\psi_k$. The choice of basis functions specifies the kind of chaos, here, to Legendre chaos. The expansion coefficients are determined by projections onto the chosen Legendre functions $\Psi_k$, i.e.,

$$f_{jk}^a(\mu_1...n) = \int \frac{f_j(p_1...n)\Psi_k(\omega_i)U(\omega_i)d\omega_i}{\Psi_k^2(\omega_i)d\omega_i} \approx \frac{1}{\int \Psi_k^2(\omega_i)U(\omega_i)d\omega_i} \sum_k \lambda_k f_j(\omega_i^k)U(\omega_i^k), \ldots, \quad (4)$$

The integral in the denominator is solved analytically and the integral in the nominator is approximated by quadrature rules. Quadrature rules provide a mesh (index $k$) and corresponding weights $\lambda_k$.

5. Results and conclusion

For the application of the surrogate model we first show the accuracy and convergence of the approximation scheme. In our study, we used prior distributions for the parameters: $U[133, 147]$ for the bottom CD, $U[74.4, 85.6]$ for the height, $U[80, 90]$ for the side-wall angle, $U[0.01, 0.1]$ for $a$ and $U[0.001, 0.1]$ for $b$ (noise parameters). We calculated the Legendre chaos expansion of the forward map up to the 10th order, drew a geometry from the prior distribution, calculated the corresponding light diffracted intensities with the rigorous FEM solver and compared these results with results from the PC surrogate model. Fig. (1b) shows a good agreement of the approximation scheme with FEM calculations for one chosen geometry. To cover a wider parameter range we considered 100 random geometries and calculated the $L^2$ norm of the difference between FEM results and PC surrogate approximations. As a measure of the overall error we took the maximum value of the deviations for chosen geometries. Fig. (1c) shows the convergence of PC surrogate approximations with increasing number of grid points. It is found to be much better than results obtained with a surrogate model based on simple nearest neighborhood interpolations (NN). For further investigations we fixed the resolution to 608 supporting points, where the relative deviation of efficiencies is lower than 1%.

Finally, we have reconstructed geometry parameters from simulated noisy data sets $y_i$. We included the PC surrogate model in a MCMC scheme and estimated geometry parameters with $10^4$ sampling steps. Fig. (1d) shows eleven reconstructions for the side-wall angle (SWA). All reconstructions are consistent with the reference value and the confidence interval of 95% is small: about $0.7^\circ$.

In this paper we introduced a polynomial chaos based surrogate model to demonstrate its suitability to speed up Bayesian inferences. We have shown the accuracy and convergence of the surrogate model. Moreover, the PC surrogate model was used for reconstructions of the geometry parameters from numerical generated noisy data sets. The reconstructions are sufficiently precise which makes the scheme applicable for solving the inverse problem in scatterometry based on MCMC methods. Typically computational times of the proposed method compared to rigorous FEM based calculations are depicted in table (1). The computational cost of pre-calculations is about 20h due to the determination of expansion coefficients. Polynomial chaos surrogate methods provide an effective method to speed up Bayesian inferences using MCMC calculations.

Figure 1. a) Cross section of one period of an EUV photomask which we used in our study. b) Polynomial chaos based surrogate model compared with rigorous finite difference (FEM) calculations. The relative error is lower than 3 percent. c) Convergence rate of polynomial chaos based surrogate model (PC) compared to nearest neighborhood method (NN). The error is given as the maximum deviation in the $L^2$ norm of the surrogate model from FEM calculations for 100 sample geometries. d) Results for eleven reconstructions of the side wall-angle from noisy data using the MCMC sampling based on PC surrogate method. The error bars show the 95% confidence interval and the dashed line shows the reference value.

<table>
<thead>
<tr>
<th>Method</th>
<th>Function evaluations</th>
<th>FEM</th>
<th>PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity Analysis</td>
<td>$3 \times 10^4$</td>
<td>41$d$</td>
<td>$20h + 0.1s$</td>
</tr>
<tr>
<td>Bayesian MCMC</td>
<td>$\sim 10^5$</td>
<td>139$d$</td>
<td>$20h + 47min$</td>
</tr>
<tr>
<td>Maximum Likelihood</td>
<td>30</td>
<td>1$h$</td>
<td>$20h + 0.84s$</td>
</tr>
<tr>
<td>Least squares</td>
<td>15</td>
<td>0.6$h$</td>
<td>$20h + 0.42s$</td>
</tr>
</tbody>
</table>

Table 1. Speed up for different methods. The PC-surrogate needs always $20h$ for the determination of expansion coefficients.