Internship report:

On the use of PCE based surrogate models in SXR Scatterometry.

by

R. Koppert

Student number:	4606760	
Project duration:	January 23, 2023 – April 21, 2023	
Supervisors:	S. Scholz	ASML
	O. El. Gawhary,	TU Delft

TU Delft

SURROGATE MODELS

REINT KOPPERT*

21-4-2023

CONTENTS

ABSTRACT

Scatterometry is a non-destructive metrology technique widely used in the semiconductor industry for the reconstruction of periodic structures from diffraction measurements. This involves solving a so-called inverse problem, which can be done by tuning the geometry parameters of a forward model such that the discrepancy between the measured diffraction pattern and the diffraction pattern computed using a Maxwell solver, are minimized. In order to meet with current semiconductor metrology demands, soft x-ray (SXR) scatterometry has been introduced. In SXR a short wavelength and broad band illumination source is used, allowing for the reconstruction of smaller and more complex structures. However, this does require the use of a computationally expensive forward model. This also complicates the assessment of the sensitivity of the measurement setup to the various grating parameters and whether the parameters can be determined independently. The current approach to this problem is strictly local. In order to address these issues, the use of a surrogate model for the Maxwell solver based on Polynomial Chaos Expansion (PCE) is investigated in this report. The performance of the PCE based surrogate model is accessed for for the SXR metrology application on a simple 1D line grating.

* *Master student Applied Physics, TU Delft.*

1 INTRODUCTION

Progress in semiconductor manufacturing has been driven by Moore's law for over half a century now. To achieve this goal, existing device designs were scaled, until physical and practical limitations forced the adaption of newer and more complex designs.This lead to overall smaller and more complex structures as outlined e.g. in [1]. This also meant that more advanced metrology techniques had to be developed and introduced to assist in the manufacturing process. The next generation devices are the so-called gate-all-around (GAA) devices. The scale and three-dimensional nature of these devices imposes new metrology challenges, as outlined e.g. in [2] and [3].

Soft x-ray scatterometry is a promising metrology solution for dealing with some of these challenges.[4] Scatterometry, in general, is a is a non-destructive metrology technique for the reconstruction of periodic structures from diffraction measurements. This technique works by solving the inverse problem of determining the geometry parameters, that characterizes the periodic structure, from the diffracted light. Solving this problem requires a forward model (Maxwell solver), that can be used to predict a measured diffraction pattern based on the illumination source used and the geometry parameters. In SXR scatterometry a small-wavelength and broadband illumination source (10-20 nm) is used to ensure that accurate estimates of all parameters can be obtained from the ill-posed inverse problem. The problem faced by scatterometry, however, is that as the complexity of the structures increases so does the complexity of the forward model, resulting in an ever increasing computational cost. In SXR scatterometry the broad-band nature of the illumination source adds to this problem.

To mitigate this problem the use of surrogate models, approximations to the forward model that are easier to evaluate in terms of computation cost, in (EUV) scatterometry was investigated in [5]. The polynomial chaos expansion (PCE) was shown to be the most promising candidate. This replaces the forward model by a (set of orhogonal) polyniomial(s), thereby drastically lifting the computational burden. This surrogate can be obtained in an non-intrusive fashion based on a number of calls to the forward model, meaning that it does not require any changes to the forward model. Over recent years effort was put in introducing more efficient and generally applicable schemes to obtain this surrogate.[6][7][8][9] In [8] a regression based approach using the optimal sampling, outlined in [10], for obtaining a PCE based surrogate in (visible light) scatterometry was introduced. This approach has been implemented in the Python package Pythia [11].

An additional benefit of the PCE based surrogate model is that it also facilitates a global sensitivity analysis, as demonstrated by [12]. The aim of global sensitivity analysis is to quantify sensitivity of a non-linear function with respect to its input parameters that does not limit itself to the local change of the function. This can be quantified by the Sobol indices, that can be efficiently approximated based on the PCE of the function.[12] In scatterometry the insight gained by the global sensitivity analysis could be useful in predicting how accurately the geometry parameters can be obtained from noisy diffraction measurements. In fact, [7] concluded that uncertainty estimates of the geometry parameters obtained using Bayesian inversion were in-line with expectations set by the Sobol coefficients. This is particularly relevant for SXR scatterometry as the current approach is strictly local and the use of naive global methods is impeded by the computational cost of the forward model.

In this work the use of PCE based surrogate models in the context of SXR scatterometry is investigated. The PCE are computed using Pythia [11]. This reports starts with briefly summarizing the relevant theory on scatterometry and polyno-

mial chaos expansions in Chapter 2. In Chapter 3 the performance of the PCE based surrogate model is accessed for a typical use case in SXR scatterometry, the Simeon after develop inspection (ADI) grating. This report ends with some concluding remarks in Chapter 4.

This investigation was carried out as part of a three month long internship (equivalent to 18 ECTs) at ASML, as part of the non-academic internship (AP3911) of the Master Applied Physics program at the TU Delft. This reports also serves as the internship report that is required to complete the course.

2 THEORY

In this chapter the relevant theory is summarized. In Section 2.1 scatterometry is briefly introduced. In Section 2.2 the polynomial chaos expansions (PCE), as used here, are introduced. For a more detailed discussion the reader is referred to [12] and [13].

2.1 Scatterometry

Scatterometry is a non-destructive metrology technique widely used in semiconductor industry for the reconstruction of periodic structures from diffraction measurements. These periodic structures are often referred to as gratings and are parametrized by a vector of grating parameters, denoted as $\mathbf{x}$. In this study, the scope is limited to 1D gratings, that are periodic in one direction and constant along the perpendicular direction.

This technique works by illuminating the grating with a light source and measuring the resulting diffraction pattern in a defined plane. The problem of determining the grating parameters $\mathbf{x}$ from the diffraction pattern is referred to as the inverse problem. Typically, this is solved by studying the forward problem, which predicts the diffraction pattern for a given set of grating parameters $\mathbf{x}$ using information on the illumination source. For scatterometry this is done by solving Maxwell's equations using a rigorous solver, e.g. [14]. In the context of this study, the forward model can be understood as a function $f_i(\mathbf{x})$ that maps the grating parameters $\mathbf{x}$ to the pixels i of the detector in the measurement plane.

Then, the inverse problem can be solved by obtaining the grating parameters for which the discrepancy between measured and predicted diffraction pattern is minimized. The inverse problem, in general, is ill-posed. Therefore, it is a priori unknown whether there is an unique set of grating parameters that explains the measured diffraction pattern. Moreover, the diffraction pattern must be sensitive to the grating parameters, so that in the presence of measurement noise accurate grating estimates can be obtained.[2] In SXR scatterometry a small-wavelength and broadband illumination source (10-20 nm) is used to ensure that accurate estimates of all parameters can be obtained from the ill-posed inverse problem.

2.2 Polynomial Chaos expansions

Polynomial Chaos Expansions are defined for functions $f(\mathbf{x})$ of a random vector $\mathbf{x}\mathbb{R}^M$ that are distributed according to $\rho(\mathbf{x})$. The used expansion functions $\Phi_j(\mathbf{x})$ are orthonormal with respect to $\rho(\mathbf{x})$, i.e.

$$\int \Phi_j(\mathbf{x})\,\Phi_i(\mathbf{x})\,\rho(\mathbf{x})\,d\mathbf{x} = \delta_{i,j}, \tag{1}$$

where $\delta_{i,j}$ is the Kronecker delta. Then, the function $f(\mathbf{x})$ can be decomposed as follows:

$$f(\mathbf{x}) = \Sigma_j c_j \Phi_j(\mathbf{x}), \tag{2}$$

where the expansion coefficients c_j follow from the orthonormality given in Eqn. 1:

$$c_j = \int \Phi_j(\mathbf{x})\,f(\mathbf{x})\,\rho(\mathbf{x})\,d\mathbf{x}. \tag{3}$$

In literature these expansions are also referred to as generalized polynomial chaos expansions (gPC). The equality given by Eqn. 2 holds for functions $f(\mathbf{x})$ that have finite variance, due to the prove by [15] as cited by e.g. [12] and [5].

Here we consider only uniform distributions on some interval $I = [\mathbf{x}_{min}, \mathbf{x}_{max}]$, as done in [6] and [9]. In this case the basis functions $\Phi_j(\mathbf{x})$ are products of normalized Legendre polynomials $L_{li}(xi)$:

$$\Phi_j(\mathbf{x}) = \Pi_{i=1}^{M} L_{li_j}(x_i). \tag{4}$$

For convenience the M dimensional multi-index $\mathbf{l}_j = (l1_j, l2_j, \ldots, lM_j)$ is introduced and used interchangeably with the index j. Then, the total order of a basis function Φ_j can be defined as : $\Sigma_{i=1}^{M} li_j$. The series given in Eqn. 2 can be truncated based on this convention, by only including all basis functions Φ_{α_j} up to order P. This truncated series contains N_c coefficients, as can be found in [12]:

$$N_c = \frac{(M+P)!}{M!P!}. \tag{5}$$

2.2.1 Computation

The expansion coefficients can be computed directly from Eqn. 3, using numerical integration. This approach, however, suffers from the dimensionality curse. Instead, we cast the problem of obtaining the coefficients as an regression problem, by introducing the following objective function:

$$\int \left(f(\mathbf{x}) - f_{PCE}(\mathbf{x})\right)^2 \rho(\mathbf{x})\, d\mathbf{x}. \tag{6}$$

The objective function in Eqn. 6 can be approximated using Monte-Carlo integration, as noted in [8], based on N_s samples of the function $f(\mathbf{x})$. Then, Eqn. 6 can be approximated (upto some constant) by the following cost function:

$$\Sigma_{k=1}^{N_s} w_k R(\mathbf{x}_k), \tag{7}$$

where the residual $R(\mathbf{x}_k) = f(\mathbf{x}_k) - f_{PCE}(\mathbf{x}_k)$ and $w_k = \rho(\mathbf{x}_k)$. Minimization of this cost function is a straight forward weighted linear least-squares problem. In order to formulate this problem in terms of a matrix equation, the vector $\mathbf{C} \in \mathbb{R}^{N_c}$ containing all expansion coefficients is introduced. Then, the expansion coefficients vector $\mathbf{C}$ follows from solving:

$$\mathbf{GC} = \mathbf{F}, \tag{8}$$

where the vector $\mathbf{F} \in \mathbb{R}^{N_c}$ and the square matrix $\mathbf{G}$ are given by:

$$\mathbf{F}_i = \Sigma_{k=1}^{N_s} w_k \Phi_i(\mathbf{x}_k) f(\mathbf{x}_k), \tag{9}$$

$$\mathbf{G}_{i,j} = \Sigma_{k=1}^{N_s} w_k \Phi_i(\mathbf{x}_k) \Phi_j(\mathbf{x}_k). \tag{10}$$

Note that $\mathbf{G}$ must be invertible in order for Eqn. 8 to have an unique solution. This means that $N_s > N_c$. In [16] it is stated that $N_s = kN_c$, with $k \geqslant 2$, samples typically is sufficient. Of course this depends on the used sampling strategy. We can investigate whether enough samples were used by computing the condition number K:

$$K = \frac{\sigma_{max}}{\sigma_{min}}, \tag{11}$$

where σ are the singular values of the matrix $\mathbf{G}$. This number indicates how well-conditioned the matrix is. In particular, this is motivated by the fact that the definition of $\mathbf{G}_{i,j}$ in Eqn. 10 indicates that $K \to 1$ as $N_s \to \infty$, since the basis functions $\Phi_i(\mathbf{x})$ satisfy Eqn. 1.

2.2.2 Sampling

A lot of sampling strategies can be found in literature, e.g. in [12] and [13]. Here the scope is limited to the strategies implemented in Pythia. The first option is to use a pre-computed data-set. The other sampling strategies follow from using importance sampling in the the Monte-Carlo integration used to approximate the objective function in Eqn. 8. Suppose that we draw N_s samples from some distribution $\mu(\mathbf{x})$, than Eqn. 8 can be approximated by:

$$\Sigma_{i=1}^{N_s} w_i R(\mathbf{x}_i), \tag{12}$$

where $w_i = \rho(\mathbf{x})\mu^{-1}(\mathbf{x})$. An option could be to sample according to $\rho(\mathbf{x})$. In [10] it was shown, as cited by [8], that sampling according to:

$$\mu(\mathbf{x}) = \rho(\mathbf{x})\frac{1}{N_c}\Sigma_{j=1}^{N_c}|\Phi_j(\mathbf{x})|^2, \tag{13}$$

is optimal in the sense of number of samples N_s required to obtain a well-condition matrix $\mathbf{G}$, defined in Eqn. 10. According to [8] the number of samples Ns must satisfy: $N_s/ln(N_s) \geqslant 4N_c$, following the results shown in [10]. In [9], however, it was observed that this condition is far to stringent.

2.2.3 Sobol indices

The aim of global sensitivity analysis is to quantify sensitivity of a non-linear function with respect to its parameters that does not limit itself to the local change of the function. In variance-based global sensitivity analysis this is quantified by evaluating the so called Sobol indices, which can be thought of as partial variances that represent the contribution of a set of input parameters to the variance of an output function. Here a short derivation is provided, following the same steps as in [12].

For a function $f(\mathbf{x})$ of a random vector $\mathbf{x} \in \mathbb{R}^M$ distributed according to $\rho(\mathbf{x})$, the variance $\mathrm{Var}(f(\mathbf{x}))$ is given by:

$$\mathrm{Var}(f) = \int \left(f^2(\mathbf{x}) - f_{mean}^2\right)\rho(\mathbf{x})\,d\mathbf{x}, \tag{14}$$

Before the Sobol indices can be defined, the Sobol decomposition must be introduced. The Sobol decomposition decomposes $f(\mathbf{x})$ in 2^M functions of the unique combinations of x_i:

$$f(\mathbf{x}) = f_0 + \Sigma_{i=1}^{M} f(x_i) + \Sigma_{j>i}\Sigma_{i=1}^{M} f(x_i, x_j) + \cdots + f(x_1, \ldots, x_M), \tag{15}$$

that are orthogonal with respect to $\rho(\mathbf{x})$. Note that $f_{mean} = f_0$ and the orthogonality of the expansion functions imply that:

$$\mathrm{Var}(f) = \Sigma_{i=1}^{M} D_i + \Sigma_{j>i}\Sigma_{i=1}^{M} D_{i,j} + \cdots + D_{1,2,\ldots,M}, \tag{16}$$

where $D_{i1,\ldots,is}$ for $s \leqslant M$ and $1 \leqslant i1 < i2 < \cdots < is \leqslant M$ is given by:

$$D_{i1,\ldots,is} = \int f^2(x_i1, \ldots, x_is)\,\rho(\mathbf{x})\,d\mathbf{x}. \tag{17}$$

This leads to the definition of the Sobol coefficients $S_{i1\ldots is}$:

$$S_{i1,\ldots,is} = \frac{D_{i1,\ldots,is}}{\mathrm{Var}(f)}, \tag{18}$$

as given in [12]. Note that the Sobol decomposition and the PCE expansion are related. Let $\mathbf{x} \in \mathbb{R}^2$ be distributed according to an uniform distribution on some interval $I = [\mathbf{x}_{min}, \mathbf{x}_{max}]$, then the PCE of truncation order $P = 2$ is given by:

$$f_{PCE}(\mathbf{x}) = c_{(0,0)}L_0(x_1)L_0(x_2) + c_{(0,1)}L_0(x_1)L_2(x_2) + c_{(1,0)}L_1(x_1)L_0(x_2) + c_{(1,1)}L_1(x_1)L_1(x_2) + c_{(2,0)}L_2(x_1)L_0(x_2) + c_{(0,2)}L_0(x_1)L_2(x_2),$$

where $L_{li}(xi)$ are the normalized Legendre polynomials. Note that the first term is a constant.

Then, we can approximate for example $f(x_1)$ given in Eqn. 15 by:

$$f(x_1) \approx c_{(1,0)} L_1(x_1) L_0(x_2) + c_{(2,0)} L_2(x_1) L_0(x_2). \tag{19}$$

Consequently, the Sobol index S_1 can be approximated as follows:

$$S_1 \approx \frac{|c_{(1,0)}|^2 + |c_{(2,0)}|^2}{|c_{(1,0)}|^2 + |c_{(0,1)}|^2 + |c_{(1,1)}|^2 + |c_{(2,0)}|^2 + |c_{(0,2)}|^2}. \tag{20}$$

Thus, we can approximate the Sobol coefficients based on the PCE expansion, as claimed in [12]. The general formula for the Sobol coefficients based on an orthonormal PCE can be found in [7].

3 SIMEON GRATINGS

In this chapter the performance of the PCE based surrogate model is tested for a particular use case. The use case considered here is the Simeon after-develop inspection grating (ADI) with four varying parameters: bottom critical dimension d_{BCD}, top critical dimension d_{TCD}, asymmetry d_{asym} and height h.

In order to limit the scope we only consider the positive first diffraction orders extracted from the diffraction pattern by taking the column wise sum (CWIS). Thus, the forward model is the mapping of the grating parameters $\mathbf{x} = (d_{BCD}, d_{TCD}, d_{asym}, h)$ to each pixel i in the CWIS: $f_i\,(h, d_{BCD}, d_{TCD}, d_{asym})$.

This forward model is approximated by the following PCE for each pixel i:

$$f^{(i)}_{PCE}(\mathbf{x}) = \Sigma_j c^{(i)}_j \Phi_j\,(\mathbf{x})\,, \tag{21}$$

for $\mathbf{x} \in \mathcal{D}$ and $\Phi_j(\mathbf{x})$ is the product of(normalized) Legendre polynomials, as given by Eqn. 4. From here on this approximation will be referred to as the PCE based surrogate model.

The number of expansion coefficients N_c in these expansions is specified by the truncation order P, as given by Eqn. 5. The expansion coefficients are obtained using Pythia [11], which implements the regression approach outlined in Section 2.2.1. In this approach the PCE based surrogate model is trained using N_s pre-computed train samples from the forward model. The accuracy of the surrogate model can be accessed by evaluating the normalized cost of the solution:

$$\chi_{train} = \frac{\sum_i \sum_{\mathbf{x}_{train}} \left| f\,(\mathbf{x}_{train}) - f_{PCE}\,(\mathbf{x}_{train}) \right|^2}{\sum_i \sum_{\mathbf{x}_{train}} \left| f\,(\mathbf{x}_{train}) \right|^2}. \tag{22}$$

In order to quantify the actual relative error of the PCE based surrogate model, the following error metric is also used:

$$e_{train} = \frac{\sum_i \sum_{\mathbf{x}_{train}} \left| f\,(\mathbf{x}_{train}) - f_{PCE}\,(\mathbf{x}_{train}) \right|}{\sum_i \sum_{\mathbf{x}_{train}} \left| f\,(\mathbf{x}_{train}) \right|}. \tag{23}$$

In order to validate the obtained surrogate model is also accurate for samples not included in the regression, the same error metrics are evaluated for N_{test} samples, $\mathbf{x}_{test}$:

$$\chi_{test} = \frac{\sum_i \sum_{\mathbf{x}_{test}} \left| f\,(\mathbf{x}_{test}) - f_{PCE}\,(\mathbf{x}_{test}) \right|^2}{\sum_i \sum_{\mathbf{x}_{test}} \left| f\,(\mathbf{x}_{test}) \right|^2}, \tag{24}$$

$$e_{test} = \frac{\sum_i \sum_{\mathbf{x}_{test}} \left| f\,(\mathbf{x}_{test}) - f_{PCE}\,(\mathbf{x}_{test}) \right|}{\sum_i \sum_{\mathbf{x}_{test}} \left| f\,(\mathbf{x}_{test}) \right|}. \tag{25}$$

We first investigate the accuracy of the PCE for a given truncation order P, by evaluating the train error e_{train} of the PCE based surrogate model for a range of P. Then, we investigate how many samples N_s are required to obtain the PCE based surrogate model for a range of P. Thereafter, we investigate how many expansion functions are actually needed, by evaluating the error metrics for a PCE based surrogate model that consists of a reduced number of expansion functions. This is repeated for the dense domain $\mathcal{D}_1$ and coarse domain $\mathcal{D}_2$, specified in table 1. Both domains $\mathcal{D}_1$ and $\mathcal{D}_2$ are sampled by a regular grid, consisting of $N_s = 101871$ and $N_s = 45056$ points, respectively.

Table 1: The upper and lower bounds of the grating parameters of the Simeon ADI-grating for the domains $\mathcal{D}_1$ and $\mathcal{D}_2$.

	$\mathcal{D}_1$		$\mathcal{D}_2$	
	Min	Max	Min	Max
d_{BCD} [nm]	40	42	30	60
d_{TCD} [nm]	40	42	30	60
d_{asym} [nm]	0	1	-5	5
h [nm]	89	91	80	110

3.1 Truncation order

First we consider the truncation order P required to obtain an accurate PCE based surrogate for both domains. Therefore, we compute the PCE based surrogate models using all available samples for a range of truncation orders P. For these surrogate models the train error e_{train} is evaluated and shown in Figure 1. It can be observed that the relative train error e_{train} decreases significantly over the considered range of truncation orders P for the PCE based surrogate model on the dense domain. Increasing P from 4 to 9 for the PCE based surrogate model on the coarse domain has no significant impact on performance, i.e. the relative train error barely decreases. Perhaps this could be related to the domain size relative to the wavelength. A simple metric to quantify this could be the typical dimension $V_{\mathcal{D}}^{1/M}$, where $V_{\mathcal{D}}$ is the volume of $\mathcal{D}$, relative to the used wavelength. Perhaps this can be used to compare the results shown here to [6] and [8].

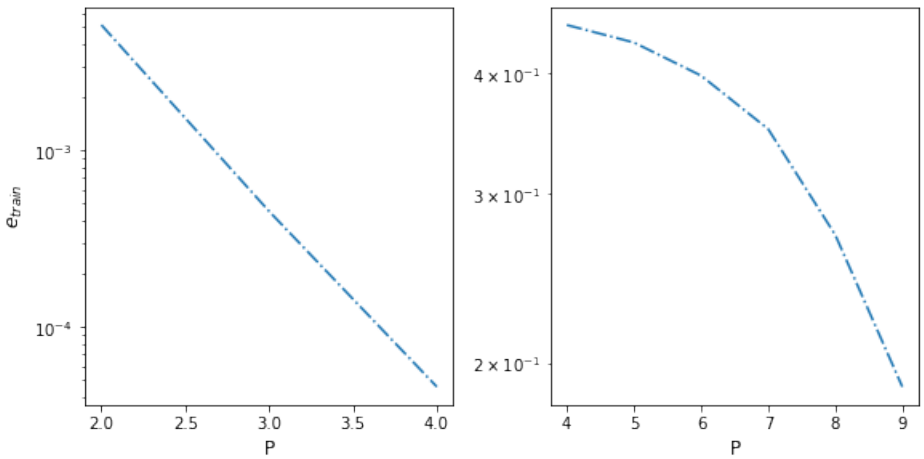

Figure 1: The train error e_{train} of the PCE based surrogate model versus the truncation order P for the dense domain $\mathcal{D}_1$ and the coarse domain $\mathcal{D}_2$ on the left and right, respectively. All samples of the forward model were used to train the surrogates.

3.2 Number of samples

Next, we investigate the number of samples required to obtain an accurate surrogate model on the dense domain D_1 for the range $P = [2, 3, 4]$. This is done by evaluating the error metrics χ_{train} and e_{train} of the PCE based surrogate model trained using N_s train samples. To validate the results, $N_t = N_s$ test samples are used to evaluate the error metrics χ_{test} and e_{test}. We repeat the process $N_{runs} = 10$ times to account for variance in these metrics that would occur when a different set of samples is used. We also evaluate the condition number K of $\mathbf{G}$, given by Eqn. 11, as motivated in Section 2.2.1.

Note that the number of expansion coefficients N_c depends on the truncation order P, as given by Eqn. 5. Therefore, we evaluate the error metrics for a fixed range of $N_s/N_c = 1.5 \ldots 6$.

In Figure 2 the results are shown. In these figures a clear discrepancy between train and test metrics can be observed, that decreases for increasing N_s/N_c as both approach the same limit. Moreover, the train metrics approach this limit from below, whereas the opposite can be observed for the test metrics. This could be explained by assuming that for the forward model there exists a unique PCE expansion, whose coefficients we approximate based on a number of samples N_s. This approximation becomes more accurate for increasing number of samples, so that eventually the discrepancy between train and test metrics vanishes. The results presented here suggest that $N_s/N_c \approx 4$ is sufficient. Note that the expansion coefficients obtained using the regression approach will be biased such that the cost function in Eqn. 2.2.1 is minimized. Meaning that error metrics based on train points will not accurately represent the performance of the PCE based surrogate model.

Furthermore, the results suggest that the condition number K is a good indicator for how well the obtained PCE based surrogate model generalizes. Note that this can be computed before the actual forward model is sampled, as $\mathbf{G}$ given in Eqn. 10 only depends on the used expansion basis.

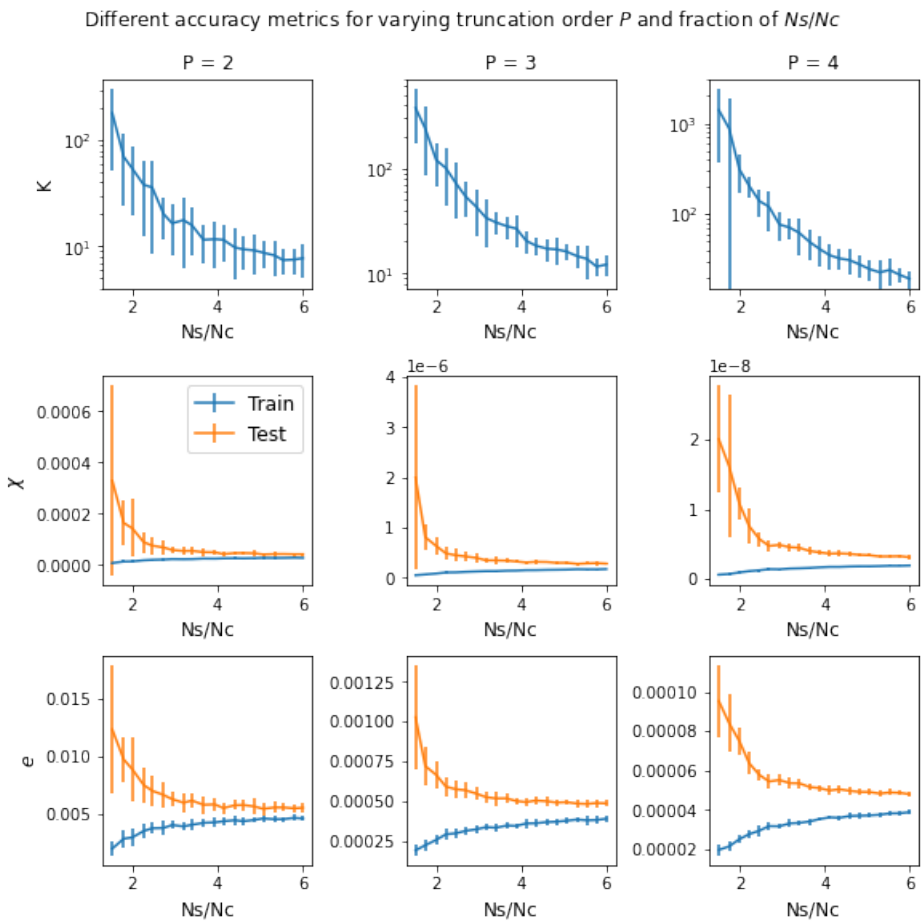

Figure 2: The different error metrics for a range of truncation orders P of the PCE based surrogate model on the dense domain $\mathcal{D}_1$ trained using N_s samples.

3.3 Sparsity PCE

The obtained PCE based surrogate model appears to be sparse for both domains $\mathcal{D}_1$ and $\mathcal{D}_2$, in the sense that a lot of expansion coefficients are negligible. To demonstrate this, we compute the PCE with truncation order $P_1 = 4$ and $P_2 = 6$ obtained based on $N_s = 5N_c$ samples on the domains $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively, for all pixels i in the CWIS. To validate the solution $N_t = N_s$ test samples are used. In order to do so, we compute the standard deviation of all coefficients $c_j^{(i)}$ along the pixel range i and sort the coefficients accordingly. Then, we evaluate the error metrics e_{train} and e_{test} for the PCE based surrogate model that includes the first n_c sorted coefficients.

In Figure 3 the results are shown for $\mathcal{D}_1$ and $\mathcal{D}_2$ on the left and right, respectively. The sparsity of the PCE based surrogate is illustrated by the rate at which the train error e_{train} decreases for increasing n_c, which falls of significantly. For the coarse domain it can be observed the test error e_{test} increases, when e_{train} starts to level of. Perhaps this suggest that the absolute error in these expansion coefficients is larger than the actual expansion coefficients themselves. Hence, could motivate the introduction of a thresholding scheme in which expansion coefficients below a set threshold are dropped from the PCE based surrogate model.

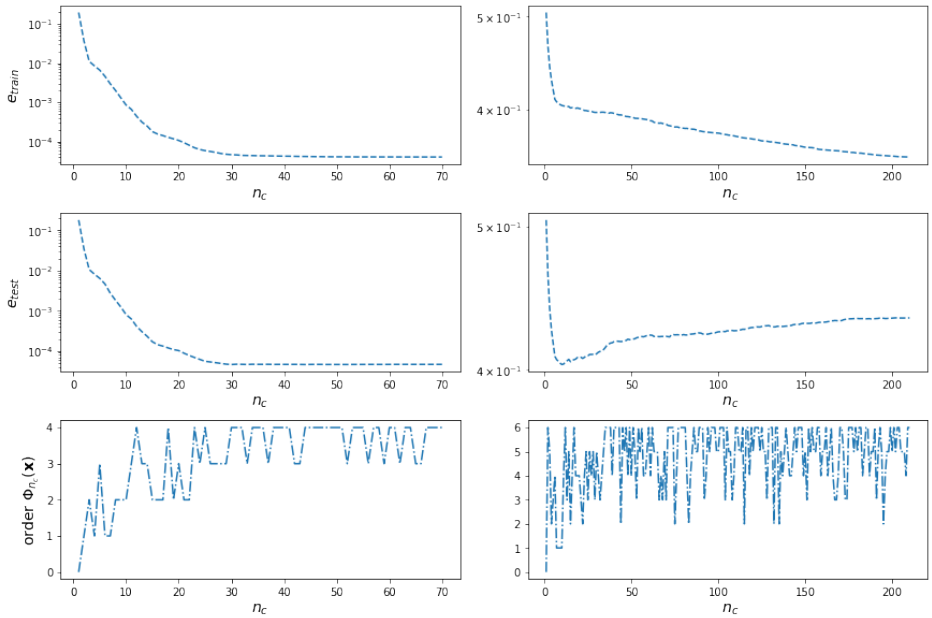

Figure 3: The train error e_{train} and test error e_{test} for the PCE based surrogate models containing an increasing number of expansion coefficients $c_j^{(i)}$, ordered by the standard deviation over the pixel range i. The results shown on the left and right correspond to the PCE based surrogate model for the dense domain $\mathcal{D}_1$ ($P_1 = 4$) and the coarse domain $\mathcal{D}_2$ ($P_2 = 6$). The surrogate models are trained using $N_s = 5N_c$ train samples and validated using $N_t = N_s$ test samples.

Based on these results we can access the truncation strategy, by plotting the order of the n_c^{th} expansion function as included in Figure 3. In general, this indicates that truncating the PCE based on total order is a good starting point. Perhaps slight improvements in performance can be obtained by introducing a truncation strategy that is tailored to the importance of the grating parameters, e.g. include basis functions $\Phi_j(\mathbf{x})$ containing higher order Legendre polynomials in the height h.

4 CONCLUDING REMARKS

In this study the use of PCE based surrogate models in SXR scatterometry was studied. It was observed that for a simple 1D line grating in soft x-ray scatterometry the forward model can be approximated by a PCE based surrogate model on a sufficiently small subdomain of the parameter space, using a reasonable number of samples of the forward model. Extending the validity of the PCE based surrogate model to larger domains proved difficult. The problem, in general, is that approximating increasingly complicated functions requires higher P and as a consequence a large number of expansion functions N_c. In turn this drives up the number of samples of the forward model, as $N_c > N_s$ in the regression approach. This is partly due to the simplistic truncation scheme used here, in which the number of expansion functions grows exponentially with P. Note that the sparsity of the of the obtained PCE based surrogate models indicates that this problem could be mitigated. Therefore, the use of sparse adaptive schemes such as presented in [16] could be interesting. It is important to note that the same problem is encountered when instead of complexity (domain size) the number of variables of the forward model is increased. Unfortunately this is inherent to PCE expansion and suggest the use of so-called low-rank approximations instead. Further research on the use of Sobol coefficients in SXR scatterometry is required.

REFERENCES

[1] Kamal Y. Kamal. The silicon age: Trends in semiconductor devices industry. *Journal of Engineering Science and Technology Review*, 15(1):110–115, 2022.

[2] N. G. Orji, M. Badaroglu, B. M. Barnes, C. Beitia, B. D. Bunday, U. Celano, R. J. Kline, M. Neisser, Y. Obeng, and A. E. Vladar. Metrology for the next generation of semiconductor devices. *Nature Electronics*, 1(10):532–547, oct 2018.

[3] Mary A. Breton, Daniel Schmidt, Andrew Greene, Julien Frougier, and Nelson Felix. Review of nanosheet metrology opportunities for technology readiness. *Journal of Micro/Nanopatterning, Materials, and Metrology*, 21(02), apr 2022.

[4] Christina Porter, Teis Coenen, Niels Geypen, Sandy Scholz, Loes van Rijswijk, Han-Kwang Nienhuys, Jeroen Ploegmakers, Johan Reinink, Hugo Cramer, Rik van Laarhoven, David O'Dwyer, Peter Smorenburg, Andrea Invernizzi, Ricarda Wohrwag, Hugo Jonquiere, Juliane Reinhardt, Omar el Gawhary, Simon Mathijssen, Peter Engblom, Heidi Chin, William T Blanton, Sury Ganesan, Brian Krist, Florian Gstrein, and Mark Phillips. Soft x-ray:novel metrology for 3d profilometry and device pitch overlay. 2023.

[5] Sebastian Heidenreich, Hermann Gross, and Markus Bar. BAYESIAN APPROACH TO THE STATISTICAL INVERSE PROBLEM OF SCATTEROMETRY: COMPARISON OF THREE SURROGATE MODELS. *International Journal for Uncertainty Quantification*, 5(6):511–526, 2015.

[6] Sebastian Heidenreich, Hermann Gross, and Markus Bär. Bayesian approach to determine critical dimensions from scatterometric measurements. *Metrologia*, 55(6):S201–S211, oct 2018.

[7] Nando Farchmin, Martin Hammerschmidt, Philipp-Immanuel Schneider, Matthias Wurm, Bernd Bodermann, Markus Bär, and Sebastian Heidenreich. An efficient approach to global sensitivity analysis and parameter estimation for line gratings. *Proc. SPIE 11057, Modeling Aspects in Optical Metrology VII, 110570J (21 June 2019)*, October 2019.

[8] Nando Farchmin, Martin Hammerschmidt, Philipp-Immanuel Schneider, Matthias Wurm, Bernd Bodermann, Markus Bär, and Sebastian Heidenreich. Efficient bayesian inversion for shape reconstruction of lithography masks. *J. Micro/Nanolith. MEMS MOEMS 19(2), 024001 (2020)*, May 2020.

[9] Nando Farchmin. *Adaptive and Non-Intrusive Uncertainty Quantification for High-Dimensional Parametric PDEs*. Technische Universitaet Berlin (Germany), 2022.

[10] Albert Cohen and Giovanni Migliorati. Optimal weighted least-squares methods. August 2016.

[11] Nando Farchmin and Sebastian Heidenreich. Pythia uq toolbox, 12 2022.

[12] Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. 93:964–979, 2008.

[13] S. Marelli, N. Lüthen, and B. Sudret. UQLab user manual – Polynomial chaos expansions. Technical report, 2022. Report UQLab-V2.0-104.

[14] Martijn C. van Beurden. Fast convergence with spectral volume integral equation for crossed block-shaped gratings with improved material interface conditions. *Journal of the Optical Society of America A*, 28(11):2269, oct 2011.

[15] R. H. Cameron and W. T. Martin. The orthogonal development of non-linear functionals in series of fourier-hermite functionals. *The Annals of Mathematics*, 48(2):385, apr 1947.

[16] Géraud Blatman and Bruno Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6):2345–2367, mar 2011.