Automatic feature selection in EUV scatterometry

Ansuinelli, Paolo; Coene, Wim M.J.; Urbach, H. P.

applied optics

Automatic feature selection in EUV scatterometry

Paolo Ansuinelli,[1,*] Wim M. J. Coene,[1,2] and H. P. Urbach[1]

[1]*Optics Research Group, Department of Imaging Physics, Faculty of Applied Sciences, Delft University of Technology, P.O. Box 5046, 2600GA Delft, The Netherlands*
[2]*ASML Netherlands B.V., De Run 6501, 5504 DR Veldhoven, The Netherlands*
Corresponding author: p.ansuinelli@tudelft.nl

Scatterometry is an important nonimaging and noncontact method for optical metrology. In scatterometry certain parameters of interest are determined by solving an inverse problem. This is done by minimizing a cost functional that quantifies the discrepancy among measured data and model evaluation. Solving the inverse problem is mathematically challenging owing to the instability of the inversion and to the presence of several local minima that are caused by correlation among parameters. This is a relevant issue, particularly when the inverse problem to be solved requires the retrieval of a high number of parameters. In such cases, methods to reduce the complexity of the problem are to be sought. In this work, we propose an algorithm suitable to automatically determine which subset of the parameters is mostly relevant in the model, and we apply it to the reconstruction of 2D and 3D scatterers. We compare the results with local sensitivity analysis and with the screening method proposed by Morris. © 2019 Optical Society of America

1. INTRODUCTION

Extreme ultraviolet (EUV) lithography is the most promising technology for the patterning of future technology nodes. One of the challenges to be tackled for its successful implementation lies in the development of actinic mask metrology tools suitable to control and monitor the lithographic process. Scanning electron microscopy (SEM) provides high lateral resolution and can be employed for critical dimension (CD) metrology, but it has low sensitivity to 3D structure height and sidewall angle (SWA) [1]. Atomic force microscopy (AFM) could be employed for CD metrology purposes, but it suffers from low throughput as it requires scanning a probe over the entire measurement area [2]. Scatterometry is a nonimaging, noncontact method for CD and overlay metrology that has been widely employed for process control in lithography [3–7]. The usual target in scatterometry is a grating. The interaction of light–target is mathematically modeled via rigorous electromagnetic solvers [8–10], and the reconstruction of the grating proceeds either by solving an inverse problem or by assembling a multidimensional library of simulations for different possible geometries [11]. The quality of the final reconstruction depends on the quality of the measured signal and also on the measurement configuration [12]. Different sensitivity analysis methods have provided the means to find suitable sets of measurement configurations, or diffraction orders to be measured, that improve the precision of the estimation [12–14].

A relevant mathematical challenge in scatterometry lies in the development of robust algorithms for the solution of the inverse problem [15,16]. Ill-defined solutions can arise and the optimization landscape is characterized by the presence of local minima [15]. These difficulties are linked to parameter correlation, and they can be allayed by means of regularization methods [17]. Even though proper regularization can stabilize the inversion and prevent trapping in local minima, the optimization time and the overall complexity of the problem depend on the number of unknowns to be retrieved. When this is a concern, methods to reduce the complexity of the problem are to be sought. A possible approach to this challenge lies in the development of methods suitable to identify which input parameters, among the many, are the most important, and to formulate the model so as to treat only those as uncertain [18]. In this paper we present an algorithm for automatic feature selection that is a nonlinear extension of the elastic net regression [19]. Its aim is to simplify the model by removing unnecessary degrees of freedom. We apply the method to 2D targets at first, and then to 3D scatterers. For the 2D case we compare the results with the screening exercise proposed by Morris [20].

2. METHODS

We are concerned with the retrieval of certain unknown parameters of an object from measured data and with the understanding of which of these unknowns are mostly relevant for an appropriate description of the object. The purpose is to simplify the model, leaving only its most important parameters as unknowns.

One could tackle the problem using sensitivity analysis methods that "allow to study how the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input or may be used to determine the most contributing input variables to an output behavior, or to ascertain some interaction effects within the model" [21]. However, for certain applications, particularly for industrial ones, it may be preferable to make this sort of modeling decision automatic, as one often wishes to keep the user from performing any kind of mathematical operation on a certain system. In 1996, Tibshirani introduced a penalized regression, the lasso [22], which is able to automatically select the most important inputs in a model by shrinking the regression coefficients of the least relevant ones to exactly zero. Further research has demonstrated that combining the ℓ_2 and ℓ_1 norm penalties in the regression can help to overcome some limitations of the lasso [19]. Even though these algorithms have been developed to solve linear regression problems, their use can be extended to the nonlinear case. In nonlinear regression the aim is to minimize the cost functional:

$$\chi^2(\boldsymbol{p}) = \frac{||\boldsymbol{y}^{\delta} - F(\boldsymbol{p})||^2}{2\sigma^2}. \quad (1)$$

In Eq. (1), $\| \cdot \|$ denotes the Euclidean norm, $\boldsymbol{y}^{\delta}$ represents the noisy data, $F(\boldsymbol{p})$ is the model evaluation given the parameter vector $\boldsymbol{p}$, and $\boldsymbol{\sigma}$ is the vector containing the uncertainties about the measured data. We assume for the variance of the measured value [5]

$$\boldsymbol{\sigma}(\lambda)^2 = [a \cdot \boldsymbol{E}^{\pm}(\lambda)]^2 + b_g^2, \quad (2)$$

where $\boldsymbol{E}^{\pm}$ are the negative or positive diffraction efficiencies, λ denotes the wavelength, a is a constant assumed to be equal to 0.05, and b_g is the background noise of the detector, assumed to be equal to 1e-5. Nonlinear least squares problems such as Eq. (1) are commonly minimized using dedicated routines rather than general optimization methods [23]. One such method is the Gauß–Newton routine. The Gauß–Newton method can be derived by computing a first-order Taylor series expansion of $F(\boldsymbol{p})$ in Eq. (1) in the neighborhood of the current iterate. At the nth iteration, given the current estimate $\boldsymbol{p}_n$ of the parameters, the improved ones are found by moving towards a decrease direction, which is identified when solving a linear system of equations:

$$\Delta\boldsymbol{p}_n = \underset{\Delta\boldsymbol{p}_n}{\arg\min}\|\boldsymbol{y}^{\delta} - F(\boldsymbol{p}_n) - \Delta\boldsymbol{p}_n^T\boldsymbol{J}_n\|^2$$
$$\boldsymbol{p}_{n+1} = \boldsymbol{p}_n + \Delta\boldsymbol{p}_n, \quad (3)$$

where $\boldsymbol{J}_n$ is the Jacobian of $F(\boldsymbol{p}_n)$ evaluated at the current iterate and $\boldsymbol{p}_n$ is the current estimate. In a case in which a sufficiently accurate prior, $\boldsymbol{p}_0$, is available, it is possible to replace $\boldsymbol{p}_n$ in Eq. (3) with $\boldsymbol{p}_0$; alternatively, $\boldsymbol{p}_0$ can be used as a first guess. When $\boldsymbol{J}_n$ is full rank, one can solve the linear equations in Eq. (3) via the ordinary least squares estimator $\Delta\boldsymbol{p}_n = (\boldsymbol{J}_n^T\boldsymbol{J}_n)^{-1}\boldsymbol{J}_n^T\Delta F$. Prior knowledge can further be enforced by adding a Tikhonov term to penalize large deviation from the known best estimate [24].

In those cases in which a large number of parameters are to be optimized, one can penalize the incremental vector $\Delta\boldsymbol{p}_n$ in Eq. (3) with a penalty term that encourages a sparse reconstruction:

$$\boldsymbol{p}_{n+1} = \boldsymbol{p}_n + \underset{\Delta\boldsymbol{p}_n}{\arg\min}\|\Delta F - \Delta p_n^T\boldsymbol{J}\| + \gamma P_{\alpha}(\Delta\boldsymbol{p}_n), \quad (4)$$

where

$$P_{\alpha}(\Delta\boldsymbol{p}_n) = (1-\alpha)\frac{1}{2}\|\Delta\boldsymbol{p}_n\|_2^2 + \alpha\|\Delta\boldsymbol{p}_n\|_1. \quad (5)$$

In Eq. (4), $\Delta F = \boldsymbol{y}^{\delta} - F(\boldsymbol{p}_n)$, γ is the regularization parameter to be determined by seeking a balance among the data fitting term and the regularization term, $\| \cdot \|_p$, with $p = 1, 2$, is the ℓ_1 or ℓ_2 norm, and $\alpha \in [0, 1]$ is a parameter that determines the relative strength among the ℓ_2 and the ℓ_1 norm. In all that follows we have chosen $\alpha = 1/2$. The penalized regression in Eq. (4), known in the literature as the "elastic net" [19], produces a more parsimonious model via a variable selection process. The output, $\Delta\boldsymbol{p}_n$, of the regression in Eq. (4) will be a vector with some entries that can be equal to zero. By adding a zero offset to some of the entries of $\boldsymbol{p}_n$, some of the parameters will be fixed to a certain value, resulting in a reduced number of unknowns to be retrieved by the estimation routine.

A further important aspect to be considered while solving Eq. (4) lies in the appropriate selection of the regularization parameter. In this work we have applied the L–curve criterion at each iteration. As the value of γ given by the L-curve can change at each iteration, this can result in an oscillatory trend. An heuristic formula that deals with this problem is [17,25]

$$\gamma_n = \begin{cases} \epsilon\gamma_{n-1} + (1-\epsilon)\gamma & \text{if } \gamma < \gamma_{n-1} \\ \gamma_{n-1} & \text{otherwise} \end{cases}. \quad (6)$$

The algorithm described above is applicable to those deterministic inverse problems for which prior information is available and for which it is possible to compute the gradient of the function to be optimized. We stop iterating when $|\chi^2(\boldsymbol{p}_{n+1}) - \chi^2(\boldsymbol{p}_n)| < 1e-3$ [26].

It is interesting to compare the results given by the algorithm presented above with the ones given by input screening methods. Input screening is "a simplified form of sensitivity analysis that allows the user to identify the most important input quantities and potentially to reformulate the model so that only the most important input quantities are treated as uncertain" [18]. In the Morris method [20], a particularly robust screening method, one aims to classify the inputs into three different categories:

- inputs having negligible effect on the output,
- inputs having significant linear effects on the output, and
- inputs having significant nonlinear and/or cross-coupling effects.

The method proceeds by discretizing the input space spanned by the parameters of interest. One explores such space by selecting a base point in the discretization grid and perturbing successively each of the inputs by a certain incremental step to define a certain trajectory. This is done in order to compute, for each input, an incremental ratio, named the "elementary effect":

$$E_i^{(j)} = \frac{F(p_1^{(j)}, p_2^{(j)}, \ldots, p_i^{(j)} + \Delta p_i^{(j)}, \ldots, p_n^{(j)}) - F(\boldsymbol{p}^{(j)})}{\Delta p_i^{(j)}}, \quad (7)$$

where $E_i^{(j)}$ is the elementary effect associated to the ith input and to the jth trajectory. This procedure is repeated a total of R times,

with R usually equal to 10–20, for R independently generated trajectories. One synthesizes the statistics of the distributions of the elementary effects using the same estimators that would be used with independent random samples:

$$\mu_i^* = \frac{1}{R}\sum_{j=1}^{R}|E_i^{(j)}| \tag{8a}$$

$$\sigma_i = \sqrt{\frac{1}{2}\sum_{j=1}^{R}\left(E_i^{(j)} - \frac{1}{R}\sum_{j=1}^{R}E_i^{(j)}\right)^2}. \tag{8b}$$

The absolute value in Eq. (8a) is used in order to keep close-valued elements of opposite sign from canceling each other out [27]. A high mean value, μ_i^*, implies an high overall effect of the ith input over the output, and a high spread, σ_i, about the mean implies that the elementary effects relative to this factor are significantly different from each other. Hence, the value of an elementary effect is strongly affected by the choice of the point in the input space at which it is computed. This indicates an input with a nonlinear effect on the output, or an input involved in interactions with other inputs. A plot of μ^* against σ allows one to examine the computed values relative to each other and to evaluate the importance of inputs in the model. If a given input has both low μ^* and low σ values, then it has low impact on the output and it is not involved in significant nonlinear interactions. Hence, it can be dropped from the model by fixing it to a certain value within its uncertainty bounds.

There are important differences among the methods described above. This is first because they explore the input space differently. Even though the Morris design is based on the computation of small steps from one point to the other, it can be considered a "global" sensitivity analysis method as it explores the entire input space. Conversely, the regression in Eq. (4) could be thought of as a "local" method, in the sense that it is looking for a solution in the neighborhood of p_n. Another difference lies in the criteria of importance. In the Morris design, a certain input is considered important when its perturbation significantly affects the output and/or when it is involved in nonlinear effects. On the other hand, the elastic net tries to remove

unimportant inputs by solving a regularized regression problem. The metric of the variable selection algorithm is determined by a trade-off among goodness of fit and complexity of the model. Oversimplistic models fail to accurately describe the data and lead to biased solutions, while overcomplicated ones are difficult to interpret and "overfit," in the sense that they are too sensitive to the noise in the data, leading to poor generalizability and applicability over future datasets [28,29]. The lasso and elastic net are regression methods able to select a simple model, starting from a complicated one, by shrinking some of the regressors to exactly zero. These methods seek to find a simple model, selected from among many, that best captures the data. A way to further understand this is by studying the algorithms that solve the minimization problem in Eq. (4). Examples include coordinate-descent algorithms, in which the update rule, besides a scaling factor, is [30]

$$\Delta p_j = \begin{cases} \sum_{i=1}^{N} r_{ij} + \alpha\gamma & \text{if } \sum_{i=1}^{N} r_{ij} < -\alpha\gamma \\ 0 & \text{if } -\alpha\gamma < \sum_{i=1}^{N} r_{ij} < \alpha\gamma \\ \sum_{i=1}^{N} r_{ij} - \alpha\gamma & \text{if } \sum_{i=1}^{N} r_{ij} > \alpha\gamma \end{cases}. \tag{9}$$

In Eq. (9), $\sum r_{ij}$ represents the sum over the residuals of the linear regression, each weighted by a certain coefficient, obtained while fitting when excluding from the model the jth input. According to Eq. (9), a certain input is excluded from the model if its presence does not improve significantly the fitting.

In what follows we have used the following Matlab packages: regtools for the L-curve regularization [31] and "Morris SU sampling," implemented by Khare and Muñoz–Carpena at the University of Florida [32]. Rigorous electromagnetic solutions of the forward problem are computed using the finite element method solver JCMsuite [33].

3. RESULTS

A. Application to EUV Gratings
We apply the algorithm described above to the problem of feature selection for EUV gratings. Figure 1(a) presents the cross section of a grating profile. The grating is parameterized with six parameters that correspond to the X and Y coordinates of
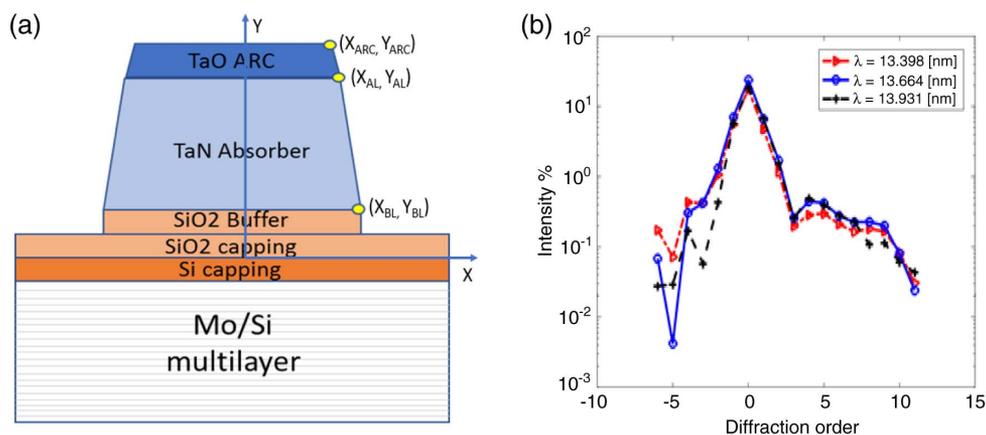

Fig. 1.   (a) Grating with parameterized profile. The independent degrees of freedom are the X and Y coordinates of the yellow points. The materials are given in Table 1. (b) Diffracted efficiencies in percentage. For the given geometry and wavelengths, only a subset of the diffraction orders can be detected.

Table 1. Layer Thicknesses and Material Properties at $\lambda = 13.5$ nm

Layer	Thickness [nm]	n	k
ARC TaO	12	0.951	0.003
Absorber TaN	54.9	0.946	0.0326
SiO$_2$ (buffer)	8	0.97352	0.01608
SiO$_2$ (oxidation)	1.246	0.97352	0.01608
Capping layer Si	12.536	0.99846	0.00184
MoSi	0.5	0.96675	0.00446
Mo	2.256	0.91872	0.00672
MoSi	0.5	0.96675	0.00446
Si	3.077	0.99846	0.00184
Substrate	6.35e6	0.97352	0.01608

the respective layers. The grating is assumed to be symmetric. The EUV radiation illuminates a Mo/Si multilayer coated reflective mask, with a patterned absorber profile on top of it. The angle of incidence is 6°, for which the multilayer is in resonance, giving a reflectance of 60%–70%. The material properties are listed in Table 1 [34]. The SWA of the SiO$_2$ layer is assumed to be equal to the SWA of the TaN layer above. The period of the grating is 420 nm, and its nominal width is 140 nm, for a line to space ratio of 1:2. For such configuration, only the diffracted orders from –6 to +11 are detectable with sufficient intensity [5]. Figure 1(b) reports the recorded diffracted intensities for the aforementioned settings and for three different wavelengths of the incoming s-polarized light field [5]: $\lambda_1 = 13.398$ nm, $\lambda_2 = 13.664$ nm, and $\lambda_3 = 13.931$ nm.

We choose our starting point for the regression, $\boldsymbol{p}_0$, by sampling a uniform prior distribution within the following intervals [18]: $X_{BL} = 70 \pm 7$ nm, $Y_{BL} = 21 \pm 5$ nm, $X_{AL} = 67 \pm 7$ nm, $Y_{AL} = 77 \pm 5$ nm, $X_{ARC} = 65 \pm 7$ nm, and $Y_{ARC} = 89 \pm 5$ nm.

In Figs. 2–4 we report the results of the presented algorithm for different starting prior vectors and different noise levels. In particular we plot:

(a) the elastic net coefficients against the strength of the regularization parameter, and

(b) the normalized local sensitivities, defined as $\sum_p \left| \frac{\partial I_p}{\partial p_i} \cdot \frac{p_i}{I_p} \right|$, where I is the computed intensity, p_i is the parameter of interest, and the summation is over the p diffraction efficiencies. They are a measure of the overall perturbation of the output due to the slight perturbation of a certain parameter.

The algorithm converges fast and, once converged, it successfully shrinks some of the entries of Δ_p to zero. However, these entries change depending on the noise level in the data and on the starting point of the optimization. In Fig. 2, the parameters that were shrunk to exactly zero were X_{AL} and Y_{AL}; in Fig. 3, the selected ones were X_{BL} and Y_{ARC}, while in Fig. 4 only Y_{BL} was exactly equal to zero. We also notice a change in the local sensitivities in Figs. 2(b), 3(b), and 4(b). Further, a comparison among Figs. 2, 3(a), and 3(b) reveals that the parameters that get fixed are not necessarily the ones for which the local sensitivity is the lowest. In other words, the parameters that locally perturb the output the most may not be the ones that the elastic net locally identifies as important for proper fitting of the data. Also, the ℓ_1 norm strongly biases toward the prior, and hence the algorithm should be used only in a start-up phase with many features, and not to carry out the estimation itself. Once the inputs are selected, the free ones can be estimated. This can be done employing the same algorithm, but retaining only the ℓ_2 norm penalty in Eq. (5). An example of such estimation, evaluated using the free parameters in Fig. 2(a), is given in Table 2, where we have approximated the covariance matrix as $(J^T \cdot J)^{-1} \sigma^2$ [26]. In what follows, we report our findings for the application of the Morris design to gratings. The input space is discretized in a 12 level grid. We have generated 1000 trajectories and have retained the 30 of them that grant the highest "spread" in the input space [27], for a total of $R \cdot (p + 1) = 210$ model evaluations. As the model produces 54 outputs—18 diffraction efficiencies per wavelength—we analyze them separately. One can then reason that if a subset of the parameters is unimportant for all of the
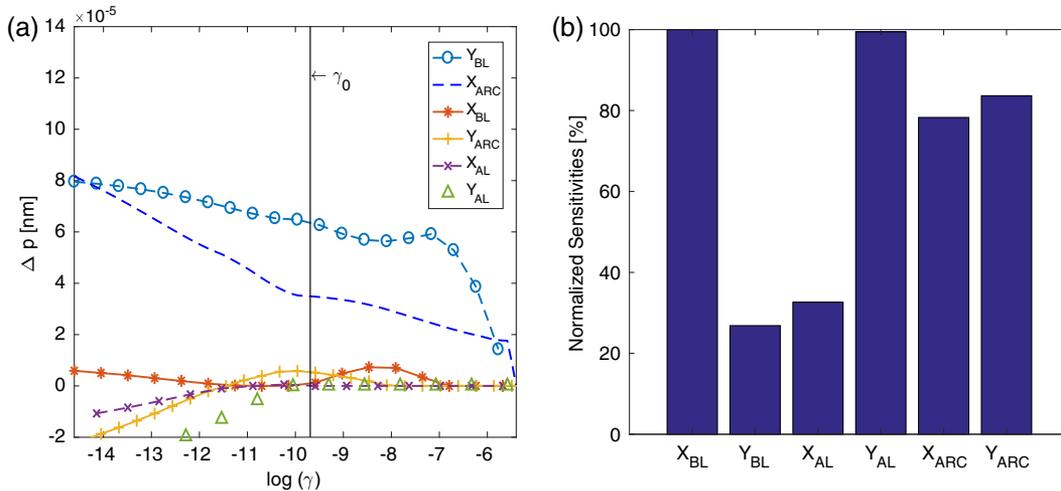

Fig. 2. Plots at the last iteration of the automatic variable selection algorithm. 5% Gaussian noise is added to the synthetic data. $\boldsymbol{p}_0$ is $[X_{BL}, Y_{BL}, X_{AL}, Y_{AL}, X_{ARC}, Y_{ARC}] = [66.9, 22.47, 73.41, 81.65, 60.21, 93.7]$ nm. (a) Elastic net coefficients as a function of regularization parameter strength. γ_0 is the regularization strength selected according to the criteria of Eq. (6). (b) Normalized local sensitivities in percentage.
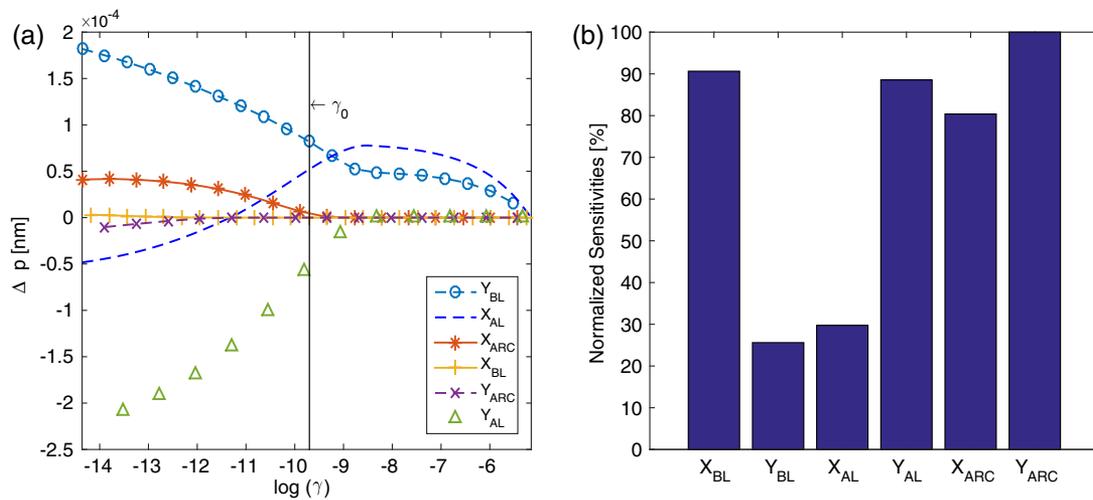
Fig. 3.   Plots at the last iteration of the automatic variable selection algorithm. 10% Gaussian noise is added to the synthetic data. $\boldsymbol{p}_0$ is $[X_{BL}, Y_{BL}, X_{AL}, Y_{AL}, X_{ARC}, Y_{ARC}] = [74.4, 26.05, 61.78, 81.13, 66.85, 84.97]$ nm. (a) Elastic net coefficients as a function of regularization parameter strength. γ_0 is the regularization strength selected according to the criteria in Eq. (6). (b) Normalized local sensitivities in percentage.

diffracted efficiencies, then it can be considered a fixed input in the model. Plots for an illustrative subset of the diffracted order at $\lambda = 13.398$ nm are shown in Fig. 4; the other two wavelengths show similar trends.

Some observation can be made about the Morris plots in Fig. 5.

• The widths of the buffer and of the absorber layers, X_{BL} and X_{AL}, respectively, which determine the CDs of the grating, and the thickness of the ARC layer, Y_{ARC}, which determines the amount of incoming power that is transmitted to the grating, are quite separated from the other inputs for most of the diffracted orders. This indicates their importance in the model.

• Y_{AL}, which determines, for a fixed thickness of the buffer layer, the thickness of the absorber, is very important for the

orders $[-3, 4]$, which are the ones that mostly propagate through the entire height of the absorber. Its importance decreases for orders diffracted at higher angles.

• X_{ARC} and Y_{BL} always appear close the origin of the plot. This indicates that they are the least important inputs in the model and can be considered fixed to a certain value within their uncertainty bounds.

• For all of the diffracted orders, some of the inputs are involved in nonlinear effects, which causes them to appear close to the diagonal in Fig. 5. It is also interesting to note that the degree of nonlinearity or correlation related to a certain input is captured by certain diffraction efficiencies rather than by others. For example, examining the plot for the order -6 [Fig. 5(a)], X_{BL} and X_{AL} appear to be involved in strong interactions or nonlinear behavior. This does not appear to be the case for the orders -1 or 4 in Figs. 5(b) and 5(c).
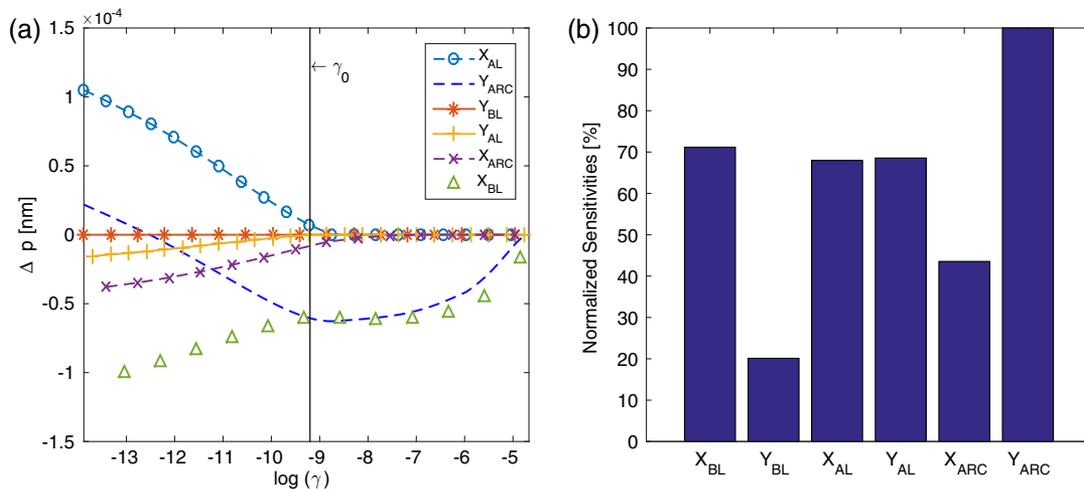

Fig. 4.   Plots at the last iteration of the automatic variable selection algorithm. 15% Gaussian noise is added to the synthetic data. $\boldsymbol{p}_0$ is $[X_{BL}, Y_{AL}, X_{AL}, Y_{AL}, X_{ARC}, Y_{ARC}] = [68.15, 23.25, 70.92, 72.81, 71, 91.7]$ nm. (a) Elastic net coefficients as a function of regularization parameter strength. γ_0 is the regularization strength selected according to the criteria in Eq. (6). (b) Normalized local sensitivities in percentage.

Table 2.   Reconstruction Results

Parameter	Reconstructed Value [nm]	Standard Deviation [nm]
X_{BL}	67.12	0.8
Y_{BL}	23.65	1.8
X_{ARC}	61.67	0.4
Y_{ARC}	87.85	0.3

The observations above are consistent with previous modeling work [5], in which the authors have retained in the model only those parameters that identified top and bottom CDs and the SWA of the grating. However, according to the Morris design, Y_{ARC} should be considered as a free degree of freedom rather than be fixed.

As the Morris design and the elastic net penalty rank the importance of parameters according to different criteria, and as they cover the input space differently, they lead to dissimilar results. For instance, in Fig. 2(a) the elastic net penalty shrank to zero the inputs X_{AL} and Y_{AL}. This, according to the Morris design, would have deprived the model of two important inputs.

In light of this, a better strategy could be to remove for the model those parameters that are identified as unimportant by both the Morris design and the penalized regression in Eq. (4). For example, for the case in Fig. 4(a), one could fix only Y_{BL}. In this way one would retain in the model those parameters that are important for proper fitting of the data and that, at the same time, have a substantial effect over the output.

B. Application to 3D Scatterers

It is interesting to apply the method developed in Section 2 to the complex case of feature selection for 3D isolated nanostructures. The model-based approach has been investigated predominantly for 2D grating profiles and 3D periodic scatterers, but its use for the reconstruction of isolated nanostructures is still to be discussed. The modeling of a 3D nanostructure is challenging, and understanding how to parametrize a given structure and which features to retain in the model is difficult. In such cases, the tools presented above can be particularly useful. We apply the algorithm described in Section 2 to the scatterer in Fig. 6(a), which is parameterized with seven parameters. We fit the diffuse scattered intensities, displayed in Fig. 6(b).

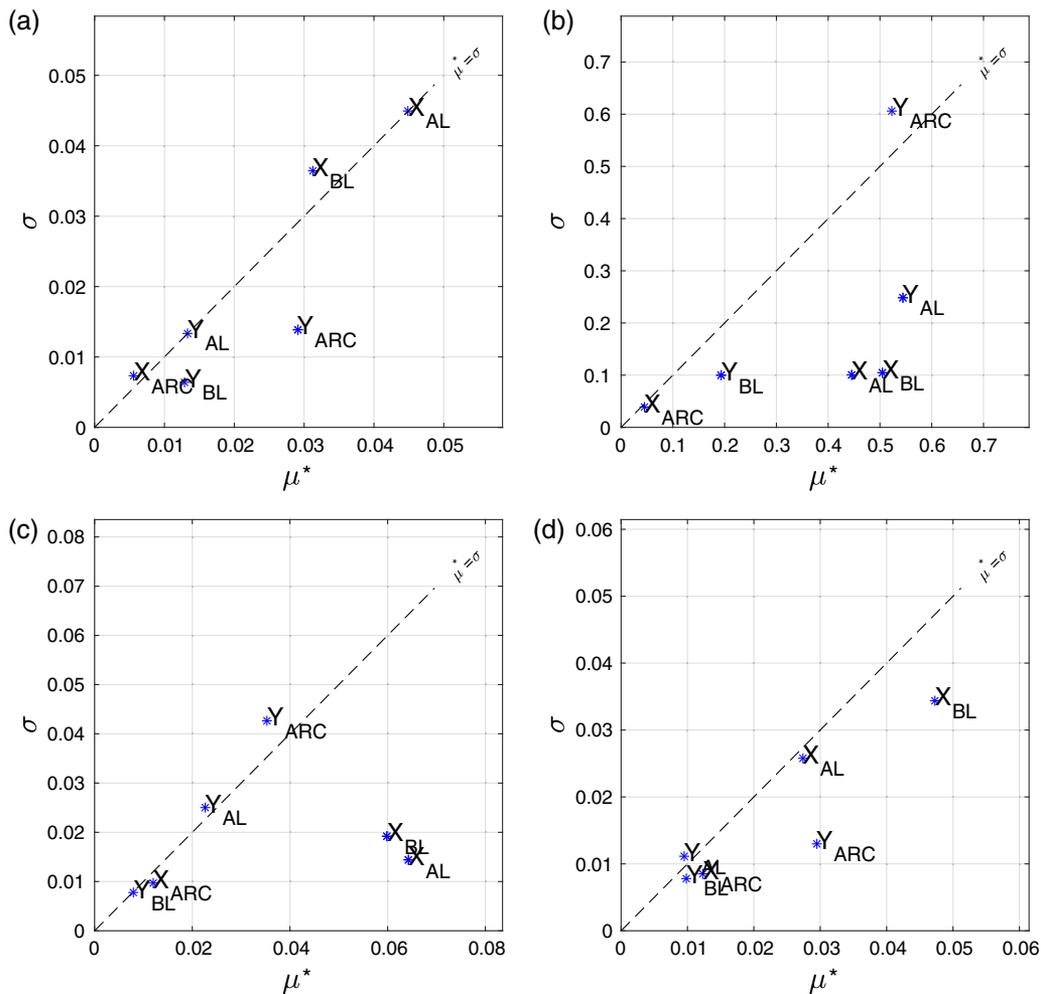For the 3D scatterer we use the following data.

Fig. 5. Morris plots for four different diffracted orders at $\lambda = 13.398$ nm: (a) order −6, (b) order −1, (c) order 4, and (d) order 9.
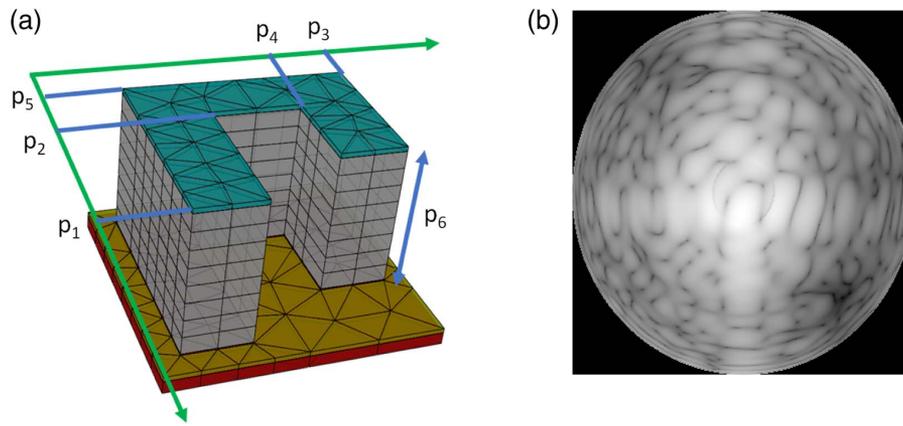
Fig. 6. (a) Scatterer with parameterized profile. The parameter p7, not indicated in the figure, is the thickness of the anti-reflective layer. (b) Diffuse scattering given by the structure in (a).
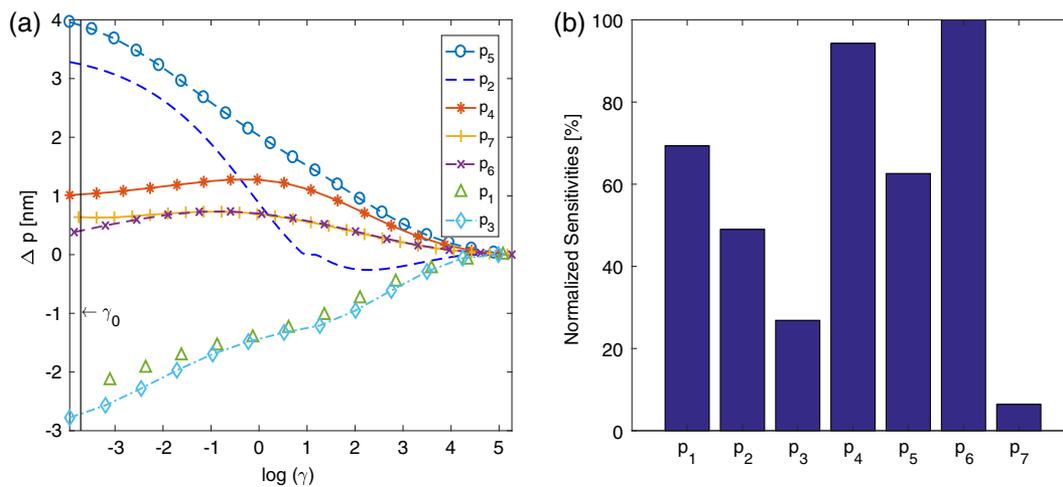

Fig. 7. (a) Elastic net coefficients as a function of the regularization parameter strength. (b) Normalized local sensitivities in percentage.

We replace the multilayer with an equivalent substrate that offers, for the given wavelength and angles of incidence, approximately the same reflectance. The incoming light field is a beam with a diameter of about 2 μm and radiating 5e11 photons/s. The detection NA is 0.5. The computational domain is truncated on all sides by the perfectly matched layers [35]. The meshing setting is such to have a relative error in the far-field evaluation of about 1%.

Figure 7 reveals that all of the parameters should be kept in the model for a proper fitting of the data, even though the contribution of p_6 and p_7 is quite limited compared to the others. A thorough study of the applicability of model-based reconstruction in 3D aperiodic case, and related modeling work, is beyond the purpose of this paper.

4. CONCLUSIONS

There are applications in which one is interested in the retrieval of unknown characteristics of an object that are enclosed within a measured signal. This is a challenging mathematical problem that suffers from ill-posedness. The use of prior information about some of these characteristics allays some difficulties, providing the means to stabilize the inversion and to look for a solution that is a deviation about the given prior. Nevertheless, the complexity of the problem depends upon the number of unknowns to be retrieved. When this is an issue, methods to reduce the complexity are to be sought. In this paper we have proposed an algorithm that is a nonlinear extension of the elastic net regression [19]. Its purpose is to identify which inputs do not contribute much to improve the fitting and to fix them to a certain value, reducing the number of unknowns to be retrieved. The algorithm can be applied to that class of deterministic inverse problems in which one can compute the gradient of the function to be optimized. We have compared the method with the Morris design [20] and with local sensitivity analysis. The comparison demonstrates that the two methods, which discern important parameters according to different viewpoints and which explore the input space differently,

Table 3. Layer Thicknesses and Material Properties at $\lambda = 13.5$ nm

Layer	Thickness [nm]	n	k
ARC TaBO	2	0.952	0.026
Absorber TaBN	58	0.95	0.031
Ru	0.5	0.88586	0.01727
Ru (capping layer)	2	0.88586	0.01727
Si	1.8968	0.99888	0.00183
$MoSi_2$	0.7986	0.96908	0.00435
Mo	2.496	0.92347	0.00649
$MoSi_2$	1.8908	0.96908	0.00435

can give different results. In view of those differences, a more robust approach consists of a joint decision that combines the results given by the methods. We have shared and discussed our findings, applying the methods to the inverse problem of EUV scatterometry.

REFERENCES

1. H. J. Wonsuk Lee and S. H. Han, "Measurement of critical dimension in scanning electron microscope mask images," J. Micro/Nanolithogr., MEMS, MOEMS **10**, 1–8 (2011).
2. G. Dahlen, M. Osborn, H.-C. Liu, R. Jain, W. Foreman, and J. R. Osborne, "Critical dimension AFM tip characterization and image reconstruction applied to the 45-nm node," Proc. SPIE **6152**, 61522R (2006).
3. H.-T. Huang and F. Terry, "Spectroscopic ellipsometry and reflectometry from gratings (scatterometry) for critical dimension measurement and in situ, real-time process monitoring," Thin Solid Films **455**-456, 828–836 (2004).
4. C. J. Raymond, M. R. Murnane, S. L. Prins, S. Sohail, H. Naqvi, J. R. McNeil, and J. W. Hosch, "Multiparameter grating metrology using optical scatterometry," J. Vac. Sci. Technol. B **15**, 361–368 (1997).
5. H. Gross, A. Rathsfeld, F. Scholze, and M. Bär, "Profile reconstruction in extreme ultraviolet (EUV) scatterometry: modeling and uncertainty estimates," Meas. Sci. Technol. **20**, 105102 (2009).
6. N. Kumar, P. Petrik, G. K. P. Ramanandan, O. E. Gawhary, S. Roy, S. F. Pereira, W. M. J. Coene, and H. P. Urbach, "Reconstruction of sub-wavelength features and nano-positioning of gratings using coherent Fourier scatterometry," Opt. Express **22**, 24678–24688 (2014).
7. Y.-S. Ku, C.-L. Yeh, Y.-C. Chen, C.-W. Lo, W.-T. Wang, and M.-C. Chen, "EUV scatterometer with a high-harmonic-generation EUV source," Opt. Express **24**, 28014–28025 (2016).
8. J. Chandezon, G. Raoult, and D. Maystre, "A new theoretical method for diffraction gratings and its numerical application," J. Opt. **11**, 235–241 (1980).
9. P. Lalanne, "Convergence performance of the coupled-wave and the differential methods for thin gratings," J. Opt. Soc. Am. A **14**, 1583–1591 (1997).
10. G. Bao, "Finite element approximation of time harmonic waves in periodic structures," SIAM J. Numer. Anal. **32**, 1155–1169 (1995).
11. X. Chen, S. Liu, C. Zhang, and H. Jiang, "Improved measurement accuracy in optical scatterometry using correction-based library search," Appl. Opt. **52**, 6726–6734 (2013).

12. Z. Dong, S. Liu, X. Chen, and C. Zhang, "Determination of an optimal measurement configuration in optical scatterometry using global sensitivity analysis," Thin Solid Films **562**, 16–23 (2014).
13. H. Gross and A. Rathsfeld, "Sensitivity analysis for indirect measurement in scatterometry and the reconstruction of periodic grating structures," Waves Random Complex Media **18**, 129–149 (2008).
14. P. C. Logofătu, "Sensitivity analysis of grating parameter estimation," Appl. Opt. **41**, 7179–7186 (2002).
15. M.-A. Henn, H. Gross, F. Scholze, M. Wurm, C. Elster, and M. Bär, "A maximum likelihood approach to the inverse problem of scatterometry," Opt. Express **20**, 12771–12786 (2012).
16. J. Zhu, S. Liu, X. Chen, C. Zhang, and H. Jiang, "Robust solution to the inverse problem in optical scatterometry," Opt. Express **22**, 22031–22042 (2014).
17. A. Doicu, T. Trautmann, and F. Schreier, Numerical Regularization for Atmospheric Inverse Problems (Springer, 2010).
18. K. Rasmussen, J. B. Kondrup, A. Allard, S. Demeyer, N. Fischer, E. Barton, D. Partridge, L. Wright, M. Bär, H. G. A. Fiebach, S. Heidenreich, M.-A. Henn, R. Model, S. Schmelter, G. Kok, and N. Pelevic, "Novel mathematical and statistical approaches to uncertainty evaluation: best practice guide to uncertainty evaluation for computationally expensive models," Tech. rep. (Euramet, 2015).
19. H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," J. R. Statist. Soc. B **67**, 301–320 (2005).
20. M. D. Morris, "Factorial sampling plans for preliminary computational experiments," Technometrics **33**, 161–174 (1991).
21. B. Iooss and P. Lemaître, "A review on global sensitivity analysis methods," in Uncertainty Management in Simulation–Optimization of Complex Systems: Algorithms and Applications, G. Dellino and C. Meloni, eds. (Springer, 2015), Chap. 5, pp. 101–122.
22. R. Tibshirani, "Regression shrinkage and selection via the lasso," J. R. Stat. Soc. Ser. B **58**, 267–288 (1996).
23. K. Madsen, H. B. Nielsen, and O. Tingleff, Methods for Non-Linear Least Squares Problems, 2nd ed. (2004).
24. K. Levenberg, "A method for the solution of certain non-linear problems in least squares," Quart. Appl. Math. **2**, 164–168 (1944).
25. J. Eriksson, "Optimization and regularization of nonlinear least squares problems," Ph.D. thesis, (Dept. of Computing Science, Umea University, Umea, Sweden, 1996).
26. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in C: The Art of Scientific Computing, 2nd ed. (Cambridge University, 1992).
27. F. Campolongo, J. Cariboni, and A. Saltelli, "An effective screening design for sensitivity analysis of large models," Environ. Model. Software **22**, 1509–1518 (2007).
28. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Series in Statistics (Springer, 2009).
29. C. Vogel, Computational Methods for Inverse Problems (Society for Industrial and Applied Mathematics, 2002).
30. J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," J. Stat. Softw. **33**, 1–22 (2009).
31. P. C. Hansen, "Regularization tools—a Matlab package for analysis and solution of discrete ill-posed problems," Numer. Algorithms **6**, 1–35 (1994).
32. Y. Khare and R. Muñoz-Carpena, "Global sensitivity analysis: elementary effects method of Morris using sampling for uniformity (SU) Matlab code manual," 2014 https://abe.ufl.edu/faculty/carpena/software/SUMorris.shtml.
33. https://jcmwave.com.
34. B. Bodermann, M. Wurm, A. Diener, F. Scholze, and H. Gross, "EUV and DUV scatterometry for CD and edge profile metrology on EUV masks," in 25th European Mask and Lithography Conference (2009), pp. 1–12.
35. J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200 (1994).