Optics EXPRESS

A maximum contributed component regression for the inverse problem in optical scatterometry

HAIPING ZHU,[1] YOUNGJOO LEE,[2] HONGMING SHAN,[1] AND JUNPING ZHANG[1,*]

[1] *School of Computer Science, Shanghai Key Lab of Intelligent Information Processing, Fudan University, Shanghai, China*
[2] *Manufacturing Technology Center, Samsung Electronics, Suwon, South Korea*
[*] *jpzhang@fudan.edu.cn*

Abstract: Scatterometry has been widely applied in microelectronic manufacturing process monitoring. As a key part in scatterometry, inverse problem uses scatter signature to determine the shape of profile structure. The most common solutions for the inverse problem are model-based methods, such as library search, Levenberg-Marquardt algorithm and artificial neural network (ANN). However, they all require a pre-defined geometric model to extract 3D profile of the structure. When facing the complex structure in manufacturing process monitoring, the model-based methods will cost a long time and may fail to build a valid geometric model. Without the assumption of the geometric model, model-free methods are developed to find a mapping between profile parameter named label Y and corresponding spectral signature X. These methods need lots of labeled data obtained from transmission electron microscopy (TEM) or cross-sectional scanning electron microscopy (XSEM) with time-consuming and highly cost, leading to the increase of production costs. To address these issues, this paper develops a novel model-free method, called *maximum contributed component regression* (MCCR). It utilizes canonical correlation analysis (CCA) to estimate the maximum contributed components from pairwise relationship of economic unlabeled data with few expensive labeled data. In MCCR, the maximum contributed components are used to guide the solution of the inverse problem based on the conventional regression methods. Experimental results on both synthetic and real-world semiconductor datasets demonstrate the effectiveness of the proposed method given small amount of labeled data.

OCIS codes: (120.3940) Metrology; (050.1950) Diffraction gratings; (290.3200) Inverse scattering.

References and links

1. G. Binnig, C. F. Quate, and C. Gerber, "Atomic force microscope," Phys. Rev. Lett. **56**, 930 (1986).
2. A. Kato, Y. Ikeda, Y. Kasahara, J. Shimanuki, T. Suda, T. Hasegawa, H. Sawabe, and S. Kohjiya, "Optical transparency and silica network structure in cross-linked natural rubber as revealed by spectroscopic and three-dimensional transmission electron microscopy techniques," JOSA B **25**, 1602–1615 (2008).
3. Y. Lepetre, J.-J. Metois, G. Rasigni, R. Rivoira, and R. Philip, "Characterization of layered synthetic microstructures using transmission electron microscopy," JOSA A **2**, 1356–1362 (1985).
4. W. Yang, E. Bricchi, P. G. Kazansky, J. Bovatsek, and A. Y. Arai, "Self-assembled periodic sub-wavelength structures by femtosecond laser direct writing," Opt. Express **14**, 10117–10124 (2006).
5. F. Qin, Z.-M. Meng, X.-L. Zhong, Y. Liu, and Z.-Y. Li, "Fabrication of semiconductor-polymer compound nonlinear photonic crystal slab with highly uniform infiltration based on nano-imprint lithography technique," Opt. Express **20**, 13091–13099 (2012).
6. 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. A **15**, 361–368 (1997).
7. X. Niu, N. Jakatdar, J. Bao, and C. J. Spanos, "Specular spectroscopic scatterometry," IEEE Trans. Semicond. Manuf. **14**, 97–111 (2001).
8. H.-T. Huang and F. L. Terry Jr, "Spectroscopic ellipsometry and reflectometry from gratings (scatterometry) for critical dimension measurement and in situ, real-time process monitoring," Thin Solid Films **455**, 828–836 (2004).
9. C. Raymond, "Overview of scatterometry applications in high volume silicon manufacturing," Characterization and Metrology for ULSI Technology 2005 **788**, 394–402 (2005).
10. 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).

#285566 https://doi.org/10.1364/OE.25.015956
Journal © 2017 Received 3 Feb 2017; revised 27 May 2017; accepted 23 Jun 2017; published 28 Jun 2017

11. N. Kumar, P. Petrik, G. K. Ramanandan, O. El Gawhary, S. Roy, S. F. Pereira, W. M. 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).
12. S. Liu, W. Du, X. Chen, H. Jiang, and C. Zhang, "Mueller matrix imaging ellipsometry for nanostructure metrology," Opt. Express **23**, 17316–17329 (2015).
13. R. Krukar, A. Kornblit, L. A. Clark, J. Kruskal, D. Lambert, E. A. Reitman, and R. A. Gottscho, "Reactive ion etching profile and depth characterization using statistical and neural network analysis of light scattering data," J. Appl. Phys. **74**, 3698–3706 (1993).
14. S. Naqvi, J. Franke, D. Haaland, R. Gottscho, A. Kornblit, T. Niemczyk, R. Krukar, and J. McNeil, "Etch depth estimation of large-period silicon gratings with multivariate calibration of rigorously simulated diffraction profiles," JOSA A **11**, 2485–2493 (1994).
15. J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Numer. Anal.* (Springer, 1978), pp. 105–116.
16. M. Hanke, "A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems," Inv. Prob. **13**, 79 (1997).
17. I. Gereige, S. Robert, S. Thiria, F. Badran, G. Granet, and J. J. Rousseau, "Recognition of diffraction-grating profile using a neural network classifier in optical scatterometry," JOSA A **25**, 1661–1667 (2008).
18. I. Kallioniemi, J. Saarinen, and E. Oja, "Optical scatterometry of subwavelength diffraction gratings: neural-network approach," Appl. Opt. **37**, 5830–5835 (1998).
19. S. Robert, A. Mure-Ravaud, and D. Lacour, "Characterization of optical diffraction gratings by use of a neural method," JOSA A **19**, 24–32 (2002).
20. J. Zhu, S. Liu, C. Zhang, X. Chen, and Z. Dong, "Identification and reconstruction of diffraction structures in optical scatterometry using support vector machine method," J. Micro/Nanolithography, MEMS, and MOEMS **12**, 013004.1–013004.10 (2013).
21. 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).
22. Y.-N. Kim, J.-S. Paek, S. Rabello, S. Lee, J. Hu, Z. Liu, Y. Hao, and W. McGahan, "Device based in-chip critical dimension and overlay metrology," Opt. Express **17**, 21336–21343 (2009).
23. N. F. Zhang, R. M. Silver, H. Zhou, and B. M. Barnes, "Improving optical measurement uncertainty with combined multitool metrology using a bayesian approach," Appl. Opt. **51**, 6196–6206 (2012).
24. H. Hotelling, "Analysis of a complex of statistical variables into principal components," J. Edu. Psychol. **24**, 417–441 (1933).
25. P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," Anal. Chim. Acta **185**, 1–17 (1986).
26. B. Thompson, "Canonical correlation analysis," Encyclopedia of Statistics in Behavioral Science (2005).
27. B. McWilliams, D. Balduzzi, and J. M. Buhmann, "Correlated random features for fast semi-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.* (2013), pp. 440–448.
28. S. M. Kakade and D. P. Foster, "Multi-view regression via canonical correlation analysis," in *Learning theory* (Springer, 2007), pp. 82–96.
29. K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn.* (ACM, 2009), pp. 129–136.
30. H. Fujiwara, *Spectroscopic ellipsometry: principles and applications* (John Wiley & Sons, 2007).

1. Introduction

In microelectronic manufacturing process, measuring 3D profiles of the structures of each profile component is extremely important since it directly influences the quality of microelectronic products. By analyzing light scattered from a periodic structure made during the wafer fabrication process, named as optical signature, optical metrology techniques can indirectly measure the 3D profiles. Different from traditional imaging techniques such as atomic force microscopy (AFM) [1], transmission electron microscopy (TEM) [2,3], and cross-sectional scanning electron microscopy (XSEM) [4,5], which are destructive, slow and expensive measuring equipment in the microelectronic manufacturing industry, scatterometry [6–8] is rapid, inexpensive, quantitative and non-destructive, making it become an attractive optical metrology technique. The scatterometry can be further decomposed into the forward problem, a process by which a scatter signature is measured, and the inverse problem [9], a process by which the geometry such as thickness, critical dimension (CD) [10–12], and angle of the scattering structure is deduced from the measured scattering signature.

To solve the inverse problem, the most common strategies are model-based analysis, which compares the measured optical scatter data to a simulated model [9]. For examples, discriminant analysis [13] and partial-lease-square method (PLS) [14] are useful under linear assumption. However, the linearized method has its inherent limitation due to the highly nonlinear relationship between the geometric structure and scattered light. Therefore, some nonlinear regression methods, such as the Levenberg-Marquardt algorithm [15, 16] and the artificial neural networks (ANN) [17–19] have been developed. For instance, library search [20, 21] generates a signature library prior to the measurement, and searches in the library to find a best match with the measured signature. All of these model-based methods require a pre-defined geometric model to extract 3D profile of the structure. Such a model can work well when handling a sort of 2D simple structure such as line and space, often seen in the laboratory environment. However, its performance will be greatly degenerated in manufacturing process monitoring as the structure of memory and logic device becomes more complex. Consequently, it takes a long time or even fail to generate a valid geometric model because it is impossible to simulate all combinations of parameters and hard to get optimal solution in complex structure. For example, in the semiconductor manufacturing process, there are dozens of channel holes of 3D NAND flash memory and we are interested in measuring of depth of those holes. The shape (e.g. CD, circularity, and bending angle) of each holes are not homogeneous because each process control is extremely difficult, and it has a significant effect on the spectral signature. Therefore, it is required to include several profile parameters of every single hole in the geometric model to get accurate depth of channel holes. However, it is impossible to build a library and fails in regression because there are too many floating parameters and it leads to curse of dimensionality. Under this case, valid geometric modeling for model-based method is not available. Moreover, we observe that for the model-based methods, the number of parameters is usually not more than five in academic research and not more than ten in industry applications in literature [6, 17, 22, 23].

Without the assumption of a geometric model, model-free methods are to find a mapping function between profile parameter, named label Y, and the corresponding spectral signature X. To restore this mapping function, it is necessary to label lots of data based on their profile parameters and corresponding spectral signatures. In general, collecting labeled data is time-consuming and costly, since it needs to use destructive imaging equipments such as TEM or XSEM. Therefore, the number of labeled data is usually exceptionally limited for the inverse problem in manufacturing process. Generally speaking, two model-free methods, principal component regression (PCR) [24] and PLS [25], can be employed to solve the inverse problem because of their effectiveness and good accuracy.

However, PLS is easy to become over-fitting when the number of labeled data and corresponding signatures are limited, as this model needs a lot of labels to train well. Similarly, PCR can select components that are not more than the number of labeled data components to fit a multiple linear regression (MLR). If the number of selected components is smaller than that of labeled data given limited labeled data, PCA cannot guarantee the selected components contain the first several highest contributed components, which represent the highest correlation between the dataset X and the reference values Y.

In this paper, we developed a novel model-free method, named as *maximum contributed component regression* (MCCR), to overcome small sample size issue. Specifically, a set of limited labeled data is used to train PCR to provide a predicted label vector for each unlabeled sample, which is economic and easy to obtain. According to the order of their predicted labels, the labeled data and unlabeled data are sorted together, followed by being divided into two disjoint groups equally based on their respective orders. Assuming that two disjoint groups are two views of data distribution, canonical correlation analysis (CCA) [26, 27] is utilized to estimate a common projection subspace. Since the subspace is obtained based on the pairwise information from two groups, it is more effective to help the subsequent regression model, PCR, to achieve higher prediction performance under this projection subspace. Experiments indicate that the proposed MCCR is effective, efficient and economic in both simulated and real-world semiconductor spectrum data given a limited collection of labeled data.

The remainder of this paper is organized as follows. Section 2 introduces analysis of the problem and details of the proposed method MCCR for the inverse problem in optical scatterometry. Section 3 presents the data description, evaluation and experimental results. Then, we draw some conclusions in Section 4.

2. The proposed method

In this section, we will analyze the issue of the limited number of labeled data, and introduce the proposed MCCR method in details.

2.1. Analysis

Traditional model-free methods usually perform unfavorable on limited labeled data. As a commonly used model-free method, PCR is a simple linear regression based on principal component analysis (PCA), which figures out the principal components by maximizing the variance of data. To avoid over-fitting, PCR selects the components that are not more than the number of labeled data to fit the linear regression. One disadvantage for this is that the chosen components from PCA will be insufficient and biased when the number of labeled data is limited. Furthermore, these selected components may not reflect or may deviate from the highest contributed components, which represent the highest correlation between the dataset X and the reference values Y. Assuming there are 100 samples but only 10 of them are labeled, each of which has 132 dimensions. PCA will select the dimensions that are not more than 10 because of the rank limitation of matrix decomposition. As a result, PCA may not capture all of the highest 10 contributed components in the first 10 principal components, as shown in the Fig. 1(a), leading to low prediction accuracy for the subsequent regression model. To address this problem, a more suitable projection subspace should be estimated, as shown in the Fig. 1(b). Fortunately, CCA could approximately achieve this purpose if we know the order of all data.

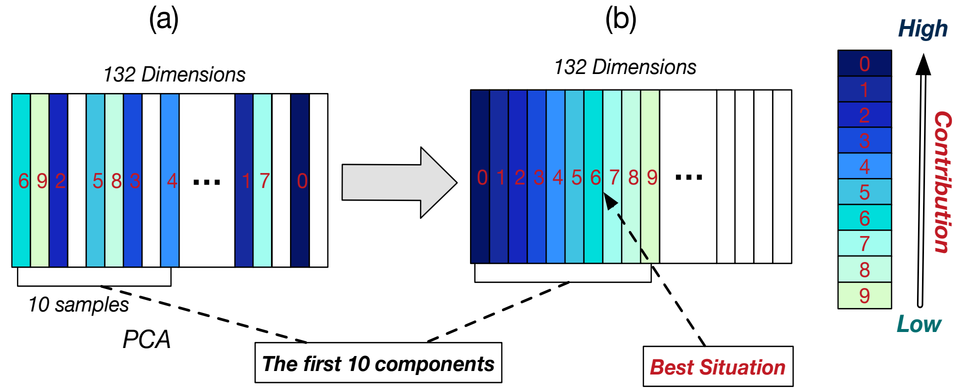


Fig. 1. Suppose there are only 10 samples, each of which has 132 dimensions. The components for data with deeper blue colors have higher contributions in regression. (a) shows the principal components from PCA; (b) shows the high contributed components in best situation.

CCA calculates two projection vectors for two heterogeneous datasets that maximize the linear correlation between these two datasets by mapping them into a common subspace [28,29]. Under this common subspace, both sets can help the predictor achieve better prediction performance compared with using only one single set. The more the number of pair-wise data is ordered in two groups according to their labels either monotonically increased or decreased, furthermore, the better predictive performance CCA can reach. If the data are completely ordered, ideally, CCA can discover the best projection components for maximizing the prediction performance, named as *the best contributed component regression* (BCCR). However, such an ideal situation is almost infeasible in practice because labeling all the spectral data by TEM or XSEM is exceptionally expensive. Fortunately, it is economic to obtain lots of unlabeled data from optical scatterometry, and we expect to utilize these economic unlabeled data to refine the predictive performance of regression. However, such homogeneous and unordered data can only provide the estimate of principal components, which is less helpful to improve the predictive performance of regression. To address this issue, we employ the limited labeled data to pre-train a simple regression model so that each unlabeled sample can be endowed with an 'approximately corrected' label. With this way, CCA can utilize these data to attain a better projection subspace. To the best of our knowledge, it is the first time to utilize CCA with a pre-training technique to capture the information of unlabeled data for the inverse problem in optical scatterometry.

2.2. Maximum contributed component regression

In this subsection, we would like to detail our proposed MCCR algorithm. Let labeled dataset be $\mathcal{L} = (\boldsymbol{X}_L, \boldsymbol{Y}_L) = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n_l}$ and unlabeled dataset be $\mathcal{U} = (\boldsymbol{X}_U) = \{\boldsymbol{x}_j\}_{j=1}^{n_u}$, where samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ reside in a d-dimensional space and y_i is the labeled value corresponding to the sample $\boldsymbol{x}_i$, n_l and n_u denote the number of labeled data and unlabeled data, respectively.

First of all, a regression model is pre-trained to label the unlabeled data $\mathcal{L}$, followed by separating the data into two disjointed data according to the order of their labels. In this paper, PCR is chosen as the pre-training model because of its effectiveness, simple and easy to implement. Specifically, we first employ PCA on the original dataset $\boldsymbol{X}_L = \{\boldsymbol{x}_i\}_{i=1}^{n_l}$ to select the first m principal components, followed by conducting regression to predict the possible labels of those unlabeled data. More concretely, assume $\boldsymbol{P}_0 \in \mathcal{R}^{d\times m}$ is the projection matrix obtained from PCA, and $\boldsymbol{x} = (x_1, x_2, \cdots, x_d)^T$ is a d-dimensional sample, then the corresponding projected sample is $\boldsymbol{z} = \boldsymbol{P}_0^T\boldsymbol{x} = (z_1, z_2, \cdots, z_m)^T$. Here multiple linear regression (MLR) model we used is given as follows,

$$y = \beta_0 + \beta_1 \times z_1 + \beta_2 \times z_2 + \cdots + \beta_m \times z_m, \tag{1}$$

where y is the predictor of profile parameter value, $(\beta_0,\beta_1, \cdots, \beta_m)^T$ is the regression coefficient vector, respectively. Let $\boldsymbol{\beta} = (\beta_0,\beta_1, \cdots, \beta_m)^T$, $\boldsymbol{z} = (1, z_1, z_2, \cdots, z_m)^T$, then Eq. (1) is simplified as $y = \boldsymbol{\beta}^T\boldsymbol{z}$. The coefficient vector $\boldsymbol{\beta}$ will be estimated as follow,

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \frac{1}{n_l}\sum_{i=1}^{n_l}(\boldsymbol{\beta}^T\boldsymbol{z}_i - y_i)^2. \tag{2}$$

where n_l denotes the number of labeled data, $\boldsymbol{z}_i$ denotes the projected variable of the i-th sample $\boldsymbol{x}_i$, and y_i denotes the corresponding profile parameter value.

Secondly, after the coefficient vector $\boldsymbol{\beta}$ is estimated, the predicted vector $\widehat{\boldsymbol{Y}}_U$ of unlabeled dataset $\boldsymbol{X}_U$ can be estimated from Eq. (1). After being sorted ascendingly according to $(\boldsymbol{Y}_L \cup \widehat{\boldsymbol{Y}}_U)$, the merging dataset $(\boldsymbol{X}_L \cup \boldsymbol{X}_U)$ are equally divided into two disjoint groups, denoted as $\boldsymbol{G}_1 \in \mathcal{R}^{k\times d}$ and $\boldsymbol{G}_2 \in \mathcal{R}^{k\times d}$, where $k = (n_l + n_u)/2$.

Thirdly, CCA is employed to compute two projection vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ such that the correlation coefficient between $\boldsymbol{w}_1^T\boldsymbol{G}_1$ and $\boldsymbol{w}_2^T\boldsymbol{G}_2$ is maximized. Each pair of the data from two groups has the same dimensionality, i.e., $\boldsymbol{w}_1 \in \mathbb{R}^{d\times 1}$ and $\boldsymbol{w}_2 \in \mathbb{R}^{d\times 1}$. The correlation coefficient ρ between $\boldsymbol{w}_1^T\boldsymbol{G}_1$ and $\boldsymbol{w}_2^T\boldsymbol{G}_2$ is calculated by,

$$\rho = \frac{\boldsymbol{w}_1^T\boldsymbol{G}_1\boldsymbol{G}_2^T\boldsymbol{w}_2}{\sqrt{(\boldsymbol{w}_1^T\boldsymbol{G}_1\boldsymbol{G}_1^T\boldsymbol{w}_1)(\boldsymbol{w}_2^T\boldsymbol{G}_2\boldsymbol{G}_2^T\boldsymbol{w}_2)}}. \tag{3}$$

Multiple projections of CCA can be simultaneously calculated by solving the following optimization function with two constraint terms:

$$\begin{aligned} &\max_{\boldsymbol{W}_1,\boldsymbol{W}_2} \operatorname{Tr}\Big[\boldsymbol{W}_1^T\boldsymbol{G}_1\boldsymbol{G}_2^T\boldsymbol{W}_2\Big], \\ &\text{s.t.} \quad \boldsymbol{W}_1^T\boldsymbol{G}_1\boldsymbol{G}_1^T\boldsymbol{W}_1 = \boldsymbol{I}, \quad \boldsymbol{W}_2^T\boldsymbol{G}_2\boldsymbol{G}_2^T\boldsymbol{W}_2 = \boldsymbol{I}, \end{aligned} \tag{4}$$

where each column of matrices $\boldsymbol{W}_1 \in \mathbb{R}^{d\times r}$ and $\boldsymbol{W}_2 \in \mathbb{R}^{d\times r}$ corresponds to a projection vector and r ($r \le d$) is the number of projection vectors to be computed. Assume that $\boldsymbol{G}_2\boldsymbol{G}_2^T$ and $\boldsymbol{G}_1\boldsymbol{G}_1^T$ are nonsingular, the projection matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are attained by calculating the first r principal eigenvectors of the following generalized eigenvalue problem, respectively:

$$\begin{cases} \boldsymbol{G}_1\boldsymbol{G}_2^T(\boldsymbol{G}_2\boldsymbol{G}_2^T)^{-1}\boldsymbol{G}_2\boldsymbol{G}_1^T\boldsymbol{w}_1 = \lambda\boldsymbol{G}_1\boldsymbol{G}_1^T\boldsymbol{w}_1, \\ \boldsymbol{G}_2\boldsymbol{G}_1^T(\boldsymbol{G}_1\boldsymbol{G}_1^T)^{-1}\boldsymbol{G}_1\boldsymbol{G}_2^T\boldsymbol{w}_2 = \lambda\boldsymbol{G}_2\boldsymbol{G}_2^T\boldsymbol{w}_2, \end{cases} \tag{5}$$

where λ is the corresponding eigenvalue. To get a common projection from both $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, the target projection matrix $\boldsymbol{W} \in \mathbb{R}^{d\times 2r}$ is concatenated as,

$$\boldsymbol{W} = [\boldsymbol{W}_1, \boldsymbol{W}_2]. \tag{6}$$

Once the projection matrix $\boldsymbol{W}$ is obtained, the spectra data are transformed from original variables $\mathcal{L} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n_l}$ to new variables $\widehat{\mathcal{L}} = \{(\boldsymbol{W}^T\boldsymbol{x}_i, y_i)\}_{i=1}^{n_l}$.

Finally, we train a target PCR model on the projected dataset $\widehat{\mathcal{L}} = \{(\boldsymbol{W}^T\boldsymbol{x}_i, y_i)\}_{i=1}^{n_l}$. Similarly, assume $\boldsymbol{P}_t \in \mathcal{R}^{2r\times m}$ is a PCA projection matrix in the t-th iterative times, it is calculated on the projected dataset $\{\boldsymbol{W}^T\boldsymbol{x}_i\}_{i=1}^{n_l}$. Then the projected variable $z_i = \boldsymbol{P}_t{}^T\boldsymbol{W}^T\boldsymbol{x}_i$ is updated, where $i = 1, 2, \cdots, n_l$. The target MLR model is calculate as Eq. (1) and Eq. (2). For better understanding, a pseudo-code and an illustration of MCCR are shown in Algorithm 1 and Fig. 2, respectively.

Algorithm 1 : *Maximum contributed component regression* (MCCR).

```
Input: labeled dataset $\mathcal{L} = (\boldsymbol{X}_L, \boldsymbol{Y}_L)$, and unlabeled dataset $\mathcal{U} = (\boldsymbol{X}_U)$.
Output: CCA projection matrix $\boldsymbol{W}_t$, PCA projection matrix $\boldsymbol{P}_t$, MLR coefficient vector $\boldsymbol{\beta}_t$.
Initialization: $\boldsymbol{W}_0 = \boldsymbol{I}$, and iterative times $T$.
Calculating PCA projection matrix $\boldsymbol{P}_0 \in \mathcal{R}^{d\times m}$ on dataset $\boldsymbol{W}_0^T\boldsymbol{X}_L$.
Estimating MLR coefficient vector $\boldsymbol{\beta}_0$ by Eq. (2).
for $t = 1$ to $T$ do
    Estimating the predicted vector $\widehat{Y}_U$ of $\mathcal{U}$ by Eq. (1).
    Sorting $(\boldsymbol{Y}_L \cup \widehat{Y}_U)$ and grouping $(\boldsymbol{X}_L \cup \boldsymbol{X}_U)$ into two groups, denoted as $\mathbf{G}_1$ and $\mathbf{G}_2$.
    Calculating $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ of CCA according to Eq. (5), then $\boldsymbol{W}_t$ is calculated by Eq. (6).
    Calculating PCA projection matrix $\boldsymbol{P}_t \in \mathcal{R}^{2r\times m}$ on dataset $\boldsymbol{W}_t^T\boldsymbol{X}_L$.
    Estimating MLR coefficient vector $\boldsymbol{\beta}_t$ on projected dataset $(\boldsymbol{P}_t^T\boldsymbol{W}_t^T\boldsymbol{X}_L, \boldsymbol{Y}_L)$ by Eq. (2).
end for
```

3. Experiments

In this section, we introduce a synthetic scatterometic semiconductor dataset and a real scatterometic semiconductor dataset. Based on these two datasets, we evaluate the performance of the proposed MCCR method by comparing with other two model-free methods PCR and PLS.

As a model-free method, it is unnecessary to compare MCCR with the mainstream solutions such as model-based nonlinear regression and look-up table methods. Model-based methods might be superior to model-free one because it utilizes lots of explicit prior knowledge on the

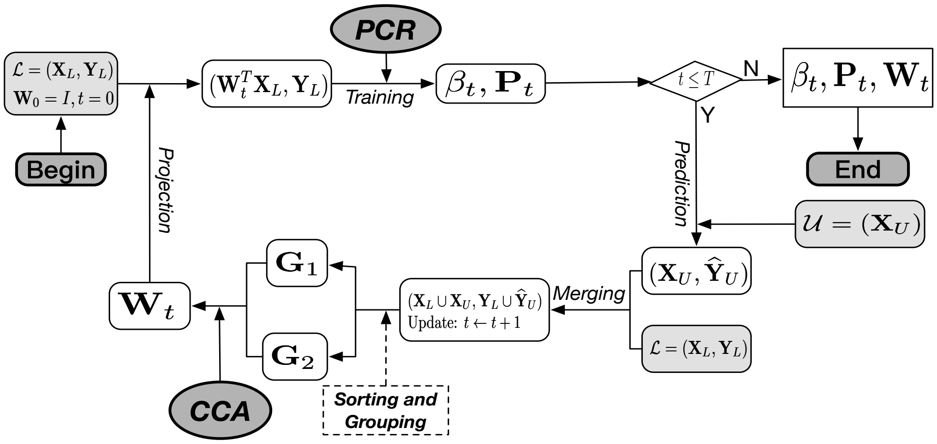

PCR
$\mathcal{L}=(\mathbf{X}_L,\mathbf{Y}_L)$
$\mathbf{W}_0=I, t=0$
$(\mathbf{W}_t^T\mathbf{X}_L,\mathbf{Y}_L)$
Training
$\beta_t,\mathbf{P}_t$
$t\le T$
N
Y
$\beta_t,\mathbf{P}_t,\mathbf{W}_t$
Begin
Projection
Prediction
End
$\mathcal{U}=(\mathbf{X}_U)$
$\mathbf{G}_1$
$\mathbf{W}_t$
$(\mathbf{X}_L\cup\mathbf{X}_U,\mathbf{Y}_L\cup\hat{\mathbf{Y}}_U)$
Update: $t\leftarrow t+1$
Merging
$(\mathbf{X}_U,\hat{\mathbf{Y}}_U)$
$\mathbf{G}_2$
$\mathcal{L}=(\mathbf{X}_L,\mathbf{Y}_L)$
CCA
Sorting and Grouping

Fig. 2. Illustration of the proposed method MCCR.

target structure such as the relationship between profile parameters. However, valid geometric modeling for model-based method is not always available, as mentioned in Section 1. Therefore, it is more reasonable to compare MCCR with the model-free methods PCR and PLS under same experimental settings.

3.1. Dataset and evaluation metric

A synthetic and a real dataset are used to demonstrate the effectiveness and accuracy of the proposed MCCR.

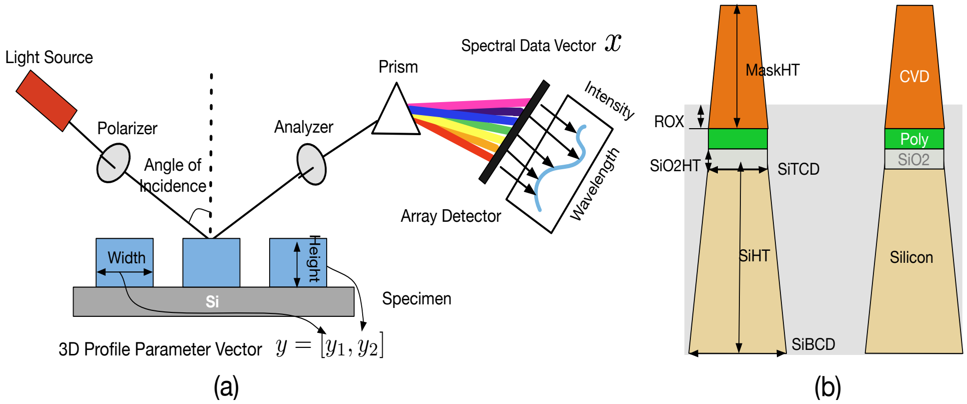


Fig. 3. (a) Measurement setup of the spectroscopic ellipsometry, and (b) the structure of the semiconductor.

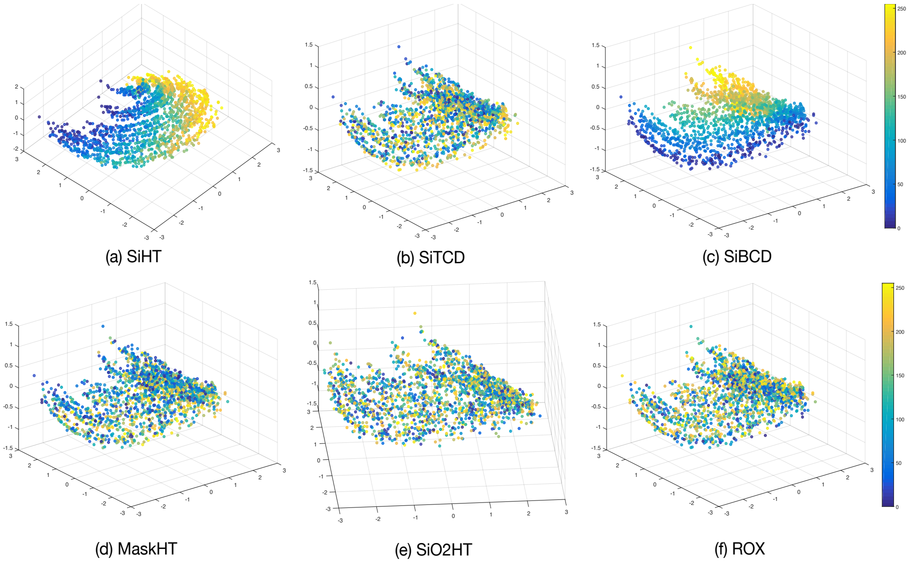
For the synthetic dataset, the spectral data consists of 2000 spectra, each of which is simulated based on a geometric model illustrated in Fig. 3(a). Spectroscopic ellipsometry (SE) [30] is an effective optical metrology tool to measure the dimensional parameter of the structure. The geometric model was made by using six floating parameters such as silicon height (SiHT), silicon top CD (SiTCD), silicon bottom CD (SiBCD), mask height (MaskHT), silicon height (SiO2HT), and recess remain oxide (ROX). The geometric model and six floating parameters are depicted in the Fig. 3(b). Then 2000 spectra were simulated from the geometric model with

Table 1. Statistical values for the six floating parameters.

Parameter (nm)	Min	Max	Range	Mean
SiHT	125	175	50	150
SiTCD	10	14	4	12
SiBCD	16	36	20	26
MaskHT	30	50	20	40
SiO2HT	5	9	4	7
ROX	0	10	10	5

randomly varying six floating parameters from -25% to 25% of their initial value for training. The signatures were simulated in terms of Psi and Delta on the 66 wavelengths ranged from 250nm to 900nm with 10nm interval. To utilize signatures in the proposed method, each signature was represented by 132 dimensional vector consisting of 66 *Psi* values and 66 *Delta* values serially. Some statistical values of the six floating parameters are listed in Table 1. Note that in semiconductor industry, most of electronic property failures are caused by improper thickness of ROX because monitoring the ROX parameter is the hardest process in this structure. Therefore, we will pay more attention to the analysis of ROX in our experimental results.

To uncover the underlying mechanism of the six parameters, we visualized the geometric structure of them by reducing its dimension from 132D to 3D using PCA, as shown in Fig. 4. Each point stands for an instance with its target value indicated by shade of color. It can be seen that SiHT and SiTCD show a stronger linear relationship than other parameters between the spectral data and the corresponding labels.



(a) SiHT (b) SiTCD (c) SiBCD

(d) MaskHT (e) SiO2HT (f) ROX

Fig. 4. Spatial structure of the six spectrum data projected to 3D space. Gradient color reflects the continous change of response variable in label values.

For the real dataset, the 165 spectra were obtained from spectroscopic ellipsometry. Only 31 labeled samples are available because of expensive measuring cost. Each spectrum data has 132 dimensions. Only ROX is studied in the experiment because it is a unique parameter measured by TEM.

R-square (RSQ) (refers to https://en.wikipedia.org/wiki/Coefficient_ of_determination) is used to evaluate the performance of the proposed method in this paper. It calculates the correlation between ground-truth and predicted values according to Eq. (7).

$$\mathcal{R}^2 = \left[\frac{1}{N-1} \sum_{i=1}^{N} \frac{(y_i - \mu)(\hat{y}_i - \hat{\mu})}{\sigma \cdot \hat{\sigma}} \right]^2, \tag{7}$$

where N is the number of test samples. Besides, parameters μ and σ are the mean and standard deviation of $Y = \{y_i\}_{i=1}^{N}$, respectively. Parameters $\hat{\mu}$ and $\hat{\sigma}$ are same to $\widehat{Y} = \{\hat{y}_i\}_{i=1}^{N}$. The greater the value it reaches, generally, the better the performance of the regression.

3.2. Experimental results

In this part, we compare the proposed MCCR method with PCR and PLS in a synthetic dataset and a real-world dataset of semiconductor. All experimental results are the average of 50 repeated trials and the iterative times T is initialized as 4 in all experiments.

3.2.1. Results in a synthetic scatterometic semiconductor dataset

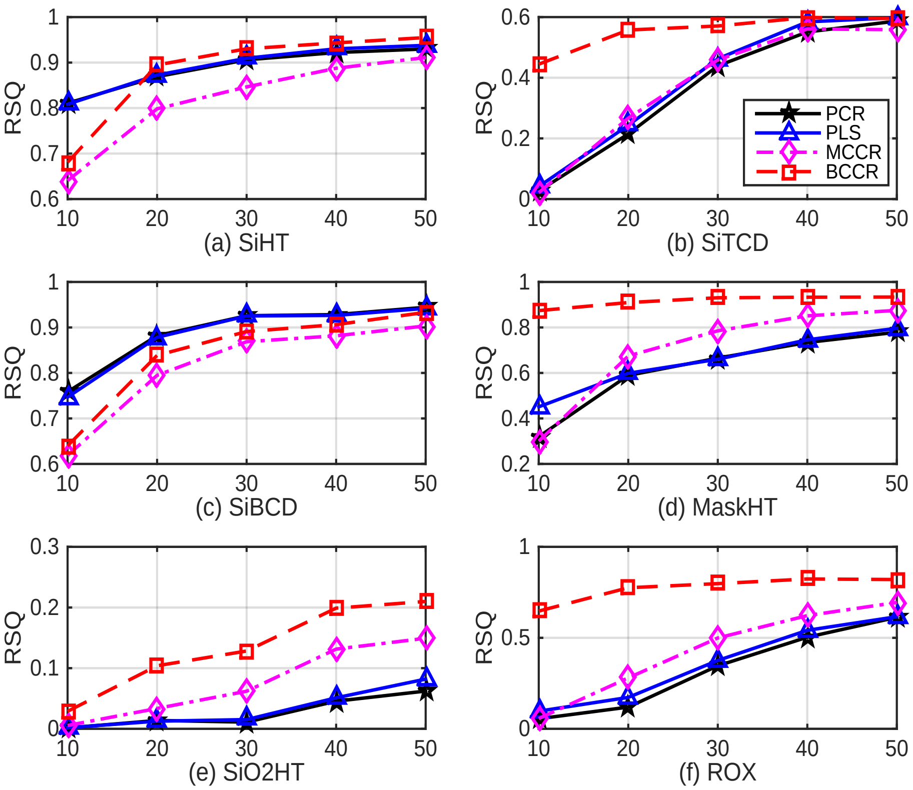


Fig. 5. Comparison of our MCCR method to PCR and PLS with different number of labeled data, which is changed from 10 to 50 in (a) SiHT, (b) SiTCD, (c) SiBCD, (d) MaskHT, (e) SiO2HT, and (f) ROX.

The number of unlabeled data and the number of test data are fixed as 200 and 1000, respectively. To the best of our experience, 200 unlabeled data is sufficient for training in our proposed

MCCR method and 1000 test data (i.e., a half of data) makes the results more convincible. Further, we test the influence of the number of labeled data that are expensive to predictive performance by changing the number of samples from 10 to 50 with 10 interval. In addition, the parameters of dimension for four methods (BCCR, MCCR, PLS, and PCR) are tuned by using 5-fold cross-validation in this experiment.

It can be seen from Fig. 5 that the proposed MCCR method outperforms PCR and PLS in the parameters of MaskHT, SiO2HT, and ROX. As expected, BCCR is always better than MCCR because it is an ideal and best situation of MCCR. In SiHT and SiBCD, MCCR also achieves a competitive performance, but slightly worse than PLS and PCR. The reason is that as shown in Fig. 4, there is a strong linear relationship between the profile parameter named label Y and corresponding spectral signature X in SiHT and SiBCD. It is reasonable that PCR and PLS perform better in these two linear datasets because they are two efficient linear regression methods. However, PCR and PLS work worse if the data are nonlinear. In contrary, the propsed method MCCR works well in linear and nonlinear datasets. For the most important parameter of ROX, the rates of RSQ are improved by 10.31% and 7% in average compared with PCR and PLS, respectively. More results about the average improvement rate of the proposed method in six parameters can be seen in Table 2.

Table 2. The average improvement rates of MCCR to PCR and PLS for six parameters.

Method	SiHT	SiTCD	SiBCD	MaskHT	SiO2HT	ROX
PCR	-7.08%	**0.96%**	-7.44%	**7.98%**	**4.96%**	**10.31%**
PLS	-7.49%	-1.26%	-7.04%	**4.69%**	**4.38%**	**7.00%**

3.2.2. Results in real-world scatterometic semiconductor dataset

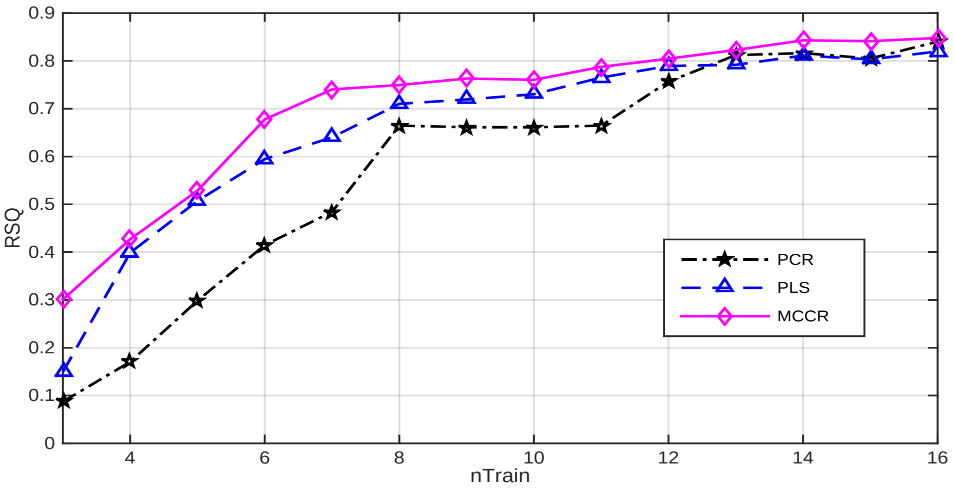


Fig. 6. Comparison of our MCCR method to PCR and PLS with different number of labeled data changing from 3 to 16 in a real dataset ROX.

We compare our MCCR method with PCR and PLS by changing the number of labeled data from 3 to 16 in a real-world dataset of ROX. In this experiment, the number of unlabeled data and the number of test data are fixed as 134 and 15, respectively. To tune the parameter of dimension, leave-one-out cross-validation (LOO-CV) is utilized in this experiment, because the number of labeled data is smaller than which of synthetic dataset. LOO-CV is a kind of cross-validation strategy by exhaustively separate the original dataset into one single sample as test sample and the remained samples training set. Figure 6 illustrates that our MCCR method outperforms PCR and PLS in all situations. Concretely, the average improvement rates of MCCR to PCR and PLS are 12.56% and 4.75%, respectively. It indicates that our proposed strategy is still effective even the number of unlabeled data is only 134, which is smaller than 200 in the experiment of synthetic dataset.

Table 3. The time of training and testing time for the proposed MCCR, here *N* is the number of labeled data, *s* is the abbreviation of seconds.

	N = 10	**N = 30**	**N = 50**
Training time (s)	2.0527	6.5967	11.5080
Testing time (s)	0.0088	0.0088	0.0088

In summary, the experimental results of the synthetic dataset and the real dataset demonstrate the proposed method is superior to two traditional model-free methods PCR and PLS, because it utilizes the structural information from pairwise unlabeled data to improve the performance. With a few expensive labeled data and a large number of economic unlabeled data, our proposed method significantly reduces the cost of semiconductor manufacturing process monitoring and meanwhile achieves high predictive accuracy. As Table 3 reported, moreover, the proposed method is highly efficient that it can be applied to monitor the quality of semiconductor elements in real-time.

4. Conclusions

In conclusion, we proposed an effective and efficient model-free regression, named MCCR, to solve the inverse problem in optical scatterometry given small amount of labeled data and lots of economic unlabeled data. With a combination of PCR and CCA, the proposed method estimates the labels of those unlabeled data, attaining a common subspace where the regressor achieves better predictive performance by using these data together with a few labeled data. The experimental results on a synthetic dataset and a real dataset demonstrate that the proposed method is effective and efficient. Without the requirement of geometric modeling, the proposed model-free method is experimentally justified that it can work well in real-time.

MCCR is a semiempirical approach and less reliable than model-based method since it can produce a “predicted” but not a “deterministic” parameter estimation to the unseen samples. However, it is a good alternative to the model-based one to monitor the profile parameter in the early stage of ramp-up where the manufacturing process is not yet stable so that pattern profile is not enough homogenous to get valid geometric modeling. “Prediction” is sort of risky when output is not controllable and estimable. To make up this weakness, in the future, it deserves further studying on prediction reliability of the proposed approach so that users or systems can automatically or interactively reject those predicted profile parameters with low reliabillity.

Funding

National Natural Science Foundation of China (NSFC 61673118); Samsung Electronics Co., Ltd (IO150831-03285-01).

Acknowledgments

We would like to express our gratitude to associate editor and two reviewers for their valuable comments.