Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-12 (2016) 083–088

Regression Methods for Predicting the Product's Quality in the Semiconductor Manufacturing Process [⋆]

Mariam Melhem [*] Bouchra Ananou [*] Mustapha Ouladsine [*]
Jacques Pinaton [**]

[*] *Aix Marseille University, CNRS, LSIS UMR 7296, 13397 Marseille, France (e-mail: firstname.name@lsis.org).*
[**] *STMicroelectronics (e-mail: jacques.pinaton@st.com).*

Abstract: The quality of production in the wafer manufacturing process cannot be always monitored by metrology tools because physical measurements are very expensive. Instead of conducting costly quality tests, it is desirable to predict the wafer quality. Regression models are useful to build such a predictor by using the production equipment data and a set of wafer quality measurements. As the semiconductor manufacturing process consists of a huge amount of data that are correlated and very few quality measurements, Ordinary Least Squares (OLS) regression fails in predicting the wafer's quality. Regression methods dealing with multicollinear high-dimensional input data are required. In this paper, a survey of regularized linear regression methods based on feature reduction and variable selection methods is presented. These methods are applied to predict the wafer quality based on the production equipment data, then compared. Regression parameter optimization and model selection are performed and evaluated via cross validation, using the Mean Squared Error (MSE). Our results indicate that reducing the predictor's dataset will improve the model robustness and the prediction accuracy.

Keywords: Quality prediction, multivariate systems analysis, regularized linear regression, model selection, semiconductor manufacturing process, yield enhancement.

1. INTRODUCTION

Semiconductor manufacturing process is a very complex process characterized by series of expensive equipments whose function is to fabricate integrated electronic circuits consisting of thousands of components. The slight failure in the process can deteriorate the wafer quality and cause a catastrophic loss in the "Yield" of manufacturing. Hence, the production equipments need to be monitored to assure a stable fabrication and a high yield rate. For monitoring the process quality, many batches are sampled at many stages of the manufacturing process, and 2 or 3 wafers within each selected batch are measured with metrology equipments. However, the sampling frequency is determined by control considerations. As there are not enough wafer measurements, a shift or drift between two measurements in the sample can be undetectable. Hence, the current metrology tools are not reliable to insure the quality performance of all the wafers. In the other hand, huge amount of data are collected from the production equipments by on-line sensors. A good alternative is to predict the wafer quality from the available equipment data.

Using standard linear methods to predict the product quality parameters often involves ill-posed problems, because the input data are high-dimensional and multicollinear. The Ordinary Least Squares (OLS) estimator fails to construct the prediction model: the predictors are difficult to be interpreted and this fact may cause overfitting. To avoid overfitting, a bias-variance trade-off must be achieved. To do so, dimension reduction and feature selection tools should be used.

Many variants of OLS can address the problems of high dimensionality and multicollinearity in data. Different multivariate statistical methodologies have been used used in the literature for different purposes like process control and monitoring Martin et al. (1996), MacGregor and Kourti (1995), Fault Detection Slama (1991) and Isolation (FDI) Kourti (2005), diagnosis Papazoglou (1998) and product quality prediction Nilsson (2005), Yu and MacGregor (2003) in industrial processes, like, for example, mineral processes Tano (2005) and pharmaceutical industry Kourti (2009), etc... A model based on the Partial Least Squares Regression (PLS) is developed in Besnard et al. (2012) to predict the Plasma Enhanced Chemical Vapor Deposition (PECVD) oxide thickness using Fault Detection and Classification (FDC) and metrology data. In Purwins et al. (2011), different approaches (simple linear regression, multiple linear regression, partial least squares regression, and ridge linear regression) are compared to predict the average Silicon Nitride cap layer thickness for the PECVD process.

The rest of the paper is organized as follows: the next section briefly reviews the regularized methods in the context of linear regression. Section III presents the proposed

[⋆] This work is part of the European project "INTEGRATE", and carried by ST-microelectronics Fab.

10.1016/j.ifacol.2016.07.554

method and its application in the semiconductor manufacturing process. In section IV, the results of our application are provided and discussed. Finally, the conclusions are summarized in section V.

2. REGULARIZED REGRESSION METHODS

We consider the problem of predicting a univariate response $y \in R$ from covariates $x = (x_1, x_2, ...x_p) \in R^p$ via a linear model. We assume that we have n observations of the response and the predictors, represented by a column vector $Y_{n\times 1}$ and a matrix $X_{n\times(p+1)}$, respectively, where the first column $e = (1, 1, ..., 1)^T$ of X is added to the matrix of the p predictors to represent the intercept of the model. The linear regression model can be represented as:

$$Y = X\beta + \varepsilon \tag{1}$$

where Y is a $(n \times 1)$ vector, X is $n \times (p + 1)$ matrix, β is a $(p+1)\times 1$ coefficient vector, and ε is an $(n \times 1)$ random error vector.

After fitting the model to the observations and estimating the coefficients, the constructed model is used to predict Y for new values of X.

The coefficients $\hat{\beta}$ can be estimated with the Oridinary Least Squares (OLS) by minimizing the Sum of Squared Residuals (RSS):

$$\hat{\beta} = argmin_{\beta} \left\| Y - X\beta \right\|_2^2 \tag{2}$$

where $\|.\|_2$ is the norm l_2.

By deriving the above criterion with respect to the coefficient vector, we obviously obtain

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \tag{3}$$

The high dimensionality in data makes the prediction model difficult to construct, especially when the number of variables p is far larger than the number of observations n in the sample ($p >> n$). Standard methods of linear regression are then inappropriate. In fact, high-dimensionality is often associated with correlations between the predictor variables, leading to singularities in the optimization problem, which makes the solutions unstable. Although the Least Squares estimator is unbiased, its variance is large in case of multicollinearity between parameters. A solution to this problem is to add a degree of bias to the estimator on condition that a low variance be guaranteed. Two strategies can handle the high-dimensional modelling problem with a tradeoff bias-variance: the first one selects the pertinent variables, based on the sparsity hypothesis, meaning that only few relevant variables contribute to constructing the model. The second is dimension reduction where the observations are projected in a low-dimensional space that captures most of the information contained in the original variables.

2.1 Dimension reduction based regression:

Principal Component Regression (PCR) and Partial Least Squares (PLS) are two techniques that combine features from Principal Component Analysis (PCA) and multiple linear regression. They can deal with a large number of predictors and small sample size, and high collinearity among predictors.

Principal Component Regression: Principal Component Regression (PCR) is an alternative to Ordinary Least Squares (OLS) whose objective is to transform the input matrix that consists of a large number of covariates into a new set of few variables that can summarize the data, and to use them to predict the response variable. Based on the Principal Component Analysis (PCA), PCR tries to produce orthogonal linear combinations of the original predictors as in (4), and then regresses the response on a subset of latent variables using the OLS. After determining the regression coefficients, the model is transformed back to the scale of original predictors by using the PCA loadings.

$$X = TP^T + E \tag{4}$$

The number of principal components chosen for regression must be able to describe large trends in data without losing too much information. Thus we require the linear combination of X ($T = Xw$) to have maximum variance:

$$w_k = argmax_w var(Xw) \tag{5}$$

We can prove that the principal components are the eigenvectors of the correlation matrix $X^T X$, and we must choose those corresponding to the highest eigenvalues to capture most variability in the original data set. As the principal components are orthogonal, the multicollinearity is identified and eliminated from the input data. However, one drawback of PCR is that it performs dimension reduction of the predictor matrix X without taking into account its correlation with the response Y. But nothing guarantees that the first few principal components are the best explanatory for the response even though they contain most of the information about the predictor variables.

Partial Least Squares Regression: Partial Least Squares or Projection on Latent Structure (PLS) Kresta et al. (1994) is an alternative to the PCR. Its benefit over the PCR is that it relates the predictor matrix X to the response variable Y by projecting both X and Y into new latent variables T and U, respectively, that maximize the covariance between X and Y. Thus, the PLS technique performs by finding linear decompositions of X and Y such that:

$$\begin{cases} X = TP^T + E \\ Y = UQ^T + F \end{cases} \tag{6}$$

Where $T_{n\times K}$ and $U_{n\times K}$ are the the X and the $Y-$scores, $P_{p\times K}$ and $Q_{1\times K}$ are the X and $Y-$loadings, $E_{n\times p}$ and $F_{n\times q}$ are the X and $Y-$residuals, respectively.

In order to specify the latent component matrix T as a linear combination of the original predictors X such that $T = XW$, having the maximal covariance with the response vector Y, PLS attempts to find the latent columns of $W = (w_1, w_2, ..., w_K)$ by optimizing the following criterion:

$$w_k = argmax_w corr^2(Y, Xw)var(Xw) \ s.t \ w'w = 1 \tag{7}$$

Hence, the PLS attempts to maximize the covariance between the latent variables T extracted from X and Y, rather than maximizing only the variance of T in case of PCR.

However, it can be shown that the weight vector w corresponds to the first eigenvector of the matrix $X^T YY^T X$.