이학석사 학위논문

Functional Classification for Semiconductor Process Data

반도체 공정 데이터 분석을 위한 함수형 분류

2023년 8월

서울대학교 대학원

통계학과

한 건 희

Functional Classification for Semiconductor Process Data
반도체 공정 데이터 분석을 위한 함수형 분류

지도교수 오 희 석

이 논문을 이학석사 학위논문으로 제출함
2023년 4월

서울대학교 대학원
통계학과
한 건 희

한건희의 이학석사 학위논문을 인준함
2023년 6월

위 원 장	**임 채 영**	**(인)**
부위원장	**오 희 석**	**(인)**
위 원	**이 권 상**	**(인)**

Functional Classification for Semiconductor Process Data

By

Geon Hee Han

A Thesis
Submitted in fulfillment of the requirement
for the degree of
Master of Philosophy
in Statistics

Department of Statistics
College of Natural Sciences
Seoul National University
August, 2023

ABSTRACT

Functional Classification for Semiconductor Process Data

Geon Hee Han
The Department of Statistics
The Graduate School
Seoul National University

In the semiconductor industry, the exponential growth of data necessitates the development of efficient analysis techniques. Optical Emission Spectroscopy (OES) dataset, originating from semiconductor process, is a complex dataset that can be considered as functional data in both the time and wavelength domains. The thesis focuses on exploring various statistical methods to effectively classify such functional data. Specifically, the thesis investigates the utilization of Functional Principal Components (FPCs) and Functional Logistic Regression as tools to address the classification challenges. The main objective of this thesis is to employ different functional classification models to classify the success of semiconductor process using OES dataset. To achieve this objective, the

thesis compares the performance of these functional classification models by using Accuracy and other relevant evaluation metrics such as Specificity and AUC. Through this comparative analysis, the thesis aims to evaluate the effectiveness of the employed functional classification models in accurately classifying the success of semiconductor processes. Additionally, the thesis extends its evaluation to another real dataset, providing a comprehensive examination of the models' performance. Finally, we discuss some limitations of this research and the need for improved performance of the models.

Keywords: Functional data analysis, Classification, Functional principal component analysis, Functional classification trees, Functional logistic regression, Fused Lasso.

Student Number: 2021-21150

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent years, there has been an increasing need for efficient analysis methods to handle the large volumes of data collected from various industries, such as semiconductor and financial companies. A lot of data involve continuous observations in terms of time, wavelength, and other parameters that can be considered as functional data. Functional Data Analysis (FDA) is a powerful statistical tool that is widely used to analyze such data. It has demonstrated a lot of success in many fields that involve high-dimensional data analysis, including finance, biology, and engineering.

The primary focus of this paper is on the classification of Optical Emission Spectroscopy (OES) dataset, a high-dimensional dataset obtained from semiconductor etching process. This data can be regarded as functional data with respect to time and wavelength. Due to the large volume and high dimensionality of the data, effective dimension reduction methods are crucial for improving the analysis performance.

To address this problem, this paper reviews recent methods that utilize Functional Principal Component Analysis (FPCA) for functional data classification. The first method is supervised FPCA, as proposed by Nie (2018) ([12]), which uses leading FPC scores to predict a binary response variable. The second method, proposed by Maturo and Verde (2023) ([11]), offers a classification method that combines FDA and tree-based models. Lastly, Chatterjee *et al.* (2023) ([3]) suggests using FPCA for each class, preserving class-specific information through projection onto functional subspaces.

Furthermore, this paper focuses on combining the fused lasso penalty with the functional logistic model. Kim, H and Kim, H (2018) ([7]) proposed this model for classifying functional data and identifying significant interval, such as time or wavelength interval. As a result, this model can be valuable for industries that utilize substantial amounts of time series data or wavelength signal data.

In this thesis, we not only focus on the classification of OES dataset but also apply the proposed methods to other dataset. Chapter 2 provides a comprehensive review of Functional Data Analysis and a brief overview of functional classification methods. Chapter 3 explains the methodology of the functional logistic model with the fused lasso penalty. In Chapter 4, we apply the proposed methods to real datasets, including OES dataset. Lastly, in Chapter 5, we present concluding remarks.

Chapter 2

Background

2.1 Functional Data Analysis (FDA)

2.1.1 Overview of FDA

Functional data analysis is a powerful methodology used to analyze complex datasets that vary over a continuum of values. This approach enables researchers to represent data as smooth curves or functions, uncovering patterns and relationships that may not be clear when using traditional statistical techniques. As a result, it has become an increasingly valuable tool across various research fields.

To perform functional data analysis, the observations Z_{it} are transformed into a functional form $X_i(t)$, where $i = 1, ..., N$ and $t = 1, ..., T$. This conversion assumes smoothness, and a basis function is used to achieve this representation. The choice of an appropriate basis function is crucial, as it can significantly impact the accuracy and interpretability of the analysis. The functional form

can be represented as:

$$X_i(t) \approx \mu_i(t) + \sum_{k=1}^{K} c_{ik}\phi_k(t) = \mu_i(t) + \mathbf{c}_i^T \boldsymbol{\phi}(t), \quad i = 1, 2, ..., N$$

where $\mathbf{c}_i = (c_{i1}, ..., c_{ip})^T$ is the vector of coefficients which defines linear combination and $\phi_k(t)$ is the k-th basis function. It is possible to assume that $\mu_i(t) = 0$ without loss of generality, so for the rest of this paper we assume $\mu_i(t) = 0$.

With this basis representation, functional data can be analyzed using various statistical techniques, such as Principal Component Analysis (PCA) or regression analysis, enabling researchers to gain valuable insights and make important decisions based on the underlying functional patterns.

2.1.2 Functional Principal Component Analysis (FPCA)

Among the many statistical methods used in FDA, Functional Principal Component Analysis (FPCA) is one of the most commonly employed techniques. FPCA is used to effectively reduce the dimensionality of functional data which inherently has an infinite number of dimensions, into a finite-dimensional representation. This is achieved by determining an orthogonal basis that captures the maximum amount of variability present in the data.

In FPCA, the functional data $X(t)$ can be represented using the Karhunen-Loève expansion as follows:

$$X(t) = \sum_{j=1}^{p} \alpha_j \xi_j(t) = \alpha^T \xi(t) \tag{1}$$

where $\xi_j(t), j = 1, \ldots, p$ are referred to as Functional Principal Components (FPCs) that are mutually orthogonal, and α_j

is called jth FPC score. Additionally, α is defined as the vector $(\alpha_1, \ldots, \alpha_p)^T$, while $\xi(t)$ is the vector $(\xi_1(t), \ldots, \xi_p(t))^T$. In practice, only the leading p FPCs are usually considered, making FPCA a computationally efficient method with broad applicability in various fields.

Through this process, we can obtain an approximation of the observed curves, effectively representing the original functional data using a reduced set of FPCs. This approach can be seen as an extension of the original PCA, adapted to handle functional data. The difference from original PCA is that FPCA assumes the smoothness, which requires the definition of the $\mathscr{L}^2$ inner product as $\langle f, g \rangle = \int_{\mathscr{T}} f(t)g(t)dt$. The norm $\|\xi\|$ is then defined as $\|\xi\| = \sqrt{\|\xi\|^2} = \sqrt{\langle \xi, \xi \rangle}$.

We summarize some formulas used in FPCA. Specifically, k-th FPC score is given by

$$\alpha_k = \int_{\mathscr{T}} X(t)\xi_k(t)dt$$

where the weight function ξ_k is an eigenfunction that satisfies the following eigenequation:

$$\hat{\mathscr{C}}\xi_k = \int_{\mathscr{T}} \hat{C}(\cdot, t)\xi_k(t)dt.$$

Here, the empirical covariance function is defined as $\hat{C}(s,t) = \frac{1}{n}\sum_{i=1}^{n} X_i(s)X_i(t)$, where $X_i(t)$ represents the i-th independent realization of $X(t)$. Moreover, $\hat{\mathscr{C}}$ is referred to as the empirical covariance operator. The specific method of estimating ξ_k will be covered in the next section.

Therefore, by utilizing the mentioned FPCA technique, we can effectively capture the main modes of variation in the functional

data, resulting in a more concise representation of the complex dataset. This dimensionality reduction can facilitate subsequent analysis and enhance the interpretability of the results.

2.2 Functional Classification Models

2.2.1 Supervised FPCA

In this subsection, we provide a review of the supervised Functional Principal Component Analysis (sFPCA) method proposed by Nie(2018) ([12]). This method is distinct in that it considers the correlation between the functional predictor and the response variable.

Our objective is to classify new unseen curves, so we employ the Functional Linear Model based on the following FPCA-based model:

$$E(Y|X(t)) = g\left(\beta_0 + \int_{\mathscr{T}} X(t)\beta(t)dt\right) = g\left(\beta_0 + \sum_{j=1}^{p} \alpha_j \int_{\mathscr{T}} \beta(t)\xi_j(t)dt\right) \tag{2}$$

by using (1). This model can also be represented as:

$$E(Y|X(t)) = g\left(\beta_0 + \alpha^T\gamma\right) \tag{3}$$

where $\beta(t) = \sum_{j=1}^{p} \gamma_j\xi_j(t) = \gamma^T\xi(t)$. To estimate $\hat{\xi}_k(t)$, we maximize the following objective function:

$$Q(\xi) = \frac{\theta\langle\xi, \hat{\mathscr{C}}\xi\rangle + (1-\theta)cov^2(Y, \langle X, \xi\rangle)}{\|\xi\|_\lambda^2}$$

subject to $\|\xi\|_\lambda = 1, \langle\xi, \hat{\xi}_j\rangle = 0$, for every $j < k$, and $0 \leq \theta \leq 1$. Here, the regularized norm is defined as $\|\xi\|_\lambda = \sqrt{\|\xi\|^2 + \lambda\|\mathscr{D}^2 f\|^2}$

where $\mathscr{D}^2 f = \int_{\mathscr{T}} f''(t)dt$ to consider a higher degree of smoothness. The weight parameter θ allows for adjusting the balance between the first and second terms in the numerator. A value of $\theta = 1$ corresponds to the conventional FPCA introduced in section 2.1.2, which does not consider the correlation. The level of smoothness is controlled by the smoothing parameter λ.

The analysis in this paper is limited to the situation where Y is binary, as the primary interest is in classification problems. In this case, $cov^2(Y, \langle X, \xi \rangle)$ in $Q(\xi)$ is replaced with $R(\xi)$, given by:

$$R(\xi) = \frac{1}{n_1}\left(\sum_{i=1}^{n} Y_i \alpha_i\right)^2 + \frac{1}{n_0}\left(\sum_{i=1}^{n} (1 - Y_i)\alpha_i\right)^2$$

where n_j is the number of occurrences of $Y_i = j$ for $j = 0, 1$. By using this term, we can estimate $\hat{\xi}_k(t)$, $k = 1, \ldots, p$.

Therefore, to classify new curves, we employ the functional logistic linear model based on (3). The model is fitted as follows:

$$\text{logit}\{P(Y = 1)\} = \beta_0 + \alpha^T \gamma \tag{4}$$

We estimate the coefficient vector $\hat{\gamma}$ by performing regression of the response variable Y on the FPC scores α in (4). Then we can obtain $\hat{\beta}(t) = \hat{\gamma}^T \hat{\xi}(t)$ and finally we can classify another set of test data by using (2).

By following these steps, we can effectively classify new unseen curves using the sFPCA method. It additionally considers the correlation between the functional predictor and binary response variable that is different from conventional FPCA. This additional consideration can contribute to the improvement of accuracy.

2.2.2 Tree-based Methods using FPCs

In a recent study by Maturo and Verde(2023) ([11]), a novel combined approach of FDA and tree-based methods was proposed to address functional classification problems. The functional data can be represented as shown in (1), where each observation $x_i(t)$ can be decomposed as a linear combination of the first p FPCs. It can be represented as:

$$x_i(t) = \sum_{k=1}^{p} v_{ik}\xi_k(t), \quad i = 1, 2, ..., n$$

To integrate the functional representation into tree-based models, the FPC score matrix V is utilized, where each row corresponds to the FPC scores of an individual curve. V is represented as:

$$V = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{np} \end{pmatrix} \tag{5}$$

The matrix V can then be used as input to tree-based models, such as classification trees, bagging, and boosting to develop a classification model for functional data. This process aims to improve classification performance by leveraging the strengths of both FDA and tree-based methods.

First, let's explore Functional Classification Trees with Functional Principal Components (FCT-FPCs). Similar to traditional classification trees, FCT-FPCs recursively split the data into smaller subsets based on the values of predictor variables, aiming to create subsets that are more homogeneous in terms of their class labels.

However, instead of using traditional predictor variables, FCT-FPCs use the scores of the FPCs obtained from the decomposition given in (5). This approach enables the use of functional data in classification problems and provides a way to identify which FPCs are more important in predicting class labels.

Second, Functional Bagging (FBAG) is a technique for improving the accuracy of FCTs by aggregating the results of multiple FCTs. Multiple FCTs are constructed using bootstrap samples of FPCs, and the results are combined by taking a majority vote of the predicted classes. This approach can help reduce overfitting and improve the generalizability of the model. Another technique for improving the performance of decision trees is Functional Boosting (FBOOST), which involves iteratively constructing a sequence of FCTs. Each subsequent tree focuses on the observations that were misclassified by the previous tree, resulting in a highly accurate classification model. However, FBOOST can be more computationally intensive and prone to overfitting than FBAG.

Finally, Functional Random Forest (FRF) is an advanced technique for classification that builds upon FBAG by constructing multiple FCTs, each based on a random subset of m FPCs out of the total p FPCs. This process helps reduce the variance when averaging FPCs, thereby reducing correlation among FCTs. Notably, when m equals k, FRF becomes equivalent to FBAG.

In summary, the proposed combined approach of FDA and tree-based methods offers a powerful framework for functional classification problems, utilizing the strengths of both methodologies to achieve accurate and interpretable classification models

for functional data.

2.2.3 Functional Classwise PCA

Functional Classwise PCA (FCPCA), proposed by Chatterjee *et al.* (2023) ([3]), is a recent method for performing classification on functional data while preserving class-specific information. The approach utilizes class-wise FPCA to eliminate non-informative subspaces of the data by mapping it to a lower-dimensional space. This approach helps identify and eliminate the non-informative subspaces of the data while retaining the most informative subspaces for classification. Consequently, it reduces the dimensionality of the data while maintaining its discriminatory power.

This method consists of two main parts: functional feature extraction and classification. In the first step, functional features are extracted using Gram-Schmidt orthonormalization. For each class ω_i, we calculate the sample mean $\hat{\mu}_i$ from a set of functional observations $X_{1,i}, \ldots, X_{n_i,i}$, where n_i is the number of observations of class ω_i. Then, we form a set $F_i = \{f_{1i}, \ldots, f_{q_i i}\}$ of empirical functional principal components of the class ω_i, and augment it with a set $G = \{\hat{\mu}(t) - \hat{\mu}_1(t), \hat{\mu}(t) - \hat{\mu}_2(t), ..., \hat{\mu}(t) - \hat{\mu}_{c-1}(t)\}$ of differences between the grand mean and the class means, where c is the number of classes. The set $\{F_i, G\}$ is then orthonormalized using the Gram-Schmidt orthonormalization process to maintain the directions of the projections. The resulting orthonormal set contains $q_i + c - 1$ functions, which generate a subspace S_i. Thus, given a test curve X^*, we can project it onto each of the c subspaces

$S_1, \ldots, S_c$ using the following mappings:

$$X^*_{(i)} = \sum_{k=1}^{q_i+c-1} \langle X^* - \hat{\mu_i}, f_{ki} \rangle f_{ki}, \quad i = 1, \ldots, c,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product.

In the second step, an unknown curve observation X^* is classified into one of c classes. The coefficient vector in the ith subspace is calculated as $(\langle X^* - \hat{\mu_i}, f_{1i} \rangle, \ldots, \langle X^* - \hat{\mu_i}, f_{q_i+c-1,i} \rangle)^T$, which we represent as $\mathbf{v}_i$. The estimated posterior probabilities $\widehat{P}(\omega_\ell | \mathbf{v}_i)$ are then obtained using Linear Discriminant Analysis (LDA). At this stage, the mean and covariance matrix of each class are estimated using the coefficient matrix of training sample curves. Finally, the class of X^* is assigned by finding the index j that maximizes $p^{(i)}$, where $p^{(i)} = \max\limits_{1 \leq \ell \leq c} \widehat{P}(\omega_\ell | \mathbf{v}_i), i = 1, \ldots, c$.

Chapter 3

Methodology

3.1 Functional Logistic Regression

The functional logistic model is a useful approach that effectively handles classification problems involving functional predictors. It is a powerful statistical framework that extends the traditional logistic regression model. The functional logistic model allows us to accommodate time-varying or wave-specific effects of continuous predictors, offering a flexible approach for modeling a binary response variable.

We consider a binary response variable, denoted as $y_i \in \{0, 1\}$, and the functional predictor, represented as $x_i(t), i = 1, 2, \ldots, N$. Assuming y_i follows a Bernoulli distribution with a success probability π_i, the functional logistic regression model can be expressed as:

$$\pi_i = P(Y_i = 1|x_i(t)) = \frac{\exp\left\{\alpha + \int_{\mathscr{T}} x_i(t)\beta(t)\,dt\right\}}{1 + \exp\left\{\alpha + \int_{\mathscr{T}} x_i(t)\beta(t)\,dt\right\}}$$

where α is the intercept term, $\beta(t)$ represents the time-varying or

wave-specific effect parameter of the functional predictor. Alternatively, we can express the model using the logit transformation:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \int_{\mathscr{T}} x_i(t)\beta(t)\,dt \tag{6}$$

This formulation allows us to interpret the model in terms of log-odds, where the left-hand side represents the logit of the success probability.

This model becomes particularly valuable when dealing with predictors that vary over time or wavelengths, such as semiconductor data or time series data. This advantage allows researchers to find complex patterns and make more accurate predictions when working with such data. Consequently, the functional logistic model becomes a powerful tool for classifying functional data in various fields.

3.2 Fused Lasso

The fused lasso method was first introduced by Tibshirani *et al.* (2005) ([14]). It is a regularization method particularly useful when dealing with situations where important variables exhibit a continuous occurrence centered around peaks, necessitating the consideration of adjacent variables simultaneously. The fused lasso method is a variant of the lasso regularization designed to promote sparsity while encouraging grouping of adjacent variables. As a result, the fused lasso effectively accounts for the smoothness and continuity of the variables.

To estimate the coefficient vector β, the fused lasso method

uses the following optimization problem:

$$\hat{\beta}^{FL} = \arg\min_{\beta} \left\{ \sum_{j=1}^{p} (y_i - x_i^T \beta)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=2}^{p} |\beta_j - \beta_{j-1}| \right\}$$

The objective function aims to minimize the squared distance between the response variable y_i and the linear combination of the predictors $x_i^T \beta$, while also considering the L1 penalty term including λ_1 for sparsity in the coefficients and λ_2 for sparsity in their differences. By solving this optimization problem, we can obtain the estimated coefficient vector $\hat{\beta}^{FL}$.

3.3 Functional Logistic Regression with Fused Lasso Penalty

Kim, H, and Kim, H (2018) ([7]) proposed a novel approach for classifying functional data. This method integrates the functional logistic regression model as a classifier with the fused lasso penalty to identify discriminant segments and this is called FLR-FLP. FLR-FLP can be viewed as an interval-based classification approach that considers the inherent characteristics of functional data, including smoothness and continuity.

The functional predictor variable $x_i(t)$ is decomposed using cubic B-spline basis function $\phi_j(t)$. The decomposition can be expressed as:

$$x_i(t) = \sum_{j=1}^{3+m+1} c_{ij}\phi_j(t) = \boldsymbol{\phi}^T \mathbf{c}_i$$

where c_{ij} is the coefficient of these basis functions, and m is the number of interior knots. Additionally, the parameter function $\beta(t)$

is decomposed using the unit step function $\psi_k(t)$. The unit step function is defined as $\psi_k(t) = \mathbb{I}(u_k \leq t < u_{k+1})$, where u_k denotes the ordered knot point. The function $\beta(t)$ can be expressed as the sum of these step functions:

$$\beta(t) = \sum_{k=1}^{K} \gamma_k \psi_k(t) = \boldsymbol{\psi}^T \boldsymbol{\gamma}$$

where K is the number of time points minus one, and γ_k is the coefficient associated with the k-th step function. As a result, the integral term in the logistic regression model can be expressed as:

$$\int_{\mathscr{T}} \beta(t) x_i(t)\, dt = \int_{\mathscr{T}} \boldsymbol{\gamma}^T \boldsymbol{\psi} \boldsymbol{\phi}^T \mathbf{c}_i\, dt = \boldsymbol{\gamma}^T \left(\int_{\mathscr{T}} \boldsymbol{\psi} \boldsymbol{\phi}^T\, dt \right) \mathbf{c}_i = \boldsymbol{\gamma}^T \mathbf{B} \mathbf{c}_i$$

where $\mathbf{B} = (b_{kj})$ and $b_{kj} = \int_{\mathscr{T}} \psi_k(t) \phi_j(t)\, dt$. Finally, (6) is equivalently expressed as:

$$l_i = \alpha + \boldsymbol{\gamma}^T \mathbf{B} \mathbf{c}_i \tag{7}$$

We only need to estimate α and $\boldsymbol{\gamma}$ in (7). To achieve this, the model uses the conditional log-likelihood function:

$$L(\alpha, \boldsymbol{\gamma}) = \sum_{i=1}^{N} [y_i l_i - \log(1 + \exp(l_i))]$$

Then, the model estimates α and $\boldsymbol{\gamma}$ by minimizing the penalized likelihood functions:

$$\min_{\alpha, \boldsymbol{\gamma}} \left(- L(\alpha, \boldsymbol{\gamma}) + \lambda_1 \sum_{k=1}^{K} |\gamma_k| + \lambda_2 \sum_{k=2}^{K} |\gamma_k - \gamma_{k-1}| \right) \tag{8}$$

where λ_1 and λ_2 are hyperparameters that can be selected through cross-validation. Equation (8) can be solved by convex optimization technique since both the negative log-likelihood function and absolute function are convex. To perform the optimization, this

paper uses CVXR, an R package specifically designed for convex optimization ([1]). The resulting non-zero coefficients obtained through this process represent significant intervals that are crucial in classifying the curves.

Chapter 4

Real Data Analysis

4.1 Optical Emission Spectroscopy (OES) dataset

4.1.1 Data preprocessing

OES dataset used in this study consists of high-dimensional data obtained from Etching process, which is an important step in semiconductor manufacturing. A total of 307 wafers were observed and the response variable, denoted as z_i, is bounded within the interval $[0, 1]$. Also, there are OES values that were measured by three-dimensional covariate X composed of j substeps, T_j time points, and m wavelength points.

For this reason, the OES dataset can be considered as functional data with respect to time, wavelength, and substep. Specifically, each observation can be expressed as $z_i = f(X_{i,j}(t_{j,l}, \lambda_{j,m})) + \varepsilon_i$ where i represents the wafer index, j denotes the substep index, l represents the time index, and m corresponds to the wavelength

index. The presence of an error term ε_i represents the noise in the measurements.

In this study, we examine the OES dataset from a classification perspective. To analyze the classification task, the continuous variable z_i is replaced with a categorical variable y_i, representing the quality of the semiconductor process. Specifically, we define two classes: class 0 and class 1. Class 0 corresponds to the situations where $z_i < 0.4$ or $z_i > 0.6$, indicating a poor quality of semiconductor process. On the other hand, class 1 represents the cases where $0.4 \leq z_i \leq 0.6$, indicating a good quality of semiconductor process.

4.1.2 Procedure

We observed that the dataset consists of 307 wafers, with 86 wafers belonging to class 0 and 221 wafers belonging to class 1. This implies that the proportion of wafers representing a good semiconductor process is approximately 72.0%, indicating that the dataset used for the classification analysis is imbalanced.

When visualizing the 307 curves over time at a fixed wavelength for each substep, we can see that the two classes are challenging to distinguish each other. Similarly, when illustrating the 307 curves over wavelength at a specific time for each substep, distinguishing the two classes is also difficult.

Furthermore, the wavelength curves show high similarity across almost all fixed time points. To address this similarity, we adopt an approach that involves calculating the mean value of the T_j time points within each wavelength point for every substep. By calcu-

lating these mean values, we can capture the overall trend of the wavelength variable while simultaneously reducing the dimensionality of the dataset. This technique allows us to summarize the wavelength domain effectively and analyze its common patterns and variations.

In Figure 4.1, we present the curve in the wavelength domain obtained by averaging over the time points for Substep 2. This illustration represents the transformed data with 221 curves of class 1 (shown in green) and 86 curves of class 0 (shown in red).

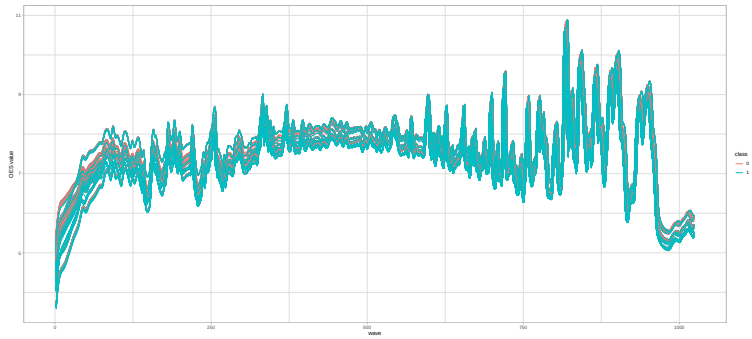

Figure 4.1: Transformed data in wavelength domian by averaging over time points

The main objective of this study is to apply various functional classification models to predict the quality of new semiconductor process based on the covariate X. We apply the classification models described in Chapter 2 and 3. To compare the performance of these models, we use traditional performance measures like Accuracy, which is commonly used for balanced dataset. Additionally, due to the imbalanced nature of the dataset with a majority of good semiconductor process, we also consider Specificity and AUC (Area Under the Curve) as evaluation metrics.

4.1.3 Results

In this study, the total of 307 wafer sets are divided into 230 training sets and 77 test sets. The focus is on Substep 2 curves, as shown in Figure 4.1. When conducting the analysis, the original data is used, not the transformed data. The evaluation of various functional classification models using three metrics is presented in Table 4.1. It is apparent that most models achieve reasonable accuracy, but their specificity and AUC are relatively low. This indicates that the proportion of bad process correctly identified as bad is much lower than the proportion of good process correctly identified as good.

Model	**Accuracy**	**AUC**	**Specificity**
sFPCA	0.77	0.58	0.21
FCTs	0.75	0.57	0.21
FBAG	0.78	0.61	0.26
FBOOST	0.74	0.54	0.11
FRF	0.77	0.55	0.10
FCPCA	0.77	0.58	0.22
FLR-LP	0.80	0.67	0.42
FLR-FLP	**0.82**	**0.69**	**0.42**

Table 4.1: Comparison of Models using Evaluation Metrics (OES dataset)

The sFPCA model, with tuning parameters $\theta = 0.1$ and $\lambda = 10$ and 4 FPCs, achieves an accuracy of 77%. However, it shows relatively low AUC and specificity scores. This can be attributed to

	Prediction	
Reference	**0**	**1**
0	2	1
1	17	57

Table 4.2: Confusion Matrix of FRF using OES dataset

the data imbalance, where there is a large proportion of good process. The same issue is encountered by tree-based models such as FCTs, FBAG, FBOOST and FRF when classifying the data using 20 FPC scores. Also, similar results are obtained using FCPCA with LDA assumptions, where 6 FPCs are used.

Among the models tested, the FLR-FLP model shows better performance compared to the others. Through cross-validation, the optimal values for the hyperparameters are determined as $\lambda_1 = 0.05$ and $\lambda_2 = 0.07$. The FLR-LP model, which only imposes general Lasso constraints by setting $\lambda_2 = 0$, shows similar (slightly worse) performance compared to the FLR-FLP model.

Also in the FLR-FLP model, the wavelength range [712nm, 717nm] is identified as having the most significant impact on the classification task. This finding is valuable as it provides useful insights into the semiconductor process. In this context, developing the FLR-FLP model with a focus on this important wavelength interval holds great promise for enhancing semiconductor production. It is anticipated to produce the superior-quality semiconductors more effectively, thereby leading to a substantial improvement in overall efficiency within the semiconductor manufacturing industry.

	Prediction	
Reference	**0**	**1**
0	8	3
1	11	55

Table 4.3: Confusion Matrix of FLR-FLP using OES dataset

Also, Table 4.2 and 4.3 show the confusion matrix of the FRF and FLR-FLP models, respectively. Overall, both models exhibit low classification performance, especially in terms of Specificity (0 to 0 classification). In Table 4.2, the FRF model correctly classifies 2 out of 19 instances as class 0, and 57 out of 58 instances as class 1. As a result, it shows very low performance in classifying instances as 0 to 0, leading to a very low Specificity and AUC.

In contrast, Table 4.3 demonstrates the improved performance of the FLR-FLP model compared to the FRF model. The FLR-FLP model correctly classifies 8 out of 19 instances as class 0 and 55 out of 58 instances as class 1. Its performance in terms of classifying instances as 0 to 0 is significantly better than that of the FRF model.

In conclusion, the FLR-FLP model shows improvement over other models, particularly in classifying instances as 0 to 0. However, it is evident that this model also still has room for enhancement in achieving higher classification performance. Further development of these models could potentially lead to better overall performance and higher AUC value.

4.2 DIPD dataset

4.2.1 Data description

The Daily Italy Power Demand (DIPD) dataset used in this study is collected from [6]. The dataset consists of one year of Italian Power Demand, represented as a time series. Similar to the classification process used for OES dataset, we consider these data as functional data with respect to time. The objective of the task is to classify two classes, distinguishing days from October to March and from April to September.The dataset was standardized for analysis.

4.2.2 Procedure

The dataset used in this study consists of a total of 67 training sets and 1029 test sets. In this dataset, class 0 represents the time period from October to March, while class 1 represents the time period from April to September. Figure 4.2 shows that the visual representation of the training sets and test sets that the red line corresponds to Class 0, and the green line corresponds to Class 1.

It is important to note that this dataset is balanced unlike OES dataset, with almost equal number of samples for both classes in the training sets. Specifically, there are 33 samples for class 1 and 34 samples for class 0 in the training sets. Similarly, in the test sets, there are 516 samples for class 1 and 513 samples for class 0. This balance between the two classes allows us to focus on accuracy rather than the specific metrics such as specificity.

This time series data is transformed into functional data with

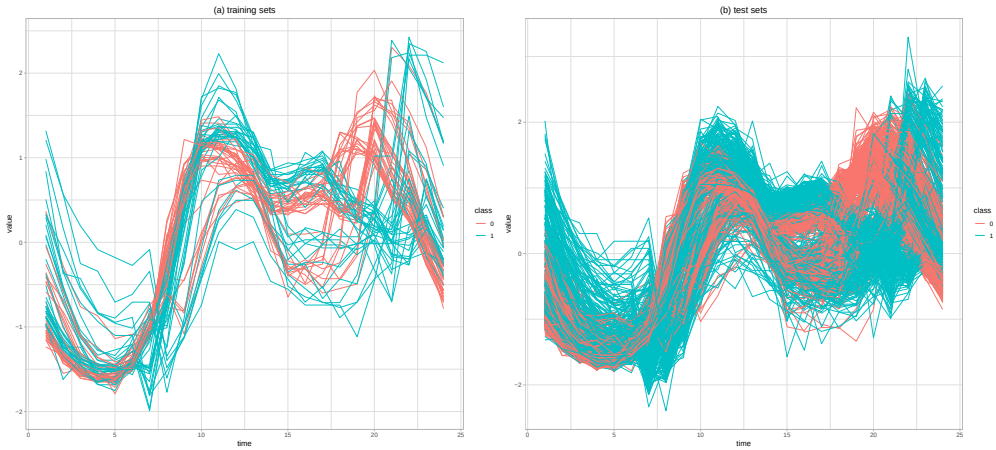

Figure 4.2: Visualization of DIPD dataset

respect to time t for analysis. Using this perspective, we apply the same functional classification methods that are used for the OES dataset. The evaluation of the various functional classification models is done using two performance metrics, accuracy and AUC. The tuning parameters for the models are determined in the same manner as we do for the OES dataset.

4.2.3 Results

Table 4.4 presents the classification performance of various classification models. It is evident that both accuracy and AUC exhibit higher values compared to the OES dataset, mainly due to the clear distinction between the two classes. Consequently, all the models demonstrate great overall model performance. The analysis procedure is consistent across all models, as is the case with analyzing OES dataset.

Model	Accuracy	AUC
sFPCA	0.7794	0.80
FCTs	0.9038	0.90
FBAG	0.9582	0.95
FRF	0.9271	0.92
FBOOST	0.9592	0.96
FCPCA	0.9495	0.95
FLR-LP	0.9312	0.93
FLR-FLP	**0.9757**	**0.98**

Table 4.4: Comparison of Models using Evaluation Metrics (DIPD dataset)

Among the models, FLR-FLP outperforms the others, showing the best performance with an accuracy of 97.57% and an AUC of 0.98. For this classification, we select the tuning parameters $\lambda_1 = 0.05$ and $\lambda_2 = 0.04$. The interval [19 to 20] is identified as the discriminative range between class 0 and class 1. Figure 4.3 visually presents the two classes of curves within the interval [19 $\sim$ 20]. It is evident that the two classes are distinctly separated in this interval. Based on these results, we can infer that this specific time interval plays a crucial role in distinguishing between the summer and winter seasons.

The classification process involves comparing the estimated probability $\widehat{\pi}_i$ to a threshold of 0.5. Observations with probability greater than 0.5 is classified as class 1, while those with probability less than or equal to 0.5 is classified as class 0. Specifically, we can examine confusion matrix of the above models to gain further

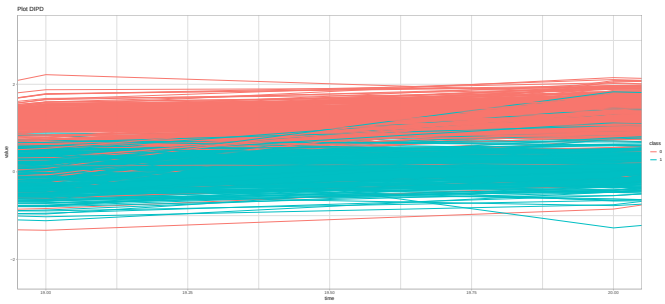

Figure 4.3: Discriminant interval [19,20] identified by FLR-FLP

	Prediction	
Reference	**0**	**1**
0	448	10
1	65	506

Table 4.5: Confusion Matrix of FRF using DIPD dataset

insights into their performance. Table 4.5 and Table 4.6 present the confusion matrix for the FRF and FLR-FLP models, respectively. In Table 4.5, the FRF model correctly classifies 448 out of 513 instances as class 0, and 506 out of 516 instances as class 1. It misclassifies 65 instances of class 1 as class 0 and 10 instances of class 0 as class 1. In Table 4.6, the FLR-FLP model shows improved performance compared to the FRF model. It correctly classifies 499 out of 513 instances as class 0, and 505 out of 516 instances as class 1. It has fewer misclassifications, with 14 instances of class 1 classified as class 0 and 11 instances of class 0 classified as class 1.

Overall, these findings indicate that the FLR-FLP model, along with the identified discriminative range and confusion matrix, is

	Prediction	
Reference	**0**	**1**
0	499	11
1	14	505

Table 4.6: Confusion Matrix of FLR-FLP using DIPD dataset

a more reliable and powerful classification model for the DIPD dataset. It achieves better accuracy in distinguishing between the two classes and exhibits a less number of misclassifications by using specific time interval compared to other models.

Chapter 5

Conclusion

In this paper, we have explored various functional classification models for effectively classifying binary response variable. We have applied recent methods that use Functional Principal Components (FPCs) and the Functional Logistic Model to address classification challenges. Through our analysis of two real datasets, we have found that the model incorporating a fused lasso penalty demonstrates better performance and provides valuable insights, such as identifying significant wavelength interval in OES dataset and time interval in DIPD dataset.

There are several important avenues for further research in the field of functional classification using OES dataset. One important aspect to consider is the development of classification models that can enhance the performance of classification, particularly in terms of Specificity. It is often crucial to accurately classify abnormal instances as abnormal, and future models should focus on achieving this goal.

Additionally, it is essential to explore the classification of mul-

tivariate functional data itself. In our study, we transform multivariate data into univariate data by averaging with respect to time. While this approach has its advantages such as dimension reduction, it also imposes limitations. Future research should concentrate on developing methods that can effectively handle and classify multivariate functional data directly, enabling a more comprehensive analysis.

Furthermore, extending the classification models to address multiclass problem is another valuable direction for future research. While our focus is on binary classification, many real-world cases require the classification of data into multiple classes. Developing efficient multiclass functional classification models would provide valuable tools for a wide range of applications.

In conclusion, we should try to address these challenges and advance traditional methods in functional classification, not only for OES dataset but also for other dataset. This will lead to the development of more accurate classification methods for complex and substantial amounts of dataset across various domains. These improved methods will enable us to effectively analyze and classify functional data, leading to enhanced decision-making process and improved outcomes in diverse fields and industries.

Bibliography

[1] Alfaro, E., Gamez, M., and García, N. (2013). adabag: An R Package for Classification with Boosting and Bagging. *Journal of Statistical Software*, **54(2)**, 1–35.

[2] Allen, G. I. (2013). Multi-way functional principal components analysis. Classification of EEG signals: An interpretable approach using functional data analysis. *2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 220–223.

[3] Chatterjee, A., Mazumder, S., and Das, K. (2023). Functional classwise principal component analysis: a classification framework for functional data analysis. *Data Mining and Knowledge Discovery*, **37**, 1–10.

[4] Febrero-Bande, M., and de la Fuente, M. O. (2012). Statistical computing in functional data analysis: The R package fda.usc. *Journal of statistical Software*, **51(1)**, 1-28.

[5] Fu, A., Narasimhan, B., and Boyd, S. (2020). CVXR: An R package for disciplined convex optimization. *Journal of statistical Software*, **94(14)**, 1–34.

[6] Keogh, E., Wei, L., Xi, A., Lonardi, S., Shieh, J., and Sirowy, S. (2006). Intelligent icons: Integrating lite-weight data mining and visualization into GUI operating systems. *IEEE International Conference on Data Mining, ICDM*, 912–916.

[7] Kim, H., and Kim, H. (2018). Functional logistic regression with fused lasso penalty. *Journal of Statistical Computation and Simulation*, **88(15)**, 2982-2999.

[8] Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, **28(5)**, 1–26.

[9] Leng, X., and Müller, H. G. (2006). Classification using functional data analysis for temporal gene expression data. *Bioinformatics*, **22(1)**, 68–76.

[10] Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. *R News*, **2(3)**, 18–22.

[11] Maturo, F., and Verde, R. (2023). Supervised classification of curves via a combined use of functional data analysis and tree-based methods. *Computational Statistics*, **38(1)**, 419-459.

[12] Nie, Y., Wang, L., Liu, B., and Cao, J. (2018). Supervised functional principal component analysis. *Statistics and Computing*, **28(3)**, 713–723.

[13] Ramsay, J. O., and Silverman, B. (2005). Functional Data Analysis, New York: Springer.

[14] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal*

of the Royal Statistical Society: Series B (Statistical Methodology), **67(1)**, 91-108.

[15] Yi, Y., Billor, N., Liang, M., Cao, X., Ekstrom, A., and Zheng, J. (2022). Classification of EEG signals: An interpretable approach using functional data analysis. *Journal of Neuroscience Methods*, **376**, 109609.

국문초록

반도체 산업에서 데이터의 폭발적인 증가로 인해 효율적인 분석 기술은 필수 요소가 되었다. 반도체 공정 과정으로부터 나오는 Optical Emission Spectroscopy(OES) 데이터는 시간 및 파장 영역에서 함수형 데이터로 간주할 수 있는 복잡한 데이터이다. 이를 고려하여 이 논문은 함수형 데이터 분류에 효과적인 다양한 통계적 방법을 탐구하고 있다. 구체적으로, 이 논문은 함수형 주성분(FPCs)과 함수형 로지스틱 회귀를 이용하여 분류 문제를 해결하는 데 초점을 맞추고 있다. 이 논문의 주요 목적은 다양한 함수형 분류 모델들을 사용하여 OES 데이터를 기반으로 반도체 공정의 성공 여부를 분류하는 것이다. 이를 위해, 이러한 함수형 분류 모델들의 성능을 정확도, 특이도, AUC 등과 같은 평가 지표를 사용하여 비교하였다. 이러한 비교 분석을 통해 위의 함수형 분류 모델들이 반도체 공정의 성공 여부를 정확하게 분류하는 데 얼마나 효과적인지를 평가하고자 한다. 추가적으로 다른 실제 데이터에 대해서도 평가를 확장하여 함수형 분류 모델들의 성능을 종합적으로 조사하였다. 마지막으로, 이 연구의 한계점과 성능이 개선된 모델의 필요성에 대해 논의하였다.

주요어 : 함수형 데이터 분석, 분류, 함수형 주성분 분석, 함수형 분류 트리, 함수형 로지스틱 회귀, fused 라쏘

학 번 : 2021-21150