Cross-Chamber Data Transferability Evaluation for Fault Detection and Classification in Semiconductor Manufacturing

Feng Zhu, Xiaodong Jia, Wenzhe Li, Min Xie, *Fellow, IEEE*, Lishuai Li, *Senior Member, IEEE*, and Jay Lee

***Abstract*—Unit-to-unit variation among the production chambers is a long-lasting challenge for Fault Detection and Classification (FDC) development in the semiconductor industry. Currently, various methods are applied for knowledge transfer among chambers and generalized FDC model development. However, the existing methods cannot give a quantitative or qualitative measure for cross-chamber data transferability evaluation. This research proposes a novel methodology for data transferability evaluation and important sensor screening, which can serve as a data quality evaluation tool for any FDC model. In this research, firstly, Time Series Alignment Kernel (TSAK) is incorporated into Multidomain Discriminant Analysis (MDA) algorithm to achieve sensor-based domain generalization. Then, domain-invariant features are directly extracted for sensor visualization. After that, Fisher's criterion ratios of the labeled good wafer samples and defective ones are computed based on the domain-invariant features of each sensor to quantitatively estimate how easy it is to transfer knowledge of each sensor among chambers, i.e., data transferability evaluation. Lastly, the proposed method develops a Recursive Feature Elimination (RFE)-based sensor selection algorithm to qualitatively analyze the importance of each sensor channel and identify the critical sensor subset. In this study, validation of the proposed method is based on two open-source datasets from real production lines.**

***Index Terms*—Semiconductor, timer series alignment kernel, domain generalization, sensor selection, fault detection and classification.**

I. Introduction

IN MANY semiconductor manufacturing processes, FDC plays a pivotal role in ensuring product quality and consistency [1], [2]. It is superior to relying on post-process metrology as it can quickly detect the defective parts based on the process data (or trace data) [3], [4]. However, chamber discrepancy (or unit-to-unit variation) has been a long-lasting challenge that limits the general application of FDC models in modern fabs because the data from different production chambers tend to form distinctive clusters [5], [6], see Fig. 1 for an example. The FDC model trained based on data from one chamber will have poor detection performance when applied to another one. Therefore, tool-based FDC models are often built as an alternative to making sure that the model can function as excepted [7], [8], [9]. Unfortunately, this is no longer a viable solution as the products are highly mixed nowadays, and hundreds of products are produced concurrently from the production line. As a result, the product combinations made in different chambers are highly diverse, and establishing a tool-based FDC model is extremely difficult for low-run products due to data scarcity. Therefore, the investigation of new solutions is required to establish the FDC model to overcome the unit-to-unit variation.

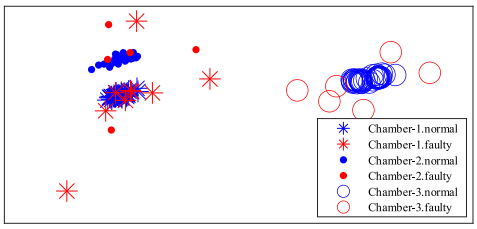


Fig. 1. Visualization of the data distribution for Eigen Vector data set [10]: three clusters are observed in the scatter plot, corresponding to the collected data from three different chambers.

This research attempts to overcome this issue using domain generalization techniques. Our objective is not to establish a Deep Neural Networks (DNNs)-based FDC model to transfer knowledge from one chamber to another because DNNs require a massive number of labeled samples. Besides, DNNs are often criticized as non-transparent due to their multi-layer nonlinear structure. Instead, we employ kernel-based domain generalization techniques to evaluate the transferability of the FDC knowledge across different chambers. Transferability is a measure that can tell us if the sensor channel contains critical information for FDC and how easy it is to transfer such

Manuscript received 30 July 2022; revised 14 October 2022; accepted 7 November 2022. Date of publication 16 November 2022; date of current version 3 February 2023. This work was supported in part by the Research Grants Council (RGC) General Research Fund under Grant CityU 11215119 and Grant CityU 11209717. *(Corresponding author: Xiaodong Jia.)*

Feng Zhu is with the Department of Advanced Design and Systems Engineering, City University of Hong Kong, Hong Kong (e-mail: fenzhu2-c@my.cityu.edu.hk).

Xiaodong Jia, Wenzhe Li, and Jay Lee are with the Center for Intelligent Maintenance Systems, University of Cincinnati, Cincinnati, OH 45221 USA (e-mail: jiaxg@ucmail.uc.edu).

Min Xie is with the Department of Advanced Design and Systems Engineering and the School of Data Science, City University of Hong Kong, Hong Kong (e-mail: minxie@cityu.edu.hk).

Lishuai Li is with the School of Data Science, City University of Hong Kong, Hong Kong (e-mail: lishuai.li@cityu.edu.hk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSM.2022.3222475.

Digital Object Identifier 10.1109/TSM.2022.3222475

0894-6507 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: CITY UNIV OF HONG KONG. Downloaded on May 06,2025 at 02:10:17 UTC from IEEE Xplore. Restrictions apply.

information across different chambers. After quantifying the data transferability of the raw data, the critical sensors contributing to such transferable FDC information are located, and hand-crafted features can be subsequently designed for FDC model training, enhancing the model generalization ability. Till now, data transferability is still a blind spot of research in the industrial area, and there is a lack of clear answers on whether the data from different processing chambers can be transferred. The presented study aims to answer this question by establishing quantitative measures for cross-chamber data transferability evaluation and qualitatively analyzing the sensor importance.

Time Series Alignment Kernel (TSAK), proposed in our previous publication [11], is used to give an objective evaluation of the cross-chamber data transferability. The traditional FDC models take features as input, and the model performance highly depends on the quality of features [9], [12], [13]. Therefore, it cannot be leveraged to give an objective evaluation of sensors in the present study. In comparison, TSAK is a positive-definite (p.d) kernel that takes multivariate time series as input. It has several friendly properties: 1) It is a kernel-based approach and thus can directly handle time variability among different wafer runs (e.g., the time duration of the trace data is different for each wafer run). Compared with the other sequence alignment methods like Dynamic Time Warping (DTW) and Sequence to Sequence (Seq2Seq) model, TSAK can align the time series meanwhile directly calculate the similarity between two time series of various lengths. 2) The kernel is compatible with almost all kernel-based classifiers like Support Vector Classifier (SVC), Gaussian Process Classifier (GPC), and more. 3) It eliminates the need for data preprocessing, subjective feature design and extraction as it can directly take multivariate trace data (raw data) as input. Thus, TSAK is a suitable method for objective evaluation of the cross-chamber data transferability. For other details about TSAK, please refer to our previous work [11].

This paper proposes a novel methodology to evaluate the cross-chamber data transferability and identify the critical sensor subset. The proposed method provides a novel data quality evaluation method for establishing FDC models. The detailed steps in the proposed method include: 1) TSAK is used to construct the kernel or gram matrix with the multivariate trace data; 2) MDA is used to generate the domain-invariant features of each sensor for visualization, addressing the unit-to-unit variation issue; 3) Based on the domain-invariant features of each sensor, Fisher's criterion ratios of the labeled good wafer samples and defective ones are computed for data transferability evaluation. 4) RFE is used to qualitatively analyze the importance of each sensor channel and identify the critical sensor subset. In summary, the contributions and novelties of this research can be summarized as follow: 1) This research considers the problem of sensor screening for multi-chamber from a global perspective for the first time and propose a novel methodology to deal with this problem; 2) The core novelty in the proposed methodology is an innovative way for domain generalization. By incorporating TSAK into MDA, the sensor-based domain generalization can directly extract domain-invariant features without any subjective feature design or data preprocessing; 3) Sensor-based domain generalization provides a novel sensor visualization method which is robust to the unit-to-unit variation among chambers; 4) We provide a measure for data transferability evaluation by quantifying the discriminative power of the domain-invariant features from every single sensor.

This article is organized as follows: Section II introduces the technical background of FDC and reviews the relevant literature about domain generalization techniques. Section III elaborates the proposed methodology for data transferability evaluation and sensor screening. In Section IV, two case studies using open-source datasets from real production lines are demonstrated to validate the proposed method. The concluding remarks are stated in Section V.

II. Technical Background

A. Literature Review

FDC models can be separated into three categories: tool-based models, product-based models, and generalized models. Tool-based models are widely investigated in the semiconductor industry. Hong et al. [14] employed Modular Neural Network (MNN) to model tool data for Fault Detection (FD) and utilized D-S theory for Fault Classification (FC), which can identify multiple faults in a plasma etching system. Chien et al. [15] proposed an FDC framework using Multiway Principal Component Analysis (MPCA) and data mining. Lee et al. [16] proposed FD-Convolutional neural network (CNN) for FC, which can extract fault features and locate the sensor channels and time intervals representing faults. Fan et al. [3] proposed a feature extraction method with supervised classification approaches to find key parameters and identify key steps during the semiconductor manufacturing process. Besides, some sensor screening methods are used to find the critical sensor channels and improve model performance, which are tool-based models. Fan et al. [4] proposed a key parameter identification method using image processing techniques and the application of Fourier transform to detect wafer defects during the manufacturing process. Fan et al. [8] proposed using the random forests algorithm to analyze the importance of equipment sensors. Then, k-means is used to filter the critical sensors. Our previous work [11] proposed a novel approach for important sensor screening by combing TSAK and minimum Redundancy Maximum Relevance (mRMR) framework. However, due to the unit-to-unit variation, the tool-based models are only applicable to their corresponding chambers because they only consider the local characteristic in each chamber. Unit-to-unit variation mainly comes from the following aspects: 1) tool state difference of the chambers over a maintenance cycle; 2) difference of the incoming materials; 3) difference of the recipe setting of the chambers [7], [17]. Because the semiconductor manufacturing process is nonlinear, time-varying, and subject to disturbances, unit-to-unit variation will lead to distribution variation or multimodal batch trajectories in the collected data [18], [19]. Therefore, the tool-based models usually cannot remain a competitive detection performance when being

Authorized licensed use limited to: CITY UNIV OF HONG KONG. Downloaded on May 06,2025 at 02:10:17 UTC from IEEE Xplore. Restrictions apply.

applied to other chambers. Even worse, if a new product line executes on the same chamber, the tool-based models will also be directly obsoleted, due to the chamber bias arose in recipe differentials, product specification, etc.

Product-based models entail transferring knowledge acquired from the previously established FDC model to aid the training of the FDC model for the new products. It avoids recourse to a time-consuming process for the FDC data accumulation of new products. Fan et al. [20] proposed a novel product-to-product (P2P) VM model to predict photoresist spacer heights in the color filter process. However, the proposed model is restricted to the different products being processed in the same chamber. It cannot address the unit-to-unit variation issue for now. There is little research about the product-based FDC models.

Generalized models consider the global characteristics of different equipment, which can be used for process monitoring of multi-chamber. However, each chamber is required to have a certain number of training samples for model development. The data accumulation of new chambers is still time-consuming. To establish a generalized FDC model, *k*-Nearest Neighbor (*k*NN) was employed as a popular algorithm to handle the unit-to-unit variation. He and Wang [18] proposed a *k*NN-based FD method that detects anomalies by evaluating the observation distance to the normal operating region. Zhou et al. [21] proposed a fault detection method based on random projection and *k*NN rule, where random projection is used for distance preservation to ensure the model robustness after dimension reduction. Zhang et al. [22] proposed a fault detection strategy based on the weighted distance of *k*NN, which provided a new statistic that can eliminate the influence of variance structure in multimodal processes and reduce the autocorrelation of statistic values. In general, the current generalized models are effective for fault detection with the unit-to-unit variation, but there are two limitations: 1) *k*NN does not perform well in the multi-class FC problem because there may be confusion among different classes in the collected data from all chambers; 2) There is still a lack of a generalized sensor screening method to consider the data transferability of each sensor and identify critical sensor subset for multi-chamber from a global perspective.

B. Domain Generalization Techniques

To address the research limitations mentioned above, this research will use domain generalization techniques to handle the unit-to-unit variation issue. Domain generalization is a generalization-related research topic to deal with distribution gaps and enhance the generalization ability of machine learning models [23]. The general idea of domain generalization is to learn a domain-invariant representation with stable distribution from all source domains [24]. Kernel-based projection is one of the main techniques to solve domain generalization problems, which has been widely investigated. Ghifary et al. [25] proposed Scatter Component Analysis (SCA) by combining Kernel Principal Component Analysis (KPCA), Kernel Fisher Discriminant Analysis (KFDA), and domain scatter in a single objective function. Then, spectral decomposition is used to solve the optimization problem to find a domain-invariant representation. Li et al. [26] proposed Conditional Invariant Domain Generalization (CIDG) under the assumption of conditional shift [27], which is an enhanced variant of SCA. It can learn a feature representation with domain-invariant class conditional distributions. Hu et al. [28] proposed Multidomain Discriminant Analysis (MDA), which further relaxes the causally motivated assumption in CIDG. It can be applicable for domain generalization under the generalized target shift [27].

Domain generalization can radically eliminate the effect of the unit-to-unit variation, which performs well in both FD and FC problems. Moreover, by incorporating TSAK into kernel-based domain generalization methods, domain-invariant features can be directly extracted from raw trace signal data of each sensor channel without any feature design or data preprocessing. Thus, the extracted features from each sensor can be used to give an objective evaluation of the cross-chamber data transferability and sensor importance. The good discriminative power of the features represents that the corresponding sensor has critical and transferable information for FDC, implying that it has high data transferability across the different chambers. Domain-invariant features offer an opportunity to assess the data transferability of each sensor, develop a robust sensor screening method and build a generalized FDC model from a global perspective.

III. Methodology

A. Overview

Before introducing the proposed methodology, the following assumptions need to be stated first: 1) The FDC tasks are identical for different chambers, which means that the chamber condition labels are shared; 2) The number and types of sensors used for process monitoring on all chambers should be the same. Supposing there are labeled trace datasets collected from several chambers with the unit-to-unit variation, this research aims to develop a sensor screening method to identify the critical sensor subset with valuable information for FDC development.

An overview of the proposed methodology is shown in Fig. 2. The proposed method includes four major stages: (1) Kernel matrix construction, (2) Domain generalization, (3) Data transferability evaluation and (4) RFE-based sensor selection. Based on the investigation and benchmark for TSAKs in [11], Global Alignment (GA) kernel is recommended as the most suitable TSAK for semiconductor FDC considering its good accuracy and the p.d property. Thus, in stage 1, GA kernel is used to construct the kernel matrix with the trace data of each sensor. In stage 2, MDA is proposed for domain generalization, considering its relaxed causally motivated assumption and better accuracy [28]. Based on the kernel matrix generated by TSAK, we obtain the domain-invariant features by using MDA for each sensor channel, which can be directly used for sensor importance visualization. The proposed sensor visualization method is robust to the unit-to-unit variation. In stage 3, Fisher's criterion ratio is

Authorized licensed use limited to: CITY UNIV OF HONG KONG. Downloaded on May 06,2025 at 02:10:17 UTC from IEEE Xplore. Restrictions apply.

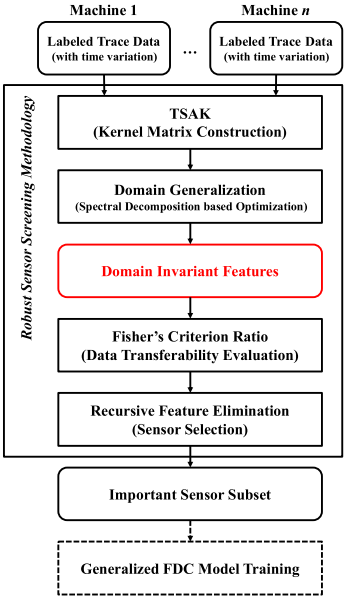


Fig. 2. Flow chart for the proposed methodology.

used for data transferability evaluation. At last, RFE framework is used for sensor selection to qualitatively analyze the sensor importance and identify the critical sensor subset.

B. Global Alignment Kernel

Let π be the alignment between two discrete time series $\mathbf{x} = (x_1, \ldots, x_a)$ and $\mathbf{y} = (y_1, \ldots, y_b)$ of lengths a and b, respectively. Then π is a pair of increasing vectors $(\pi^{\mathbf{x}}, \pi^{\mathbf{y}})$ of length $p \le a + b - 1$, then $1 = \pi_1^{\mathbf{x}} \le \cdots \le \pi_p^{\mathbf{x}} = a$ and $1 = \pi_1^{\mathbf{y}} \le \cdots \le \pi_p^{\mathbf{y}} = b$ are the warping path for $\mathbf{x}$ and $\mathbf{y}$ respectively. Based on the alignment π, GA kernel can be calculated as below:

$$\begin{aligned} k_{GA}(\mathbf{x}, \mathbf{y}) &\overset{\text{def}}{=} \sum_{\pi \in \mathcal{A}(\mathbf{x},\mathbf{y})} e^{-\sum_{i=1}^{p} \varphi(\mathbf{x}(\pi_i^{\mathbf{x}}), \mathbf{y}(\pi_i^{\mathbf{y}}))} \\ &= \sum_{\pi \in \mathcal{A}(\mathbf{x},\mathbf{y})} \prod_{i=1}^{p} e^{-\varphi(\mathbf{x}(\pi_i^{\mathbf{x}}), \mathbf{y}(\pi_i^{\mathbf{y}}))} \\ &= \sum_{\pi \in \mathcal{A}(\mathbf{x},\mathbf{y})} \prod_{i=1}^{p} k_L\big(\mathbf{x}(\pi_i^{\mathbf{x}}), \mathbf{y}(\pi_i^{\mathbf{y}})\big), \end{aligned} \tag{1}$$

where $\mathcal{A}(\mathbf{x}, \mathbf{y})$ is the set of all possible alignment paths, φ is defined as the squared Euclidean distances $\varphi(x, y) = \|x - y\|^2$, then the local kernel function k_L induced from the φ as $k_L \overset{\text{def}}{=} e^{-\varphi}$. According to (1), k_{GA} is defined as the summation of exponentiated distance over all possible alignment paths. Then, there will not be any cost or risk of the bad warping path. Moreover, the theoretical discussion in [29] indicates the p.d property of GA kernel. The computation of GA kernel can be stated as:

$$M_{i,j} = \big(M_{i-1,j} + M_{i-1,j-1} + M_{i,j-1}\big) \cdot k_L\big(x_i, y_j\big), \tag{2}$$

where $M_{a,b}$ is GA distance between $\mathbf{x}$ and $\mathbf{y}$. GA kernel directly takes trace data as algorithm inputs and measures the discrepancy between two time series with different lengths. Then, we can construct the kernel matrix with time series data in most kernelized machine learning algorithms by using GA kernel. It is essential to note the following details for the implementation of GA kernel: 1) If one sequence is more than two times longer than another sequence, GA kernel may produce a diagonal dominant Gram matrix. Triangular GA (TGA) kernel can better address this issue. Interested readers can refer to [29] for more details; 2) Regarding the parameter tuning strategy for GA kernel, readers can follow the instruction given in [29]; 3) In the following discussion, the normalized counterpart $\tilde{k}(\mathbf{x}, \mathbf{y})$, which can be described by $k(\mathbf{x}, \mathbf{y})/\sqrt{k(\mathbf{x}, \mathbf{x}) \cdot k(\mathbf{y}, \mathbf{y})}$, will be employed for further algorithm development.

C. Multidomain Discriminant Analysis

Before introducing MDA algorithms, we briefly revisit the kernel mean embedding, an important mathematical tool for representing and comparing probability distributions. Kernel mean embedding represents distributions as elements of a reproducing kernel Hilbert space (RKHS) [30], [31]. An RKHS $\mathcal{H}$ on domain $\mathcal{X}$ associated with a kernel $k(\cdot, \cdot)$ on $\mathcal{X} \times \mathcal{X}$ is a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$. Denoting its inner product by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, RKHS $\mathcal{H}$ meets the reproducing property $\langle f(\cdot), k(\mathbf{x}, \cdot) \rangle_{\mathcal{H}} = f(\mathbf{x})$, where $\phi(\mathbf{x}) \coloneqq k(\mathbf{x}, \cdot)$ represents the canonical feature map of $\mathbf{x}$ such that $\phi(\mathbf{x}) \in \mathcal{H}$. Given two observations $\mathbf{x}_1^s \in \mathcal{X}$ and $\mathbf{x}_2^s \in \mathcal{X}$ from domain s, we have $\langle \phi(\mathbf{x}_1^s), \phi(\mathbf{x}_2^s) \rangle = k(\mathbf{x}_1^s, \mathbf{x}_2^s)$. Then, the kernel embedding of a distribution $\mathbb{P}(\mathbf{X})$ is defined as:

$$\mu_{\mathbf{X}} \coloneqq E_{\mathbf{X} \sim \mathbb{P}(\mathbf{X})}[\phi(\mathbf{X})] = E_{\mathbf{X} \sim \mathbb{P}(\mathbf{X})}[k(\mathbf{X}, \cdot)]. \tag{3}$$

If the kernel k is characteristic, $\mu_{\mathbf{X}}$ will be injective and can capture all information of the distribution $\mathbb{P}(\mathbf{X})$, which means that $\|\mu_{\mathbf{X}} - \mu_{\mathbf{X}'}\|_{\mathcal{H}} = 0$ if and only if $\mathbb{P}(\mathbf{X})$ and $\mathbb{P}(\mathbf{X}')$ are the same [31]. Although the kernel mean embedding cannot be directly computed, it can be estimated by using observations. Given $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^n \in \mathbb{P}(\mathbf{X})$, where n is the sample size of the domain $\mathbb{P}(\mathbf{X})$, the kernel mean embedding can be empirically estimated by:

$$\widehat{\mu}_{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} \phi(\mathbf{x}_i) = \frac{1}{n} \sum_{i=1}^{n} k(\cdot, \mathbf{x}_i). \tag{4}$$

Four regularization measures have been derived based on the kernel mean embedding to formulate a feature learning algorithm referred to as MDA, including 1) Average domain discrepancy, 2) Average Class Discrepancy, 3) Multidomain between-class scatter, 4) and Multidomain within-class scatter.

Average domain discrepancy, Ψ^{add}, is used to measure the discrepancy of the class-conditional distributions. Given a set of all class-conditional distributions $\mathcal{P} = \{\mathbb{P}_j^s\}$ for $s \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, c\}$, where s is the domain index, and j is the class index, denote the kernel mean embedding of $\mathbb{P}_j^s$ by μ_j^s. Then, Ψ^{add} is defined as

$$\Psi^{\text{add}} = \frac{1}{c\binom{m}{2}} \sum_{j=1}^{c} \sum_{1 \le s < s' \le m} \left\| \mu_j^s - \mu_j^{s'} \right\|_{\mathcal{H}}^2, \tag{5}$$

Authorized licensed use limited to: CITY UNIV OF HONG KONG. Downloaded on May 06,2025 at 02:10:17 UTC from IEEE Xplore. Restrictions apply.

where $\binom{m}{2}$ is the number of 2-combinations from a set of m elements.

Average class discrepancy, Ψ^{acd}, is used to avoid that the means of class-conditional distribution of different classes are close, which is defined as

$$\Psi^{\text{acd}} = \frac{1}{\binom{c}{2}} \sum_{1\le j<j'\le c} \left\| \mu_j - \mu_{j'} \right\|_{\mathcal{H}}^2, \tag{6}$$

where $\mu_j = \sum_{s=1}^{m} \mathbb{P}_j^s \mu_j^s$ is the mean representation of class j in RKHS $\mathcal{H}$.

Multidomain between-class scatter, Ψ^{mbs}, and within-class scatter, Ψ^{mws}, are used to measure the discriminative power in RKHS $\mathcal{H}$, which are derived from Kernel Fisher Discriminant Analysis (KFDA). Ψ^{mbs} is defined as

$$\Psi^{mbs} = \frac{1}{n} \sum_{j=1}^{c} n_j \left\| \mu_j - \overline{\mu} \right\|_{\mathcal{H}}^2, \tag{7}$$

where n_j is the total number of instances in class j, and $n = \sum_{j=1}^{c} n_j$. $\overline{\mu} = \sum_{j=1}^{c} \mathbb{P}_j \mu_j$ is the mean representation of the entire set in RKHS $\mathcal{H}$. Ψ^{mws} is defined as

$$\Psi^{mws} = \frac{1}{n} \sum_{j=1}^{c} \sum_{s=1}^{m} \sum_{i=1}^{n_j^s} \left\| \phi\left(\mathbf{x}_{i\in j}^s\right) - \mu_j \right\|_{\mathcal{H}}^2, \tag{8}$$

where $\mathbf{x}_{i\in j}^s$ denotes the feature vector of ith instance of class j in domain s.

Based on these four regularization measures, MDA aims to search a transformation from RKHS $\mathcal{H}$ to a q-dimensional space $\mathbb{R}^q$, $\mathbf{W} : \mathcal{H} \to \mathbb{R}^q$, to eliminate the distribution variation among different domains. Let $\boldsymbol{DS} = \{\mathbf{x}_i, y_i\}_{i=1,\ldots,n}$ be the dataset from all m domains ($n = \sum_{s=1}^{m} n^s$, where n^s is the number of samples in the s-th domain). Kernel function can give a feature mapping $\phi : \mathbb{R} \to \mathcal{H}$, and define a set of functions arranged in a column vector $\boldsymbol{\Phi} = [\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_n)]^{\mathrm{T}}$. Then, $\mathbf{W}$ can be expressed as a linear combination of all canonical feature maps in $\boldsymbol{\Phi}$, i.e., $\mathbf{W} = \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{B}$, where $\mathbf{B} \in \mathbb{R}^{n\times q}$ is the coefficient matrix of canonical feature maps [32].

Considering the property of norm in RKHS, Ψ^{add} can be computed as

$$\Psi^{\text{add}} = \text{tr}\left(\frac{1}{c\binom{m}{2}} \sum_{j=1}^{c} \sum_{1\le s<s'\le m} \left(\mu_j^s - \mu_j^{s'}\right)\left(\mu_j^s - \mu_j^{s'}\right)^{\mathrm{T}} \right), \tag{9}$$

where $\text{tr}(\cdot)$ denotes the trace operator. After applying the transformation $\mathbf{W}$, the Ψ^{add} in the transformed feature space in $\mathbb{R}^q$ can be computed as

$$\Psi_{\mathbf{B}}^{\text{add}} = \text{tr}\left(\mathbf{B}^{\mathrm{T}}\mathbf{G}\mathbf{B}\right), \tag{10}$$

where

$$\mathbf{G} = \frac{1}{c\binom{m}{2}} \sum_{j=1}^{c} \sum_{1\le s<s'\le m} \boldsymbol{\Phi}\left(\mu_j^s - \mu_j^{s'}\right)\left(\mu_j^s - \mu_j^{s'}\right)^{\mathrm{T}} \boldsymbol{\Phi}^{\mathrm{T}}$$

Similarly, the other three regularization measures in the transformed feature space in $\mathbb{R}^q$ are given by

$$\begin{aligned} \Psi_{\mathbf{B}}^{acd} &= \text{tr}\left(\mathbf{B}^{\mathrm{T}}\mathbf{F}\mathbf{B}\right), \\ \Psi_{\mathbf{B}}^{mbs} &= \text{tr}\left(\mathbf{B}^{\mathrm{T}}\mathbf{P}\mathbf{B}\right), \\ \Psi_{\mathbf{B}}^{mws} &= \text{tr}\left(\mathbf{B}^{\mathrm{T}}\mathbf{Q}\mathbf{B}\right), \end{aligned} \tag{11}$$

where $\mathbf{F} = \frac{1}{\binom{c}{2}} \sum_{1\le j<j'\le c} \boldsymbol{\Phi}(\mu_j - \mu_{j'})(\mu_j - \mu_{j'})^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}$, $\mathbf{P} = \frac{1}{n}\sum_{j=1}^{c} n_j \boldsymbol{\Phi}(\mu_j - \overline{\mu})(\mu_j - \overline{\mu})^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}$, $\mathbf{Q} = \frac{1}{n}\sum_{j=1}^{c}\sum_{s=1}^{m}\sum_{i=1}^{n_j^s} \boldsymbol{\Phi}(\phi(\mathbf{x}_{i\in j}^s) - \mu_j)(\phi(x_{i\in j}^s) - \mu_j)^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}$.

At last, MDA unifies all the measure and seek the transformation by solving an optimization problem in the form of the following expression:

$$\underset{\mathbf{B}}{\text{argmax}} \frac{\Psi_{\mathbf{B}}^{acd} + \Psi_{\mathbf{B}}^{mbs}}{\Psi_{\mathbf{B}}^{add} + \Psi_{\mathbf{B}}^{mws}} \tag{12}$$

It is noted that the maximization of the objective function can preserve the discriminative power among different classes while improving the overall compactness of distributions of all classes to make the class-conditional distributions of the same class as close as possible. By substituting (10) and (11), adding the trade-off parameters to control the significance of each measure, and adding $\mathbf{W}^{\mathrm{T}}\mathbf{W} = \mathbf{B}^{\mathrm{T}}\mathbf{K}\mathbf{B}$ for regulation, the objective function (12) can be reformulated as

$$\underset{\mathbf{B}}{\text{argmax}} \frac{\text{tr}\left(\mathbf{B}^{\mathrm{T}}(\beta\mathbf{F} + (1-\beta)\mathbf{P})\mathbf{B}\right)}{\text{tr}\left(\mathbf{B}^{\mathrm{T}}(\gamma\mathbf{G} + \alpha\mathbf{Q} + \mathbf{K})\mathbf{B}\right)}, \tag{13}$$

where α, β and γ are the trade-off parameters. Because the objective (13) is consistent with the re-scaling of $\mathbf{B}$, it can be transformed as the following generalized eigenvalue problem based on its Lagrangian

$$(\beta\mathbf{F} + (1-\beta)\mathbf{P})\mathbf{B} = (\gamma\mathbf{G} + \alpha\mathbf{Q} + \mathbf{K})\mathbf{B}\boldsymbol{\Gamma} \tag{14}$$

where $\boldsymbol{\Gamma} = \text{diag}(\lambda_1, \ldots, \lambda_q)$ is the diagonal matrix collecting q leading eigenvalues, and $\mathbf{B}$ contains the corresponding eigenvectors. Interested readers can refer to [28, Algorithm 1 and Appendix B] for more details.

As a summary, this subsection revisited MDA algorithm. Before moving to the next subsection, it is essential to note the following details for implementation: 1) GA kernel perfectly matches with MDA. By incorporating GA kernel, MDA can directly take the multivariate trace data as the input; 2) Because there is no testing dataset in the sensor selection problem, we only need to generate a domain-invariant feature space for all existing datasets. Therefore, the trade-off parameters are selected by using kNN-based Leave-one-out Cross-Validation (kNN-LOOCV) based on the extracted features from MDA; 3) Regarding the scope of all trade-off parameters tuning, readers can follow the instructions given in [28, Appendix E].

D. Data Transferability Evaluation and Sensor Selection

After the domain generalization, the domain-invariant features of each sensor can be obtained. As mentioned above, the good discriminative power of the domain-invariant features

Authorized licensed use limited to: CITY UNIV OF HONG KONG. Downloaded on May 06,2025 at 02:10:17 UTC from IEEE Xplore. Restrictions apply.

represents the corresponding sensor can contribute transferable or domain invariant information for generalized FDC modeling. Thus, the high data transferability is implied by the good discriminative power. Fisher's criterion ratio is used to quantify the discriminative power of the domain-invariant features extracted from every single sensor, which refers to the data transferability score. It can be calculated by

$$\mathbf{S}_w = \sum_{j=1}^{c} \sum_{i=1}^{n_j} \left(\mathbf{x}_{i\in j} - \mu_j\right)\left(\mathbf{x}_{i\in j} - \mu_j\right)^{\mathrm{T}},$$

$$\mathbf{S}_b = \sum_{j=1}^{c} n_j \left(\mu_j - \mu\right)\left(\mu_j - \mu\right)^{\mathrm{T}},$$

$$\text{Data transferability score} = \mathrm{tr}(\mathbf{S}_b)/\mathrm{tr}(\mathbf{S}_w), \tag{15}$$

where $\mathbf{x}_{i\in j}$ denotes the feature vector of i-th instance of class j, μ_j represents the mean vector of class j, μ represents the mean vector of all samples, n_j represents the number of samples in class j. Data transferability score serves as a new metric for sensor importance evaluation in a multi-chamber environment.

However, the data transferability score cannot be directly used for sensor selection because it has not considered the performance of features or sensors in the model development and may eliminate part of the information. RFE is a robust feature selection framework for classification problems that utilize a specific classifier to qualitatively evaluate the feature subset and remove the weakest feature until the specified number of features is reached [33]. Features are ranked by the model's attributes and recursively eliminate a small number of features per loop. RFE attempts to eliminate dependencies and collinearity in the feature set and keep all the crucial information for the classification model. In this study, we preserve the idea of RFE and adapt it for sensor selection.

Let us denote the feature set obtained by FeaSet $= \{\mathbf{Z}_i\}_{i=1}^{g}$, where $\mathbf{Z}_i$ represents the domain-invariant features of i-th sensor, g represents the total number of installed sensors. Based on the notation, the proposed algorithm for important sensor selection is listed in Algorithm 1.

In Algorithm 1, each loop will remove the features corresponding to one useless sensor, which can be found by exhaustive search. In the exhaustive search, all the features of sensors in set $\mathbf{ss}'$ will be combined as a single feature matrix, and kNN-based LOOCV is used to evaluate its quality for model performance. Also, other simple classifier and validation methods can be used for evaluation, such as SVM, logistic regression, etc. Then, the sensor subset with the highest accuracy can enter the next round of screening. Theoretically, as the sensor continues to be removed, the accuracy will gradually increase. The feature elimination process will continue until the accuracy decreases, and then the final critical sensor subset can be identified.

Algorithm 1 RFE for Sensor Selection

Input: Feature Set $FeaSet = \{\mathbf{Z}_i\}_{i=1}^{g}$
Output: Sensor set **ss**
$\mathbf{ss} \leftarrow \{1, 2, \ldots, g\}$
$acc^1 = 0$
$acc^2 = 0$
while $acc^1 \geq acc^2$ **do**
  $acc^2 = acc^1$
  for $i = 1$ **to** length(**ss**) **do**
    $\mathbf{ss}' \leftarrow \mathbf{ss} \backslash \mathbf{ss}(i)$
    $acc_i = \text{knnModelEvaluation}(\mathbf{ss}')$
  end for
  $i\_remove = \text{argmax}_i(acc_i)$
  $\mathbf{ss} \leftarrow \mathbf{ss} \backslash \mathbf{ss}(i\_removed)$
  $acc^1 = \max(acc_i)$
end while
return ss

TABLE I
THE MACHINE STATES USED FOR PROCESS MONITORING

1. BCl_3 flow	8. RF tuner	15. TCP impedance
2. Cl flow	9. RF load	16. TCP top power
3. RF bottom power	10. Phase error	17. TCP reflected power
4. RF bottom reflected power	11. RF power	18. TCP load
5. Endpoint A detector	12. RF power	19. Vat valve
6. Helium pressure	13. TCP tuner	
7. Chamber pressure	14. TCP phase error	

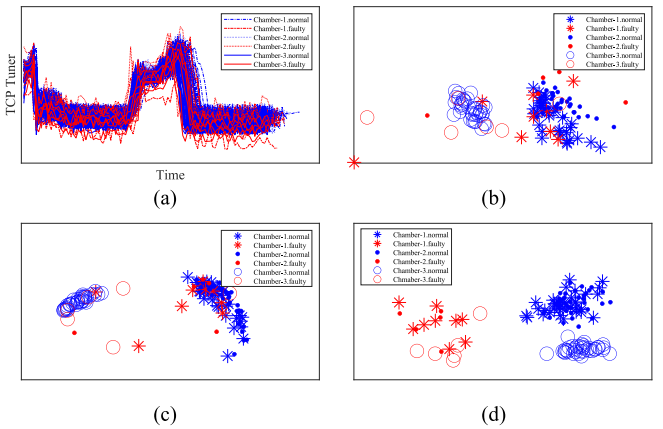


Fig. 3. Demonstration of a single useful sensor (Sensor #13): (a) Raw trace signal; (b) Feature visualization; (c) Sensor visualization based on TSAK+KPCA; (d) Robust sensor visualization based on TSAK+MDA (Data transferability score: 0.84).

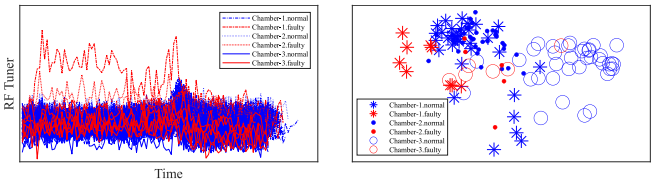
IV. Results and Discussion

A. Case Study 1: Eigen Vector Dataset

The case study will validate the proposed method in the FD problem. The dataset is collected on a commercial scale LAM 9600 TCP metal etcher at Texas Instruments, which consists of the engineering variables over the course of etching 129 wafers [10]. Nineteen machine state signals were used for process monitoring, as listed in Table I. The experiment was performed three times with the widely spread interval, and then the process drift is apparent in the data. This data is commonly studied in the literature for FD model validation and benchmarking.

Fig. 3 proves the effectiveness of the proposed method for sensor-based domain generalization, which does not require any feature design and extraction. Sensor #13 is taken as an example. Fig. 3 (a), (b) and (c) show the raw trace signal and scatter plot in PC space based on the traditional PCA and TSAK+KPCA we proposed in [11]. Even if this sensor contains useful information for FDC, it is challenging to

Authorized licensed use limited to: CITY UNIV OF HONG KONG. Downloaded on May 06,2025 at 02:10:17 UTC from IEEE Xplore. Restrictions apply.



(a) (b)

Fig. 4. Demonstration of a single useless sensor (Sensor #8): (a) Raw trace signal; (b) Robust sensor visualization based on TSAK+MDA (Data transferability score: 0.06).

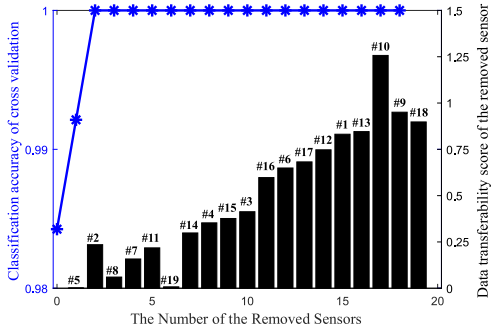
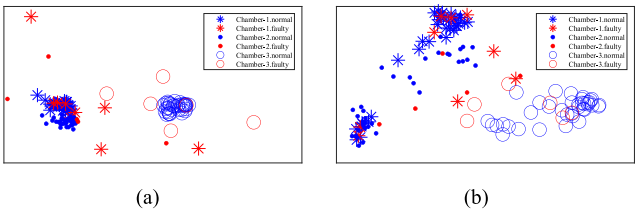


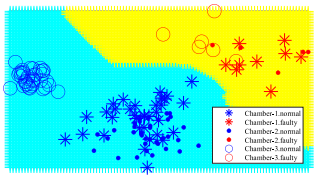
Fig. 5. RFE results and data transferability score of the removed sensor channels.

distinguish between normal and faulty samples using these two visualization methods. Thus, they are not practical for sensor importance evaluation and further FDC modeling in this scenario. Fig. 3 (d) show the scatter plot of the domain-invariant features extracted by TSAK+MDA. One can see that the samples from different classes are separate. Sensor #13 is therefore considered to have high data transferability considering the good discriminative power. For comparison, Fig. 4 shows the raw trace signal and the scatter plot obtained by TSAK+MDA of a useless sensor (Sensor #8). The extracted features cannot show a good separation if the sensor has no helpful and transferable information, making it easy to distinguish between useful and useless sensors. We can evaluate the data transferability of each sensor by observing the discriminative power of their domain-invariant features, which provides a sensor importance visualization way. Also, we can directly quantify the discriminative power by using Fisher's criterion ratio, obtaining the data transferability score of each sensor. The scores of these two sensors are listed in captions for reference.

Fig. 5 verifies the effectiveness of the proposed RFE-based sensor selection method, which is used to remove the useless sensor channels. The classification accuracy changes during the process of the RFE, and the data transferability score of the sensor channels removed per loop are also shown. One can see the classification accuracy keeps increasing as the sensor channels are removed, which demonstrates that the proposed sensor selection methods can remove useless sensors to improve model performance. The sensors with low data transferability scores will be removed first. After removing two sensor channels, the classification accuracy has achieved 100%. After reaching the best performance, users can further decide the



(a) (b)



(c)

Fig. 6. Demonstration of selected sensor group: (a) Feature visualization; (b) Sensor visualization based on TSAK+KPCA; (c) Robust sensor visualization based on TSAK+MDA.

TABLE II
CLASSIFICATION ERROR USING LOOCV

Before Sensor Selection			
	Total error	Type I error	Type II error
SVM	12.75%	0%	12.75%
PCA	14.96%	0%	14.96%
KPCA	15.75%	1.57%	14.17%
MDA	1.57%	0%	1.57%
After Sensor Selection			
	Total error	Type I error	Type II error
SVM	14.96%	3.94%	11.02%
PCA	16.54%	3.15%	13.39%
KPCA	13.39%	4.72%	8.66%
MDA	0%	0%	0%

selected sensor set based on the data transferability score of each sensor, considering the scores of part sensors are still low.

Fig. 6 demonstrates that the domain-invariant features extracted by TSAK+MDA can be used for generalized model development, handling the unit-to-unit variation. Based on the data transferability evaluation and RFE-based sensor selection, sensor #13, #10, #9, and #18 are selected for sensor subset visualization and model validation. Similarly, Fig. 6 (a) and (b) have not shown a clear separation between different classes. However, Fig. 6 (c) indicates a good separation between healthy and faulty, where the healthy area was marked by blue and the faulty area was marked by yellow. The boundary was obtained by using the MAP estimation in kNN model [34]. Based on the domain-invariant features of the selected sensors, a standard kNN model can achieve high accuracy. Table II lists the classification accuracy before and after sensor selection, which further proves the performance of TSAK+MDA for generalized modeling and the effectiveness of the proposed data transferability evaluation and sensor selection method.

B. Case Study 2: CMU Dataset

The second case study will validate the proposed method in the wafer dataset donated by Carnegie Mellon University (CMU). The dataset is a collection of in-process control measurements recorded from an etching process. Six

Authorized licensed use limited to: CITY UNIV OF HONG KONG. Downloaded on May 06,2025 at 02:10:17 UTC from IEEE Xplore. Restrictions apply.

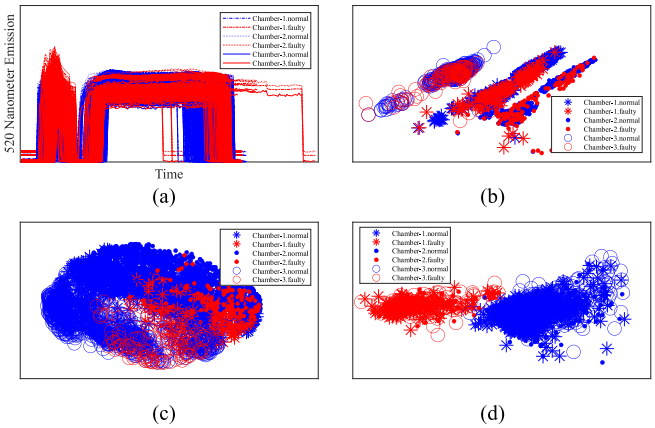


Fig. 7. Demonstration of a single useful sensor (Sensor #5): (a) Raw trace signal; (b) Feature visualization; (c) Sensor visualization based on TSAK+KPCA; (d) Robust sensor visualization based on TSAK+MDA (Data transferability score: 0.6037).

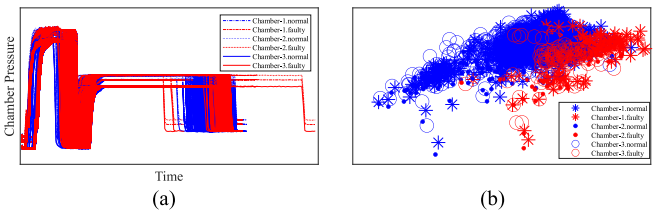


Fig. 8. Demonstration of a single useless sensor (Sensor #3): (a) Raw trace signal; (b) Robust sensor visualization based on TSAK+MDA (Data transferability score: 0.1237).

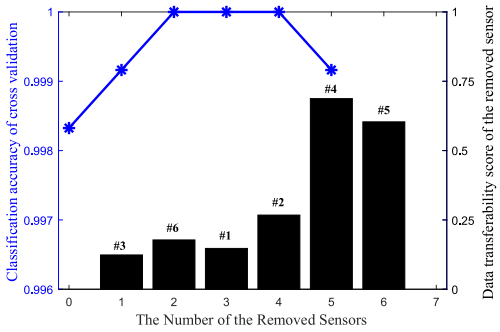


Fig. 9. RFE results and data transferability score of the removed sensor channels.

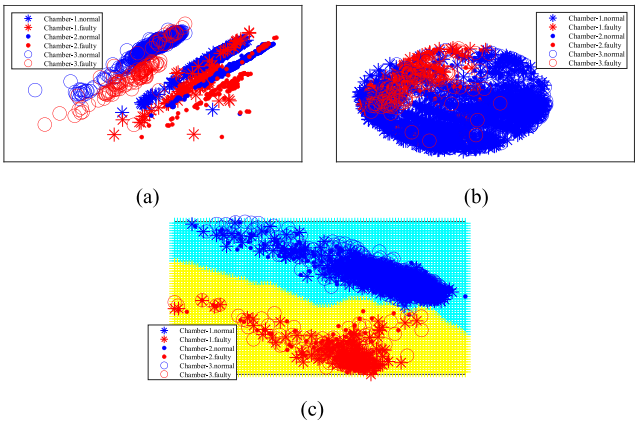


Fig. 10. Demonstration of selected sensor group: (a) Feature visualization; (b) Sensor visualization based on TSAK+KPCA; (c) Robust sensor visualization based on TSAK+MDA.

sensor channels are given for process monitoring, including 1) Radio frequency forward power; 2) Radio frequency reflected power; 3) Chamber pressure; 4) 405 nanometer (nm) emission; 4) 520 nanometer (nm) emission; 6) Direct current bias. More details about this dataset can be found in the research of [35].

Because this dataset only contains process measurements collected from one chamber, data augmentation is necessary to simulate the unit-to-unit variation issue. Unit-to-unit variation always leads to distribution variation or multimodal batch trajectories in the collected data. Thus, we choose to add amplitude drift with the white noise to the raw trace signal to generate a new domain. Suppose $\mathbf{t} = (t_1, \ldots, t_k)$ of length k is one of the sensor channel readings in the one wafer samples in CMU dataset. We can add amplitude drift to all the trace data by $t_i^{new} = t_i + e_i$ to simulate the multimodal batch trajectories, where $e_i \in \mathbb{R} \sim ([\sigma, \Sigma^2])$. For different sensor channels, the σ and Σ^2 of the amplitude drift could be varied. We generate two additional domains for the method validation. Fig. 7 (a) and Fig. 8 (a) show the raw trace signal of two sensor channels from the CMU dataset (Chamber 1) and two simulation datasets (Chamber 2 and Chamber 3). Comparing to the first case, the data volume significantly increased.

Fig. 7 proves the effectiveness of the proposed method for sensor-based domain generalization and robust sensor visualization in the CMU data set. Sensor #5, which has already been selected as a useful sensor for investigation in [35], is taken as an example. Fig. 7 (a), (b), and (c) demonstrate the raw trace signal and scatter plot in PC space based on PCA and TSAK+KPCA. In the case of increased data volume, it is more difficult to distinguish between normal and faulty samples. Fig. 7 (d) shows the scatter plot obtained by the proposed sensor-based domain generalization based on TSAK+MDA. The results show a clear separation between the two classes. Similarly, the domain generalization results of a useless sensor (Sensor #3), as shown in Fig. 8, have not shown a good separation. Fig. 7 and Fig. 8 prove the effectiveness of the proposed sensor-based domain generalization, and the domain-invariant features help visualize the sensor importance and compute the data transferability score. The data transferability scores of these two sensors are given, and Sensor # 5 has the higher data transferability naturally.

Fig. 9 verifies the effectiveness of the proposed RFE sensor selection method for this dataset. The classification accuracy keeps increasing as the sensor channels are removed, which demonstrates that the proposed sensor selection method still works in this case. After removing two sensors, the classification accuracy has achieved 100%. However, unlike case study 1, when removing 5 sensors, the classification accuracy begins to decrease, which means that the information of Sensor #5 is not enough. Therefore, Sensor #4 and Sensor #5 are selected as the final sensor subset, which is used for further visualization and modeling.

Fig. 10 demonstrates that the domain-invariant features extracted by TSAK+MDA can also be used for generalized

Authorized licensed use limited to: CITY UNIV OF HONG KONG. Downloaded on May 06,2025 at 02:10:17 UTC from IEEE Xplore. Restrictions apply.

TABLE III
CLASSIFICATION ERROR USING LOOCV

Before Sensor Selection			
	Total error	Type I error	Type II error
SVM	2.79%	0.50%	2.29%
PCA	2.23%	0.50%	1.73%
KPCA	3.04%	0.87%	2.18%
MDA	0.17%	0	0.17%
After Sensor Selection			
	Total error	Type I error	Type II error
SVM	3.70%	1.20%	2.50%
PCA	2.48%	0.87%	1.62%
KPCA	4.80%	2.12%	2.68%
MDA	0%	0%	0%

model development in this case. Based on the data transferability evaluation and RFE sensor selection, two sensors are selected for sensor subset visualization and model validation. A conclusion consistent with case study 1 can be obtained. Fig. 10 (a) and (b) have not shown a clear separation, while Fig. 10 (c) indicates a good separation between two classes. Then, high accuracy can be guaranteed by using a common *k*NN model. Table III listed the classification accuracy before and after sensor selection, which further proves the performance of TSAK+MD for generalized modeling and the effectiveness of the proposed data transferability evaluation and sensor selection method.

As a summary, these two cases can draw the following conclusions: 1) The proposed methodology provides a way for sensor-based domain generalization without feature design and extraction; 2) The domain-invariant features can be used for robust sensor importance visualization and data transferability evaluation; 3) A robust sensor selection method is provided to remove the useless sensor channels; 4) The domain-invariant features extracted by TSAK+MDA can be used for the generalized model development, and good performance is guaranteed.

V. CONCLUSION

This paper proposes a novel approach for data transferability evaluation and robust sensor screening for FDC development in semiconductor manufacturing, locating the critical sensor channels containing valuable and transferable FDC information. By incorporating GA kernel into MDA, domain-invariant features are directly extracted from each sensor and used for sensor importance visualization and data transferability evaluation. The good discriminative power of the features represents that the corresponding sensor has critical information for FDC and high transferability across different chambers. Then, based on the RFE framework, the proposed sensor selection algorithm can effectively locate and remove the useless sensor channels. At last, the domain-invariant features can be used for the generalized FDC model development. The effectiveness of the proposed methodology is demonstrated in two case studies by using public datasets. The results show that the proposed sensor selection algorithm can quantify the data transferability of each sensor and remove the useless sensor channels while guaranteeing the model performance. Besides, the high accuracy of the generalized model demonstrates the model robustness to the unit-to-unit variation.

However, the proposed methodology belongs to the generalized FDC models, which require that each chamber have enough training samples for sensor selection and model development. Sometimes insufficient data in some chambers may affect the model performance. In the future investigation, we plan to address this issue by combining the idea of generalized models and product-based models. We plan to extend cross-chamber data transferability evaluation to cross-product data transferability evaluation and then build an FDC model which can work for different product lines being processed in different chambers. The characteristics of product-based models will avoid recourse to a time-consuming process for the FDC data accumulation of new products or new chambers. Besides, the number of in-situ sensors may vary from chamber to chamber, due to the differences in working ways and conditions. Currently, the proposed methodology cannot handle this sensor variation issue. There are two potential solutions. Firstly, data-driven virtual metrology can be used to replace the missing sensor channels in certain chambers to ensure the sensors in all chambers are the same. However, it requires additional computational and time resources. Secondly, one-pass learning is another potential solution, which attempts to compress important information of missing features into functions of existing features, and then expand to include the augmented features [36]. It brings an opportunity to solve the sensor variation issue, and we will further investigate it in the future.

REFERENCES

[1] J. R. Moyne, H. Hajj, K. Beatty, and R. Lewandowski, "SEMI E133—The process control system standard: Deriving a software interoperability standard for advanced process control in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 408–420, Nov. 2007, doi: 10.1109/TSM.2007.907617.

[2] J. Moyne, J. Samantaray, and M. Armacost, "Big data capabilities applied to semiconductor manufacturing advanced process control," *IEEE Trans. Semicond. Manuf.*, vol. 29, no. 4, pp. 283–291, Nov. 2016, doi: 10.1109/TSM.2016.2574130.

[3] S.-K. S. Fan, S.-C. Lin, and P.-F. Tsai, "Wafer fault detection and key step identification for semiconductor manufacturing using principal component analysis, AdaBoost and decision tree," *J. Ind. Prod. Eng.*, vol. 33, no. 3, pp. 151–168, 2016, doi: 10.1080/21681015.2015.1126654.

[4] S.-K. S. Fan, D.-M. Tsai, F. He, J.-Y. Huang, and C.-H. Jen, "Key parameter identification and defective wafer detection of semiconductor manufacturing processes using image processing techniques," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 544–552, Nov. 2019.

[5] T.-H. Pan, D. S.-H. Wong, and S.-S. Jang, "Chamber matching of semiconductor manufacturing process using statistical analysis," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 4, pp. 571–576, Jul. 2012, doi: 10.1109/TSMCC.2011.2161669.

[6] X. Li, W. Zhang, N.-X. Xu, and Q. Ding, "Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places," *IEEE Trans. Ind. Electron.*, vol. 67, no. 8, pp. 6785–6794, Aug. 2020, doi: 10.1109/TIE.2019.2935987.

[7] M. A. Djeziri, B. Ananou, M. Ouladsine, and J. Pinaton, "Health index extraction methods for batch processes in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 3, pp. 306–317, Aug. 2015, doi: 10.1109/TSM.2015.2438642.

[8] S.-K. S. Fan, C.-Y. Hsu, D.-M. Tsai, F. He, and C.-C. Cheng, "Data-driven approach for fault detection and diagnostic in semiconductor manufacturing," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 4, pp. 1925–1936, Oct. 2020, doi: 10.1109/TASE.2020.2983061.

Authorized licensed use limited to: CITY UNIV OF HONG KONG. Downloaded on May 06,2025 at 02:10:17 UTC from IEEE Xplore. Restrictions apply.

[9] P. Li et al., "A novel method for deposit accumulation assessment in dry etching chamber," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 2, pp. 183–189, May 2019, doi: 10.1109/TSM.2019.2904889.

[10] B. M. Wise, N. B. Gallagher, S. W. Butler, D. D. White, and G. G. Barna, "A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process," *J. Chemom.*, vol. 13, nos. 3–4, pp. 379–396, 1999.

[11] F. Zhu et al., "Methodology for important sensor screening for fault detection and classification in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 34, no. 1, pp. 65–73, Feb. 2021, doi: 10.1109/TSM.2020.3037085.

[12] J. Feng, J. Iskandar, J. Moyne, M. Armacost, and J. Lee, "Pattern-based trace segmentation and feature extraction for semiconductor manufacturing and application to fault detection," in *Proc. 30st Annu. SEMI Adv. Semicond. Manuf. Conf.*, 2018.

[13] H. Cai et al., "A framework for semi-automated fault detection configuration with automated feature extraction and limits setting," in *Proc. 31st Anal. SEMI Adv. Semicond. Manuf. Conf. (ASMC)*, 2020, pp. 1–6, doi: 10.1109/ASMC49169.2020.9185395.

[14] S. J. Hong, W. Y. Lim, T. Cheong, and G. S. May, "Fault detection and classification in plasma etch equipment for semiconductor manufacturing e-diagnostics," *IEEE Trans. Semicond. Manuf.*, vol. 25, no. 1, pp. 83–93, Feb. 2012, doi: 10.1109/TSM.2011.2175394.

[15] C.-F. Chien, C.-Y. Hsu, and P.-N. Chen, "Semiconductor fault detection and classification for yield enhancement and manufacturing intelligence," *Flex. Services Manuf. J.*, vol. 25, no. 3, pp. 367–388, 2013, doi: 10.1007/s10696-012-9161-4.

[16] K. B. Lee, S. Cheon, and C. O. Kim, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 2, pp. 135–142, May 2017, doi: 10.1109/TSM.2017.2676245.

[17] M. Azamfar, X. Li, and J. Lee, "Deep learning-based domain adaptation method for fault diagnosis in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 3, pp. 445–453, Aug. 2020, doi: 10.1109/TSM.2020.2995548.

[18] Q. P. He and J. Wang, "Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 345–354, Nov. 2007, doi: 10.1109/TSM.2007.907607.

[19] J. Yu, "Fault detection using principal components-based Gaussian mixture model for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 432–444, Aug. 2011, doi: 10.1109/TSM.2011.2154850.

[20] S.-K. S. Fan, X.-W. Chang, and Y.-Y. Lin, "Product-to-product virtual metrology of color filter processes in panel industry," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 4, pp. 3496–3507, Oct. 2022, doi: 10.1109/TASE.2021.3124157.

[21] Z. Zhou, C. Wen, and C. Yang, "Fault detection using random projections and k-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 1, pp. 70–79, Feb. 2015, doi: 10.1109/TSM.2014.2374339.

[22] C. Zhang, X. Gao, Y. Li, and L. Feng, "Fault detection strategy based on weighted distance of k nearest neighbors for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 1, pp. 75–81, Feb. 2019, doi: 10.1109/TSM.2018.2857818.

[23] J. Wang et al., "Generalizing to unseen domains: A survey on domain generalization," *IEEE Trans. Know. Data Eng.*, early access, May 26, 2022, doi: 10.1109/TKDE.2022.3178128.

[24] K. Muandet, D. Balduzzi, and B. Schölkopf, "Domain generalization via invariant feature representation," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 10–18.

[25] M. Ghifary, D. Balduzzi, W. B. Kleijn, and M. Zhang, "Scatter component analysis: A unified framework for domain adaptation and domain generalization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1414–1430, Jul. 2017, doi: 10.1109/TPAMI.2016.2599532.

[26] Y. Li, M. Gong, X. Tian, T. Liu, and D. Tao, "Domain generalization via conditional invariant representations," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, 2018, pp. 1–16.

[27] K. Zhang, B. Schölkopf, K. Muandet, and Z. Wang, "Domain adaptation under target and conditional shift," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 819–827.

[28] S. Hu, K. Zhang, Z. Chen, and L. Chan, "Domain generalization via multidomain discriminant analysis," in *Proc. 35th Uncertainty. Artif. Intell. Conf.*, 2020, pp. 292–302.

[29] M. Cuturi, "Fast global alignment kernels," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 929–936.

[30] L. Song, K. Fukumizu, and A. Gretton, "Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric inference in graphical models," *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 98–111, Jul. 2013, doi: 10.1109/MSP.2013.2252713.

[31] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. Lanckriet, "Hilbert space embeddings and metrics on probability measures," *J. Mach. Learn.Res.*, vol. 11, pp. 1517–1561, Apr. 2010.

[32] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998, doi: 10.1162/089976698300017467.

[33] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," in *Data Classification: Algorithms Applications*. Boca Raton, FL, USA: CRC Press, 2014, p. 37.

[34] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[35] R. T. Olszewski, "Generalized feature extraction for structural pattern recognition in time-series data," Ph.D. dissertation, School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, 2001.

[36] C. Hou and Z.-H. Zhou, "One-pass learning with incremental and decremental features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 11, pp. 2776–2792, Nov. 2018, doi: 10.1109/TPAMI.2017.2769047.

Authorized licensed use limited to: CITY UNIV OF HONG KONG. Downloaded on May 06,2025 at 02:10:17 UTC from IEEE Xplore. Restrictions apply.