A Wavelet-Based Penalized Mixed-Effects Decomposition for Multichannel Profile Detection of In-Line Raman Spectroscopy

Xiaowei Yue, *Member, IEEE*, Hao Yan, Jin Gyu Park, Zhiyong Liang, and Jianjun Shi

***Abstract*— Modeling and analysis of profiles, especially high-dimensional nonlinear profiles, is an important and challenging topic in statistical process control. Conventional mixed-effects models have several limitations in solving the multichannel profile detection problems for in-line Raman spectroscopy, such as the inability to separate defective information from random effects, computational inefficiency, and inability to handle high-dimensional extracted coefficients. In this paper, a new wavelet-based penalized mixed-effects decomposition (PMD) method is proposed to solve the multichannel profile detection problem in Raman spectroscopy. The proposed PMD exploits a regularized high-dimensional regression with linear constraints to decompose the profiles into four parts: fixed effects, normal effects, defective effects, and signal-dependent noise. An optimization algorithm based on the accelerated proximal gradient (APG) is developed to do parameter estimation efficiently for the proposed model. Finally, the separated fixed effects coefficients, normal effects coefficients, and defective effects coefficients can be used to extract the quality features of fabrication consistency, within-sample uniformity, and defect information, respectively. Using a surrogated data analysis and a case study, we evaluated the performance of the proposed PMD method and demonstrated a better detection power with less computational time.**

***Note to Practitioners*—This paper was motivated by the need of implementing multichannel profile detection for Raman spectra to realize in-line process monitoring and quality control of continuous manufacturing of carbon nanotube (CNT) buckypaper. Existing approaches, such as the mixed-effects model or the smooth-sparse decomposition method, cannot separate defective information in random effects effectively. This paper develops a penalized mixed-effects decomposition which decomposes Raman spectra into four components: fixed effects, normal effects, defective effects, and signal-dependent noise, respectively. The first three components can be applied to monitor the fabrication consistency, degree of uniformity, and defect information of buckypaper, respectively. With this new approach, several quality features can be monitored simultaneously and the algorithm based on the accelerated proximal gradient (APG) method can satisfy the computation speed requirement of in-line monitoring. This paper provides a solid foundation for in-line process monitoring and quality control for scalable nanomanufacturing of CNT buckypaper. Furthermore, the developed methodology can be applied in the decomposition of other signal systems with fixed, normal, and defective effects.**

***Index Terms*— Detection, mixed-effects model, nanomanufacturing, profile monitoring, raman spectroscopy, wavelet.**

I. Introduction

PROFILE data, also called functional data or waveform data, has been commonly observed in various industrial sensing and monitoring systems. Examples of such profile data include tonnage signals in stamping process [1] and valve seat pressing data in the engine assembly process [2]. As a result, it is an important, yet challenging problem to model and analyze profile data for process monitoring and diagnosis. This problem is especially true in the continuous CNT buckypaper manufacturing process, where in-line Raman spectroscopy instrument is used to provide Raman spectra, which is a special type of profile data, to indicate quality characteristics and properties of buckypaper.

CNT buckypaper is an important multifunctional platform material with great potential for creating lightweight, high-performance materials for various applications due to its superior mechanical and electrical characteristics. One of the critical roadblocks to scale-up the production of high-quality buckypaper is on-line quality monitoring and control of its production system. As an effective off-line characterization method for nanostructure information, Raman spectroscopy has been widely used in batch-to-batch nanomanufacturing of CNT buckypaper [3], [4]. Recently, in-line Raman spectroscopy has been developed to monitor biochemical reactions and pharmaceutical crystallizations [5], because of its nondestructive nature, fast data acquisition, and ability to provide detailed material information [6]. However, the in-line Raman spectroscopy techniques of monitoring biochemical reactions and pharmaceutical crystallizations cannot be used to monitor nanomanufacturing of CNT buckypaper majorly due to two reasons.

1) Only specific intensity peaks are considered in the monitoring of biochemical reactions and pharma-

Manuscript received July 24, 2016; revised June 4, 2017; accepted October 31, 2017. Date of publication December 12, 2017; date of current version July 2, 2018. This paper was recommended for publication by Associate Editor L. Zhu and Editor M. Y. Wang upon evaluation of the reviewers' comments. This work was supported by the National Science Foundation Scalable Nanomanufacturing Program under Grant SNM 1344672. *(Corresponding author: Jianjun Shi.)*

X. Yue and J. Shi are with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: xwy@gatech.edu; jianjun.shi@isye.gatech.edu).

H. Yan is with the School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: haoyan@asu.edu).

J. G. Park and Z. Liang are with the High-Performance Materials Institute, Florida State University, Tallahassee, FL 32310 USA (e-mail: jgpark@fsu.edu; zliang@fsu.edu).

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. The Supplementary Material contains material that is not included within the paper itself. This material is 22.6 MB in size.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2017.2772218

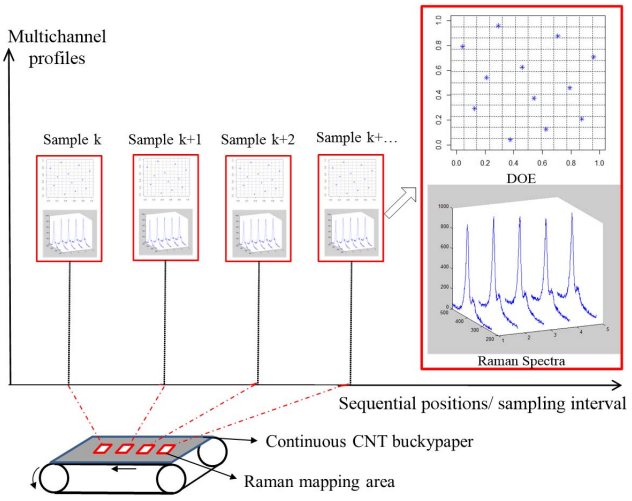


Fig. 1. Schematic of in-line measurement procedure based on Raman spectroscopy in a continuous CNT buckypaper production process.

ceutical crystallizations, while in the monitoring of nanomanufacturing, the whole profiles need to be analyzed.

2) The techniques used in the monitoring of biochemical reactions and pharmaceutical crystallizations cannot separate defect information from the normal random effects.

Thus, we focus on developing a profile modeling and detection approach based on in-line Raman spectroscopy for the roll-to-roll CNT buckypaper fabrication process.

Raman analysis is critical for detecting the quality information of the product such as impurity, the degree of chemical functionalization, and the alignment of CNTs in buckypaper. For example, the radial breathing mode ($<300\ \text{cm}^{-1}$) is used to determine the diameter of a single-walled CNT (SWCNT); D-band (1250–1400 cm^{-1}) is related to the disorder or molecular defects in the CNT structure; and D-band to G-band intensity ratio is an effective indicator of CNT quality and functionalization [7]. In addition, polarized Raman provides angular dependence of the Raman intensity, and then the degree of CNTs alignment can be estimated [8].

A schematic of in-line measurement procedure based on Raman spectroscopy is shown in Fig. 1. In a roll-to-roll fabrication process of CNT buckypaper, samples of Raman spectra are measured sequentially with Raman mapping technique. Each sample is shown as a red square in Fig. 1. Within each sample, a specific number of Raman spectra are collected in the positions predetermined by design of experiment to represent the quality information in the sample area. Usually, the maximin Latin hypercube design [9], which demonstrates good space-filling properties and first-dimension projection properties, is implemented to obtain the representative sampling positions, as shown in the drawing of partial enlargement in Fig. 1. If each representative sampling position can be regarded as a sensor channel to collect Raman spectra, the process modeling and detection for buckypaper fabrication process can be formulated as a multichannel profile modeling problem.

Three quality characteristics are of critical importance during the statistical process control of buckypaper fabrication process as follows:

1) the fabrication consistency of quality features measured by Raman spectroscopy, which indicates whether there is a gradual mean shift in the sequentially roll-to-roll fabrication of CNT buckypaper;
2) within-sample uniformity, which quantifies the uniformity of quality features in the sample area (shown as red square in Fig. 1), such as degree of alignment, degree of functionalization, nanotube distribution, and dispersion;
3) within-sample defect information, which indicates whether there are defects in the CNT buckypaper.

In addition, a specific band of the Raman spectrum denotes corresponding defective information of the product. In our previous publication [6], signal-dependent property of noise in Raman spectra has been identified and validated with the experimental data. A generalized wavelet shrinkage method was proposed for denoising by applying individual adaptive wavelet thresholds. However, the denoised Raman spectra cannot be directly used in multichannel profile monitoring because all components associated with different quality characteristics (e.g., fabrication consistency, uniformity, and defects) are all mixed in the signals. Therefore, profile modeling and signal decomposition for multichannels signals are essential.

In order to monitor the aforementioned fabrication consistency, within-sample uniformity, and defect information for the in-line Raman spectroscopy in the sequentially buckypaper fabrication process, we propose a new multichannel profile modeling and analysis approach with consideration of the following two characteristics.

1) Signal-dependent noise in Raman spectra, which has been validated in [6].
2) Multichannel profile modeling and detection based on profiles in a sample reveal within-sample uniformity and defect information in the sample area, while changes among different samples illustrate the quality shift in the sequential long-term manufacturing process.

In this paper, a new penalized mixed-effects decomposition (PMD) method is proposed for the multichannel profile detection based on in-line Raman spectroscopy in the continuous nanomanufacturing process. The proposed PMD can separate the profiles into the following four parts:

1) fixed effects, which measure the fabrication consistency among samples over time;
2) normal effects, which measure the within-sample uniformity of quality features;
3) defective effects, which reveal the existence and location of the defects within a sample area;
4) signal-dependent noise, which is proportional to the Raman intensity in terms of specific Raman shift.

To model waveform signals of Raman spectroscopy, wavelet basis is applied to generate design matrices in the proposed PMD. The merits of using wavelet basis will be discussed in Section II. Another advantage is that random effects coefficients extracted by the PMD is sparse and has a much lower dimension than the original data due to the L_1 norm regularization term for defective effects coefficients. An APG algorithm is developed to effectively and efficiently estimate the parameters in the proposed PMD method. Sensitivity

analysis is also conducted to determine the bounds of normal effects based on product tolerance.

The remainder of this paper is organized as follows. Section II provides a literature review of linear and nonlinear profile modeling and analysis. Section III describes the proposed PMD formulation and compares it with the linear mixed-effects model (LMM) and the smooth sparse decomposition (SSD) method. In addition, an APG algorithm for parameter estimation is introduced, and a sensitivity analysis of the bounds and the penalized parameter is conducted. Section IV presents a surrogated data analysis to test the performance of the PMD approach, followed with a real in-line Raman case study that is implemented in Section V. Finally, a brief summary is discussed in Section VI.

II. Literature Review

This section provides a detailed literature review for profile modeling and analysis. Conventional methods of linear and nonlinear profile modeling and analysis are discussed first. Nonparametric techniques to handle nonsmooth profiles are then introduced. Finally, mixed-effects model-based profile modeling approaches and their limitations are discussed.

In [10]–[12], most of the researchers focused on linear profile modeling and monitoring. Nonlinear profile monitoring was also investigated to address the nonlinearity of profiles in the different manufacturing processes. Different techniques, such as multiple and polynomial regression, cubic spline, and B-spline, were used to monitor nonlinear profile data [13]–[17]. However, the aforementioned approaches only focus on the modeling of smooth and differentiable profiles, which are not applicable when the profiles include spikes or peaks such as those in the Raman spectroscopy of CNTs. To address this challenge, a piecewise method is proposed to divide a profile into several segments and fit them independently by using techniques mentioned above. However, these techniques require engineering background and estimation to decide all the break points and segment intervals. Besides, different segmentation approaches will also affect the modeling and monitoring results for a given application.

Some other techniques are investigated to develop nonparametric methods to model nonsmooth signals directly. Among all these techniques, principal components analysis (PCA), wavelet analysis, and empirical model decomposition (EMD) are the most common techniques that directly model nonlinear and nonstationary profiles.

PCA is a powerful unsupervised statistical procedure to reduce the dimension directly by orthogonal transformation such that the selected principal components account for most of the variability. Therefore, PCA is widely used in dimension reduction and statistical process control [18], [19]. However, there are some inherent limitations when the PCA method is used. One limitation is its weak interpretability since all the random variables are included in the principal components. Another limitation is the inconsistency when applying PCA to process high-dimensional data with noise, proved by Johnstone and Lu [20] by showing that the sample eigenvectors do not converge to the population eigenvectors and the noise does not average out if the dimension is much higher compared to the sample size. Sparse PCA was developed to improve the interpretability by imposing the sparsity constraint on the principal components [21]. However, sparse principal components after dimension reduction usually do not cover all the bands in profiles, which lead to the loss of detection power in the band corresponding to the zero loadings. In addition, the defects tend to be correlated in the neighbor bands of profiles, and the sparse loadings ignore this kind of correlation.

Wavelet analysis is another powerful and widely applied nonparametric tool for time–frequency transformation because of the following three advantages.

1) It is an effective nonparametric approach that can model nonlinear profiles with sharp jumps and nondifferentiable points.
2) It provides multiresolution time–frequency localized representation, which overcomes the limitations of traditional Fourier transform (FT) and short-time (STFT).
3) Many natural images and signals have sparse representations using a wavelet basis, which is useful in denoising and data compression.

Due to the popularity of time-dependent nonlinear profile data in system monitoring and diagnosis, wavelet-based methods are widely used to model the profile data for process modeling and monitoring problems [2], [22], [23].

In addition, EMD is a powerful time–frequency analysis technique. It can decompose the signal into a group of orthogonal components called intrinsic mode function (IMF) based on identification and interpolation of peaks and troughs in the signal. EMD has been widely utilized in various areas including process control, detection, and diagnosis [24], [25]. However, EMD suffers from three disadvantages.

1) The modes extracted by the method are not always physically meaningful, and the decomposition is limited to decompose each mode based on the varying frequency.
2) It is noise sensitive, which is not suitable for Raman spectra with signal-dependent noise.
3) It has mode mixing (identifiability) issue, which is defined as a single IMF including oscillations of dramatically disparate scales, or a component of a similar scale residing in different IMFs.

Considering all these factors of PCA, wavelet analysis, and EMD, wavelet-based methods are preferred to model the profiles in continuous nanomanufacturing.

In profile modeling and analysis, two kinds of variations are commonly considered. One is within-profile variation which is normally due to the random noise in the process such as environmental noise and measurement errors; the other is between-profile variation due to the inevitable profile-to-profile variability including process condition variations and fixture or tooling tolerance. Both within-profile variation and between-profile variation should be modeled to account for the total variability. This desire motivated to develop a mixed-effects model.

Researchers developed parametric or nonparametric mixed-effects model for profile monitoring. For example, parametric mixed-effects models were developed by treating a few model parameters as random effects [13], [15]; this kind of model is suitable for the data with complete domain knowledge

and clear model structure; A nonparametric mixed-effects model with B-spline and PCA was developed to monitor nonlinear profiles [16], which is usually achievable for smooth profiles. Paynabar and Jin [2] used the wavelet-based mixed-effects model to characterize nonlinear profiles. However, the existing mixed-effects model cannot be directly used to profile modeling and analysis in the continuous nanomanufacturing for the following three reasons.

1) A mixed-effects model can only measure between-profile variation by random effects, instead of separating normal between-profile variation and defective between-profile variation. The normal between-profile variation results from measurements and product uncertainty within the tolerance, while the defective between-profile variation is due to the defects in the buckypaper, which is not acceptable per engineering specifications.
2) There are too many coefficients used in a mixed-effects model, especially for high-dimensional profiles. This will result in both high false alarm rate (FAR) and low detection rate (DR) in the multivariate process control.
3) The algorithms to do a parameter estimation of the mixed-effects model are not computationally efficient enough for in-line detection.

Thus, we proposed a PMD method that overcomes these three limitations from the conventional mixed-effects models. To address these challenges, the random effects are separated into normal effects and defective effects in our proposed model. Meanwhile, to reduce the dimensionality of the coefficients, wavelet basis and L_1 norm regularization are applied since it can achieve a sparse representation of the data. The proposed methodology will be discussed in detail in the following section.

III. Penalized Mixed-Effects Decomposition

In this section, we will first review the conventional LMM [26] and the SSD model [27]. We will then introduce the PMD for multichannel profile detection. After that the PMD is compared with the LMM and the SSD. An APG algorithm is developed to estimate parameters of the PMD. Finally, a sensitivity analysis about the bounds of normal effects and the penalized parameter is discussed.

A. Review of the Mixed-Effects Models and the Smooth-Sparse Decomposition

An LMM [26] is formulated for multichannel profile modeling as

$$\boldsymbol{y_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{\varepsilon_i} \tag{1}$$

where $\boldsymbol{y_i}$ is a measurement profile and $\boldsymbol{X_i}$ and $\boldsymbol{Z_i}$ are the design matrices associated with the fixed effects and the random effects, respectively. $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$ are the coefficients vectors associated with the fixed effects and the random effects. $\boldsymbol{\varepsilon_i}$ is a noise vector that depends on the fixed effects and the random effects intensity because of the signal-dependent property.

The mixed-effects model has been widely used in functional data analysis [2], [13], [15], [16]. However, this conventional mixed-effects model has the following limitations when it is applied to multichannel profile modeling based on in-line Raman spectra.

1) It can only separate the fixed effects and the random effects, but cannot separate the normal random effects and the defective random effects. As a consequence, it cannot be directly used in the monitoring of fabrication consistency, within-sample uniformity, and within-sample defect information for the Raman spectra simultaneously.
2) The coefficients extracted by a mixed-effects model are still high dimensional, which brings challenges to achieve both a low FAR and a high DR in the multivariate process monitoring problem.
3) In general, a restricted maximum likelihood (REML) method or an expectation–maximization method can be used to estimate the model parameters. However, those methods are computationally expensive, which does not meet the need for real-time detection in a continuous roll-to-roll nanomanufacturing process, especially for profiles with heterogeneous noise, such as Raman spectra.

To separate the defect information from the smooth functional mean profile, an SSD method [27] is proposed to decompose each profile $\boldsymbol{y}$ into a smooth functional mean $\boldsymbol{B\theta}$, a sparse anomaly $\boldsymbol{B}_a\boldsymbol{\theta}_a$, and a noise vector $\boldsymbol{e}$ by solving the following optimization problem:

$$\begin{aligned} &\underset{\boldsymbol{\theta},\boldsymbol{\theta}_a}{\arg\min}\ ||\boldsymbol{e}||^2 + \lambda\boldsymbol{\theta}^T\boldsymbol{R}\boldsymbol{\theta} + \gamma\,|\boldsymbol{\theta}_a|_1 \\ &\text{Subject to } \boldsymbol{y} = \boldsymbol{B\theta} + \boldsymbol{B}_a\boldsymbol{\theta}_a + \boldsymbol{e} \end{aligned} \tag{2}$$

where $||\cdot||$ and $|\cdot|_1$ are L_2 operator and L_1 norm operator, respectively. $\boldsymbol{B}$, $\boldsymbol{B}_a$, and $\boldsymbol{R}$, are the design matrices for smooth part, sparse part, and a roughness matrix, respectively. λ and γ are tuning parameters to realize regularization for smooth coefficients $\boldsymbol{\theta}$ and sparse coefficients $\boldsymbol{\theta}_a$. An accelerated proximal algorithm is applied to efficiently estimate $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_a$ simultaneously.

The SSD method is effective and efficient to separate the background, anomalies, and defect for both smooth functional data and images with the smooth background. However, it cannot be applied to the multichannel profile detection problem based on in-line Raman spectra for the following two reasons.

1) The SSD is designed for anomaly detection under smooth background, while Raman spectrum is inherently nonsmooth. Therefore, a spline basis, which is used in the SSD, is not suitable for the nonsmooth profiles.
2) The loss function of the noise term does not consider the covariance matrix, which is not appropriate for signal-dependent noise in Raman spectra.

Inspired by the SSD, we propose to decompose the defect random effect and the normal random effect by using wavelet basis to describe the waveform signal of Raman spectra and using a covariance matrix to consider the signal-dependent noise.

B. Penalized Mixed-Effects Decomposition

A typical in-line Raman process monitoring problem is shown in Fig. 1. The sequentially measured Raman profiles

are grouped into N samples, and each sample has n_i profiles containing n_{ij} observations for different Raman shifts. The samples are indexed by $i = 1, \ldots, N$ and the profiles in the ith sample are indexed by $j = 1, \ldots, n_i$. The PMD for multichannel profile detection of Raman spectra is formulated as

$$\boldsymbol{y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{W}\boldsymbol{\theta}_{ij} + \boldsymbol{W_a}\boldsymbol{\delta}_{ij} + \boldsymbol{e}_{ij}$$

where $\boldsymbol{y}_{ij}$ is a measurement profile with dimension n, corresponding to the jth profile in the ith sample; $\boldsymbol{\mu}_i$ denotes fixed effects in the ith sample; $\boldsymbol{W}$ and $\boldsymbol{W_a}$ are wavelet-based design matrices with dimension $n \times p$ and $n \times q$ for normal effects and defective effects, respectively; $\boldsymbol{\theta}_{ij}$ and $\boldsymbol{\delta}_{ij}$ are coefficients vectors associated with normal effects and defective effects with regards to the jth profile in the ith sample. $\boldsymbol{e}_{ij}$ represents a signal-dependent noise vector for the jth profile in the ith sample.

In the problem formulation, a weighted sum of square loss function is applied to model the signal-dependent noise and the L_1 norm regularization is implemented to encourage the sparsity of the defective random coefficients. Aside from this, the normal effects will be bounded by $\boldsymbol{B}_L$ and $\boldsymbol{B}_U$, which can be determined by the product tolerance and a sensitivity analysis. Therefore, the proposed PMD is formulated as

$$\begin{aligned}
&\arg\min_{\boldsymbol{\theta}_{ij}, \boldsymbol{\delta}_{ij}} \ \boldsymbol{e}_{ij}^T \boldsymbol{\Omega}^{-1} \boldsymbol{e}_{ij} + \gamma \left|\boldsymbol{\delta}_{ij}\right|_1 \\
&\text{subjectto } \boldsymbol{y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{W}\boldsymbol{\theta}_{ij} + \boldsymbol{W_a}\boldsymbol{\delta}_{ij} + \boldsymbol{e}_{ij} \\
&\quad \boldsymbol{B}_L \preccurlyeq \boldsymbol{W}\boldsymbol{\theta}_{ij} \preccurlyeq \boldsymbol{B}_U \\
&\quad \boldsymbol{B}_U = \sqrt{\operatorname{diag}(\boldsymbol{\Omega})} \circ \boldsymbol{\omega}_1, \quad \boldsymbol{B}_L = -\sqrt{\operatorname{diag}(\boldsymbol{\Omega})} \circ \boldsymbol{\omega}_2
\end{aligned} \tag{3}$$

where $\boldsymbol{\Omega}$ is a variance–covariance matrix of Raman spectra noise; $\preccurlyeq$ denotes the componentwise inequality; $\boldsymbol{B}_L$ and $\boldsymbol{B}_U$ are a lower bound and a upper bound for normal effects; $\operatorname{diag}(\boldsymbol{\Omega})$ denotes the vector of diagonal elements of $\boldsymbol{\Omega}$; $\circ$ represents the hadamard product (element-wise multiplication for two vectors). Other detailed introduction for each notation in the model (3) can be illustrated as follows.

1) $\boldsymbol{\mu}_i$ denotes the fixed effects within each sample. Monitoring $\boldsymbol{\mu}_i (i = 1, \ldots, N)$ reveals the fabrication consistency of the samples, which is the long-term mean shift among samples as the sequentially roll-to-roll CNT buckypaper fabrication process goes on. Usually, it can be calculated by the median vector of multiple profiles in each sample for robust estimation.
2) $\boldsymbol{W}\boldsymbol{\theta}_{ij}$ represents the normal effects within each sample. $\boldsymbol{\theta}_{ij}$ can be used to quantify the uniformity of quality features in the sample area, such as degree of alignment, degree of functionalization, nanotube distribution, and dispersion. Because of the sparse representation property of most waveform signals in the wavelet analysis, the dimension of $\boldsymbol{\theta}_{ij}$ is much lower than the original dimension of Raman spectra, which can improve the pecision and efficiency of profile modeling. In addition, $\boldsymbol{W}\boldsymbol{\theta}_{ij}$ is bounded by $\boldsymbol{B}_U$ and $\boldsymbol{B}_L$. $\boldsymbol{B}_U$ and $\boldsymbol{B}_L$ determine the monitoring robustness for the uniformity. If the degree of uniformity requirement is strict, the bounds will be tighter. We can estimate the bounds by the corresponding weighted standard deviations of the Raman intensity for a specific Raman shift. The weight vectors $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ can be determined by the product tolerance and sensitivity analysis, which will be discussed in Section III-E.
3) $\boldsymbol{W_a}\boldsymbol{\delta}_{ij}$ describes the defective effects within each sample, which is to model the defect information. It reveals whether there are specific sampling points with defective quality features occurred in certain bands of a Raman spectrum. L_1 norm regularization encourages the sparsity of the anomalous coefficients, which synergizes with the inherent sparsity of wavelet coefficients. Penalized parameter γ can be determined by the cross validation described in Section III-F. Furthermore, according to the mathematical link between wavelet coefficients and original signal segments [22], we can derive which specific bands of a Raman spectrum have defects based on the location of the nonzero coefficients.
4) $\boldsymbol{e}_{ij}$ denotes the signal-dependent noise. The detailed noise source and its signal-dependent characteristic are validated in [6]. A weighted sum of square loss function for the signal-dependent noise can separate the noise accurately, which will improve the effect and efficiency of profile monitoring based on the random effect. The variance–covariance matrix can be obtained by the real Raman spectra data collected during phase-I analysis [6].
5) $\boldsymbol{W}$ and $\boldsymbol{W_a}$ are the wavelet-based design matrices for the normal effects and the defective effects, which are determined by the selected wavelet basis and its coefficients. $\boldsymbol{W}$ is required to be orthogonal. The wavelet basis usually can be selected upon a defined criterion [6]. It is worth noting that similar as the design matrices determination in the mixed-effects model [26]; usually, $\boldsymbol{W_a}$ is chosen as a subset of columns of $\boldsymbol{W}$ based on random effects pattern from engineering knowledge. If we do not know the design matrix columns associated with random effects, we can choose $\boldsymbol{W_a}$ equals to $\boldsymbol{W}$. Even $\boldsymbol{W}$ and $\boldsymbol{W_a}$ are the same, the identifiability can also be ensured. We realize the identifiability of coefficients by two settings: a) the $\boldsymbol{W}\boldsymbol{\theta}_{ij}$ is bounded by $\boldsymbol{B}_U$ and $\boldsymbol{B}_L$ and b) both $\boldsymbol{e}_{ij}^T \boldsymbol{\Omega}^{-1} \boldsymbol{e}_{ij}$ and $\gamma \left|\boldsymbol{\delta}_{ij}\right|_1$ are considered in the loss function. For example, first, $\boldsymbol{\mu}_i$ can be isolated by the median profile. Since $\boldsymbol{\delta}_{ij}$ is in the loss function, while $\boldsymbol{\theta}_{ij}$ is not in the loss function. The model will distribute as many components as possible into $\boldsymbol{\theta}_{ij}$ until $\boldsymbol{W}\boldsymbol{\theta}_{ij}$ is bounded. If $\boldsymbol{W}\boldsymbol{\theta}_{ij}$ is bounded, other components in the profiles will be distributed into $\boldsymbol{\delta}_{ij}$ and $\boldsymbol{e}_{ij}$ according to the tradeoff between $\boldsymbol{e}_{ij}^T \boldsymbol{\Omega}^{-1} \boldsymbol{e}_{ij}$ and $\left|\boldsymbol{\delta}_{ij}\right|_1$.

C. Comparison Between the PMD and the LMM/SSD

In this section, the PMD is compared with the conventional LMM and the SSD, as shown in Fig. 2. In comparison with the LMM, fixed effects obtained by the PMD correspond to fixed effects separated by the LMM, while random effects in the LMM can be further decomposed into normal effects and defective effects by the PMD. The normal effects are generated from the uncertainty of the spectra measurement and normal production conditions. Usually, the tolerance of product design

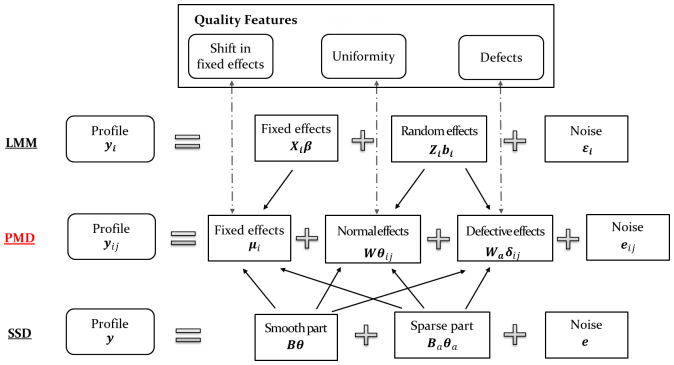


Fig. 2. Schematic of decomposition comparison among the PMD, the LMM, and the SSD.

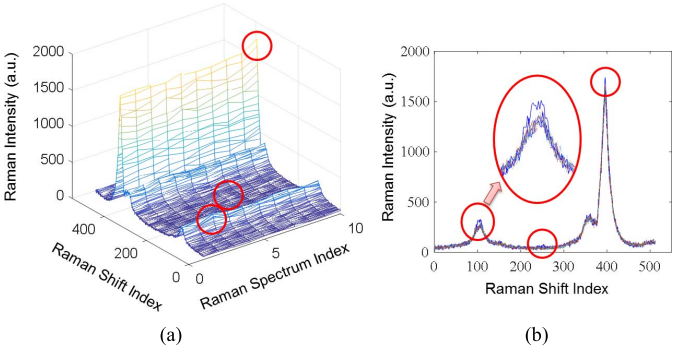


Fig. 3. Ten Raman spectra with three defects in different bands of different profiles. (a) 3-D mesh plot of ten Raman spectra. (b) 2-D plot of ten Raman spectra. (Note that defects are marked by the red circles.)

is used to control the uncertainty of manufacturing and measurement process. However, defective effects are generated from defects such as impurities mixed in the CNT buckypaper or large interspace hole in aligned CNTs, which should be detected during the manufacturing process. Thus, the PMD could be regarded as an improved mixed-effects model that can achieve more detailed decomposition.

The SSD decomposes each profile $\boldsymbol{y}$ into a smooth functional mean $\boldsymbol{B\theta}$, a sparse anomaly $\boldsymbol{B}_a\boldsymbol{\theta}_a$, and a noise vector $\boldsymbol{e}$, which works well for the data with smooth background and sparse defects. However, when applied to the nonsmooth Raman spectra, the defective effects can also be very smooth like the other band (e.g., band between 200 and 300 of Raman shift index in Fig. 3), not associated with D-band or G-band, in a Raman spectrum. In this situation, the defective information will be mistakenly regarded as the smooth background by using the SSD. On the other hand, the sparse peak regarded as an anomaly in the SSD may also be the fixed effects or normal effects, such as the G-band peak in a Raman spectrum. The PMD combines the advantage of the LMM and the SSD and decomposes a profile into fixed effects, normal effects, defective effects, and signal-dependent noise, which can be used to monitoring different quality features like shift in fixed effects, degree of uniformity, and defects.

In summary, Fig. 2 provides a schematic of decomposition comparison between the PMD and LMM/SSD. Intuitively, two facts are conveyed by Fig. 2: 1) the PMD can realize deeper decomposition than the LMM and 2) the PMD and the SSD decompose the profiles from different point of views.

D. Algorithm for Parameter Estimation

In this section, the PMD (3) is simplified to a weighted least absolute shrinkage and selection operator (LASSO) [28] problem. After that the APG algorithm to solve this convex optimization problem will be derived.

It can be shown that the variance–covariance matrix $\boldsymbol{\Omega}$ is a diagonal matrix with all the diagonal entries equal to the variance of Raman intensity corresponding to the related Raman shift. Thus, $\boldsymbol{\Omega}$ is invertible and each diagonal entry is positive. Equation (3) can be transformed into the following equation:

$$\begin{aligned}\min_{\boldsymbol{\theta}_{ij},\boldsymbol{\delta}_{ij}} \quad & (\boldsymbol{y}_{ij}-\boldsymbol{\mu}_i-\boldsymbol{W\theta}_{ij}-\boldsymbol{W}_a\boldsymbol{\delta}_{ij})^T\boldsymbol{\Omega}^{-1} \\ & \times(\boldsymbol{y}_{ij}-\boldsymbol{\mu}_i-\boldsymbol{W\theta}_{ij}-\boldsymbol{W}_a\boldsymbol{\delta}_{ij})+\gamma\,|\boldsymbol{\delta}_{ij}|_1 \\ \text{subjectto}\ & \boldsymbol{B}_L \preccurlyeq \boldsymbol{W\theta}_{ij} \preccurlyeq \boldsymbol{B}_U \\ & \boldsymbol{B}_U=\sqrt{\text{diag}(\boldsymbol{\Omega})}\circ\boldsymbol{\omega}_1, \quad \boldsymbol{B}_L=-\sqrt{\text{diag}(\boldsymbol{\Omega})}\circ\boldsymbol{\omega}_2.\end{aligned} \tag{4}$$

Thus, (4) is a weighted LASSO problem with linear constraints. We can further prove the PMD shown in (4) is equivalent to a weighted LASSO problem, and a detailed description is shown in Proposition 1.

Proposition 1: Assume $\boldsymbol{W}$ is orthogonal, the PMD in (4) is equivalent to minimize a weighted sum of square loss function plus L_1 norm regularization in the form of

$$\underset{\boldsymbol{\delta}_{ij}}{\text{argmin}}\, F(\boldsymbol{\delta}_{ij}) := \left(\boldsymbol{z}_{ij}-\boldsymbol{W\theta}^*_{ij}\right)^T\boldsymbol{\Omega}^{-1}\left(\boldsymbol{z}_{ij}-\boldsymbol{W\theta}^*_{ij}\right)+\gamma\,|\boldsymbol{\delta}_{ij}|_1 \tag{5}$$

where $\boldsymbol{\theta}^*_{ij} = \boldsymbol{W}^T\boldsymbol{z}_{ij}\circ\boldsymbol{I}(\boldsymbol{B}_L \preccurlyeq \boldsymbol{z}_{ij} \preccurlyeq \boldsymbol{B}_U) + \boldsymbol{W}^T\boldsymbol{B}_L\circ\boldsymbol{I}(\boldsymbol{z}_{ij} \preccurlyeq \boldsymbol{B}_L) + \boldsymbol{W}^T\boldsymbol{B}_U\circ\boldsymbol{I}(\boldsymbol{z}_{ij} \succcurlyeq \boldsymbol{B}_L)$, $\boldsymbol{z}_{ij} := \boldsymbol{y}_{ij}-\boldsymbol{\mu}_i-\boldsymbol{W}_a\boldsymbol{\delta}_{ij}$, $\boldsymbol{I}(\cdot)$ is an indicator operator. The formulation above is the weighted LASSO formulation.

The proof can be found in Appendix A.

Conventional LASSO solvers, such as least-angle regression, cannot efficiently solve the weighted LASSO problem for high-dimensional data [27]. Consequently, an effective and efficient algorithm should be developed to solve the problem to satisfy the in-line monitoring requirement. A popular efficient optimization algorithm, the proximal gradient method [27], [29] can be used to solve (5). The proximal gradient method is focused on the minimization of summation of a group of convex function. The function in (5) can be regarded as two subfunctions: $f(\boldsymbol{\delta}_{ij}) := (\boldsymbol{z}_{ij}-\boldsymbol{W\theta}^*_{ij})^T\boldsymbol{\Omega}^{-1}(\boldsymbol{z}_{ij}-\boldsymbol{W\theta}^*_{ij})$ and $g(\boldsymbol{\delta}_{ij}) := \gamma\,|\boldsymbol{\delta}_{ij}|_1$. Since $\boldsymbol{\Omega}^{-1}$ is a positive semidefinite matrix, $f(\boldsymbol{\delta}_{ij})$ is a convex differentiable function. Besides, $g(\boldsymbol{\delta}_{ij})$ is a nondifferentiable convex function. We also proved in Proposition 2 that $f(\cdot)$ is Lipchiz continuous, which can guarantee the proximal gradient method to converge.

Proposition 2: $f(\cdot)$ is Lipchiz continuous, which implies there is a constant L that makes gradient function $\nabla f(\cdot)$ satisfy $\|\nabla f(\boldsymbol{a})-\nabla f(\boldsymbol{b})\| \le L\|\boldsymbol{a}-\boldsymbol{b}\|$ for any $\boldsymbol{a},\boldsymbol{b}\in\boldsymbol{R}$ with $L=(2/\min(\text{diag}(\boldsymbol{\Omega})))\|\boldsymbol{W}_a\|_2^2$, where $||||_2^2$ denotes square of matrix spectral norm.

The proof can be found in Appendix B.

Therefore, the proximal gradient algorithm can optimize $F(\boldsymbol{\delta}_{ij})$ by using an iterative algorithm given by

$$\boldsymbol{\delta}_{ij}^{(k)} = \underset{\boldsymbol{\delta}_{ij}}{\operatorname{argmin}} \left\{ f\left(\boldsymbol{\delta}_{ij}^{(k-1)}\right) + \left\langle \boldsymbol{\delta}_{ij} - \boldsymbol{\delta}_{ij}^{(k-1)}, \nabla f\left(\boldsymbol{\delta}_{ij}{}^{(k-1)}\right)\right\rangle + \frac{L}{2}\|\boldsymbol{\delta}_{ij} - \boldsymbol{\delta}_{ij}^{(k-1)}\|^2 + \gamma\,|\boldsymbol{\delta}_{ij}|_1 \right\}$$

where super-indices (k) and $(k-1)$ denote the iteration numbers and $\langle\cdot,\cdot\rangle$ means the inner product operator.

Proposition 3: The proximal gradient algorithm for the PMD in (5), given by

$$\boldsymbol{\delta}_{ij}^{(k)} = \underset{\boldsymbol{\delta}_{ij}}{\operatorname{argmin}} \left\{ f\left(\boldsymbol{\delta}_{ij}^{(k-1)}\right) + \left\langle \boldsymbol{\delta}_{ij} - \boldsymbol{\delta}_{ij}^{(k-1)}, \nabla f\left(\boldsymbol{\delta}_{ij}{}^{(k-1)}\right)\right\rangle + \frac{L}{2}\|\boldsymbol{\delta}_{ij} - \boldsymbol{\delta}_{ij}^{(k-1)}\|^2 + \gamma\,|\boldsymbol{\delta}_{ij}|_1 \right\}$$

has a closed-form solution in each iteration k, in the form of a soft-thresholding function as follows:

$$\boldsymbol{\delta}_{ij}^{(k)} = S_{\frac{\gamma}{L}}\left(\boldsymbol{\delta}_{ij}^{(k-1)} - \frac{2}{L}\boldsymbol{W}_{\boldsymbol{a}}^{\boldsymbol{T}}\boldsymbol{\Omega}^{-1}\left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_i - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*} - \boldsymbol{W}_{\boldsymbol{a}}\boldsymbol{\delta}_{ij}^{(k-1)}\right)\right)$$

with $L = (2/\min(\operatorname{diag}(\boldsymbol{\Omega})))\|\boldsymbol{W}_{\boldsymbol{a}}\|_2^2$.

$S_{(\gamma/L)}(x) = \operatorname{sgn}(x)(|x| - (\gamma/L))_+$ is the soft-thresholding operator, and sgn(x) is the sign function and $x_+ = \max(x, 0)$. The proof of the Proposition 3 can be found in Appendix C.

Therefore, the convergence of the proximal gradient algorithm is guaranteed [29]. In order to improve the convergence speed of the optimization algorithm, an APG method [30] can be applied to adjust the original proximal gradient algorithm. The final algorithm can be summarized as follows.

Algorithm 1: Accelerated Proximal Gradient Based Algorithm for the Penalized Mixed-Effects Decomposition

While $\boldsymbol{j = 1 : N}$
  Initialization:
    $\boldsymbol{\delta}_{ij}^{(0)} = \mathbf{0}$, $\boldsymbol{L} = \frac{2}{\min(\operatorname{diag}(\boldsymbol{\Omega}))}\|\boldsymbol{W}_{\boldsymbol{a}}\|_2^2$,
    $\boldsymbol{\mu}_i = \mathbf{median}(\boldsymbol{y}_{i\cdot})$, $\boldsymbol{x}^{(0)} = \mathbf{0}$, $\boldsymbol{t_0} = \mathbf{0}$
  End
  While $\left|\boldsymbol{\delta}_{ij}^{(k-1)} - \boldsymbol{\delta}_{ij}^{(k)}\right| > \boldsymbol{\epsilon}$
    Let
    $\boldsymbol{\theta}_{ij}^{*} = \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{z}_{ij} \circ \boldsymbol{I}\left(\boldsymbol{B}_{\boldsymbol{L}} \preccurlyeq \boldsymbol{z}_{ij} \preccurlyeq \boldsymbol{B}_{\boldsymbol{U}}\right) - \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{B}_{\boldsymbol{L}} \circ \boldsymbol{I}\left(\boldsymbol{z}_{ij} \preccurlyeq \boldsymbol{B}_{\boldsymbol{L}}\right) - \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{B}_{\boldsymbol{U}} \circ \boldsymbol{I}\left(\boldsymbol{z}_{ij} \succcurlyeq \boldsymbol{B}_{\boldsymbol{L}}\right)$
    Update
    $\boldsymbol{\delta}_{ij}^{(k)} = \boldsymbol{S}_{\frac{\gamma}{L}}(\boldsymbol{x}^{(k-1)} - \frac{2}{L}\boldsymbol{W}_{\boldsymbol{a}}^{\boldsymbol{T}}\boldsymbol{\Omega}^{-1}\left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_i - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*} - \boldsymbol{W}_{\boldsymbol{a}}\boldsymbol{x}^{(k-1)}\right))$
    Update
    $\boldsymbol{t_k} = \frac{1+\sqrt{1+4t_{k-1}^2}}{2}$
    Update
    $\boldsymbol{x}^{(k)} = \boldsymbol{\delta}_{ij}^{(k-1)} - \frac{t_{k-1}-1}{t_k}\left(\boldsymbol{\delta}_{ij}^{(k-1)} - \boldsymbol{\delta}_{ij}^{(k-2)}\right)$
  End
End

E. Sensitivity Analysis of the Bounds

In the modeling and analysis, how to determine the bounds $\boldsymbol{B}_L$ and $\boldsymbol{B}_U$ is a very important topic. The bounds are related to the product uniformity tolerance, which is a tradeoff between manufacturing cost and product quality. A sensitivity analysis is conducted to study how the optimal value changes with respect to perturbations of the constraints.

According to the analysis in Section III-D, the optimization problem (4) is convex. Meanwhile, Slater's condition holds for linear box constraints in this optimization problem [31]. Thus, the strong duality does hold for this optimization problem, which indicates that the best bound that can be obtained from the Lagrange dual function is tight.

In order to investigate the sensitivity of the optimal value with respect to the perturbations of the constraints, a perturbed version of the optimization is formulated as

$$\begin{aligned} \min_{\boldsymbol{\theta}_{ij},\boldsymbol{\delta}_{ij}} \; & (\boldsymbol{z}_{ij} - \boldsymbol{W}\boldsymbol{\theta}_{ij})^T\boldsymbol{\Omega}^{-1}(\boldsymbol{z}_{ij} - \boldsymbol{W}\boldsymbol{\theta}_{ij}) + \gamma\,|\boldsymbol{\delta}_{ij}|_1 \\ \text{subjectto } & \boldsymbol{B}_L - \boldsymbol{W}\boldsymbol{\theta}_{ij} \preccurlyeq \boldsymbol{\Delta B}_L \\ & \boldsymbol{W}\boldsymbol{\theta}_{ij} - \boldsymbol{B}_U \preccurlyeq \boldsymbol{\Delta B}_U \end{aligned} \tag{6}$$

where $\boldsymbol{\Delta B}_L$ and $\boldsymbol{\Delta B}_U$ are perturbated values of the bounds of normal effects, and $\boldsymbol{z}_{ij} = \boldsymbol{y}_{ij} - \boldsymbol{\mu}_i - \boldsymbol{W}_{\boldsymbol{a}}\boldsymbol{\delta}_{ij}$.

Let $G(\boldsymbol{\theta}_{ij}, \boldsymbol{\delta}_{ij}) := (\boldsymbol{z}_{ij} - \boldsymbol{W}\boldsymbol{\theta}_{ij})^T\boldsymbol{\Omega}^{-1}(\boldsymbol{z}_{ij} - \boldsymbol{W}\boldsymbol{\theta}_{ij}) + \gamma\,|\boldsymbol{\delta}_{ij}|_1$. Let $p^*(\boldsymbol{\Delta B}_L, \boldsymbol{\Delta B}_U)$ be the optimal value of the perturbed problem in (6). Thus, $p^*(0, 0)$ is the optimal value of the original unperturbed problem in (4). Let $(\boldsymbol{\lambda_1^*}, \boldsymbol{\lambda_2^*})$ be optimal for the dual equation, then we obtain [31]

$$p^*(\boldsymbol{\Delta B}_L, \boldsymbol{\Delta B}_U) \geq p^*(0, 0) - \boldsymbol{\lambda_1^{*T}}\boldsymbol{\Delta B}_L - \boldsymbol{\lambda_2^{*T}}\boldsymbol{\Delta B}_U.$$

We can draw the conclusions that: 1) if λ_{1k}^* or λ_{2k}^* is large and we tighten the kth constraint by choosing the negative $\boldsymbol{\Delta B}_{Lk}$ or $\boldsymbol{\Delta B}_{Uk}$, then the optimal value $p^*(\boldsymbol{\Delta B}_L, \boldsymbol{\Delta B}_U)$ will increase greatly and 2) if λ_{1k}^* or λ_{2k}^* is small and we lossen the kth constraint by choosing the positive $\boldsymbol{\Delta B}_{Lk}$ or $\boldsymbol{\Delta B}_{Uk}$, then the optimal value $p^*(\boldsymbol{\Delta B}_L, \boldsymbol{\Delta B}_U)$ will not decrease too much. Based on the Karush–Kuhn–Tucker condition of the optimality [31], the constraint $\boldsymbol{B}_{Lk} = (\boldsymbol{W}\boldsymbol{\theta}_{ij})_k$ will be active if λ_{1k}^* is nonzero. In practice, the selection of bounds can be done by running the modeling and analysis based on phase-I data sets, and checking whether the detected defective part is acceptable under the corresponding uniformity tolerance. If some small peaks detected as defective parts are acceptable under uniformity tolerance, we can loose the corresponding bounds to realize better decomposition. The sensitivity analysis of the bounds will be discussed in the surrogated data analysis in Section IV-B.

F. Criterion for Selection of the Penalized Parameter

The penalized parameter γ directly controls the tradeoff of weighted sum of square loss function for the signal-dependent noise $\boldsymbol{e}_{ij}^T\boldsymbol{\Omega}^{-1}\boldsymbol{e}_{ij}$ and L_1 norm regularization of the anomalous coefficients $|\boldsymbol{\delta}_{ij}|_1$. Thus, γ has a significant impact on the performance of the PMD. In this paper, we use the cross validation to determine the value of the penalized parameter γ. The optimal penalized parameter γ suffices minimum mean

sum of square of difference between the original signals and estimated signals

$$\gamma^* = \underset{\gamma \in R^+}{\operatorname{argmin}} \frac{1}{n\left(\sum_{i=1}^{N} n_i\right)} \sum_{i=1}^{N} \sum_{j=1}^{n_i} (\boldsymbol{y}_{ij} - \bar{\hat{\boldsymbol{y}}}_{i[j]})^T (\boldsymbol{y}_{ij} - \bar{\hat{\boldsymbol{y}}}_{i[j]})$$

where $\boldsymbol{y}_{ij}$ is the original jth measurement profile with dimension n in the ith sample; $\bar{\hat{\boldsymbol{y}}}_{\boldsymbol{i[j]}}$ is the estimated jth measurement profile in the ith sample. $\bar{\hat{\boldsymbol{y}}}_{\boldsymbol{i[j]}}$ is estimated by the mean profile of other reconstructed profiles except the jth one in the ith sample. $\bar{\hat{\boldsymbol{y}}}_{\boldsymbol{i[j]}} = \sum_{h=1,h\neq j}^{n_i} (\hat{\boldsymbol{\mu}}_i + \boldsymbol{W}\hat{\boldsymbol{\theta}}_{ih} + \boldsymbol{W_a}\hat{\boldsymbol{\delta}}_{ih})$.

This criterion makes full use of the uniformity of different profiles in the same sample. In-control profiles should be used when the optimal penalized parameter γ^* is selected.

IV. Surrogated Data Analysis of Raman Spectra

In this section, the performance of the PMD is evaluated through the surrogated Raman spectra from a practical experiment. Twenty Raman spectra are collected by measuring a piece of SWCNTs buckypaper sample. A Renishaw inVia confocal micro-Raman microscopy is used in the experiment, which is set with 785-nm laser source and 0.5-s exposure time. Ensemble averaging is applied to improve the signal-to-noise ratio and generate the mean vector as the ideal Raman spectrum for further surrogate profiles generation. Ten surrogated Raman spectra are generated based on the ideal Raman spectrum. Specifically, we consider three types of defects in different bands of the profiles: D-band in the third profile, a middle band between D-band and G-band in the sixth profile, and G-band in the tenth profile, respectively. Moreover, signal-dependent noise is generated by $\boldsymbol{e} \sim \boldsymbol{N}(\boldsymbol{0}, \text{diagonal}(\boldsymbol{S}) + 42)$, and $\boldsymbol{S}$ is the corresponding spectrum with dimension 512. The diagonal($\boldsymbol{S}$) denotes the matrix with diagonal vector $\boldsymbol{S}$ and zero nondiagonal elements. The signal-dependent noise distribution has been validated by the real Raman spectra data in [6]. All the profiles with defective information and signal-dependent noise are shown in Fig. 3, and the defective parts are marked by the red circles. We can observe that the defects are hidden in the noise, and it is not feasisble to pick up defective profiles mannually.

A. Performance Comparison Between the PMD and the LMM/SSD

We compare our proposed PMD with two benchmark methods: the LMM and the SSD as introduced in Section III. For the LMM, the median vector of all the profiles are extended to generate the design matrix of the fixed effects; and the design matrix of the random effects are generated based on the linear spline regression settings [26]. For the SSD, two cubic spline basis matrices are generated as the design matrices with 100 and 128 kn; and the tuning parameters are selected automatically based on the generalized cross validation criterion and Otsu's method [27]. The SSD is implemented to analyze all ten profiles. In the PMD, the penalized parameter γ can be determined by the cross validation in Section III-F, and the selected penalized parameter γ^* is 0.15 in this surrogate data analysis. The weight vectors $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ were adjusted carefully to make that the sensitivity of detecting defects in the D-band and the G-band regions is higher than other regions since the D-band and the G-band are closely related to the degree of functionalization and degree of alignment of CNT buckypaper. According to the comparison among Coiflet, Daubechies, and Symmlet families, Daubechies 18 is chosen as the wavelet basis upon a criterion defined by the optimal estimated peak intensity [6]. Corresponding wavelet design matrices $\boldsymbol{W}$ and $\boldsymbol{W_a}$ are generated based on the first level of wavelet coefficients. The variance–covariance matrix is $\boldsymbol{\Omega} = \text{diagonal}(\boldsymbol{S}) + 42$.

TABLE I
Detect Rate, FAR, Detected Intensity Difference, and Computation Time for Different Methods

	LMM	SSD	PMD
Detection Rate (DR)	100.00%	13.33%	100.00%
False Alarm Rate (FAR)	85.69%	2.21%	2.63%
Detected Peak Intensity Difference (DPID)	77.73%	74.10%	30.95%
Mean Square Error (MSE)	9.93	10.06	9.18
Computation Time	1203.27s	0.55s	0.94s

In order to evaluate the performance of the proposed PMD and the benchmark methods, we have defined the following criteria for comparison:

1) detection rate (DR) defined as the proportion of the real defect points detected as the defect points;
2) false alarm rate (FAR), defined as the percentage of the false classification of defect points in all the nondefective points, which is the smaller, the better;
3) detected peak intensity difference (DPID), defined as $1/n_d \sum_{i=1}^{n_d} |\max(I_i) - \max(I_i^*)|/|\max(I_i^*)|$, where $\mathrm{n_d}$ denotes the number of real defect bands, $\max(I_i)$ denotes the detected peak intensity in band i, $\max(I_i^*)$ denotes the real peak intensity in the band I, which is the smaller, the better;
4) mean square error (MSE);
5) computation time.

The criterion (1) is the larger the better. The criteria (2–5) are the smaller the better.

The decomposition results by the LMM, SSD, and PMD are shown in Fig. 4 and Table I. The fixed effects, random effects, and noise separated by the LMM are shown in Fig. 4(a)–(c), respectively. We can observe that the random effects separated by the LMM include both normal and defective effects. Also, part of defective effects remains in the noise. The DR is 100%, but the FAR is as high as 85.69%, and the DPID is 77.73%. Thus, it is difficult to quantify the uniformity and detect the defects based on the LMM. Fig. 4(d)–(f) shows the decomposition results by the SSD, including smooth part, sparse part, and noise. The FAR is only 2.21% but the DR is very low (13.33%), and the DPID is huge (74.10%). The reason for such a low DR and a large DPID is that the SSD only applied to curves with smooth functional mean, and it mistakenly detected the spikes existed in the G-band of Raman spectra in the normal situation as anomalies due to the spline basis, while the simulated defects remain in the smooth part with fixed effects and normal effects. Therefore, we cannot realize the quantification of uniformity and detection only based on the smooth part. With the PMD, the result

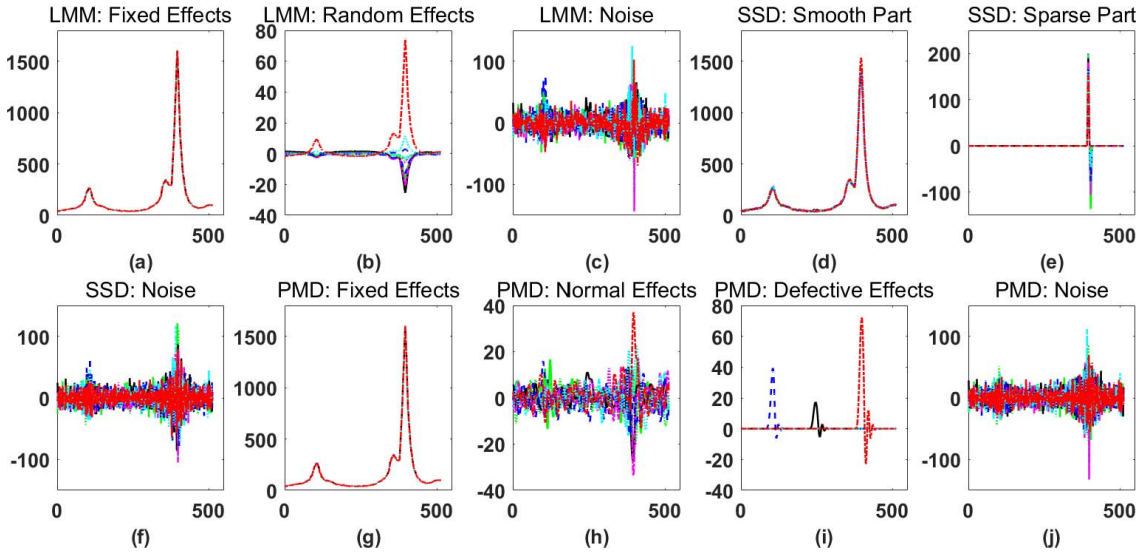


Fig. 4. Decomposition comparison among the LMM, the SSD, and the PMD. [Note that all the x-axis denotes Raman shift index, and all the y-axis denotes the Raman intensity (a.u.).] (a)–(c) Fixed effects, random effects, and noise separated by the LMM method, respectively. (d)–(f) Smooth part, sparse part, and noise separated by the SSD method, respectively. (g)–(j) Fixed effects, normal effects, defective effects, and noise separated by the proposed PMD method, respectively.

of successfully separating the whole profiles into four parts including fixed effects within each sample, normal effects within each sample, defective effects within each sample, and signal-dependent noise are shown in Fig. 4(g)–(j). Monitoring the fixed effects in Fig. 4(g) for different samples can reveal fabrication consistency of quality features, which is the long-term mean shift among samples as the sequentially buckypaper fabrication process goes on. Meanwhile, the normal effects, which are usually due to the inherent uncertainty of regular product fabrication and measurement, are separated and shown in Fig. 4(h). As shown in Fig. 4(i), the defects information in the third, sixth, and tenth profiles, can be detected in the defective effects without false detection in other profiles. The DR is 100%. The FAR is 2.63% because of the boundary effect when applying wavelet basis as shown in Fig 4(i). The DPID is 30.95%.

In addition, the signal-dependent noise is shown in Fig. 4(j). According to these four decomposed effects, we can monitor fabrication consistency, within-sample uniformity, and within-sample defect information of quality features based on in-line Raman spectroscopy.

Table I shows that the MSE of the LMM, SSD, and PMD are 9.93, 10.06, and 9.18, respectively. The MSE of the PMD is the smallest one, which makes sense because the normal effects and defective effects are separated thoroughly. The computation time of the LMM, SSD, and PMD are 1203.27, 0.55, and 0.94 s. The computation time of the LMM is very large since the estimation based on REML estimation is very slow while the APG algorithm for the SSD and the PMD is very quick. The computation efficiency of the SSD and the PMD meet the requirements of real-time process monitoring based on in-line Raman spectroscopy.

The dimension of the original Raman spectra is 512 while only 64 wavelet coefficients in the first level are extracted for normal effects. The dimension of $\boldsymbol{\theta}_{ij}$ is much lower than the original dimension of Raman spectra, which improves the

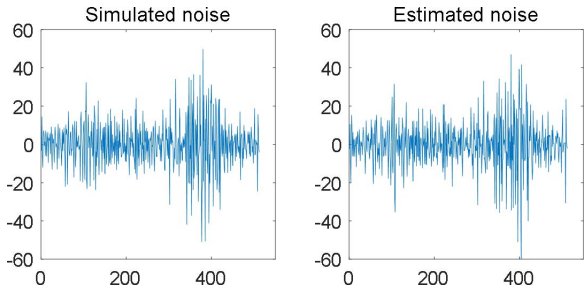


Fig. 5. Comparison between the simulated noise and the estimated noise.

efficacy and efficiency of profile modeling. Similarity measure can then be used to quantify the uniformity among ten profiles. Another 64 wavelet coefficients of defective effects $\boldsymbol{\delta}_{ij}$ includes the defective information in corresponding profiles. Due to the property of wavelet transformation, each defective coefficient is related to a specific band in the Raman spectra, which provides useful information for root cause analysis of the defect.

In terms of noise, the simulated noise (underlying truth) and the estimated noise by the PMD are shown in Fig. 5. We observe that the estimated noise after applying the PMD is quite consistent with the original simulated noise during generating simulation profiles. We introduce an index to quantify the difference between the simulated noise and the estimated noise, $D_e = \|e_{\text{sim}} - e_{\text{est}}\|_F^2 / \|e_{\text{sim}}\|_F^2$, where e_{sim} and e_{est} are the simulated noise and the estimated noise, respectively, and $\|\cdot\|_F^2$ is square of the Frobenius norm. In the surrogated data analysis, $D_e = 0.3029$, which implies the nonuniform noise is modeled well in the PMD.

B. Sensitivity Analysis of the Bounds and the Penalized Parameter

In order to evaluate detection accuracy of the PMD with different bounds choice, we fix the penalized parameter to 0.15. Then, we compare the simulated defects (underlying

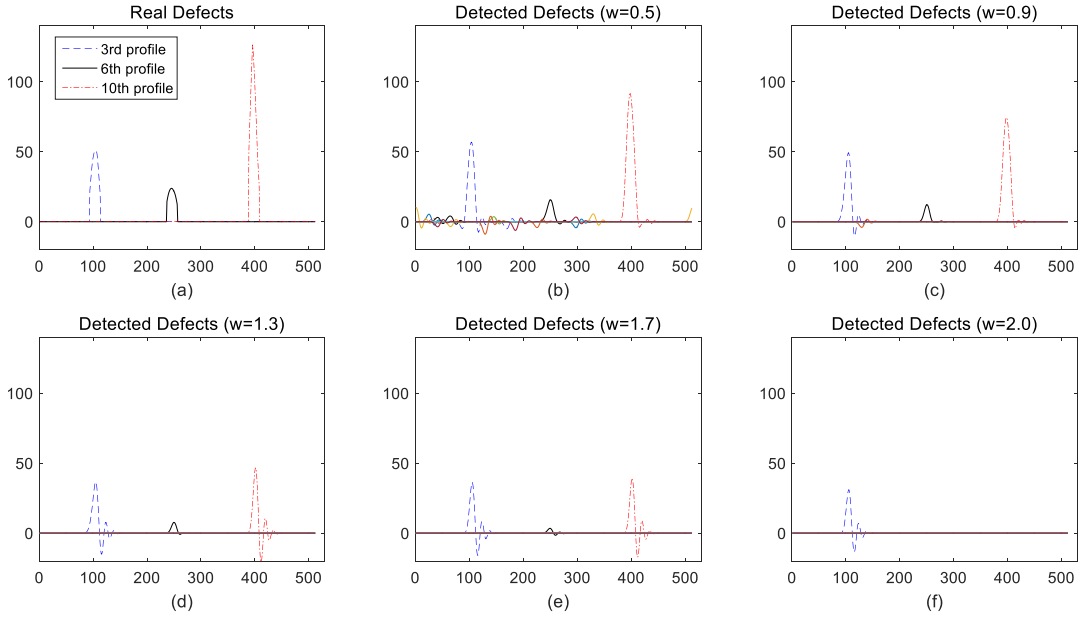


Fig. 6. Comparison between real defects and detected defects in different profiles under different bounds. [Note that all the x-axis denotes Raman shift index, and all the y-axis denotes the Raman intensity (a.u.).] (a) Real defects in three profiles. (b)–(f) Detected defects by the proposed PMD method under different bound parameter w.

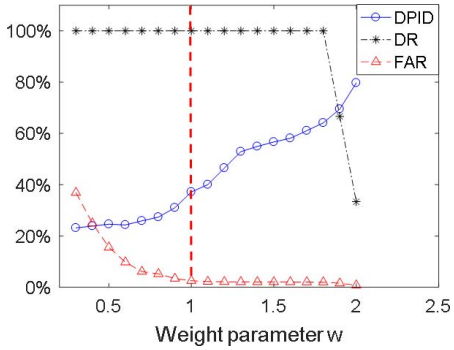


Fig. 7. DPID, DR, and FAR at different weighting coefficients w. (Note that weighting coefficients correspond to different bounds by $\boldsymbol{\omega}_1 = -\boldsymbol{\omega}_2 = w\mathbf{1}$; the red-dashed vertical line denotes the selected weight parameter.)

truth) introduced in the surrogated data and the detected defects in the decomposed defective effects under different bounds. Let bounds $\boldsymbol{\omega} = \boldsymbol{\omega}_1 = -\boldsymbol{\omega}_2 = w\mathbf{1}$, where w is the weighting coefficient. The visual comparison can be found in Fig. 6(a)–(f).

The comparison based on the criteria, including DPID, DR, and FAR are shown in Fig. 7. Fig. 6(a) corresponds to the simulated defects in the surrogated data. Fig. 6(b)–(f) is associated with decomposed defective effects under $w = 0.5, 0.9, 1.3, 1.7,$ and 2, respectively. From Fig. 7, it is obvious that the DPID tends to become larger as the weighting coefficient w increases. The DR is lower than 100% when $w > 1.8$. The FAR becomes lower as the weighting coefficient w increases, and FAR is close to zero when $w \geq 1$. The DR is high and meanwhile the FAR is low when $1 \leq w < 1.8$. Also the result is consistent with the detected defects shown visually in Fig. 6.

All the defects in different profiles can be detected accurately when $w = 1.3$ or 1.7. The detected Raman intensities tend to be a little smaller than the Raman intensities of the true defects. This can be explained from the following three aspects.

1) The most important one is due to the use of L_1 norm regularization, which leads to the soft-thresholding of the true defective information.
2) The reduced amount of detected defects can also be explained as pure normal effects within the bound. The detection sensitivity is controlled by the setting of bounds. The real defective information exceeds the bounds cannot be interpreted by the pure normal effects and it will be recognized by the defective effects. Thus the reduced amount of detected defects tend to be large when the bounds change from 0.5 to 2, which is consistent with the trend of DPID in Fig. 7. If the weighting vector is $w = 0.5$ and 0.9, there are false detection in other normal profiles shown in Fig. 6(b) and (c), which will increase the type-I error during monitoring. While applying large weight vector such as $w = 2$ results in under-detection in the sixth and the tenth profiles shown in Fig. 6(f), which will increase the type-II error during monitoring. A reasonable setting of bounds can realize the tradeoff between the detection power and the FAR in the monitoring. Usually we choose tighter bounds under the D-band, the G-band, or other important zones in a Raman spectra since they are closely related to the degree of functionalization and the degree of alignment for the quality of buckypaper.
3) Although we choose oracle bounds for the PMD, the detected defects are smaller than the simulated defects. One reason of this is that the signal-dependent noise compensates part of defective change in the profiles and the compensation results in the reduced amount of Raman intensity in defective effects.

We also explore the sensitivity analysis of the penalized parameter γ in the surrgoated analysis. In this case, we fix the bounds by $\boldsymbol{\omega}_1 = -\boldsymbol{\omega}_2 = \mathbf{1}$. The change of quantitative indices, including DPID, DR, and FAR, over the penalized parameter γ is shown in Fig. 8. From Fig. 8, it is obvious that

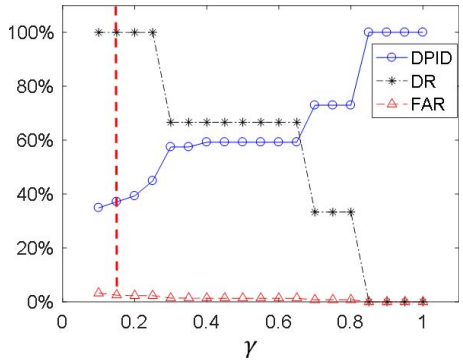


Fig. 8. DPID, DR, and FAR at different penalized parameters γ. (Note that the red-dashed vertical line denotes the selected penalized parameter γ^*.)

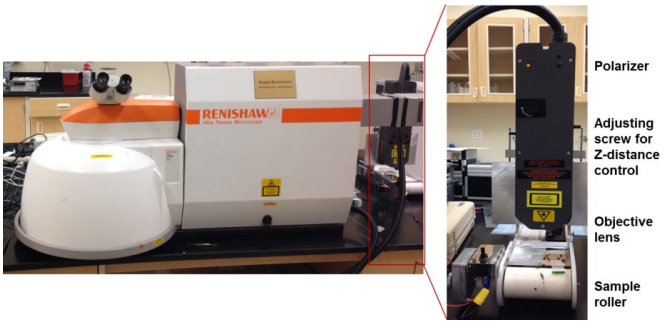


Fig. 9. Renishaw inVia micro-Raman system with custom-designed remote optical probe and roller sample stage.

the DPID tends to become larger as the penalized coefficient γ increases, while the DR tends to become smaller. The reason of aforementioned properties is that larger penalized parameter results in more regularization on the L_1 norm of the anomalous coefficients. The pattern of DR and DPID has step/ladder patterns because there are three types of defects in different bands of the profiles in the surrogated data. Once a defect is missed, the DR has a step decrease. For the FAR, we can find that it decreases slightly as the increase of penalized parameter. That means the FAR is not highly dependent on the penalized parameter, and it is mainly influenced by the bounds ($\boldsymbol{B}_L$ and $\boldsymbol{B}_U$). Furthermore, the selected optimal penalized parameter $\gamma^* = 0.15$ obtained via the cross validation is shown in the red dashed line in Fig. 8. The selection of γ^* can achieve very high DR with low FAR and DPID.

V. Case Study

In this section, the proposed PMD is applied to a case study of in-line Raman spectra modeling and analysis. The setup of in-line Raman spectroscopy is shown in Fig. 9. In the experimental setup of the case study, near infrared laser with a wavelength of 785 nm and a power of 150 mW was used to eliminate the effect of ambient light. Low magnification lens was used to achieve a larger focus tolerance.

Raman mapping technique is used to record Raman spectra in the measurement zone with a rectangular shape and a grid pattern. The measurement points and part of the quality information obtained by the PMD are shown in Fig. 10.

As shown in Fig. 10, ten samples with each sample containing ten measurement points are tested sequentially with a distance of 12 μm in the x-axis, which are marked with the

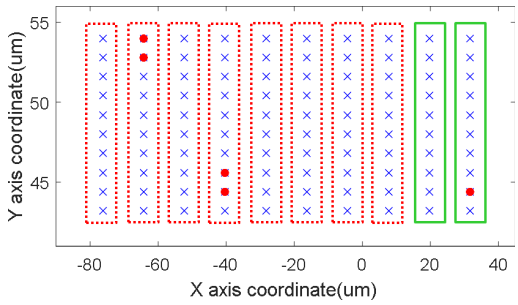


Fig. 10. In-line Raman spectroscopy measurement zone and quality information obtained by the PMD.

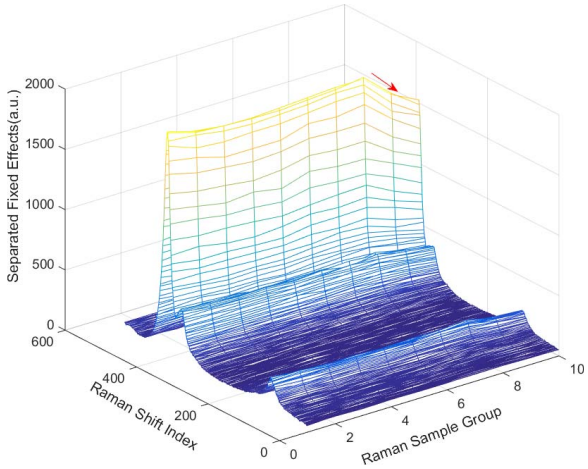


Fig. 11. Fixed effects separated by the PMD in different Raman samples.

red rectangular (dotted line) or green rectangular (line). All measurement locations are marked with a blue cross or a red circle. Raman spectrum with 512 Raman shifts (dimension) is collected for each measurement point. Within the same Raman sample, the distance between two measurement points is 1.2 μm in the y-axis, which is much smaller than the distance (12 μm) between samples in the x-axis. All the Raman spectra are collected based on a piece of SWCNTs buckypaper in the Renishaw confocal research Raman microscopy with 785-nm laser source and 0.5-s exposure time for each measurement point.

The PMD is implemented to decompose the in-line Raman spectra into fixed effects, normal effects, defective effects, and signal-dependent noise. The fixed effects of these ten Raman samples are shown in Fig. 11. We can observe that there is a mean shift started from the ninth sample. The samples after the mean shift are marked by the green rectangulars in Fig. 10. The Raman intensity tends to be smaller than other normal samples, which likely comes from the measurement equipment (the focus depth change due to sample local deformation) or product quality (the degree of functionalization, the degree of alignment changes, nanotube distribution, and dispersion).

The defective effects separated by the PMD show that the first two measurement points in the second sample, the eighth and the ninth points in the fourth sample, and the ninth point in the tenth sample are defective points, which are marked as red circles in Fig. 10. The original Raman spectra and defective effects by the PMD in the second, the fourth, and the tenth samples are shown in Fig. 12. The sixth sample,

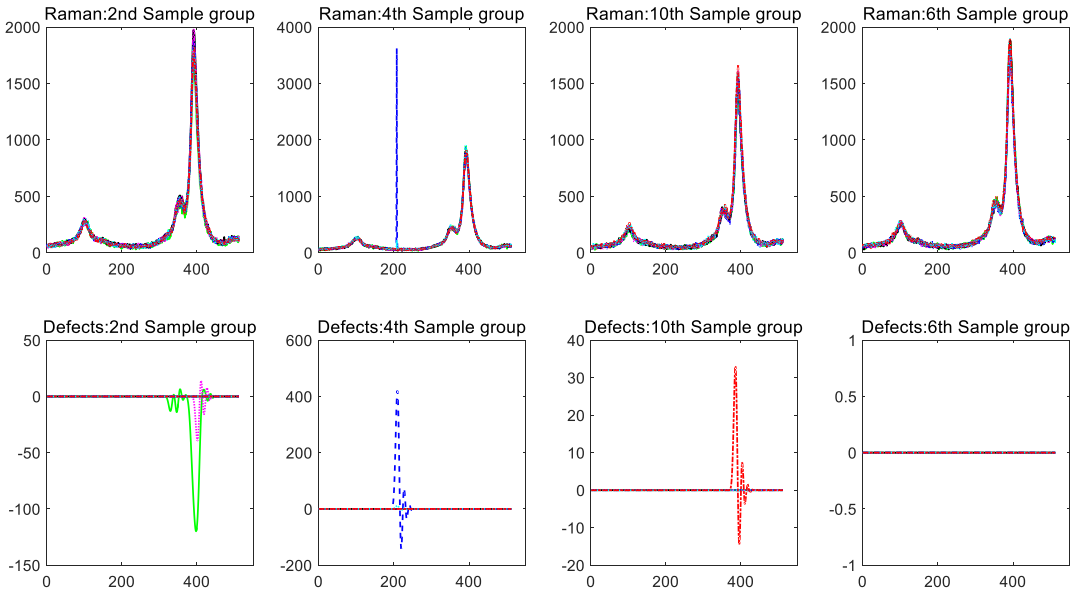


Fig. 12. Original Raman spectra and defective effects separated by the PMD in the second, fourth, sixth, and tenth samples.

which is a normal Raman sample, is also shown in Fig. 12 for the benchmark. We can observe that the defective effects remain constant for the sixth sample. We find that the defective effects occur in the G-band zone for the second and tenth samples. The defects are due to impurity of raw material, different degree of functionalizations, different alignment of CNTs, or bad nanotube dispersion. Moreover, a large spike is detected as defective random effect between the D-band and the G-band for the fourth sample, which may result from some measurement errors, such as external light.

VI. Summary

In the continuous CNT buckypaper manufacturing process, in-line Raman spectroscopy is applied in collecting Raman spectra, which is a complicated profile data, to monitor the quality characterization and quantification of buckypaper. Three quality characteristics are primary concerns during the monitoring and control of buckypaper fabrication process: 1) quality consistency; 2) within-sample uniformity; and 3) within-sample defect information.

The main contribution of this paper is to propose a new PMD method to achieve multichannel profile detection and analysis based on in-line Raman spectroscopy in the continuous nanomanufacturing process. The proposed PMD can decompose the profiles into four parts: fixed effects, normal effects, defective effects, and signal-dependent noise. Wavelet basis is applied as a design matrix in the penalized mixed model because its ability for sparse representation of the Raman spectrum and its ability to handle nonsmooth profiles. A weighted sum of square loss function is applied to model the signal-dependent noise and the L_1 norm regularization is implemented to induce the sparsity of the defective random coefficients. The PMD combines the advantage of the LMM and SSD, and performs better in multichannel profile detection based on the in-line Raman spectroscopy. An APG-based algorithm is developed to efficiently handle the parameter estimation of the PMD with linear constraints. The developed algorithm is computationally efficient and meets the real-time monitoring requirement during the CNT buckypaper fabrication process. Furthermore, the sensitivity analysis of the bounds is conducted to demonstrate how to determine proper bounds to achieve a tradeoff between the detection power and the FAR.

A surrogated data analysis is conducted to compare the performance of the PMD with the LMM and the SSD. The PMD decompose the simulated profiles into four parts successfully. The suitable bounds can be determined by the sensitivity analysis and the product tolerance. We have defined the five criteria for comparison: 1) detection rate (DR); 2) false alarm rate (FAR); 3) detected peak intensity difference (DPID); 4) mean square error (MSE); and 5) computation time. The PMD can get a higher DR with a lower FAR and a lower DPID compared to the benchmark methods. The PMD can realize 100% DR with 2.63% FAR and 30.95% DPID. In addition, the MSE and the computation time are lower than the benchmark methods. The case study shows that the decomposition based on the PMD is able to detect the mean shift of fixed effects and detect the defects in different bands of different profiles. All these studies indicate that the PMD works well for the modeling and analysis of multichannel in-line Raman spectroscopy.

Appendix A

Proposition 1 Proof: Given $\boldsymbol{\delta}_{ij}, \boldsymbol{\theta}_{ij}$ can be derived by the optimization problem as follows:

$$\boldsymbol{\theta}_{ij} = \underset{\boldsymbol{\delta}_{ij}}{\operatorname{argmin}} (\boldsymbol{z}_{ij} - \boldsymbol{W}\boldsymbol{\theta}_{ij})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{z}_{ij} - \boldsymbol{W}\boldsymbol{\theta}_{ij}) + \gamma \, |\boldsymbol{\delta}_{ij}|_1$$
$$\text{subjectto} \quad \boldsymbol{B}_L \preccurlyeq \boldsymbol{W}\boldsymbol{\theta}_{ij} \preccurlyeq \boldsymbol{B}_U$$

where $\boldsymbol{z}_{ij} = \boldsymbol{y}_{ij} - \boldsymbol{\mu}_i - \boldsymbol{W}_a \boldsymbol{\delta}_{ij}$.

Since $\boldsymbol{\Omega}$ is the diagonal matrix with each entry is positive, we can solve the optimization componentwise by univariate quadratic programming with linear constraints. (Note that if $\boldsymbol{\Omega}$ is not diagonal, the optimization can be solved by the quadratic programming with linear constraints. However, the computation speed will be slower.) Considering the orthogonal

wavelet basis matrix $\boldsymbol{W}$, the problem can be solved by a closed-form solution

$$\boldsymbol{\theta}_{ij}^{*} = \boldsymbol{W}^{T}\boldsymbol{z}_{ij} \circ \boldsymbol{I}(\boldsymbol{B}_{L} \preccurlyeq \boldsymbol{z}_{ij} \preccurlyeq \boldsymbol{B}_{U}) + \boldsymbol{W}^{T}\boldsymbol{B}_{L} \circ \boldsymbol{I}(\boldsymbol{z}_{ij} \preccurlyeq \boldsymbol{B}_{L}) + \boldsymbol{W}^{T}\boldsymbol{B}_{U} \circ \boldsymbol{I}(\boldsymbol{z}_{ij} \gg \boldsymbol{B}_{L})$$

It can be shown that

$$\boldsymbol{\delta}_{ij} = \underset{\boldsymbol{\delta}_{ij}}{\operatorname{argmin}} F(\boldsymbol{\delta}_{ij}) = \left(\boldsymbol{z}_{ij} - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*}\right)^{T} \boldsymbol{\Omega}^{-1} \left(\boldsymbol{z}_{ij} - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*}\right) + \gamma \left|\boldsymbol{\delta}_{ij}\right|_{1}$$

which is the weighted LASSO formulation.

APPENDIX B

Proposition 2 Proof: It is readily to show that $\boldsymbol{\Omega}^{-1}$ is a diagonal, positive definite matrix with all diagonal entries equal to the variance of Raman intensity corresponding to the related Raman shift. Thus, the spectrum norm of matrix $\boldsymbol{\Omega}^{-1}$ is the maximum eigenvalue of the matrix $\boldsymbol{\Omega}^{-1}$, which is $1/\min(\operatorname{diag}(\boldsymbol{\Omega}))$

$$\|\boldsymbol{\Omega}^{-1}\| = 1/\min(\operatorname{diag}(\boldsymbol{\Omega})).$$

We can write down the gradient function

$$\begin{aligned}
&\nabla f(\boldsymbol{a}) \\
&= \nabla\left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{i} - \boldsymbol{W}_{\boldsymbol{a}}\boldsymbol{a} - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*}\right)^{T} \boldsymbol{\Omega}^{-1} \left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{i} - \boldsymbol{W}_{\boldsymbol{a}}\boldsymbol{a} - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*}\right) \\
&= -2\boldsymbol{W}_{\boldsymbol{a}}^{\boldsymbol{T}} \boldsymbol{\Omega}^{-1} \left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{i} - \boldsymbol{W}_{\boldsymbol{a}}\boldsymbol{a} - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*}\right)
\end{aligned}$$

Then, we can get the difference between the gradient function for $\boldsymbol{a}$ and the one for $\boldsymbol{b}$ satisfies

$$\begin{aligned}
&\|\nabla f(\boldsymbol{a}) - \nabla f(\boldsymbol{b})\| \\
&\quad = \left\|2\boldsymbol{W}_{\boldsymbol{a}}^{\boldsymbol{T}} \boldsymbol{\Omega}^{-1} \boldsymbol{W}_{\boldsymbol{a}} (\boldsymbol{a} - \boldsymbol{b})\right\| \\
&\quad \le \left\|2\boldsymbol{W}_{\boldsymbol{a}}^{\boldsymbol{T}} \boldsymbol{\Omega}^{-1} \boldsymbol{W}_{\boldsymbol{a}}\right\|_{2} \cdot \|\boldsymbol{a} - \boldsymbol{b}\| \\
&\quad \le 2\|\boldsymbol{\Omega}^{-1}\| \|\boldsymbol{W}_{\boldsymbol{a}}\|_{2}^{2} \cdot \|\boldsymbol{a} - \boldsymbol{b}\| \\
&\quad = (2/\min(\operatorname{diag}(\boldsymbol{\Omega})))\|\boldsymbol{W}_{\boldsymbol{a}}\|_{2}^{2} \cdot \|\boldsymbol{a} - \boldsymbol{b}\|.
\end{aligned}$$

Thus, $f(\cdot)$ is the Lipchiz continuous, which implies that there is a constant L that makes gradient function $\nabla f(\cdot)$ satisfing $\|\nabla f(\boldsymbol{a}) - \nabla f(\boldsymbol{b})\| \le L\|\boldsymbol{a} - \boldsymbol{b}\|$ for any $\boldsymbol{a}, \boldsymbol{b} \in \boldsymbol{R}$ with $L = (2/\min(\operatorname{diag}(\boldsymbol{\Omega})))\|\boldsymbol{W}_{\boldsymbol{a}}\|_{2}^{2}$.

APPENDIX C

Proposition 3 Proof: The proximal gradient algorithm for the PMD problem in (5), given by

$$\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k)} = \underset{\boldsymbol{\delta}_{ij}}{\operatorname{argmin}} \left\{ f\left(\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right) + \left\langle \boldsymbol{\delta}_{\boldsymbol{ij}} - \boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}, \nabla f\left(\boldsymbol{\delta}_{\boldsymbol{ij}}{}^{(k-1)}\right)\right\rangle + \frac{L}{2}\left\|\boldsymbol{\delta}_{ij} - \boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right\|^{2} + \gamma \left|\boldsymbol{\delta}_{ij}\right|_{1} \right\}$$

has a closed-form solution in each iteration k, in the form of a soft-thresholding function as follows:

$$\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k)} = S_{\frac{\gamma}{L}}\left(\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)} - \frac{2}{L}\boldsymbol{W}_{\boldsymbol{a}}^{\boldsymbol{T}} \boldsymbol{\Omega}^{-1}\left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{i} - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*} - \boldsymbol{W}_{\boldsymbol{a}}\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right)\right)$$

with $L = (2/\min(\operatorname{diag}(\boldsymbol{\Omega})))\|\boldsymbol{W}_{\boldsymbol{a}}\|_{2}^{2}$.

$S_{(\gamma/L)}(x) = \operatorname{sgn}(x)(|x| - (\gamma/L))_{+}$ is the soft-thresholding operator, and sgn(x) is the sign function and $\mathrm{x}_{+} = \max(x, 0)$.

To prove the proposal above, we write down the gradient function

$$\begin{aligned}
\nabla f(\boldsymbol{a}) &= -2\boldsymbol{W}_{\boldsymbol{a}}^{\boldsymbol{T}} \boldsymbol{\Omega}^{-1}\left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{i} - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*} - \boldsymbol{W}_{\boldsymbol{a}}\boldsymbol{a}\right) \\
\nabla f\left(\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right) &= -2\boldsymbol{W}_{\boldsymbol{a}}^{\boldsymbol{T}} \boldsymbol{\Omega}^{-1}\left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{i} - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*} - \boldsymbol{W}_{\boldsymbol{a}}\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right).
\end{aligned}$$

For each iteration given by $(k-1)$ step

$$\begin{aligned}
&\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k)} \\
&= \underset{\boldsymbol{\delta}_{ij}}{\operatorname{argmin}} \left\{ f\left(\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right) + \left\langle \boldsymbol{\delta}_{\boldsymbol{ij}} - \boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}, \nabla f\left(\boldsymbol{\delta}_{\boldsymbol{ij}}{}^{(k-1)}\right)\right\rangle + \frac{L}{2}\left\|\boldsymbol{\delta}_{ij} - \boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right\|^{2} + \gamma \left|\boldsymbol{\delta}_{ij}\right|_{1} \right\} \\
&= \underset{\boldsymbol{\delta}_{ij}}{\operatorname{argmin}} \left\{ \frac{1}{2L}\left\|\nabla f\left(\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right)\right\| + \left\langle \boldsymbol{\delta}_{\boldsymbol{ij}} - \boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}, \nabla f\left(\boldsymbol{\delta}_{\boldsymbol{ij}}{}^{(k-1)}\right)\right\rangle + \frac{L}{2}\left\|\boldsymbol{\delta}_{ij} - \boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right\|^{2} + \gamma \left|\boldsymbol{\delta}_{ij}\right|_{1} \right\} \\
&= \underset{\boldsymbol{\delta}_{ij}}{\operatorname{argmin}} \left\{ \frac{L}{2}\left\|\boldsymbol{\delta}_{ij} - \boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)} - \frac{1}{L}\nabla f\left(\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right)\right\|^{2} + \gamma \left|\boldsymbol{\delta}_{ij}\right|_{1} \right\} \\
&= \underset{\boldsymbol{\delta}_{ij}}{\operatorname{argmin}} \left\{ \frac{L}{2}\left\|\boldsymbol{\delta}_{ij} - \boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)} - \frac{2}{L}\boldsymbol{W}_{\boldsymbol{a}}^{\boldsymbol{T}} \boldsymbol{\Omega}^{-1}\left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{i} - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*} - \boldsymbol{W}_{\boldsymbol{a}}\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right)\right\| + \gamma \left|\boldsymbol{\delta}_{ij}\right|_{1} \right\} \\
&= S_{\frac{\gamma}{L}}\left(\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)} - \frac{2}{L}\boldsymbol{W}_{\boldsymbol{a}}^{\boldsymbol{T}} \boldsymbol{\Omega}^{-1}\left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{i} - \boldsymbol{W}\boldsymbol{\theta}_{ij}^{*} - \boldsymbol{W}_{\boldsymbol{a}}\boldsymbol{\delta}_{\boldsymbol{ij}}^{(k-1)}\right)\right)
\end{aligned}$$

which is the exact soft-thresholding solution.

REFERENCES

[1] J. Jin and J. Shi, "Automatic feature extraction of waveform signals for in-process diagnostic performance improvement," *J. Intell. Manuf.*, vol. 12, pp. 257–268, Jun. 2001.

[2] K. Paynabar and J. Jin, "Characterization of non-linear profiles variations using mixed-effect models and wavelets," *IIE Trans.*, vol. 43, no. 4, pp. 275–290, 2011.

[3] H. H. Gommans, J. W. Alldredge, H. Tashiro, J. Park, J. Magnuson, and A. G. Rinzler, "Fibers of aligned single-walled carbon nanotubes: Polarized Raman spectroscopy," *J. Appl. Phys.*, vol. 88, pp. 2509–2514, Aug. 2000.

[4] N. R. Raravikar, P. Keblinski, A. M. Rao, M. S. Dresselhaus, L. S. Schadler, and P. M. Ajayan, "Temperature dependence of radial breathing mode Raman frequency of single-walled carbon nanotubes," *Phys. Rev. B, Condens. Matter*, vol. 66, p. 235424, Dec. 2002.

[5] G. Févotte, "*In situ* Raman spectroscopy for in-line control of pharmaceutical crystallization and solids elaboration processes: A review," *Chem. Eng. Res. Des.*, vol. 85, no. 7, pp. 906–920, 2007.

[6] X. Yue *et al.*, "Generalized wavelet shrinkage of inline Raman spectroscopy for quality monitoring of continuous manufacturing of carbon nanotube buckypaper," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 1, pp. 196–207, Jan. 2017.

[7] Q. Cheng, B. Wang, C. Zhang, and Z. Liang, "Functionalized carbon-nanotube sheet/bismaleimide nanocomposites: Mechanical and electrical performance beyond carbon-fiber composites," *Small*, vol. 6, no. 6, pp. 763–767, 2010.

[8] D. A. Jack, C.-S. Yeh, Z. Liang, S. Li, J. G. Park, and J. C. Fielding, "Electrical conductivity modeling and experimental study of densely packed SWCNT networks," *Nanotechnology*, vol. 21, no. 19, p. 195703, May 2010.

[9] V. R. Joseph and Y. Hung, "Orthogonal-maximin Latin hypercube designs," *Stat. Sinica*, vol. 18, no. 1, pp. 171–186, 2008.

[10] L. Kang and S. L. Albin, "On-line monitoring when the process yields a linear profile," *J. Quality Technol.*, vol. 32, no. 4, pp. 418–426, 2000.

[11] K. Kim, M. A. Mahmoud, and W. H. Woodall, "On the monitoring of linear profiles," *J. Quality Technol.*, vol. 35, no. 3, p. 317, 2003.

[12] M. A. Mahmoud, P. A. Parker, W. H. Woodall, and D. M. Hawkins, "A change point method for linear profile data," *Quality Rel. Eng. Int.*, vol. 23, no. 2, pp. 247–268, 2007.

[13] S. A. Mosesova, H. A. Chipman, R. J. MacKay, and S. H. Steiner, "Profile monitoring using mixed-effects models," Dept. Statist. Actuarial Sci., Univ. Waterloo, Waterloo, ON, Canada, Tech. Rep. 06-06, 2006.
[14] C. Zou, F. Tsung, and Z. Wang, "Monitoring profiles based on nonparametric regression methods," *Technometrics*, vol. 50, no. 4, pp. 512–526, 2008.
[15] W. A. Jensen and J. B. Birch, "Profile monitoring via nonlinear mixed models," *J. Quality Technol.*, vol. 41, no. 1, pp. 18–34, 2009.
[16] J.-J. H. Shiau, H.-L. Huang, S.-H. Lin, and M.-Y. Tsai, "Monitoring nonlinear profiles with random effects by nonparametric regression," *Commun. Stat.-Theory Methods*, vol. 38, no. 10, pp. 1664–1679, 2009.
[17] S. I. Chang and S. Yadama, "Statistical process control for monitoring non-linear profiles using wavelet filtering and B-spline approximation," *Int. J. Prod. Res.*, vol. 48, no. 4, pp. 1049–1068, 2010.
[18] S. Zhou and J. Jin, "Automatic feature selection for unsupervised clustering of cycle-based signals in manufacturing processes," *IIE Trans.*, vol. 37, no. 6, pp. 569–584, 2005.
[19] Y. Ding, L. Zeng, and S. Zhou, "Phase I analysis for monitoring nonlinear profiles in manufacturing processes," *J. Quality Technol.*, vol. 38, no. 3, pp. 199–216, 2006.
[20] I. M. Johnstone and A. Y. Lu, "On consistency and sparsity for principal components analysis in high dimensions," *J. Amer. Stat. Assoc.*, vol. 104, no. 486, pp. 682–693, 2012.
[21] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Statist.*, vol. 15, no. 2, pp. 265–286, Jun. 2006.
[22] J. Jin and J. Shi, "Feature-preserving data compression of stamping tonnage information using wavelets," *Technometrics*, vol. 41, no. 4, pp. 327–339, 1999.
[23] E. Chicken, J. J. Pignatiello, Jr., and J. R. Simpson, "Statistical process monitoring of nonlinear profiles using wavelets," *J. Quality Technol.*, vol. 41, no. 2, pp. 198–212, 2009.
[24] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
[25] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 35, nos. 1–2, pp. 108–126, 2013.
[26] A. Gałecki and T. Burzykowski, *Linear Mixed-Effects Models Using R: A Step-by-Step Approach*. New York, NY, USA: Springer, 2013.
[27] H. Yan, K. Paynabar, and J. Shi, "Anomaly detection in images with smooth background via smooth-sparse decomposition," *Technometrics*, vol. 59, no. 1, pp. 102–114, 2017.
[28] H. Zou, "The adaptive lasso and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006.
[29] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.
[30] P. Tseng. *On Accelerated Proximal Gradient Methods for Convex-Concave Optimization*. Accessed: Oct. 30, 2017. [Online]. Available: https://www.mit.edu/~dimitrib/PTseng/papers.html
[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

Xiaowei Yue (S'15) received the B.S. degree in mechanical engineering from the Beijing Institute of Technology, Beijing, China, in 2011, the M.S. degree in power engineering and engineering thermophysics from Tsinghua University, Beijing, in 2013, and the M.S. degree in statistics from the Georgia Institute of Technology, Atlanta, GA, USA, in 2016, where he is currently pursuing the Ph.D. degree with the H. Milton Stewart School of Industrial and Systems Engineering.

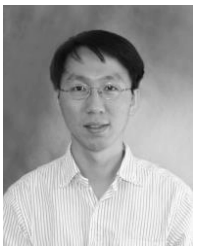
His research interests include data analytics for advanced manufacturing.

Mr. Yue is a member of the American Society of Mechanical Engineers (ASME), ASQ, the Institute for Operations Research and the Management Sciences (INFORMS), and the Institute of Industrial and Systems Engineers (IISE).

Hao Yan received the B.S. degree in physics from Peking University, Beijing, China, in 2011, the M.S. degree in statistics from the Georgia Institute of Technology, Atlanta, GA, USA, in 2015, and the M.S. degree in computational science and engineering and the Ph.D. degree in industrial and systems engineering from the Georgia Institute of Technology in 2017.

He is currently an Assistant Professor with the School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA. His research interests include the development of machine learning algorithms for modeling large-scale high-dimensional data in the manufacturing system.

Dr. Yan is a member of the Institute for Operations Research and the Management Sciences (INFORMS) and the Institute of Industrial and Systems Engineers (IISE).



Jin Gyu Park received the B.S., M.S., and Ph.D. degrees in physics from Seoul National University, Seoul, South Korea, in 1997, 1999, and 2003, respectively.

He is currently a Research Faculty Member with the High-Performance Materials Institute, Florida State University, Tallahassee, FL, USA. His main research areas are the physical properties characterization of carbon-related materials including electrical/thermal transport, Raman spectroscopy of carbon materials, and electron microscopy of carbon nanotube and its composites.

Zhiyong Liang received the B.S., M.S., and Ph.D. degrees in materials science and engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1987, 1990, and 2000, respectively.

He was an Associate Director of the Polymer and Composite Program at the Beijing University of Aeronautics and Astronautics. He is currently a Professor of Industrial and Manufacturing Engineering with FAMU-FSU College of Engineering, and the Director of the High-Performance Materials Institute, Florida State University, Tallahassee, FL, USA. His research interests include carbon nanotube and its nanocomposite for multifunctional applications with enhanced structural and electrical/thermal properties, the characterization of carbon nanotube/resin interface and packing structures, and scale-up nanomanufacturing.

Jianjun Shi received the B.S. and M.S. degrees in electrical engineering from the Beijing Institute of Technology, Beijing, China, in 1984 and 1987, respectively, and the Ph.D. degree in mechanical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1992.

He is currently the Carolyn J. Stewart Chair Professor with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA. His research interests include the fusion of advanced statistical and domain knowledge to develop methodologies for modeling, monitoring, diagnosis, and control for manufacturing systems.

Dr. Shi is a Fellow of the Institute of Industrial and Systems Engineers (IISE), the American Society of Mechanical Engineers (ASME), and the Institute for Operations Research and the Management Sciences (INFORMS), an Academician of the International Academy for Quality, an elected member of the ISI, and a Life Member of ASA.