

Discovering the Nature of Variation in Nonlinear Profile Data

Zhenyu Shi and Daniel W. Apley¹

Department of Industrial Engineering & Management Sciences, Northwestern University
Evanston, Illinois 60208-3119

zhenyushi2013@u.northwestern.edu, apley@northwestern.edu

George C. Runger

School of Computing, Informatics, and Decision Systems Engineering, Arizona State University
Tempe, AZ 85287

george.runger@asu.edu

Abstract

Profile data have received substantial attention in the quality control literature. Most of the prior work has focused on the profile monitoring problem of detecting sudden changes in the characteristics of the profiles, relative to an in-control sample set of profiles. In this paper we present an approach for exploratory analysis of a sample of profiles, the purpose of which is to discover the nature of any profile-to-profile variation that is present over the sample. This is especially challenging in big data environments in which the sample consists of a stream of high dimensional profiles, such as image or point cloud data. We use manifold learning methods to find a low-dimensional representation of the variation, followed by a supervised learning step to map the low-dimensional representation back into the profile space. The mapping can be used for graphical animation and visualization of the nature of the variation, in order to facilitate root cause diagnosis. Although this mapping is related to a nonlinear mixed model sometimes used in profile monitoring, rather than assuming some prespecified parametric profile model and monitoring for variation in those specific parameters, our focus is on discovering an appropriate characterization of the profile-to-profile variation. We illustrate with two examples and include an additional example in the online supplement to this article on the *Technometrics* website.

Keywords: Profile data, manifold learning, independent component analysis, principal component analysis, visualization.

¹ Corresponding author.

1. Introduction

Profile data collected for quality control purposes is increasingly common in many industries. For the discrete parts industries, these include spatially dense surface measurement data from machine vision imaging and/or tracing across the part surface via either contact stylus probes or noncontact laser probes. More general image data, which can be viewed as profiles over a 2D domain, are commonplace too (Megahed, et al. 2011). Profiles can also represent machine signatures that are related to product quality, such as stamping press tonnage measured over the time duration of a stamping cycle (Jin and Shi, 1999). Profile data are also common in chemical, pharmaceutical, and agricultural industries, for example as dose-response profiles (Williams, Birch, Woodall, and Ferry, 2007). Profile data can be extremely high dimensional (e.g., images with millions of pixels, surface point cloud measurement data with tens of millions of points per part, etc.), and a steady stream of such data may be available for analysis. How to effectively reduce dimensionality and identify fundamental phenomena that cause profile-to-profile variation is a challenging problem.

There is a growing body of literature on monitoring image (Megahed, et al., 2011) and other profile data, including nonlinear profile monitoring, (e.g., Williams, Birch, Woodall, and Ferry, 2007, Williams, Woodall, and Birch, 2007, Jensen and Birch, 2009, Noorossana, et al., 2011). The nonlinear mixed (NLM) model used in Jensen and Birch (2009) is of the form $x_i(\mathbf{u}) = f(\mathbf{u}, \boldsymbol{\theta}_i) + \text{error}$, where $f(\mathbf{u}, \boldsymbol{\theta})$ is some known function of a set of domain variables \mathbf{u} (e.g., \mathbf{u} is a 2D vector of pixel coordinates for an image profile; a scalar time-index for a tonnage signature profile; a 2D vector of spatial surface coordinate indices for a point cloud of data representing a scanned part surface; a scalar dosage amount in a dose-response profile; etc.) parameterized by a vector of unknown parameters $\boldsymbol{\theta}$, and $\{x_i(\mathbf{u}): \mathbf{u} \in \mathbf{U}_i\}$ are the measurements (e.g., of pixel grayscale level for an image; of stamping tonnage; etc.) constituting the i th profile ($i = 1, 2, \dots, N$) in a sample of N measured profiles, where \mathbf{U}_i denotes the discrete set of known domain values at which the i th profile is measured. The parameters $\boldsymbol{\theta}_i$ are assumed to vary randomly from profile to profile. In the NLM model, "nonlinearity" refers to the situation that $f(\mathbf{u}, \boldsymbol{\theta})$ is a nonlinear function of $\boldsymbol{\theta}$, so that

nonlinear regression methods must be used to estimate each θ_i . Throughout, we assume $U_i = U$ is the same set of domain values for each profile, a situation that has been referred to as a common fixed design (e.g., Zou, et al., 2008, Zhang and Albin, 2009). For situations in which this is not the case, one should interpolate (over the domain variables \mathbf{u}) the data for each profile to transform it into a common fixed design structure. If such interpolation is not meaningful or reliable, and the domain values U_i differ for each profile, then our method is not applicable.

The NLM-based profile monitoring work assumes the parametric model $f(\mathbf{u}, \theta)$ is known in advance, and the systematic profile-to-profile variation is accounted for by variation in θ_i over the sample of profiles. Their objective is to monitor for changes in θ_i over time using some form of control chart, e.g., a multivariate T^2 chart. The objective that we address in this paper is substantially different and much more exploratory in nature, in line with the exploratory aspects of Phase I control charting (Jones-Farmer, et al., 2014). Suppose we have collected a sample of measured profiles from the product or manufacturing process, but we have no prepostulated parametric model to represent the profile-to-profile variation. Using only the sample of profiles, our objective is to identify, visualize, and understand the nature of the profile-to-profile variation that is occurring over the sample. The intent is that this will facilitate the quality improvement process of identifying and eliminating the root causes of the variation and/or suggest an appropriate parametric model to adopt in further studies. In essence, our objective is to discover and visualize the nature of the variation, whereas the objective of the NLM-based profile monitoring work is to monitor for changes in specific sources of variation whose nature is known in advance (via a parametric model $f(\mathbf{u}, \theta)$ that has been identified in advance).

There has also been prior work on nonparametric profile monitoring that, rather than assuming a specific parametric model, fits some nonparametric smoother to each profile (e.g., Walker and Wright, 2002, Zou, Tsung, and Wang, 2008, Qiu, Zou, and Wang, 2010). Like the NLM-based work, the objective of this body of work is largely control chart monitoring for changes in the profiles over time, with little applicability to identifying the nature of the variation. Walker and Wright (2002) is an exception, as their objective is to assess the sources of profile-to-profile

variation, but their approach and scope are substantially different from ours. They used an ANOVA-type variance decomposition to quantify the contributions of a few general characteristics, such as differences in the average profile level (averaged across \mathbf{u} , for each profile), differences between pre-specified groups of profiles, etc. Their approach is not intended to discover the nature of variation sources that have not been previously identified.

Our paradigm is closer to that of Apley and Shi (2001), Apley and Lee (2003), Lee and Apley (2004) and Shan and Apley (2008). They assumed that across a sample of measured parts (or more generally, a sample of multivariate data) there are a few major, systematic sources of part-to-part variation, each of which leaves distinct spatial variation pattern in the data. Their objective, like ours, is to discover the nature of the variation sources and the patterns left by the sources, based only on a sample of measured parts, with no prior knowledge or parametric models of the variation. The primary distinction is that their approach is only applicable to identifying linear variation patterns, whereas ours is more generally applicable to complex nonlinear variation patterns that are common in profile data.

To illustrate this paradigm and what we mean by nonlinear variation patterns, consider the example depicted in Fig. 1, which shows a 1D profile obtained by scanning across a bead on the flat surface of an automotive head gasket. The bead is a slightly raised and rounded ridge that extends around the perimeter of each cylinder hole in the gasket, the purpose of which is to provide a tight seal around each combustion chamber. The profile shown in Fig. 1(a) has been discretized into $n = 50$ points and the measurement data are the surface height [the vertical direction in Fig. 1(a), with 0 taken to be the nominal height of the flat gasket surface] at the 50 discrete locations labeled $j = 1, 2, \dots, 50$. Hence, the measurement data for profile i can be represented as a vector $\mathbf{x}_i = [x_{1,i}, x_{2,i}, \dots, x_{50,i}]^T$. To connect this to the NLM model, \mathbf{x}_i is a stacked version of the data $\{x_i(\mathbf{u}): \mathbf{u} \in \mathbf{U}\}$ for profile i , \mathbf{u} is a scalar domain variable, and \mathbf{U} is the set of 50 spatial locations [corresponding to the horizontal axis in Fig. 1(a)] at which the surface profile height is measured. In our example, we have 400 profiles representing the gasket bead surface over a sample of $N = 400$ different gaskets. Because this work focuses on variation, $\{\mathbf{x}_i; i = 1, 2, \dots, N\}$ can be translated

in whatever manner is convenient, e.g., by subtracting the multivariate sample mean vector or the nominal/target profile.

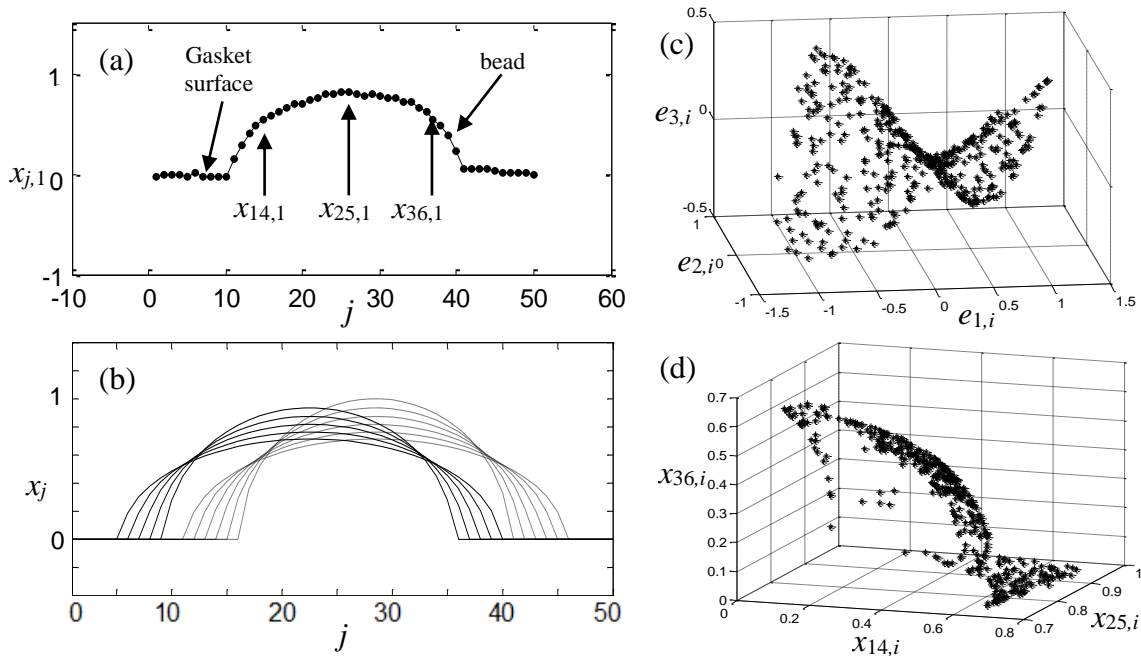


Figure 1. Gasket bead example of 1D profile data containing nonlinear variation patterns: (a) a scanned profile across the gasket bead on a single gasket # 1, discretized into 50 points; (b) illustration of two different variation patterns, a translation and a flattening of the bead; (c) a scatter plot of the first three PCA scores over the 400 profiles; and (d) a scatter plot of the 14th vs 25th vs 36th component of \mathbf{x} over the 400 profiles. Both (c) and (d) show a strong nonlinear pattern in the data.

Figure 1(b) illustrates two profile-to-profile variation patterns by plotting a set of idealized noiseless profiles for ten different gaskets. The first pattern represents the bead flattening (and simultaneously elongating) by differing amounts on different gaskets, which can be visualized as the profile-to-profile differences within the group of five solid profiles and/or within the group of five dashed profiles. The second pattern represents the position of the bead translating left/right by differing amounts on different gaskets, which can be visualized as the differences between the solid and dashed profiles. The actual profile measurement data are a mixture of these two systematic patterns, plus unsystematic random noise. From the sample of 400 such profiles, Figure 1(c) is a scatter plot of the first three principal component scores obtained by applying linear principal component analysis (PCA) to the data. The scatter plot exhibits a distinct nonlinearity

that appears as a saddle shaped surface, which clearly indicates the presence of nonlinear variation patterns in the data. Figure 1(d) is a scatter plot of three elements of the measurement vector \mathbf{x} (the 14th, 25th and 36th elements, the positions of which are indicated by arrows in Fig. 1(a)), which also shows a distinct nonlinearity, although perhaps not quite as clear as in the scatterplot of the PCA scores.

Loosely speaking (see Section 2 for a more formal definition), linear variation patterns are those that cause the measurement vector \mathbf{x} to vary over a linear subspace of n -dimensional space, the dimension of which is the same as the number of distinct variation sources; and nonlinear variation patterns are variation patterns that are not linear, i.e., that cause \mathbf{x} to vary over a nonlinear manifold in n -dimensional space. If only linear variation patterns are present, scatterplots of elements of \mathbf{x} show data that are distributed over a linear subspace (aside from noise). If nonlinear variation patterns are present, the scatterplots show data that fall on more complex nonlinear contours, e.g., as in Figure 1(d). Figure 2, discussed later, illustrates this distinction. It should be noted that whether a variation pattern is nonlinear is not directly related to whether the nominal or average shape of the profile is a nonlinear function of \mathbf{u} . Ding, et al. (2006) and Woodall, et al. (2004) contain examples of profile data for which the average shape is nonlinear but the variation about the average is predominantly linear and well accounted for by linear PCA or linear independent components analysis (ICA). The distinction between linear and nonlinear variation patterns is important, because the nature of linear patterns can be discovered and visualized using linear PCA, ICA, or factor rotation methods (see Apley and Shi, 2001, Apley and Lee, 2003, Lee and Apley, 2004, and Shan and Apley, 2008). For example, the eigenvectors and scores from PCA provide a parameterization of the lower-dimensional subspace in which the major variation patterns lie.

This paper addresses the same objective of discovering and visualizing the nature of profile-to-profile variation, but we extend the concepts to the case of nonlinear variation patterns. We use manifold learning methods to find and parameterize the lower dimensional manifold in which the nonlinear variation patterns lie. We also develop a graphical visualization tool that uses the

manifold parameterization, together with a follow-up supervised learning step in which the manifold parameters are mapped back to the original n -dimensional measurement space, to help visualize the nature of the variation patterns. Apley and Zhang (2007) used principal curves to identify and visualize a nonlinear variation pattern when only a single variation source is present. This work can be viewed as an extension to situations in which multiple sources of variation are present. The remainder of the paper is organized as follows. In Section 2, we describe the model that we use to represent nonlinear profile-to-profile variation and draw a connection to manifold learning. In Section 3, we provide background on manifold learning methods. In Section 4, we present our approach for identifying and visualizing the nonlinear variation patterns and illustrate with two examples. In Section 5, we further discuss the relationship with the NLM model and the dose-response profile example considered in Williams, et al. (2007) and Jensen and Birch (2009). In Section 6, we provide guidelines for selecting the dimension p of the manifold, and contrast our approach with image and profile registration methods. Section 7 concludes the paper.

2. A Model for Nonlinear Variation and Connection to Manifold Learning

Before introducing the nonlinear model, it is helpful to revisit the model for linear variation patterns. Let n denote the number of discrete points measured on each profile, \mathbf{x}_i denote the n -length vector of measurements for profile number i , and p denote the number of major variation sources present over the sample. The linear model assumed in Apley and Shi (2001) and subsequent work on identifying linear variation patterns is $\mathbf{x}_i = \mathbf{C}\mathbf{v}_i + \mathbf{w}_i$, where $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p]$ is an $n \times p$ constant matrix, $\mathbf{v}_i = [v_{1,i}, v_{2,i}, \dots, v_{p,i}]^T$ quantifies the contributions of the p variation sources on profile i , and $\mathbf{w}_i = [w_{1,i}, w_{2,i}, \dots, w_{n,i}]^T$ represents measurement noise and the effects of additional minor, less systematic variation. The p variation sources are assumed to be statistically independent. The column vector \mathbf{c}_j represents the spatial nature of the j th variation source and can be interpreted as the pattern caused by the source. The effects $\{\mathbf{C}\mathbf{v}_i: i = 1, 2, \dots, N\}$ of the variation sources lie in the p -dimensional subspace spanned by the pattern vectors $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p\}$. With noise included, the data $\{\mathbf{x}_i: i = 1, 2, \dots, N\}$ lie close to this p -dimensional subspace, as

illustrated by the scatterplot of $\{\mathbf{x}_i: i = 1, 2, \dots, N\}$ in Figure 2(a) for the case of $p = 2$ variation sources in $n = 3$ dimensional measurement space. In this case, the data are represented as $\mathbf{x}_i = \mathbf{c}_1 v_{1,i} + \mathbf{c}_2 v_{2,i} + \mathbf{w}_i$, where \mathbf{c}_1 and \mathbf{c}_2 are vectors that span the 2-dimensional subspace in Figure 2(a), and $v_{1,i}$ and $v_{2,i}$ can be viewed as the coordinates of the variation pattern component of \mathbf{x}_i in this subspace.

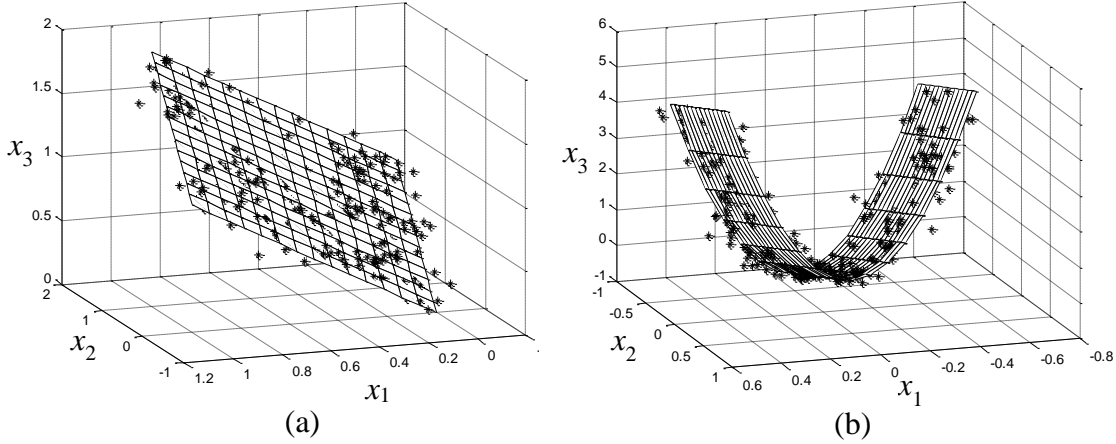


Figure 2 Illustration of the differences between (a) linear variation patterns and (b) nonlinear variation patterns, for 3-dimensional data with two variation sources.

The nonlinear counterpart of the preceding linear model, which we will assume throughout this work, is $\mathbf{x}_i = \mathbf{f}(\mathbf{v}_i) + \mathbf{w}_i$, where $\mathbf{f}(\mathbf{v}_i) = [f_1(\mathbf{v}_i), f_2(\mathbf{v}_i), \dots, f_n(\mathbf{v}_i)]^T$ denotes some nonlinear function that maps the p -dimensional space of variation sources to the n -dimensional space of the n discrete measured points on each profile. The linear model is a special case of the nonlinear model with $\mathbf{f}(\mathbf{v}_i) = \mathbf{C}\mathbf{v}_i$. Whereas the pattern components $\mathbf{C}\mathbf{v}_i$ due to the variation sources in the linear model lie in a linear subspace, the analogous pattern components $\mathbf{f}(\mathbf{v}_i)$ in the nonlinear model lie in a p -dimensional manifold in n -dimensional space, as illustrated in Figure 2(b). Notice that the common fixed design assumption implies that the matrix \mathbf{C} in the linear model is the same for each profile (i.e., \mathbf{C} does not depend on i). Similarly, the function $\mathbf{f}(\cdot)$ in the nonlinear model is the same for each i , although its argument \mathbf{v}_i obviously depends on i .

To further illustrate, we return to the gasket bead example introduced in Section 1 and depicted Figure 1. The $p = 2$ systematic variation patterns can be parameterized by the two-dimensional vector $\mathbf{v}_i = [v_{1,i}, v_{2,i}]^T$, where $v_{1,i}$ represents the amount of bead flattening on gasket number i of

the sample, and $v_{2,i}$ represents the amount of left/right bead translation on gasket number i . From the scatterplots in Figures 1(c) and 1(d), the variation patterns are clearly nonlinear and cannot be represented by the linear model. However, because the variation due to these two sources can be parameterized by two scalars, it must be the case that we can represent the data as $\mathbf{x}_i = \mathbf{f}(\mathbf{v}_i) + \mathbf{w}_i$, for *some* nonlinear function $\mathbf{f}(\bullet)$ that maps two-dimensional space to n -dimensional space. The set $\{\mathbf{f}(\mathbf{v}): \mathbf{v} \in \mathcal{D}\}$, with \mathcal{D} some subset of p -dimensional space, constitutes a p -dimensional manifold in n -dimensional space.

Figure 3 illustrates this by plotting $\mathbf{f}(\bullet)$ over the gasket surface in Fig. 3(a) and also in the n -dimensional measurement space in Fig. 3(b), the latter for only three (14th, 25th, and 36th) of the n elements of \mathbf{x} . More specifically, Fig. 3(a) shows ten different gasket bead profiles, which are plots of $\mathbf{f}(\mathbf{v})$ across the surface of the gasket for ten different values of $\mathbf{v} = [v_1, v_2]^T$. The values of \mathbf{v} are indicated for four of the ten profiles. The ten bullet points plotted in Fig. 3(b) correspond to the same ten profiles in Fig. 3(a). The surface in Fig. 3(b) is the manifold $\{\mathbf{f}(\mathbf{v}): \mathbf{v} \in \mathcal{D}\}$ for some domain \mathcal{D} , and the profiles in Fig. 3(a) are the same as those shown in Fig. 1(b). A larger v_1 value corresponds to a more flattening of the bead, and a larger v_2 value corresponds to a bead that is translated further to the right. The group of dashed profiles are for $v_2 = 0.5$, and the group of solid profiles are for $v_2 = -0.5$. Within each group, the five profiles are for $v_1 = \{-1, -0.5, 0, 0.5, 1\}$.

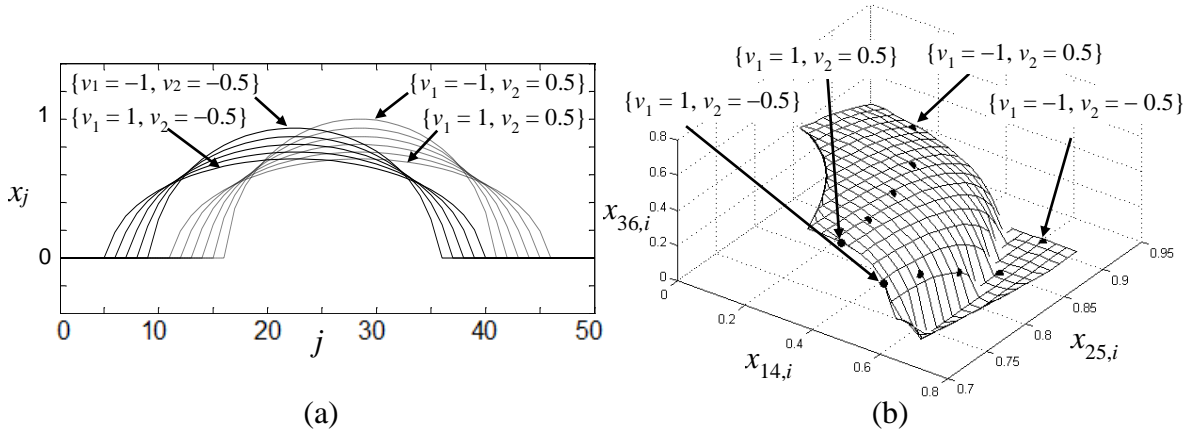


Figure 3. For the gasket bead example, illustration of (a) the physical nature of the two variation patterns via plots of the profile $\mathbf{f}(\mathbf{v})$ across the surface of the gasket in the profile geometry space for ten different values of $\mathbf{v} = [v_1, v_2]^T$; and (b) the corresponding two-dimensional manifold

$\{\mathbf{f}(\mathbf{v}): \mathbf{v} \in \mathcal{D}\}$ plotted in (three components of) the n -dimensional measurement space. The ten bullets in (b) correspond to the same ten profiles in (a).

It is important to note that for problems of this nature, there may be no clean mathematical representation of the function $\mathbf{f}(\bullet)$ or the corresponding manifold. Moreover, attempting to visualize the geometry of the manifold in the n -dimensional measurement space, as in Fig. 3(b), may offer little insight into the nature of the variation patterns. The salient point in Fig. 3(b) is that the effects of the variation sources can be represented as a p -dimensional manifold in n -dimensional space. If we can empirically find a p -dimensional parameterization of the manifold using manifold learning techniques applied to a sample of measured profiles, and then empirically learn the mapping from this p -dimensional space back to the n -dimensional measurement space, we can plot the results in the original profile geometry space to graphically visualize the nature of the variation. In the remainder of the paper, we discuss our approach for accomplishing this. As a matter of terminology, we refer to plots like Fig. 3(a) as being in the *profile geometry space*, whereas plots like Fig. 3(b) are in the n -dimensional measurement space.

3. Background on Manifold Learning

For the linear model $\mathbf{x}_i = \mathbf{C}\mathbf{v}_i + \mathbf{w}_i$, the variation pattern component $\mathbf{C}\mathbf{v}_i$ lies in the p -dimensional manifold $\{\mathbf{C}\mathbf{v}: \mathbf{v} \in \mathcal{D}\}$, which (for $\mathcal{D} = \mathbb{R}^p$) is a p -dimensional linear subspace. Hence, linear PCA can be used to find and parameterize this linear subspace. The span of the p dominant eigenvectors from PCA constitutes an estimate of the subspace, and the corresponding p PCA scores provide a p -dimensional parameterization.

For the nonlinear model $\mathbf{x}_i = \mathbf{f}(\mathbf{v}_i) + \mathbf{w}_i$, identifying the nonlinear manifold $\{\mathbf{f}(\mathbf{v}): \mathbf{v} \in \mathcal{D}\}$ is typically a more challenging problem. In this section, we review a number of existing manifold learning methods that, given a sample of n -dimensional observations $\{\mathbf{x}_i: i = 1, 2, \dots, N\}$, attempt to find an underlying p -dimensional ($p \ll n$ typically) manifold $\{\mathbf{f}(\mathbf{v}): \mathbf{v} \in \mathcal{D}\}$ on which the data approximately lie. Manifold learning can be viewed as a nonlinear form of multidimensional scaling (MDS), as the algorithms attempt to map each \mathbf{x}_i to a corresponding point in some implicitly defined p -dimensional space. The set $\{\mathbf{z}_i \in \mathbb{R}^p: i = 1, 2, \dots, N\}$ of mapped points in

this p -dimensional space represent the coordinates of $\{\mathbf{x}_i: i = 1, 2, \dots, N\}$ on the p -dimensional manifold lying in the original n -dimensional space. In this sense, the p coordinates from manifold learning can be viewed as an implicit p -dimensional parameterization of the manifold. Traditional MDS (Torgerson, 1952) does precisely this, although the structure of the mapping has a certain linearity imposed. For nonlinear manifold learning, a number of algorithms have been developed, including ISOMAP (Tenenbaum, de Silva and Langford, 2000), geodesic nonlinear mapping (GNLM) (Yang, 2004), Sammon mapping (SM) (Sammon, 1969), local linear embedding (LLE) (Roweis and Saul, 2000), Hessian local linear embedding (HLLE) (Donoho and Grimes, 2005), and local tangent space alignment (LTSA) (Zhang and Zha, 2004). A survey of manifold learning algorithms can be found in Huo, et al. (2007). Principal surface algorithms (Hastie, 1984, Delicado, 2001) may be viewed as a form of manifold learning, although formal manifold learning algorithms have emerged as more popular implementations with more software packages available, and the so-called principal surface algorithms seem primarily intended for low-dimensional (small n) applications. In the remainder of this section, we briefly review the ISOMAP algorithm. Diverse applications of the various manifold learning algorithms can be found in the preceding references and in Ekins, et al. (2006), Duraiswami and Raykar (2005), and Patwari and Hero (2004).

ISOMAP is almost a direct nonlinear generalization of MDS. The classical MDS algorithm attempts to ensure that the pairwise Euclidean distances between the p -dimensional representation $\{\mathbf{z}_i: i = 1, 2, \dots, N\}$ are as close as possible to the pairwise Euclidean distances between the original $\{\mathbf{x}_i: i = 1, 2, \dots, N\}$. Specifically, MDS finds the set $\{\mathbf{z}_i: i = 1, 2, \dots, N\}$ to minimize the objective function $\sum_{i < j} (d_{ij} - \|\mathbf{z}_i - \mathbf{z}_j\|)^2$, where $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ is the Euclidean distance between \mathbf{x}_i and \mathbf{x}_j . Preserving Euclidean distances is reasonable when $\{\mathbf{x}_i: i = 1, 2, \dots, N\}$ approximately lie in a p -dimensional linear subspace, as in Fig. 2(a). However, when $\{\mathbf{x}_i: i = 1, 2, \dots, N\}$ approximately lie on a p -dimensional nonlinear manifold, as in Fig. 2(b), it makes more sense to preserve geodesic distances. As illustrated in Fig. (4), the geodesic distance between two points is the shortest path connecting them along the manifold, which may be much different than their Euclidean distance if the manifold is highly nonlinear.

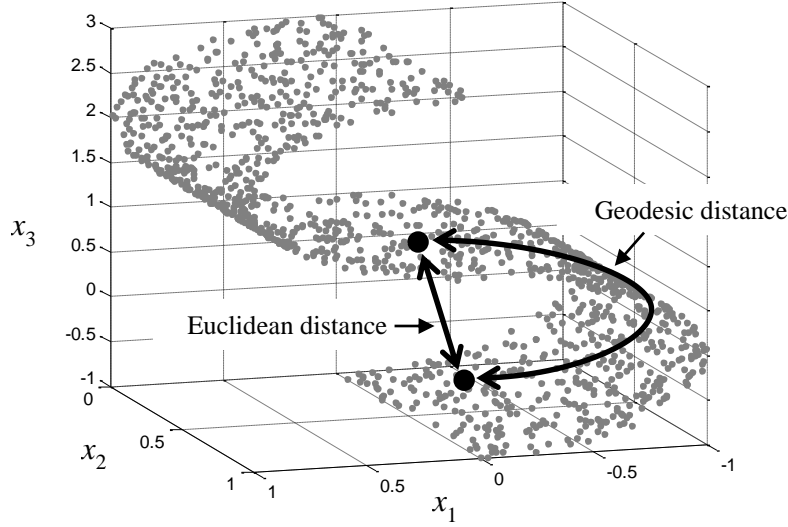


Figure 4. Illustration of geodesic distance along a manifold and the difference between geodesic and Euclidean distances.

Noting this, Tenenbaum, et al. (2000) proposed their ISOMAP algorithm to find the p -dimensional representation $\{\mathbf{z}_i: i = 1, 2, \dots, N\}$ to minimize the objective function $\sum_{i < j} (d_{ij}^g - \|\mathbf{z}_i - \mathbf{z}_j\|)^2$, which is the same as MDS but with Euclidean distance d_{ij} replaced by the geodesic distance d_{ij}^g between \mathbf{x}_i and \mathbf{x}_j . To approximate the geodesic distance between data points, ISOMAP first builds a graph G with N vertices corresponding to $\{\mathbf{x}_i: i = 1, 2, \dots, N\}$ and edges that link each \mathbf{x}_i with its K nearest neighbors for some chosen K . The edges are weighted by the Euclidean distances d_{ij} . The geodesic distance between \mathbf{x}_i and \mathbf{x}_j is then taken to be the length of the shortest path between \mathbf{x}_i and \mathbf{x}_j on graph G .

We note that del Castillo, et al. (2014) used the ISOMAP algorithm to analyze 3D point cloud data, although they used it in a completely different manner than we do. Their original data are 3D surface measurement data for a single measured part, i.e., $\{\mathbf{x}_i \in R^3: i = 1, 2, \dots, N\}$ are the 3D coordinates for N points measured on the surface of a single part. They use manifold learning to map the point cloud data for a *single* part from 3D space to a 2D space of manifold coordinates that represent the surface coordinates of each measured point on that part. In contrast, we use manifold learning to analyze part-to-part (or profile-to-profile) variation across a sample of N parts, and each $\mathbf{x}_i \in R^n$ represents the collection of all n measurements on part number i . We use

manifold learning to map the high n -dimensional space to a much lower p -dimensional space of manifold coordinates that represent the sources of variation.

4. Using manifold learning to identify and visualize profile-to-profile variation

For the profile data $\{\mathbf{x}_i \in R^n: i = 1, 2, \dots, N\}$ that behaves according to the model $\mathbf{x}_i = \mathbf{f}(\mathbf{v}_i) + \mathbf{w}_i$ with p distinct variation sources present, the coordinate representation $\{\mathbf{z}_i \in R^p: i = 1, 2, \dots, N\}$ from manifold learning serves as an estimate of (some transformed version of) the sources $\{\mathbf{v}_i \in R^p: i = 1, 2, \dots, N\}$. In order to gain insight into the nature of the variation, however, it is helpful to visualize the manifold $\mathbf{f}(\mathbf{v})$ as the variation sources \mathbf{v} are varied. In light of the relationship $\mathbf{x}_i = \mathbf{f}(\mathbf{v}_i) + \mathbf{w}_i$, we accomplish the latter by fitting an appropriate supervised learning model with $\{\mathbf{x}_i: i = 1, 2, \dots, N\}$ as the multivariate output (aka response) variables and the estimated $\{\mathbf{v}_i: i = 1, 2, \dots, N\}$ from the manifold learning algorithm as the input (aka predictor) variables. The fitted supervised learning model then serves as an estimate of $\mathbf{f}(\mathbf{v})$, which can be interactively visualized in the profile geometry space using a graphical user interface. Details of our procedure are described in the following.

Step 1. Apply linear PCA to $\{\mathbf{x}_i: i = 1, 2, 3, \dots, N\}$ to reduce dimensionality and noise.

As discussed in Apley and Zhang (2007), the p -dimensional manifold $\mathbf{f}(\mathbf{v})$ will typically lie, at least approximately, in some P -dimensional linear subspace of the n -dimensional measurement space with $p < P < n$. Thus, applying linear PCA to $\mathbf{x}_i = \mathbf{f}(\mathbf{v}_i) + \mathbf{w}_i$ will tend to filter out the noise component \mathbf{w}_i , while preserving much of the signal component $\mathbf{f}(\mathbf{v}_i)$. The idea is to apply the manifold learning algorithm to the PCA score vectors $\{\mathbf{e}_i \in R^P: i = 1, 2, 3, \dots, N\}$ associated with the largest P eigenvalues. Although manifold learning could be applied to the original n -dimensional data, there are two benefits of using linear PCA as a preprocessing step. One is computational, and the other is performance-related. Regarding the latter, all of the manifold learning algorithms rely heavily on the local structure in the data and must identify the sets of neighbors that lie on the same local portion of the manifold. Because this is more difficult to accomplish reliably in higher dimensions and/or with higher noise levels, the performance of

manifold learning methods can be improved using PCA to reduce the dimensionality and/or the noise. Some manifold learning methods (e.g., HLLE, Donoho and Grimes, 2005) are more sensitive to high dimensionality/noise than others. We have observed this in our simulations, by comparing plots of the estimated $\hat{\mathbf{f}}(\mathbf{v})$ (as in Figures 9 and 11) with and without linear PCA. PCA also reduces the computational expense of the neural network fitting (see Step 3 below) via the dimensionality reduction, although it has little impact on the computational expense of the manifold learning in Step 2.

Step 2. Apply a manifold learning algorithm to the P -dimensional PCA scores $\{\mathbf{e}_i: i = 1, 2, 3, \dots, N\}$.

This step is a direct application of an existing manifold learning algorithm. The p -dimensional coordinate representation produced by the manifold learning algorithm is taken to be the estimates $\{\hat{\mathbf{v}}_i = [\hat{v}_{1,i}, \hat{v}_{2,i}, \dots, \hat{v}_{p,i}]^T: i = 1, 2, 3, \dots, N\}$ of the variation sources. To be more precise, these should be viewed only as estimates of some transformed version of the variation sources. Any of the manifold learning algorithms can be used for this step. In the examples of this paper, we used the ISOMAP algorithm.

Step 3. Fit a single layer, multi-output neural network model to approximate the manifold $\mathbf{f}(\cdot)$.

One could consider any appropriate supervised learning method for this step. We focus on a neural network model in our examples, because of its ability to handle multiple output variables that share similar dependence on the input variables. In our case we have an n -variate output \mathbf{x}_i and a p -variate input $\hat{\mathbf{v}}_i$ for $i = 1, 2, \dots, N$. To estimate the manifold $\mathbf{f}(\cdot)$ when n is small, we could fit a single neural network model with n nodes in the output layer, p nodes in the input layer, and a single hidden layer. Denoting the fitted neural network model as $\hat{\mathbf{f}}(\mathbf{v}) = [\hat{f}_1(\mathbf{v}), \hat{f}_2(\mathbf{v}), \dots, \hat{f}_n(\mathbf{v})]^T$, $\hat{\mathbf{f}}(\cdot)$ represents the estimated manifold, and $\hat{f}_j(\mathbf{v})$ represents the functional dependence of the j th element of the measurement vector on the variation sources \mathbf{v} . A single neural network with n nodes in the output layer can efficiently represent the situation that some elements of $\hat{\mathbf{f}}(\mathbf{v})$ share similar functional dependencies on \mathbf{v} (which is clearly the case for

the neighboring elements of the gasket profile patterns illustrated in Figure 1), but it also allows other elements of $\hat{\mathbf{f}}(\mathbf{v})$ to have very different dependencies on \mathbf{v} .

However, when the data are high dimensional (i.e., large n), fitting a neural network with n nodes in the output layer is computationally intensive and can also be unstable. Consequently, instead of using the n -dimensional vectors $\{\mathbf{x}_i: i = 1, 2, 3, \dots, N\}$ as the output of the neural network, we use the r -dimensional vectors $\{\mathbf{e}_i: i = 1, 2, 3, \dots, N\}$ of the first r PCA scores (r need not be the same as P). If we denote the fitted neural network for predicting \mathbf{e} (as a function of \mathbf{v}) by $\hat{\mathbf{g}}(\mathbf{v})$, an estimate of $\mathbf{f}(\cdot)$ can be constructed via $\hat{\mathbf{f}}(\mathbf{v}) = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_r]\hat{\mathbf{g}}(\mathbf{v})$, where $\{\mathbf{q}_j \in \mathbb{R}^n: j = 1, \dots, r\}$ denote the r eigenvectors associated with the r PCA scores that comprise \mathbf{e} . To fit the neural network model, we used the R package `nnet` with cross-validation to choose the number of nodes in the hidden layer and the shrinkage parameter.

Step 4. Use a graphical user interface (GUI) to visualize the estimated $\hat{\mathbf{f}}(\mathbf{v})$ in the profile geometry space.

In our examples, we use a Matlab GUI to visualize the estimated manifold, which ultimately helps visualize the nature of the variation patterns and identify their root causes. We use p slide bar controls to represent the p elements of \mathbf{v} , which represent the p variation sources. As an user varies the elements of \mathbf{v} using the slide bars, the n elements of $\hat{\mathbf{f}}(\mathbf{v})$ are plotted dynamically in the profile geometry space (see Figure 10, discussed shortly, for an illustration of the GUI). Interactively dragging a slide bar back-and-forth serves to animate and visualize the nature of the profile-to-profile variation caused by the corresponding variation source. The intent is to provide the user with insight on the nature of the variation, which, when coupled with engineering knowledge of the manufacturing process, helps to identify the root causes of the variation.

Before revisiting the gasket bead example, we next illustrate the preceding steps with an example involving image profile data.

Example with Image Data. Image data are increasingly common in manufacturing quality control (Megahed, et al., 2011). In this example, we analyze a sample of 400 penny images. To simulate variation patterns we translated and rotated the pennies in each image, and we also added

random noise to each image. Figure 5 shows a random sample of 25 of the 400 penny images. The size of each image is $32 \times 32 = 1024$ pixels, so that the measurement vector \mathbf{x}_i for penny image i is a 1024-length vector, each element of which represents the grayscale level (scaled to the interval $[0,1]$) of the corresponding pixel. The two simulated variation patterns were a translation of the penny over an arc-shaped trajectory and a rotation of the penny. It is somewhat difficult to discern the nature of the two patterns from the raw image data in Figure 5. As we will demonstrate shortly, the patterns become much clearer using the manifold learning and visualization approach.

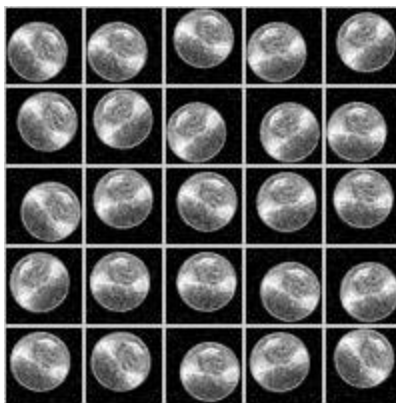


Figure 5. A subsample of 25 noisy penny images with two underlying variation patterns.

We begin (Step 1) by applying linear PCA to the sample $\{\mathbf{x}_i: i = 1, 2, 3, \dots, 400\}$ of the 1024-dimensional data. Figure 6 shows a scatterplot of the first 3 PCA scores, which demonstrates a pronounced nonlinearity. Figure 7 is a scree plot of the first 20 eigenvalues, which account for 90.02% of the total variance in 400 images. We retained $P = 20$ principal components and worked with the sample $\{\mathbf{e}_i: i = 1, 2, 3, \dots, 400\}$ of 20-dimensional PCA score vectors. We then applied (Step 2) the ISOMAP manifold learning algorithm to $\{\mathbf{e}_i: i = 1, 2, 3, \dots, 400\}$ with $p = 2$, for which the estimated manifold coordinates $\{\hat{\mathbf{v}}_i = [\hat{v}_{1,i}, \hat{v}_{2,i}]^T: i = 1, 2, 3, \dots, 400\}$ are plotted in Figure 8.

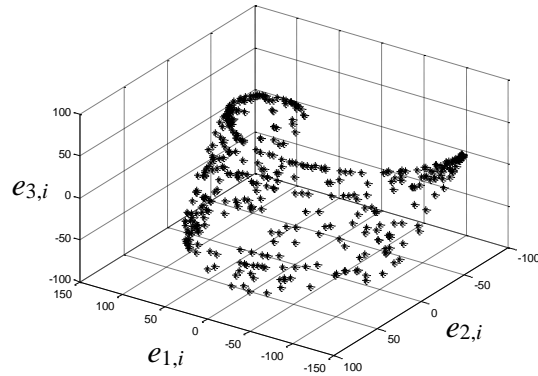


Figure 6. 3D scatterplot of the first 3 linear PCA scores for the penny image data

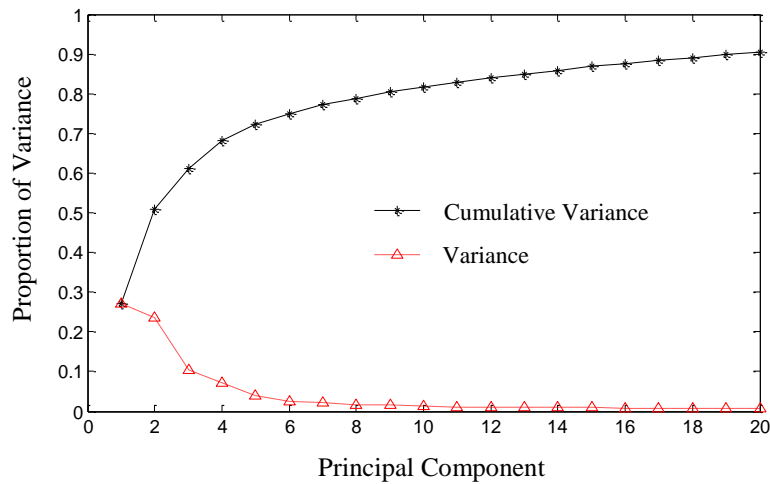


Figure 7. Scree plot showing the proportion and cumulative proportion of variance for the first 20 linear principal components for the penny image data.

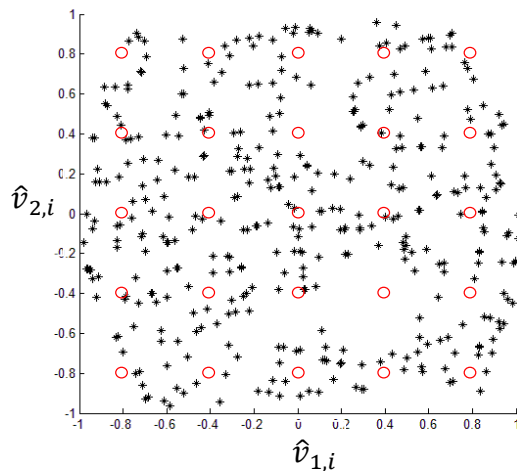


Figure 8. Estimated coordinates for the 2-dimensional representation of the manifold for the penny image data using the ISOMAP algorithm. The 25 open red circles correspond to the 25 penny images plotted in Figure 9.

The manifold coordinates in Figure 8 provide little direct insight into the nature of the variation patterns. To understand the nature of the variation patterns, we next fit (Step 3) a multi-response neural network model using the two-dimensional data $\{\hat{\mathbf{v}}_i = [\hat{v}_{1,i}, \hat{v}_{2,i}]^T : i = 1, 2, \dots, 400\}$ as the input and the PCA score vectors $\{\mathbf{e}_i \in R^{100} : i = 1, 2, 3, \dots, 400\}$ as the output. We chose $r = 100$, because it accounted for almost all (99.7%) of the variation in the data. Note that the data to which the neural network model was fit were comprised of $N = 400$ rows/observations, $p = 2$ input variable columns, and $r = 100$ output variable columns. Denoting the fitted neural network by $\hat{\mathbf{g}}(\mathbf{v})$, we took the estimated manifold to be $\hat{\mathbf{f}}(\mathbf{v}) = [\hat{f}_1(\mathbf{v}), \hat{f}_2(\mathbf{v}), \dots, \hat{f}_{1024}(\mathbf{v})]^T = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{100}] \hat{\mathbf{g}}(\mathbf{v})$. As a static visualization of the nature of the two variation patterns represented by the two-dimensional manifold, Figure 9 plots 25 different penny images. Each image in Figure 9 is a plot of $\hat{\mathbf{f}}(\mathbf{v})$ in the profile geometry space for \mathbf{v} corresponding to one of the 25 open red circles in Figure 8. Scanning across each row of Figure 9 indicates that the variation pattern corresponding to v_1 is an arc-shaped spatial translation of the pennies. Similarly, scanning down each column of Figure 9 indicates that the variation pattern corresponding to v_2 is a rotation of the pennies. These agree quite closely with the two variation patterns that were introduced to the image data. Compared to the random sample of raw penny images in Figure 5, the nature of the patterns is much easier to visualize in Figure 9.

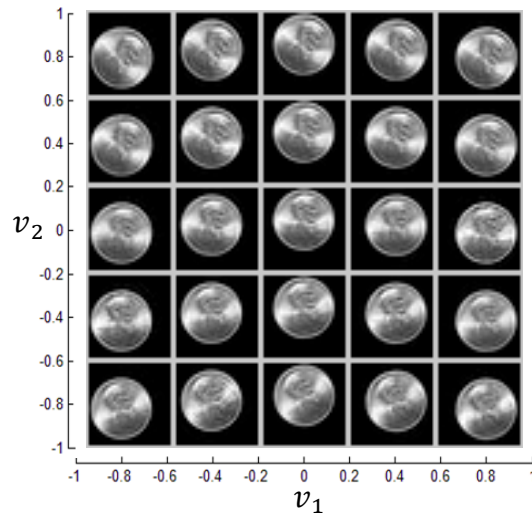


Figure 9. Static visualization of the estimated variation patterns represented by the two-dimensional manifold. The 25 images are plots of $\hat{\mathbf{f}}(\mathbf{v})$ for values of \mathbf{v} coinciding with the 25

open red circles in Figure 8. Scanning across each row, the variation pattern corresponding to v_1 is an arc-shaped spatial translation of the pennies; scanning down each column, the variation pattern corresponding to v_2 is a rotation of the pennies.

While the visualization in Figure 9 is static (suitable for published paper format), a dynamic visualization is typically more effective at conveying the nature of the profile-to-profile variation patterns. Figure 10(a) depicts a Matlab GUI that we have used for interactive dynamic visualization of the profile-to-profile variation patterns. The GUI dynamically plots the estimated $\hat{\mathbf{f}}(\mathbf{v})$ as the user varies the elements of \mathbf{v} using slide bar controls. Because $p = 2$ for the penny image data, there are two slide bars in Figure 10(a) representing v_1 and v_2 . As the user moves the v_1 slide bar back-and-forth, the arc-shaped translation pattern is animated as shown in the top panel of Figure 10(b). Likewise, as the user moves the v_2 slide bar back-and-forth, the rotational pattern is animated as shown in the bottom panel of Figure 10(b).

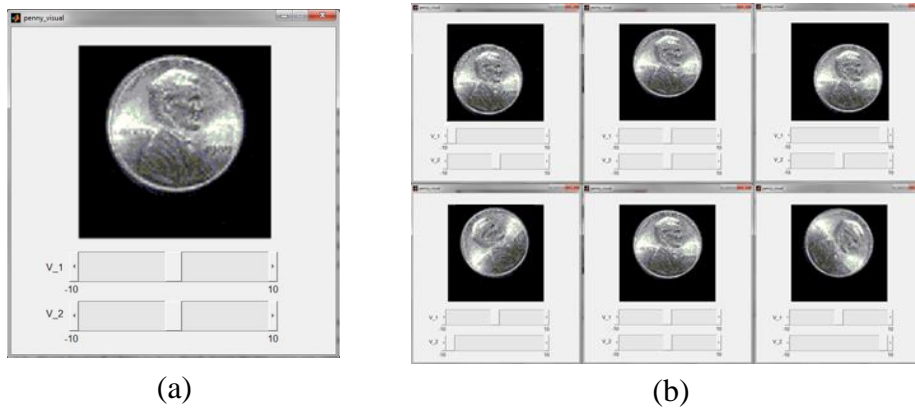


Figure 10. (a) Interactive graphical visualization tool interface for conveying the nature of the variation patterns estimated from manifold learning and (b) six different snapshots of the visualization corresponding to six different combinations of v_1 and v_2 specified by the slide bars. Compare to the static visualization in Figure 9.

It should be noted that if one knew in advance that the image-to-image variation consisted exclusively of translations and rotations, then a parametric modeling approach in which one used image processing techniques specifically to estimate the amount of rotation and translation of each penny would obviously estimate the patterns more effectively than manifold learning, which does not take into account any advance knowledge of the nature of the variation. The strength of the manifold learning approach is its generality and applicability to situations in which one has no

advance knowledge or parametric models of the variation and, instead, seeks to discover the nature of the variation based only on a sample of profiles. We discuss this further in Section 6.2.

Gasket Bead Example. The results of applying the manifold learning approach to the gasket bead data are shown in Figure 11. Figure 11(a) is a plot of the estimate manifold coordinates $\{\hat{\mathbf{v}}_i = [\hat{v}_{1,i}, \hat{v}_{2,i}]^T : i = 1, 2, 3, \dots, 400\}$ for a sample of 400 gasket profiles. The ISOMAP algorithm was applied to the linear PCA scores with the top $P = 3$ scores retained. Figure 11(b) is a plot of the estimated manifold $\hat{\mathbf{f}}(\mathbf{v}) = [\hat{f}_1(\mathbf{v}), \hat{f}_2(\mathbf{v}), \dots, \hat{f}_{50}(\mathbf{v})]^T$ in the profile geometry space (analogous to Figure 9 for penny images) for the 10 values of $\mathbf{v} = [v_1, v_2]^T$ indicated by the open red circles in Figure 11(a). The nature of the estimated patterns from Figure 11(b) is quite close to that of the patterns shown in Figure 3(a), with v_1 representing the bead flattening pattern and v_2 representing the left-right bead translation pattern. Similar to what is illustrated in Figure 10, a GUI with slide bar controls for v_1 and v_2 could be used to animate the flattening and translation patterns.

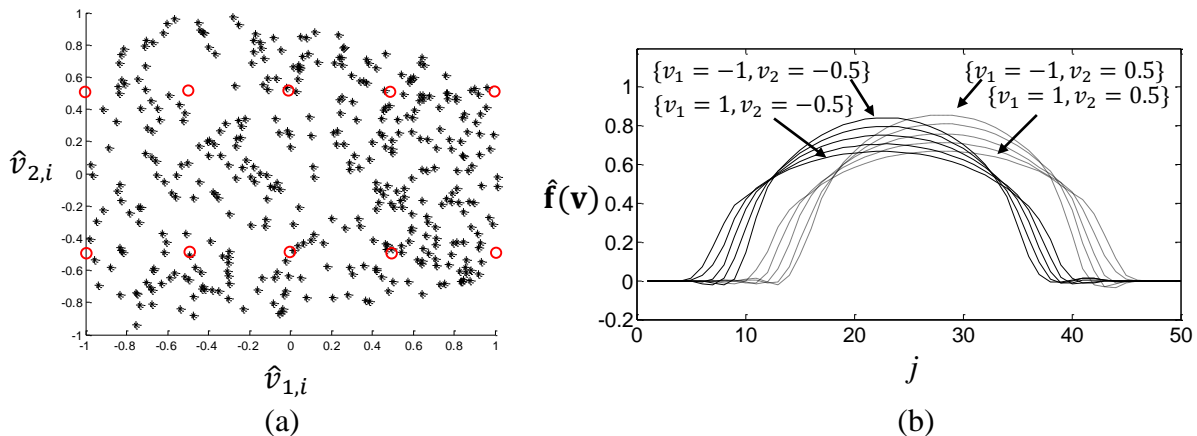


Figure 11. For the gasket bead example, (a) scatterplot of the estimated two-dimensional manifold coordinates $\hat{\mathbf{v}}_i = [\hat{v}_{1,i}, \hat{v}_{2,i}]^T$; and (b) plot of the estimated manifold $\hat{\mathbf{f}}(\mathbf{v})$ for ten different values of $\mathbf{v} = [v_1, v_2]^T$ corresponding to the open red circles in (a).

5. Dose-response profile example

As an example of the NLM model $x_{ji} = x_i(\mathbf{u}_{ji}) = f(\mathbf{u}_{ji}, \boldsymbol{\theta}_i) + w_{ji}$ (i is the profile index; \mathbf{u}_{ji} is the j th domain location at which profile i is measured; x_{ji} and w_{ji} are the profile measurement and random error at location j on profile i , respectively), Williams, Birch, Woodall, and Ferry (2007),

Williams, Woodall, and Birch (2007), and Jensen and Birch (2009) considered the drug dose-response model

$$x_{ji} = \theta_{1i} + \frac{\theta_{2i} - \theta_{1i}}{1 + (u_{ji}/\theta_{3i})^{\theta_{4i}}} + w_{ji}, \quad (1)$$

where u_{ji} a scalar drug dosage, x_{ji} is the response of a subject to dosage x_{ji} for batch i of the drug, and $\boldsymbol{\theta}_i = [\theta_{1i}, \theta_{2i}, \theta_{3i}, \theta_{4i}]^T$ are unknown parameters that characterize the dose-response profile for batch i . The parameters $\boldsymbol{\theta}_i$ vary randomly from batch-to-batch. If we restrict $u_{ji} = u_j$ (all profiles are measured at the same domain locations) their stacked version of the NLM model can be written as $\mathbf{x}_i = \mathbf{f}(\boldsymbol{\theta}_i) + \mathbf{w}_i$, which is the same as our profile variation model with the parameters $\boldsymbol{\theta}_i$ acting as the parameterized variation sources \mathbf{v}_i .

The aforementioned papers assumed the parametric model (1) is known in advance, which is equivalent to assuming the nature of the variation is known in advance, and their objective was to monitor for changes in the profile parameters $\boldsymbol{\theta}_i$. In contrast, our manifold learning and visualization approach could be used to discover an appropriate parameterization that captures the nature of the major sources of profile-to-profile variation (i.e., profile-to-profile variation in the parameters $\boldsymbol{\theta}_i$), which we illustrate in the online supplement to this article on the *Technometrics* website.

6. Discussion

6.1 A method for choosing p

In Step 2, we use a manifold learning algorithm to estimate the p -dimensional representation of the original n -dimensional data. The parameter p is the intrinsic dimension of the manifold (in our context, the number of distinct variation sources), and it must be chosen by the user. In this section, we present a simple graphical approach for choosing an appropriate value for p .

We can view the output $\hat{\mathbf{f}}(\hat{\mathbf{v}}_i)$ of our algorithm as a lower-dimensional approximation of \mathbf{x}_i . Specifically, $\hat{\mathbf{f}}(\hat{\mathbf{v}}_i)$ represents a p -dimensional approximation (using the nonlinear manifold coordinates $\hat{\mathbf{v}}_i$) of the n -dimensional vector \mathbf{x}_i . Consequently, $\|\mathbf{x}_i - \hat{\mathbf{f}}(\hat{\mathbf{v}}_i)\|$ represents the error in the approximation, which should theoretically decrease to zero as p increases to n . If a small value

is selected for p , below the true dimension of the manifold, the error should decrease markedly as p is increased. On the other hand, if a large value is selected for p , above the true dimension of the manifold, then further increasing p should decrease the error by only a small amount, consistent with including some of the noise component in manifold approximation. In light of this, our procedure for selecting p is to plot the relative error

$$RE = \frac{\sum_{i=1}^N \|\mathbf{x}_i - \hat{\mathbf{f}}(\hat{\mathbf{v}}_i)\|^2}{\sum_{i=1}^N \|\mathbf{x}_i - \bar{\mathbf{x}}\|^2} \quad (2)$$

versus p and look for the elbow in the curve. In (2), $\bar{\mathbf{x}} = \sum_{i=1}^N \mathbf{x}_i$ is the sample average, and RE represents the proportion of total variance in $\{\mathbf{x}_i: i = 1, 2, 3, \dots, N\}$ that is *not* accounted for by its p -dimensional manifold representation. This approach is analogous to looking for the elbow in a cumulative scree plot to choose the appropriate number of eigenvalues in linear PCA, because $\hat{\mathbf{f}}(\hat{\mathbf{v}}_i)$ is analogous to the projection of \mathbf{x}_i onto the space spanned by the first p eigenvectors in PCA.

Figure 12 illustrates the approach for the gasket bead example, for which there were truly $p = 2$ variation sources present. There is a distinct elbow at $p = 2$ in Figure 12, which indicates that the two-dimensional estimated manifold accounted for most of the variation in the original data.

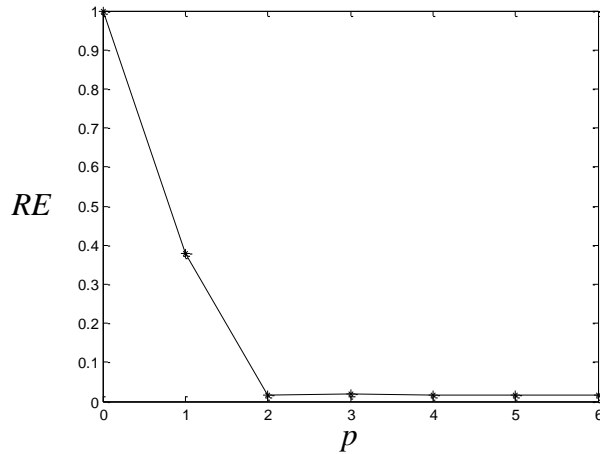


Figure 12. For the gasket bead example, illustration of how a plot of RE versus p is used to select an appropriate value for the dimension p of the manifold.

6.2 Feature discovery versus known feature extraction and profile registration

There is a large body of work on parametric registration and feature extraction for profile and image data (e.g., Dryden and Mardia, 1998, Ramsay and Silverman, 2005). Such work assumes

essentially the same NLM-type model $x_i(\mathbf{u}) = f(\mathbf{u}, \boldsymbol{\theta}_i) + \text{error}$ discussed in the introduction. Like the NLM work reviewed in the introduction, it also assumes the function $f(\bullet, \bullet)$ is of known, predetermined form, and the parameters $\boldsymbol{\theta}_i$ (which vary from profile to profile) have predetermined meaning. Given the data for profile number i , the profile registration objective is to estimate/extract the value of the parameters $\boldsymbol{\theta}_i$ so that the functional representation $f(\mathbf{u}, \boldsymbol{\theta}_i)$ best matches the profile data. This is fundamentally different than the manifold learning method of this paper, for the same reason that the NLM work reviewed in the introduction is fundamentally different: Our objective is to discover the nature of the function $f(\bullet, \bullet)$ and a corresponding parameterization $\boldsymbol{\theta}$ that represents the profile-to-profile variation, with no prior knowledge of what this function or parameterization might be.

The difference boils down to whether one knows in advance what one is looking for, regarding the nature of the profile-to-profile variation. In our penny image example, if one knew in advance that image-to-image variation was comprised entirely of translations and rotations of the pennies, then image registration methods could have been used to identify the amount of translation and rotation in each image. In this case, the function $f(\bullet, \bullet)$ would be predetermined based on rigid body kinematics and computer vision modeling, $\boldsymbol{\theta}$ would be a 3-dimensional vector representing the rigid body motion of the penny in a plane, and the profile analysis objective would have been to find the value of $\boldsymbol{\theta}$ for which the function $f(\bullet, \bullet)$ best matches the observed data for each profile.

In order to use such parametric image registration methods, clearly one must know the nature of the variation and its mathematical model $f(\bullet, \bullet)$ in advance. Without using any prior knowledge of this function, the manifold learning approach was able to discover that the image-to-image variation was a translation and a rotation, as illustrated clearly in Figures 9 and 10. Taking into account the knowledge of the nature of the variation, the parametric image registration approach obviously would have estimated the translations and rotations more accurately and efficiently. However, the results in Figures 9 and 10 still demonstrate the power of the manifold learning approach to discover the nature of previously unknown and unmodeled variation.

It should be noted that there is another body of work on *nonparametric* profile/image registration (Qiu and Xing 2013a, Xing and Qiu, 2011), but this also is fundamentally different than our manifold learning approach. Parametric registration methods use a low-dimensional parametric transformation (via θ) of known, predetermined form to best match the profile data to its mathematical model or some template profile. In contrast, nonparametric registration methods use some smoothly varying local transformation (e.g., a locally linear stretching with the amount of stretching varying smoothly over the profile domain variable \mathbf{u}) to transform or warp each profile to best match some template or model. Because the transformations are nonparametric and do not represent each profile registration via a small set of parameters θ , it would be difficult to use them as a basis for discovering and visualizing the nature of the underlying variation via developing low-dimensional visualization controls like the ones that we use with our manifold learning method (e.g., the two slide bar controls in Figure 10 associated with the two learned manifold coordinates v_1 and v_2).

However, using the manifold learning approach in conjunction with a profile registration method may be an attractive strategy for situations in which known variation sources are present and their effects are easily modeled (e.g., the penny image translation and rotation or the translational pattern for the gasket bead in Figure 1(b)), but additional unknown variation sources are also present. The profile registration method could be used as a preprocessing step to first estimate the parameters that represent the variation sources of known, predetermined form. Then, the manifold learning approach could be used on the residual errors from the registration step, in order to discover the nature of any additional variation sources beyond the known ones. Apley and Lee (2010) developed an analogous method for simultaneously identifying premodeled and additional unmodeled variation sources, but for the situation in which all patterns are linear. For the nonlinear situation, when profiles have complex shape and/or are very high-dimensional, extracting features from the profile to be used in the registration process (e.g., as in Qiu and Xing, 2013b) may also be useful.

7. Conclusions

This paper has presented a manifold learning based approach to identify and visualize the nature of profile-to-profile variation in a sample of profile data. Relative to plotting the raw profile data $\{\mathbf{x}_i: i = 1, 2, \dots, N\}$, plotting the estimated manifold $\hat{\mathbf{f}}(\mathbf{v})$ as \mathbf{v} varies generally provides much clearer visualization of the nature of the variation patterns. For example, compare the raw penny images plotted in Figure 5 versus the estimate manifold for the pennies plotted in Figure 9. In general, the manifold learning approach facilitates visualization of the variation for at least three reasons. First, manifold learning inherently reduces the level of noise in the estimated $\hat{\mathbf{f}}(\mathbf{v})$, relative to noise present in the original data, leaving the systematic variation patterns more visible. This is analogous to how linear PCA reduces noise levels. Second, the p coordinates of the p -dimensional manifold are kept distinct, so that the individual patterns can be visualized as each element of \mathbf{v} is varied individually. In contrast, when plotting the raw profiles $\{\mathbf{x}_i: i = 1, 2, \dots, N\}$, comparing any sequence of profiles will reflect the combined effects of all p variation sources mixed together. Third, and perhaps most importantly, the manifold learning visualization orders the profiles in terms of increasing values of each element of \mathbf{v} . For example, v_1 smoothly increases as one scans left-to-right in Figure 9, and v_2 smoothly increases as one scans bottom-to-top. This smooth ordering of the profiles according to the effects of each variation source allows the nature of the patterns to be seen much more clearly than simply plotting a sequence of raw profiles, for which the effects of the variation sources are in random, or partially random, order.

Finally, our approach has a distinctly different objective than most profile monitoring work. Namely, our objective is exploratory (Phase I) discovery of the nature of the variation across a sample of profiles, as opposed to monitoring (Phase II) to detect changes in the profiles. However, the manifold learning approach could also be used to discover an appropriate parametric model structure, for example, to serve as the basis for the Phase II NLM monitoring. This may be particularly beneficial for complex, high-dimensional profile data, for which it is difficult to postulate appropriate parametric mathematical models and for which extreme dimensionality reduction is helpful.

Acknowledgments

The authors gratefully acknowledge NSF Grant #CMMI-1265709 for support of this work.

Supplementary Materials

The supplementary material for this article on the *Technometrics* website consists of the results of applying the manifold learning approach to the dose-response profile example introduced in Section 5.

References

- Apley, D. W., and Lee, H. Y. (2003), "Identifying Spatial Variation Patterns in Multivariate Manufacturing Processes: A Blind Separation Approach," *Technometrics*, 45, 220-234.
- (2010), "Simultaneous Identification of Premodeled and Unmodeled Variation Patterns," *Journal of Quality Technology*, 42, 36—51.
- Apley, D. W., and Shi, J. (2001), "A Factor-Analysis Methods for Diagnosing Variability in Multivariate Manufacturing Processes," *Technometrics*, 43, 84-95.
- Apley, D. W., and Zhang, F. (2007), "Identifying and Visualizing Nonlinear Variation Patterns in Multivariate Manufacturing Data," *IIE Transactions*, 39, 691-701.
- del Castillo, E., Colosimo, B. M., Tajbakhsh, S. (2014), "Geodesic Gaussian Processes for the Parametric Reconstruction of a Free-Form Surface," *Technometrics* [online], Available at <http://amstat.tandfonline.com/doi/abs/10.1080/00401706.2013.879075#.VFTVY762tg0>
- Delicado, P. (2001), "Another Look at Principal Curves and Surfaces," *Journal of Multivariate Analysis*, 77(1), 84-116.
- Ding, Y., Zeng, L., and Zhou, S. (2006), "Phase I Analysis for Monitoring Nonlinear Profiles in Manufacturing Processes," *Journal of Quality Technology*, 38, 199-216.
- Donoho, D. L., and Grimes, C. (2005), "Hessian Eigenmaps: New Locally Linear Embedding Techniques for High-dimensional Data," *Proceedings of the National Academy of Sciences*, 102, 7426-7431.
- Dryden, I. L., and Mardia, K.V. (1998), *Statistical Shape Analysis*, New York: Wiley.
- Duraiswami, R., and Raykar, V. C. (2005), "The Manifolds of Spatial Hearing," *2005 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3, 285-288.
- Ekins, S., Balakin, K. V., Savchuk, N., and Ivanenkov, Y. (2006), "Insights for Human Ether-a-Go-Go-Related Gene Potassium Channel Inhibition Using Recursive Partitioning and Kohonen and Sammon Mapping Techniques," *Journal of Medicinal Chemistry*, Vol. 49, No. 17, 5059-5071.

- Hastie, T. (1984), “*Principal Curves and Surfaces*,” Ph.D, Dissertation, Stanford University, Stanford, CA.
- Huo, X., Ni, S., and Smith, A. K. (2007), “A Survey of Manifold-Based Learning Methods,” *Recent Advances in Data Mining of Enterprise Data*, eds. T.W. Liao, and E. Triantaphyllou , Singapore: World Scientific, 691-745.
- Jensen, W. A. and Birch, J. B., (2009), “Profile Monitoring via Nonlinear Mixed Models,” *Journal of Quality Technology*, 41, 18-34.
- Jones-Farmer, L. A., Woodall W. H., Steiner, S. H., and Champ, C. W. (2014), “An Overview of Phase I Analysis for Process Improvement and Monitoring,” *Journal of Quality Technology* 46(3), 265-280.
- Jin, J., and Shi, J. (1999), “Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelets,” *Technometrics*, 41, 327-339.
- Lee, H. Y., and Apley, D. W. (2004), “Diagnosing Manufacturing Variation Using Second-Order and Fourth-Order Statistics,” *International Journal of Flexible Manufacturing Systems*, 16, 45-64.
- Megahed, F. M., Woodall, W. H., and Camelio, J. A. (2011), “A Review and Perspective on Control Charting with Image Data,” *Journal of Quality Technology*, 43, 83-98.
- Noorossana, R., Saghaei, A., and Amiri, A. (eds) (2011). *Statistical Analysis of Profile Monitoring*. Hoboken, NJ: Wiley.
- Patwari, N., and Hero, A. O. (2004), “Manifold Learning Algorithms for Localization in Wireless Sensor Networks,” In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 3, 857-860.
- Qiu, P., and Xing, C. (2013a), “Feature Based Image Registration Using Non-Degenerate Pixels,” *Signal Processing*, 93, 706-720.
- (2013b), “On nonparametric Image Registration,” *Technometrics*, 55, 174-188.
- Qiu, P., Zou, C., and Wang, Z. (2010), “Nonparametric Profile Monitoring By Mixed Effects Modeling,” *Technometrics*, 52, 265-277.
- Ramsay, J., and Silverman, B.W. (2005) *Functional Data Analysis*, (2nd ed.), Springer Series in Statistics, New York: Springer.
- Roweis, S. T., and Saul, L. K. (2000), “Nonlinear Dimensionality Reduction by Locally Linear Embedding,” *Science*, 290, 2323-2326.
- Sammon, J. W. (1969), “A Nonlinear Mapping for Data Structure Analysis,” *IEEE Transactions on Computers*, Vol, C-18, No. 5, 401-409.
- Shan, X., and Apley, D. W. (2008), “Blind Identification of Manufacturing Variation Patterns by Combining Source Separation Criteria,” *Technometrics*, 50, 332-343.

- Tenenbaum, J. B., de Silva, V., and Langford, J.C. (2000), "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, 290, 2319-2323.
- Torgerson, W. S. (1952), "Multidimensional Scaling: I. Theory and Method," *Psychometrika*, 17, 401-419.
- Walker, E., and Wright, S. P. (2002), "Comparing Curves Using Additive Models," *Journal of Quality Technology*, 34, 118-129.
- Williams, J. D., Woodall, W. H., and Birch, J. B. (2007), "Statistical Monitoring of Nonlinear Product and Process Quality Profiles," *Quality and Reliability Engineering International*, 23, 925-941.
- Williams, J. D., Birch, J. B., Woodall, W. H., and Ferry, N. M. (2007), "Statistical Monitoring of Heteroscedastic Dose-Response Profiles from High-Throughput Screening," *Journal of Agricultural, Biological, and Environmental Statistics*, 12, 216-235.
- Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004), "Using Control Charts to Monitor Process and Product Quality Profiles," *Journal of Quality Technology*, 36, 309-320.
- Xing, C., and Qiu, P. (2011), "Intensity Based Image Registration by Nonparametric Local Smoothing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33, 2081-2092.
- Yang, L. (2004), "Sammon's Nonlinear Mapping Using Geodesic Distances," *Proceedings of the 17th International Conference on Pattern Recognition*, 2, 303-306.
- Zhang, Z., and Zha, H. (2004), "Principal Manifolds and Nonlinear Dimensionality Reduction via Local Tangent Space Alignment," *SIAM Journal of Scientific Computing*, 25, 313-338.
- Zhang, H., and Albin, S. (2009). "Detecting Outliers in Complex Profiles Using a X2 Control Chart Method," *IIE Transactions* 41(4), 335-345.
- Zou, C., Tsung, F., and Wang, Z. (2008), "Monitoring Profiles Based on Nonparametric Regression Methods," *Technometrics*, 50, 512-526.