

The Effect of Pre-smoothing Functional Data on Cluster Analysis

David B. Hitchcock James G. Booth
University of South Carolina Cornell University

George Casella
University of Florida

June 23, 2006

Abstract

We investigate the possible benefits of pre-smoothing functional data before performing a cluster analysis. A simulation study compares the accuracy of clustering results based on using unsmoothed functional data—and two smoothed versions of the data—as the inputs in a clustering algorithm. Smoothing is usually found to produce a more accurate clustering, with the best results arising from a novel James-Stein-type shrinkage adjustment to the standard linear smoother. Two real functional data sets are clustered using the competing methods to illustrate the procedure.

KEY WORDS: Cluster analysis; Classification; Distance; Dissimilarity measures; Stein estimation; Shrinkage estimation.

2000 Mathematics Subject Classification Codes. 62-07; 62H30.

1 Introduction

Functional data may be defined as data for which the measurements on each observation are part of a single underlying continuous process (e.g., a time or space process). One example of such data are the Canadian weather data discussed by Ramsay and Silverman [14, Sec. 7.4], in which the measurements are the daily temperatures in Montreal over a period of 34 years. Ramsay and Silverman [13] generally label such data as *functional data*, since the underlying data are thought to be intrinsically smooth, continuous curves having domain \mathcal{T} .

Although the underlying processes generating the data (denoted $\mu_1(t), \dots, \mu_N(t)$ for a set of N data curves) are often assumed to be smooth, the data we observe likely contain random noise. In practice, we observe these curves at a grid of n points, t_1, \dots, t_n , so that we observe N independent vectors, each $n \times 1$.

Our model for the generation of the functional data can be summarized as follows (for $i = 1, \dots, N$):

$$\mu_i(t) \rightarrow y_i(t) \rightarrow \mathbf{y}_i.$$

Here, $\mu_i(t)$ is the i th smooth, underlying functional process; $y_i(t) = \mu_i(t) + \epsilon_i(t)$ is the i th functional observation, which contains random noise $\epsilon_i(t)$; and $\mathbf{y}_i = (y_i(t_1), \dots, y_i(t_n))'$ is the i th observed data vector, which represents a discretized version of $y_i(t)$, recorded at the n measurement points t_1, \dots, t_n .

Typically, in functional data analysis (a term attributed to Ramsay and Dalzell [12]), the primary goal is to discover something about the smooth curves that underlie the functional observations, and to analyze the entire set of functional data (consisting of many curves). Cluster analysis is a very common type of analysis of functional data. The goal of cluster analysis is to find groups, or clusters, in data. The objects in a data set should be grouped so that objects in the same cluster

are similar and objects in different clusters are dissimilar [3, 7].

The simplest approach to clustering functional data is to cluster the observed data curve without any pre-smoothing. This may be done using, for example, a clustering method based on the pairwise dissimilarities among the curves. The most common such methods are k-means [10] and its robust analogue k-medoids [6]. To use such clustering methods we must specify an appropriate dissimilarity measure for functional data. A natural dissimilarity is the squared L_2 distance between the i th and j th curves:

$$\int_{\mathcal{T}} [y_i(t) - y_j(t)]^2 dt. \quad (1)$$

Rather than using the raw data, we may wish to input some pre-smoothed version of the curves (denoted $\hat{\mu}_1(t), \dots, \hat{\mu}_N(t)$) into the clustering algorithm. The pairwise squared L_2 distances would then be

$$\int_{\mathcal{T}} [\hat{\mu}_i(t) - \hat{\mu}_j(t)]^2 dt. \quad (2)$$

for the N “smooths.” (Since in practice, the data are recorded as discretized functions, the actual dissimilarity measure used is a trapezoidal-rule approximation to the integral, given in Section 2.) A natural question is whether this pre-smoothing benefits the resulting cluster analysis and how much smoothing is optimal.

Note that sometimes, when the observed data are, say, hourly or daily averages, the very act of averaging, in a sense, smooths the measurements somewhat. In other cases, the recorded data do represent genuine snapshots of the functional observation at each t_j value. For our purposes, whether the data represent snapshots or averages over some short interval, we treat the observed data as “unsmoothed.”

Some methods for clustering functional data recently have been presented in the statistical literature. Serban and Wasserman [15] develop a method for clustering

curves after transforming and smoothing them and screening out flat curves. They also estimate the clustering error rate. James and Sugar [5] introduce a model-based clustering method that is especially useful when the points measured along the sample curves are sparse and irregular. Tarpey and Kinader [17] discuss representing each curve with basis functions and clustering via k-means, to estimate the “principal points” (underlying cluster means) of the data’s distribution. Abraham et al. [1] propose representing the curves via a B-spline basis and clustering the estimated coefficients with a k-means algorithm, and they derive consistency results about the convergence of the algorithm. Tarpey, Petcova and Ogden [18] apply methods for clustering functional data to the analysis of a pharmaceutical study.

It seems intuitive that in the analysis of functional data, some form of smoothing of the observed data is appropriate, and some previous methods [15, 5, 17, 1] do involve smoothing. Tarpey and Kinader [17, p. 113] pose the question of the effect of smoothing on the clustering of functional data, citing the need “to study the sensitivity of clustering methods on the degree of smoothing used to estimate the functions.”

In this paper, we investigate the effect pre-smoothing functional data has on the cluster analysis of that data. A simulation study will demonstrate that a shrinkage type of pre-smoothing leads to a clearly improved performance in correctly grouping objects into their proper clusters. Two real data sets are analyzed to further illustrate the effect of pre-smoothing on clustering.

2 Competing Methods

We will compare several approaches to clustering functional data via a simulation study and two real data examples.

Method 1 clusters the observed data curves based on the pairwise dissimilarities (1) using the k-medoids algorithm implemented by the R function `pam`. In the simulation study, the actual dissimilarity measure used is the trapezoidal-rule approximation d_{ij} [8]. For the observed data, measured n times along the domain $[t_1, t_n]$, this is

$$d_{ij} = \frac{t_n - t_1}{n - 1} \left\{ [y_i(t_1) - y_j(t_1)]/2 + \sum_{k=2}^{n-1} [y_i(t_k) - y_j(t_k)] + [y_i(t_n) - y_j(t_n)]/2 \right\},$$

$$i, j \in \{1, \dots, N\}, i \neq j.$$

The other methods cluster a smoothed version of the data and also employ the k-medoids algorithm, with dissimilarities d_{ij} approximating (2), i.e.,

$$d_{ij} = \frac{t_n - t_1}{n - 1} \left\{ [\hat{\mu}_i(t_1) - \hat{\mu}_j(t_1)]/2 + \sum_{k=2}^{n-1} [\hat{\mu}_i(t_k) - \hat{\mu}_j(t_k)] + [\hat{\mu}_i(t_n) - \hat{\mu}_j(t_n)]/2 \right\},$$

$$i, j \in \{1, \dots, N\}, i \neq j.$$

Method 2 smooths each data curve via a cubic spline basis in which the estimated curve is a linear combination of cubic splines joined at m knots [13, pp. 48-49]. In particular, we use a B-spline basis, which consists of compactly supported cubic splines that aid computation. A higher value of m uses more basis functions and yields a less smooth (more wiggly) estimated curve. We will implement the smoothing method for various levels of m to examine the effect of different degrees of smoothing.

Method 3 uses a James-Stein-type shrinkage smoother which is a weighted

average of the observed data vector \mathbf{y} and a linear smooth $\mathbf{S}\mathbf{y}$:

$$\mathbf{S}\mathbf{y}_i + \left(1 - \frac{a}{\|\mathbf{y}_i - \mathbf{S}\mathbf{y}_i\|^2}\right)_+ (\mathbf{y}_i - \mathbf{S}\mathbf{y}_i)$$

for $i = 1, \dots, N$, where \mathbf{S} is the smoothing matrix of a linear smoothing method (in our case, a B-spline smoother), $\|\cdot\|$ denotes the Euclidean norm, $(\cdot)_+$ denotes the positive part, and a is a user-chosen parameter. Following Lehmann and Casella [9, p. 367], we let $a = n - r - 2$, where $r = \text{rank}(\mathbf{S})$. For a cubic spline basis with m knots, $r = m + 4$. We will investigate varying choices of m and thus r . It has been shown in Hitchcock, Casella and Booth [4] that a similar Stein-type dissimilarity estimator leads to an improvement in estimating dissimilarities for noisy functional data, and it is reasonable to investigate whether the shrinkage smoother also leads to improved clustering results.

We will measure the “correctness” of the clustering method by the proportion of all possible pairs of objects that are correctly matched in the resulting grouping of objects. (A correct match for two objects means correctly putting the two objects in the same cluster *or* correctly putting the two objects in different clusters, depending on the “truth.”) This proportion serves as a measure of concordance between the clustering of the data set and the underlying clustering structure.

3 Simulation Study

In this section we summarize a simulation study to ascertain the nature of clustering improvement obtained by pre-smoothing. In each of 5000 replications, a sample of $N = 40$ noisy functional data were generated. For each sample of curves, several (four) “clusters” were built into the data by creating each functional observation from one of four distinct mean curves, to which the random noise was added.

Within the program the curves were represented in a discretized form, with values generated at n equally spaced points along $\mathcal{T} = [0, 5]$. Consider the following four mean curves, such that a simulated data set consists of 10 curves derived from each of them:

$$\mu_1(t) = -\sin(t-1)\ln(t+0.5), t \in [0, 5] \quad (3)$$

$$\mu_2(t) = \cos(t)\ln(t+0.5), t \in [0, 5] \quad (4)$$

$$\mu_3(t) = -0.25 - 0.1\cos(0.5(t-1))t^{1.5}\sqrt{5t^{1/2}+0.5}, t \in [0, 5] \quad (5)$$

$$\mu_4(t) = 0.6\cos(t)\ln(t+0.5)\sqrt{t+0.5}, t \in [0, 5] \quad (6)$$

These four curves, shown in Figure 1, were intentionally chosen to be similar enough to provide a good test for the clustering methods that attempted to group the curves into the correct clustering structure, yet different enough that they represented four clearly distinct processes. In short, the curves were chosen so that, when random noise was added to them, they would be difficult but not impossible to distinguish.

From these mean curves we generated noisy functional data (observed at $n = 50$ points in $[0, 5]$), by adding one of two types of random noise, one independent across measurement points and the other dependent. The first noise structure is independent $N(0, \sigma^2)$ with varying σ^2 . The second is a stationary Ornstein-Uhlenbeck process (in which the covariance between measurement points t_i and t_m is $\sigma^2(2\beta)^{-1}\exp\{-\beta|t_i - t_m|\}$) with varying σ^2 and $\beta = 1$. The Ornstein-Uhlenbeck error process results in an autoregressive covariance structure for the equally-spaced discretized data in the simulation. In either case, we refer to the unsmoothed noisy data as the “observed data.”

The simulated noisy data were smoothed using a cubic B-spline basis with a

Four mean curves for simulation study

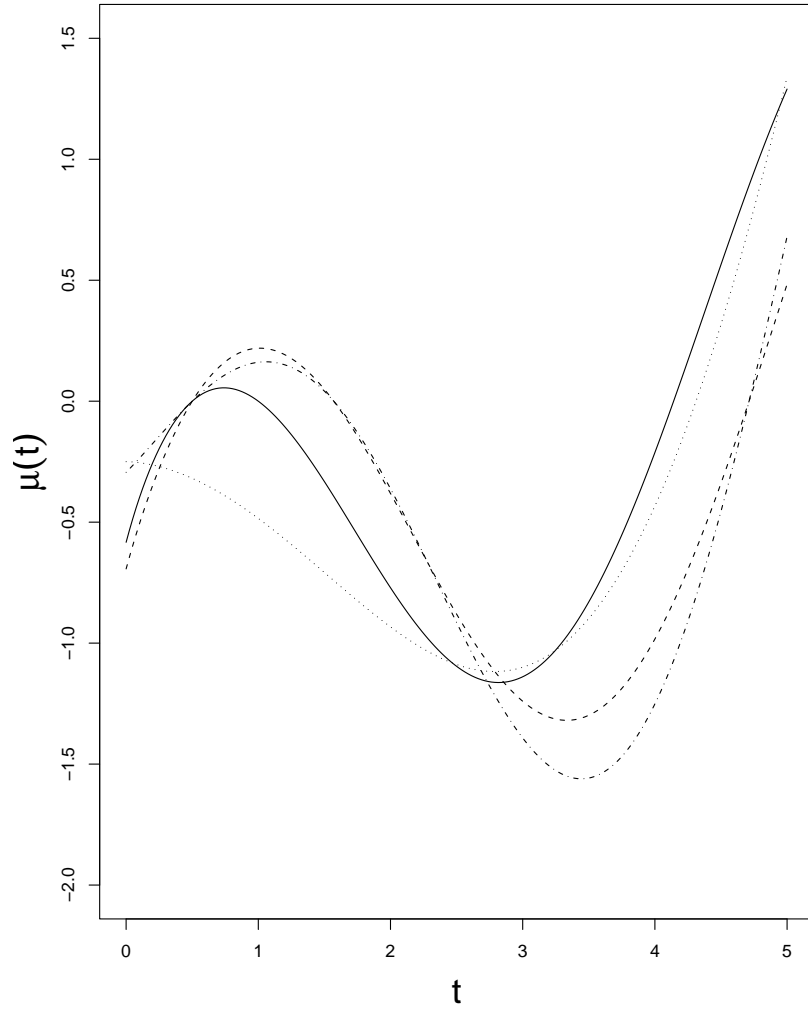


Figure 1: The four curves given by (3)-(6). Solid line: $\mu_1(t)$. Dashed line: $\mu_2(t)$. Dotted line: $\mu_3(t)$. Dot-dashed line: $\mu_4(t)$.

varying number of knots interspersed evenly through the data. To study how the degree of smoothing affected clustering success, we used, variously, $m = 2$ knots (a large amount of smoothing), $m = 6$ (moderate smoothing), and $m = 16$ (less smoothing).

We also implemented the James-Stein smoothing method by adjusting the B-spline smoother via the shrinkage method described earlier. With the James-Stein method we again investigated the effect of the degree of smoothing on the clustering results by varying the number of knots as described above.

We defined the pairwise dissimilarity between any two curves as the squared L_2 distance (1) (or (2) for the smoothed data). We calculated these dissimilarities for our simulated curves (actually using an approximation to account for the discretization of the simulated data). We then used the resulting dissimilarity matrix to cluster the observed data, and we similarly calculated dissimilarities and clustered each set of smoothed data. The clustering algorithm used was the k-medoids method, implemented by the `pam` function in R. We examined the resulting clustering structure of both the observed data and the smoothed data to determine which clustering better captures the structure of the underlying clusters, as defined by the four mean curves.

The simulated data curves were generated at eight equally spaced levels of σ^2 : 0.25, 0.5, ..., 2.0 (or nine levels: 0.5, 0.75, ..., 2.5, for the dependent-error data). The proportions of pairs of curves correctly matched are plotted in Figure 2 and Figure 3 for the observed data, the B-spline smooths, and the James-Stein-adjusted B-spline smooths.

For the independent-error data, Figure 2 indicates the clustering results are more accurate with either form of the smoothed data than with the observed

data, with the James-Stein adjustment increasing the improvement compared to the straightforward smoother. At low noise levels, the B-spline smoother and the shrinkage smoother coincide and their performances are equal, while for large σ values, the James-Stein smoother outperforms the other. Note also that the proportions are highest for $m = 2$ knots (and lowest for $m = 16$), indicating that rather severe smoothing (adjusted via the James-Stein shrinkage) provides the most accurate clustering.

Very similar results are seen from the simulations with the dependent error structure (Figure 3), with the major difference being that for the dependent-error data, the proportions of correct matches are generally higher for all methods than in the independent-error case. Again, the key conclusion is that smoothing produces a more accurate clustering than using the raw data, with the James-Stein adjustment further improving the smoothing for moderate-to-large noise levels. As with the independent-error data, a large amount of smoothing leads to a more accurate clustering.

Note that the Monte Carlo standard error for each of these proportions is at most 0.007 (when the true proportion of correct matches is 0.5). Based on this standard error, the clustering improvement from smoothing is quite clear, although the additional improvement from the James-Stein adjustment is only significant for higher levels of σ , e.g., $\sigma \geq 1.25$ for the independent-error data and $\sigma \geq 2$ for the dependent-error data.

The k-medoids procedure is a *partitioning* method of clustering, in which for a fixed number of clusters, the data objects are iteratively rearranged until a certain criterion is optimized. Another approach to cluster analysis is *hierarchical* clustering, in which, for example, objects are merged one at a time in a stepwise

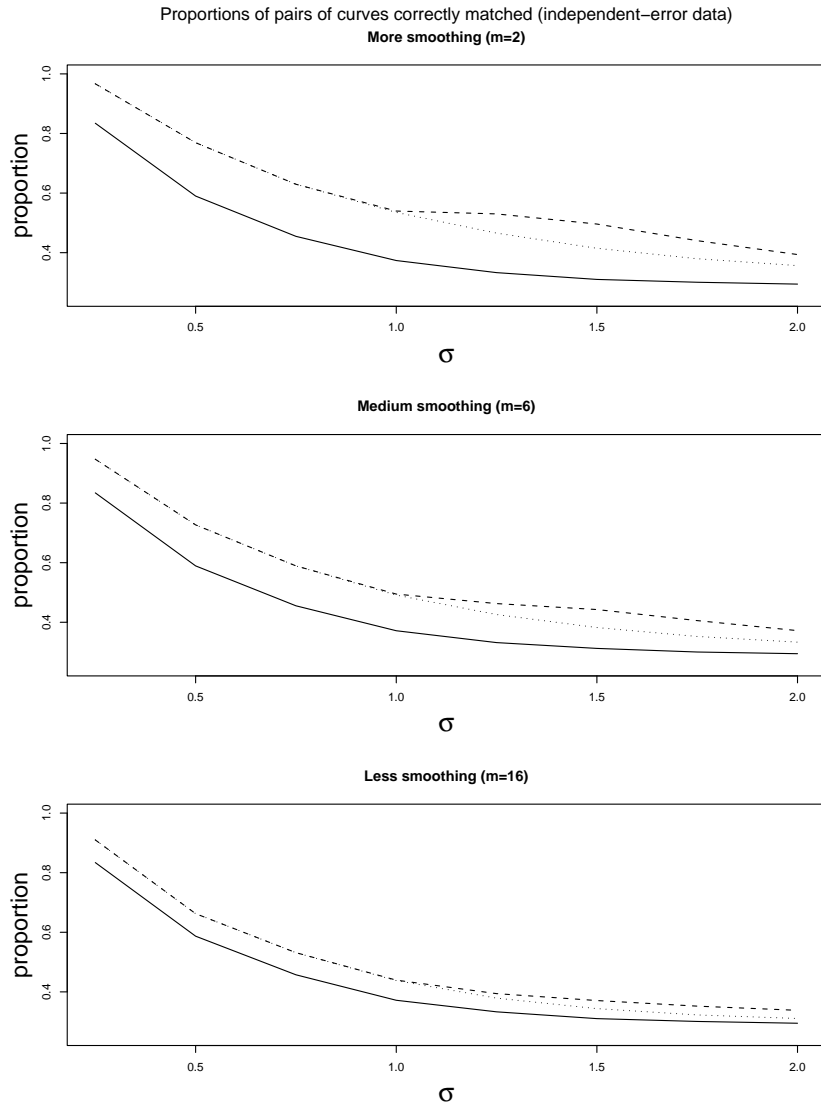


Figure 2: Solid line: Proportions based on observed data. Dotted line: Proportions based on B-spline approach. Dashed line: Proportions based on James-Stein approach.

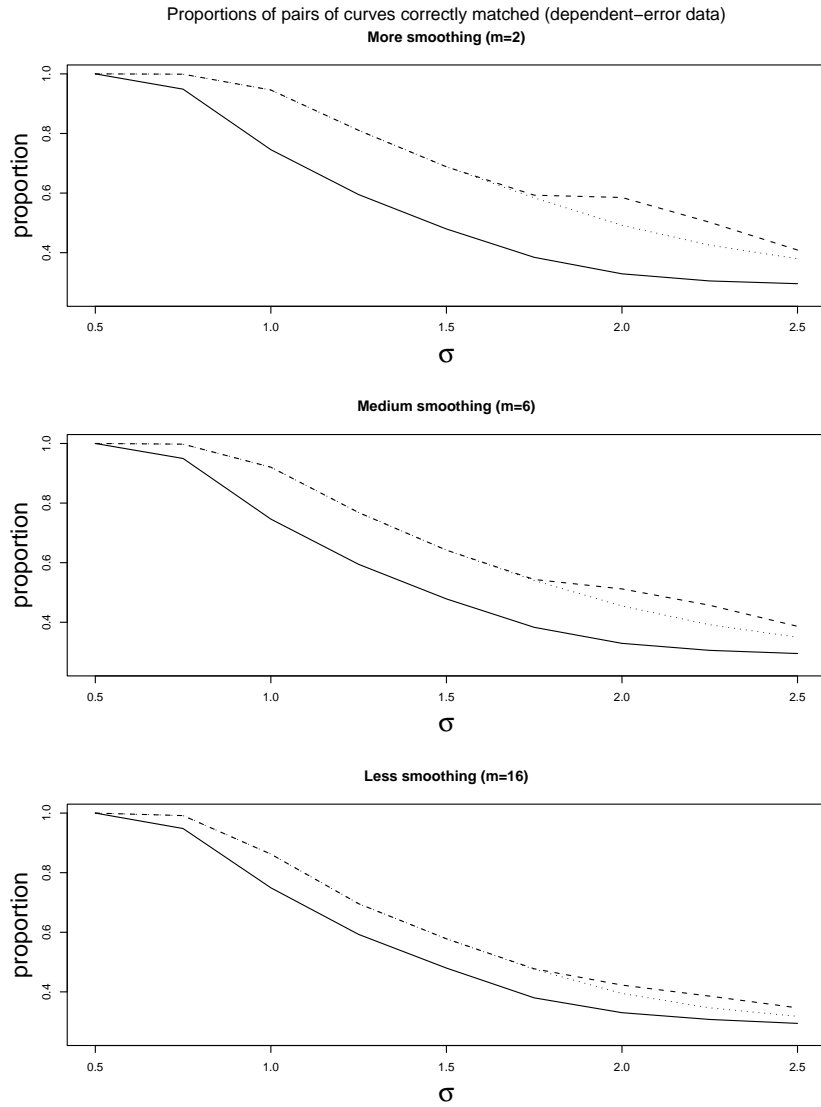


Figure 3: Solid line: Proportions based on observed data. Dotted line: Proportions based on B-spline approach. Dashed line: Proportions based on James-Stein approach.

optimal manner, forming a treelike structure that may be cut off after any step to yield a separation of the objects into a specified number of clusters. To determine the effect of smoothing on hierarchical clustering, we conducted a similar simulation study in which the curves were clustered using an agglomerative hierarchical method implemented with the R function `hclust`, with the average-linkage option, and with the clustering cut to form four clusters.

For small values of σ , no significant difference in performance was seen among the three methods. For larger values of σ , the James-Stein smoothing method again yielded the most accurate clustering, but here clustering the observed data did better than clustering the unadjusted smooths. Figure 4 shows the proportions of correct pairings (for both independent- and dependent-error data) for the medium-smoothing ($m = 6$) case; the results for $m = 2$ and $m = 16$ were similar. The performance of clustering the unadjusted smooths was relatively less poor when the data were smoothed less ($m = 16$), indicating that oversmoothing was possibly detrimental here. In general, however, the user cannot be certain of the proper degree of smoothing needed, and the James-Stein shrinkage adjustment provides an automated procedure to alleviate this uncertainty. Also, while our intention was not to compare the merits of hierarchical and partitioning methods of clustering, it is interesting that the hierarchical method, for this example, produces notably higher proportions of correct pairings than the k-medoids method does.

4 Two Real Data Examples

We now investigate the effect of smoothing on the clustering of functional data by analyzing two real data sets.

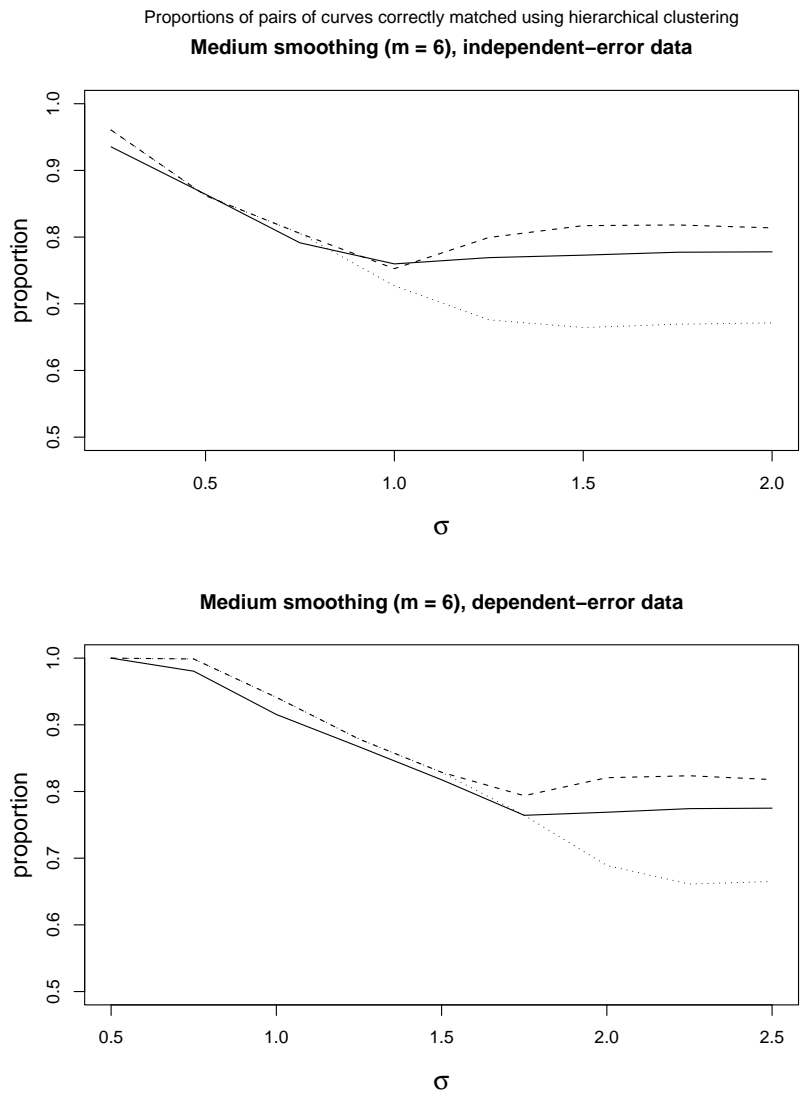


Figure 4: Solid line: Proportions based on observed data. Dotted line: Proportions based on B-spline approach. Dashed line: Proportions based on James-Stein approach.

Table 1: The 18 sampled sites, categorized according to their regions.

Region	NBDC Site Number
Northeast	44004, 44005, 44007, 44008, 44013, 44018
Southeast	40014, 41001, 41002, 41004, 41008
Florida/Eastern Gulf	41009, 41010, 42003
Western Gulf	42001, 42019, 42038, 42041

4.1 Example 1: Analysis of Ocean Wind Speed Curves

First, we analyze wind speed data gathered at buoys in the Atlantic Ocean and Gulf of Mexico. The data are available at the National Buoy Data Center (NBDC) historical data page <http://www.ndbc.noaa.gov/hmd.shtml> at the National Oceanic and Atmospheric Administration (NOAA) web site. We focus on wind speeds at 18 sites during the first week of 2005. Since wind is a continuous process, the speed patterns over time can be thought of as curves. These wind speeds (in meters per second) are recorded every hour (in fact the average wind speed for each hour is recorded). Thus for the week there are 168 measurements. Each site is located in one of four different regions (Northeast, Southeast, Florida/Eastern Gulf, and Western Gulf), as classified on the NBDC web site (see Table 1). Therefore we attempt to group the curves into four different clusters, with the “correctness” of the clustering being measured against the known grouping. For example, the sample curves for NBDC sites 44004 (from the Northeast region) and 41004 (from the Southeast Region) are given in Figure 5. (The solid curves are the functional observations, while the dotted curves represent the B-spline smooths described in the next paragraph.)

In the data set analyzed, curves 1 through 6 belonged to the Northeast cluster,

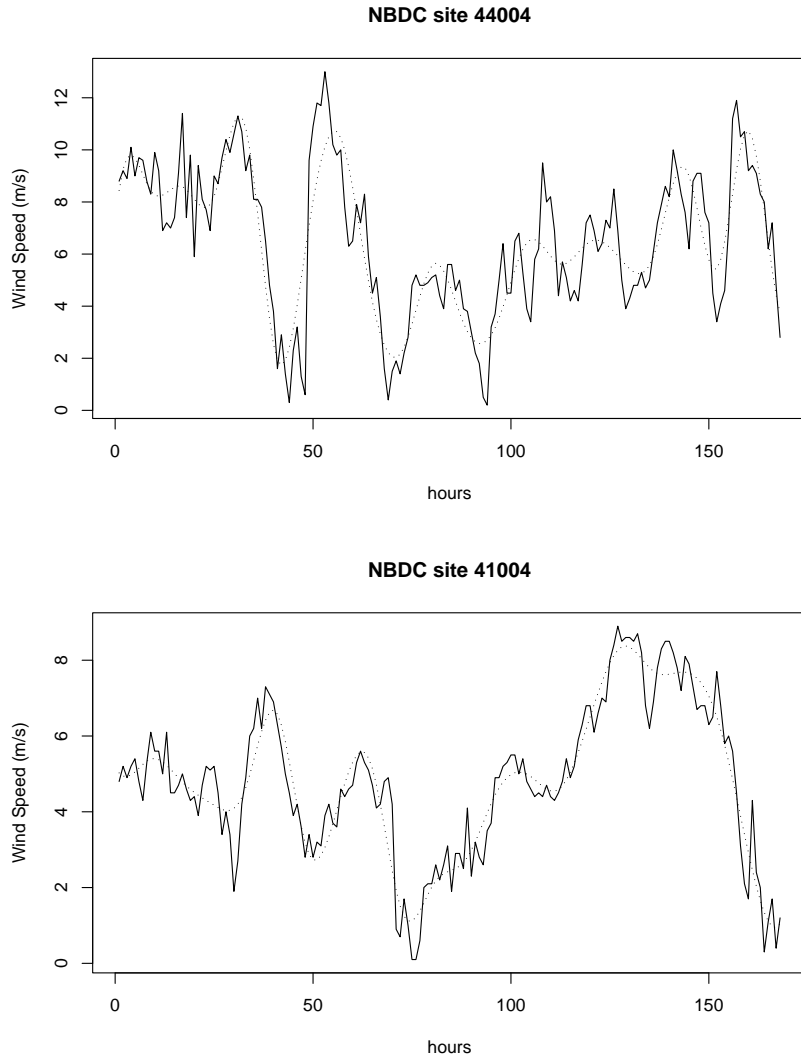


Figure 5: Solid line: Functional wind speed observations (meters per second) for two sites. Dotted line: Cubic B-spline smooths of the functional observations.

Table 2: The classification of the 18 sites into four clusters, based on observed data, smoothed data, and James-Stein-smoothed data.

Cluster	Observed Data	Smoothed Data	J-S Smoothed Data	“True” Clustering
1	1,2,3,4,5	1,2,3,4,5,6	1,2,3,4,5,6	1,2,3,4,5,6
2	6,8,18	7,8,18	7,8,18	7,8,9,10,11
3	7,9,10,11,13	9,10,11,12,13	9,10,11,12,13	12,13,14
4	12,14,15,16,17	14,15,16,17	14,15,16,17	15,16,17,18

curves 7-11 to the Southeast cluster, curves 12-14 to Florida/Eastern Gulf, and curves 15-18 to Western Gulf. We clustered the observed wind speed curves into four groups using the k-medoids method, implemented by the R function `pam`. We then clustered smoothed versions of these curves, using the two different smoothing methods: First we smoothed the data using a cubic B-spline basis having 20 knots interspersed evenly through the data. Secondly, we then adjusted the smooths via the James-Stein shrinkage method. Using the k-medoids algorithm, we clustered into four groups the unadjusted smooths and the adjusted smooths. The results of the three clusterings are shown in Table 2. The James-Stein smoothing method and the unadjusted B-spline smooths came closest to capturing the “true” grouping of the curves: 67.6% of the possible pairs of curves were correctly “matched” by each of these methods. Clustering the observed data resulted in 58.8% of the pairs of curves being correctly matched. Note that while all three methods “misclassified” site 18, the methods using the smooths put site 6, site 7, and site 12 into more appropriate clusters than did the method using the observed data.

4.2 Example 2: Analysis of Expression Ratios of Yeast Genes

As a second example, we consider yeast gene data analyzed in Alter, Brown and Botstein [2]. The curves in this situation correspond to 78 genes, and the measured responses are (log-transformed) expression ratios measured 18 times (at 7-minute intervals) $t_j = 7j$ for $j = 0, \dots, 17$. As described in Spellman et al. [16], the yeast was analyzed in an experiment involving “alpha-factor-based synchronization,” after which the mRNA expression for each gene was measured over time. (In fact, there were three separate synchronization methods used, but here we focus on the measurements produced by the alpha-factor synchronization.)

Biologists believe that these genes fall into five clusters—labeled G1, S, S/G2, G2/M, and M/G1—according to the cell cycle phase corresponding to each gene. (The letters S, M, and G denote Synthesis, Mitosis, and Gap, respectively.) For the data set analyzed, genes 1 through 13 are believed to belong to the M/G1 group, genes 14-52 to the G1 group, genes 53-60 to the S group, genes 61-67 to the S/G2 group, and genes 68-78 to the G2/M group.

To analyze these data, we treated them as 78 separate functional observations. Initially, each multivariate observation on a gene was centered by subtracting the mean response (across the 18 measurements) from the measurements. Then we clustered the genes into 5 clusters using the k-medoids method, implemented by the R function `pam`.

To investigate the effect of smoothing on the cluster analysis, first we simply clustered the unsmoothed observed data. Then we smoothed each observation and clustered the smooth curves. A cubic B-spline smoother, with six interior knots interspersed evenly within the given timepoints, was chosen. The smooths were

adjusted via the James-Stein shrinkage procedure described in Section 2, with $a = 6$ here, since $n - r - 2 = 6$, where $r = 10$ is the rank of the smoothing matrix for this cubic spline smoother.

None of the methods captured the supposed clustering structure extremely well: Again, the clustering of the smooths came closest to capturing the “true” grouping of the curves: 49.4% of the possible pairs of curves were correctly “matched” by this method, both with and without the James-Stein adjustment. Clustering the observed data resulted in 44.5% of the pairs of curves being correctly matched. For this data set, we obtained higher proportions (over 70%) of correct matches when grouping the curves into four (or even three) clusters, indicating that perhaps the ostensible classification into five distinct groups of genes is questionable. The S group is the only cluster of the five that is particularly well-defined, rendering the formation of the five-cluster partition difficult at best.

For each of the previous two examples, we also clustered the data using the agglomerative hierarchical method, with the clustering tree cut to four clusters (for the wind data) and five clusters (for the yeast gene data). For both examples, the hierarchical clustering results were virtually identical across the three methods. For the wind data, the three sets (observed, smoothed, and James-Stein smoothed) yielded identical clusterings, with 79.4% of pairs of curves correctly matched. The three clusterings of the genes were nearly identical, with the observed data having 91.3% of pairs correctly matched, and the other two methods 90.8%. Although by this measure, the hierarchical clustering performed very well, a close examination of the partitions showed some strange results: For the wind data, site 11 was in a cluster by itself under all three methods. For the gene data, all three methods produced a huge cluster containing the great majority of the genes, while another

cluster contained but a single gene. In short, conclusions based on the real data must be tempered by the uncertainty about the true number of clusters and the form of the “true” clustering.

5 Conclusion

We have found that some type of smoothing (most notably the James-Stein smoothing) does improve clustering results. This is clearly apparent in the findings of our simulation study, and less clear in the analysis of the two real data sets. It should be noted, however, that the “true” clustering structure to which the competing clusterings are compared is a far more solid gold standard in the case of the simulations than in the real data examples. When one judges the competing methods, the simulation study offers more reliable evidence about their comparative performances.

It should be noted that often the appropriate number of clusters is unknown *a priori* and must be determined from the data. In the simulations and data analyses presented here, there was evidence specifying the correct number of clusters. In other problems, however, investigators may need to use a method such as those in [11] to help determine an appropriate number of clusters.

The evidence in this paper suggests that when clustering functional data, investigators are usually justified in pre-smoothing the data curves, particularly when a partitioning method of clustering is chosen. Among the approaches considered here, the James-Stein-adjusted smoother provides the best clustering results. In terms of the optimal degree of smoothing, the most accurate clustering resulted from smoothing the data substantially, and then adjusting via the James-Stein method, which has the effect of shrinking the severe smooth back toward the ob-

served data.

A natural reason that the pre-smoothed data are often clustered more correctly is that the smoothed curves, rather than the noisy data, more accurately reflect the underlying processes generating the functional observations—and implicitly generating the clusters. A danger for investigators is that an improperly chosen smoother may oversmooth the data. This concern is alleviated by using the James-Stein adjustment, which offers a data-driven compromise between the smooths and the observed data curves.

In this study, a B-spline basis smoother was chosen for its simplicity and its natural representation as a linear smoother. Other choices of smoothing methods, such as smoothing splines or kernel-based methods, may provide different levels of improvement, although applying the James-Stein shrinkage might be computationally less straightforward in those cases. Regardless of the choice of smoother, the practice of pre-smoothing functional data appears beneficial in many situations.

Acknowledgements

This work was partially supported by National Science Foundation grants DMS-0405543 and DMS-9971586. The authors thank an associate editor and a referee for their helpful comments which improved the article.

References

- [1] Abraham, C., Cornillon, P.A., Matzner-Lober, E. and Molinari, N., 2003, Unsupervised curve clustering using B-splines. *The Scandinavian Journal of Statistics*, **30**, 581–595.

- [2] Alter, O., Brown, P.O. and Botstein, D., 2000, Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences*, **97**, 10101–10106.
- [3] Everitt, B., Landau, S. and Leese, M., 2001, *Cluster Analysis* (London: Edward Arnold Publishers).
- [4] Hitchcock, D.B., Casella, G. and Booth, J.G., 2006, Improved estimation of dissimilarities by presmoothing functional data. To appear in *Journal of the American Statistical Association*, **101**.
- [5] James, G.M. and Sugar, C.A., 2003, Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, **98**, 397–408.
- [6] Kaufman, L. and Rousseeuw, P.J., 1987, Clustering by Means of Medoids. In: *Statistical Data Analysis Based on the L_1 Norm*, pp. 405–416.
- [7] Kaufman, L. and Rousseeuw, P.J., 1990, *Finding Groups in Data: An Introduction to Cluster Analysis* (New York: Wiley).
- [8] Larson, R.E., Hostetler, R.P. and Edwards, B.H., 1990, *Calculus* (Heath).
- [9] Lehmann, E.L. and Casella, G., 1998, *Theory of Point Estimation* (New York: Springer-Verlag).
- [10] MacQueen, J., 1967, Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1*, pp. 281–297.
- [11] Milligan, G.W. and Cooper, M.C., 1985, An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, **50**, 159–179.

- [12] Ramsay, J.O. and Dalzell, C.J., 1991, Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B, Methodological*, **53**, 539–561.
- [13] Ramsay, J.O., and Silverman, B.W., 1997, *Functional Data Analysis* (New York: Springer-Verlag).
- [14] Ramsay, J. O. and Silverman, B. W., 2002, *Applied Functional Data Analysis: Methods and Case Studies* (New York: Springer-Verlag Inc)
- [15] Serban, N. and Wasserman, L., 2005, CATS: Clustering after transformation and smoothing. *Journal of the American Statistical Association*, **100**, 990–999.
- [16] Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D. and Futcher, B., 1998, Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*, **9**, 3273–3297.
- [17] Tarpey, T. and Kinateder, K., 2003, Clustering functional data. *The Journal of Classification*, **20**, 93–114.
- [18] Tarpey, T., Petkova, E. and Ogden, R.T., 2003, Profiling placebo responders by self-consistent partitioning of functional data. *Journal of the American Statistical Association*, **98**, 850–858.