Depth-Based Nonparametric Description of Functional Data, with Emphasis on Use of Spatial Depth

Robert Serfling^{a,*}, Uditha Wijesuriya^b

^a*Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75080, USA*

^b*Department of Mathematics, University of Southern Indiana, Evansville, IN 47712, USA*

Abstract

Statistical depth and related quantile functions, originally introduced for nonparametric description and analysis of multivariate data in a way sensitive to inherent geometry, are in active development for functional data and in this setting offer special options since the data may be visualized regardless of dimension. Here depth-based methods are provided for revealing the structure of the data set in terms of relevant sample quantile curves displayed at selected levels, and for constructing and displaying confidence bands for corresponding “population” versions. Also, the functional boxplot is enhanced, by adding inner fences to flag possible shape outliers, along with the outer fences that flag location outliers. This enables the boxplot to serve as a stand-alone tool for functional data, as with univariate and multivariate data. Further, the spatial depth approach, well-established for multivariate data, is investigated for nonparametric description of functional data. In comparison with four other depth-based approaches for functional data over a range of actual and simulated data sets, this approach is seen to offer a very competitive combination of robustness, efficiency, computational ease, simplicity, and versatility.

Keywords: Functional data, Nonparametric, Quantiles, Outlier detection, Boxplot

2000 MSC: Primary 62G99, Secondary 62H99

1. Introduction and preliminaries

Functional data, where each point is a curve, arises in all areas of modern science, engineering, and industry. Nonparametric description of such data sets, in terms of a median curve, quantile curves at various levels, the middle 50% region of curves, and identification of outliers, is fundamental to understanding structure. A natural approach is provided by depth functions, which order data points according to some given notion of centrality or, equivalently, of outlyingness, and which generate related quantile functions. With functional data, which may be visualized regardless of dimension, depth methods offer special options beyond their standard role with multivariate data.

Here we present three depth-based methodological tools for nonparametric description of functional data. The first concerns definition and display of sample quantile curves at arbitrary levels. The second concerns construction of confidence bands for the corresponding “population” quantile curves being estimated by the sample versions, under the technical condition that asymptotic normality holds for the discretized sample quantile curves based on the depth function in use. The third concerns nonparametric identification of outliers,

*Corresponding author

Email address: serfling@utdallas.edu (Robert Serfling)

URL: www.utdallas.edu/~serfling (Robert Serfling)

for which we introduce a “double-fence functional boxplot”, enhancing the Sun and Genton (2011) functional boxplot by incorporating additional fences in order to better flag “shape” outliers, which are not envisioned in the classical boxplot design oriented only to location outliers.

Further, we focus on a particular depth function – the spatial depth – and establish its effectiveness and special appeal in the functional data setting. Despite its success with multivariate data, this approach has received but limited treatment with functional data, even though Chaudhuri (1996) indicated its potential for natural extension from Euclidean spaces to infinite-dimensional Hilbert and Banach spaces. One attractive feature is that it satisfies an asymptotic normality condition required for our confidence band procedure. Also, for outlier identification using the double-fence functional boxplot, we compare the spatial depth approach with four leading depth-based approaches for functional data, over a range of actual and simulated data sets and a variety of outlier types, and find that the spatial approach exhibits superior overall performance.

Our setting consists of a data set of curves $X_k(t), t \in [a, b]$, for $1 \leq k \leq n$, defined over a finite interval $[a, b]$ as elements of a Hilbert space $\mathcal{H}$. (Confining to an interval $[a, b]$ is arbitrary, for convenience.) We can work entirely within $\mathcal{H}$, but it is technically convenient and computationally efficient to make use also of associated discretized versions of the curves as vectors in $\mathbb{R}^d$ for some suitable d . Indeed, in many typical applications, the curves are actually secondary constructions from data that has arisen initially in discretized form. In any case, for $1 \leq k \leq n$, we associate with the curve $X_k(\cdot)$ the vector $\mathbf{X}_k = (X_k(t_1), \dots, X_k(t_d))'$ in $\mathbb{R}^d$, corresponding to discretization using d equally spaced points

$$a \leq t_1 < t_2 < \dots < t_d \leq b. \quad (1)$$

Thus we effectively use $\mathbb{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ as our basic data set for analytical work, while for plots we use the curves $X_k(\cdot)$, $1 \leq k \leq n$. Here no restriction such as $d \leq n$ or $d \geq n$ is imposed. Of course, in cases where n is very large, the displays of the data curves might be carried out just for a subsample, in order not to obscure visualization of the typical structure of the curves. Even so, the analytical work defining quantile curves and confidence bands to add to such plots is based on the full sample or at least a large subsample. Also, for technical reasons, the initial computation of the confidence band procedure is based on discretization to $d_0 = \min\{d, 20, n - 1\}$ as explained in Section 3.1.3.

Remark 1. ON DISCRETIZATION OF FUNCTIONAL DATA CURVES. *In discretizing into $\mathbb{R}^d$ to enable use of standard multivariate analysis techniques, it is helpful to keep in mind five important differences between the usual multivariate data setting and that of discretized functional data.*

(i) Besides “location” outliers lying apart from the main body of data, a functional data set can also have “shape” outliers differing in structure or pattern from the typical cases, as well as outliers deviating in both location and shape. Shape outliers embedded within the main body of curves make outlier detection considerably more challenging.

(ii) In the functional data setting, full affine invariance of outlier detection procedures is neither required nor desired. In particular, invariance under heterogeneous scale changes of $X(t), t \in [a, b]$, is irrelevant, it sufficing to have invariance merely under homogeneous scale changes and under orthogonal transformations of discretized versions. Similar comments apply regarding equivariance of the median and other quantile curves. Thus, fortuitously, invariance and equivariance requirements are less stringent in the functional data setting.

(iii) The component variables $X(t)$ defining a curve are ordered in the functional data setting, and this carries over to the components of their associated discretized versions. Therefore, in the functional data setting, we have the very attractive feature that each data point, discretized or not, may be represented and viewed as a curve in a 2-dimensional plot, $X(t)$ versus t , regardless of the inherent dimensionality of the data.

(iv) Further, since such plots naturally have an “upper” region and a “lower” region, the selection of representative quantile curves at various levels may accordingly be confined to “upper” and “lower” types relative to the location of the median curve. In comparison, in the standard multivariate data setting, the quantiles at a given level form a contour that is associated with all directions from the “center”, and there is no simplifying “upper” versus “lower” orientation. (See Sections 3.1.1 and 3.1.2.)

(v) Finally, moreover, the directions that index the quantile representations in $\mathbb{R}^d$ of the discretized data curves also have ordered components and, therefore, like the data curves themselves, may be represented and viewed as curves in a 2-dimensional plot, regardless of the dimensionality of the discretization (see Figure 1). $\square$

Formally, the following objectives typify nonparametric description of functional data:

1. Identification of the sample median curve and selected sample quantile curves, for example a representative “90th percentile curve”;
2. Identification and designation of “outlier” curves, both location and shape types;
3. Display of the data as a 2-dimensional functional boxplot exhibiting the “middle” 50% region of sample curves and “fences” demarking outlier regions;
4. Determination and display of confidence bands for the population median curve and selected other population quantile curves, when the data is a sample from some parent population, or a representative subsample from a very large data set.

Such goals are pursued using a selected depth function suitably designed for functional data, along with its associated outlyingness and quantile functions. Section 2 introduces the several functional depth functions involved in our treatment, with Section 2.1 discussing depth and related functions in general in the setting of $\mathbb{R}^d$.

General methodology is presented and illustrated in Section 3. Regarding the *first* goal, the “median” curve as simply the curve produced by minimizing the relevant outlyingness function. Other quantile curves are defined in correspondence with quantile levels of the outlyingness function, but there are infinitely many possibilities at any level. In order to accommodate simplicity of display, a method to select at any level a single representative “canonical quantile curve” is introduced in Sections 3.1.1-3.1.2.

Apropos to the *fourth* goal, a procedure for placing confidence bands on population quantile curves estimated by sample versions is developed in Section 3.1.3. This requires an assumption of asymptotic normality for the sample quantile vectors in $\mathbb{R}^d$ associated with the chosen depth function, a condition satisfied by the spatial depth approach emphasized here but not by other typical functional depths.

Regarding the *second and third* goals, one must distinguish both “location” outliers lying apart from the main body of data and “shape outliers” embedded within the main body, as well as sample curves deviating in both location and shape. For *location outliers*, the “functional boxplot” of Sun and Genton (2011), displays the “50% central region” or “middle half” of the data consisting of the curves with the 50% lowest outlyingness values and demarks location outlier thresholds by “fence” curves placed above and below the 50% central region, at distances 1.5 times the pointwise interquartile range of this region. Curves which cross or lie beyond the lower and upper “fence” curves so defined are candidates for designation as location outliers.

However, the identification of *embedded shape outliers* poses a different kind of challenge not anticipated by the classical boxplot design. A shape outlier misclassified into the “middle half” can greatly distort this central region and consequently also the fences defined for identification of location outliers, causing the functional boxplot to fail completely. Some depth-based methods with functional data avoid such an outcome by using, in conjunction with the chosen depth and the associated functional boxplot, some additional *ad hoc* tool

formulated especially for identification of shape outliers. However, when the chosen depth function *reliably avoids classifying serious shape outliers into the 50% central region*, a simpler resolution is available, namely augmentation of the functional boxplot by adding strategically placed (additional) lower and upper fence curves defined using the factor 0.5 instead of 1.5. The two additional fences lie closer to the main body of data than the fences for location outlier identification and thus flag candidate shape outliers, i.e., curves which intersect an inner fence, or lie between the two upper inner and outer fences, or between the two lower inner and outer fences. This “*double-fence functional boxplot*” is treated in Section 3.2.

Relative to the outlier detection goal, the performance of the spatial depth approach (“*Spatial*”) is compared with four leading depth-based approaches for functional data, the Mahalanobis distance depth (*Foutliers*), two versions of the integrated depth (*FM(trim)* and *FM(pond)*), and the modified band depth (*MBD*). Based on our double-fence functional boxplot as the tool for sample outlier identification, percentages of correct decisions are compared. Section 2 formally introduces all five functional depth methods. In Section 4, the methods are compared over 175 contaminated data sets created from seven representative actual data sets by adding a single artificial shape outlier chosen from one of five types and given one of five locations. In Section 5, they are compared across four simulated functional data models with contamination, three from Arribas-Gil and Romo (2014) and a fourth that we formulate to cover the case of spike outliers.

Section 6 provides concluding discussion. Further perspective on use of the double-fence functional boxplot is presented. Also, a strong recommendation of *Spatial* as a highly competitive approach for nonparametric description of functional data is developed, based on both methodological considerations and our empirical performance studies.

Concluding this introduction, we mention some related work and further pursuits. (i) An artifact of nonparametric description of functional data that we have not included in our quantile-focused treatment is the *mean function*, which in fact is more straightforward to treat than the median and other quantiles, which require notions of depth function and have numerous competing formulations. For treatment of the mean curve, see Cardot and Josserand (2011) and Cardot, Goga, and Lardin (2013), including approaches for placing confidence bands on the mean function of the full data set when the data at hand is a large subsample. (ii) Besides the particular depth functions for functional data that we consider here, several other choices are available. See Cuevas, Febrero, and Fraiman (2007) for projection-based functional depth methods, Cuesta-Albertos and Nieto-Reyes (2008) for a random Tukey depth, and Fraiman and Zvarc (2013) for data-adaptive robust estimation based on random projections. (iii) Nieto-Reyes and Battey (2015) review several functional depths relative to selected criteria and develop a theoretical foundation for the formulation of functional depth functions. (iv) It is desirable to extend the present contributions to *multivariate* functional data, as formulated, for example, by Claeskens, Hubert, Slaets, and Vakili (2014). (v) It is desirable to extend also to *spatial* functional data, i.e., to functions defined over both time and space, as, for example, arise with neuroimaging data (see Zhu, Fan, and Kong, 2014).

2. The spatial and other depth approaches for functional data

2.1. Depth, outlyingness, quantile, and rank functions in $\mathbb{R}^d$

Let F be a given distribution F on $\mathbb{R}^d$. A *depth function* $D(\mathbf{x}, F)$ on $\mathbb{R}^d$ provides a center-outward ordering of the points $\mathbf{x}$ in $\mathbb{R}^d$ relative to F , desirably with nested contours of equal depth and satisfying an invariance requirement such as affine invariance, or at least orthogonal invariance. The point of maximal depth is interpreted both as “center” and as “median”. An equivalent *outlyingness function* $O(\mathbf{x}, F)$ is given by some one-to-one

mapping such as $O = 1 - D$ or $O = 1/(1 + D)$. The contours of equal depth (or outlyingness) generate an associated *quantile function* $\mathbf{Q}(\mathbf{u}, F)$ in $\mathbb{R}^d$, indexed by $\mathbf{u}$ in the unit ball $\mathbb{B}^d$ in $\mathbb{R}^d$, whereby each $\mathbf{x}$ has a quantile representation $\mathbf{x} = \mathbf{Q}(\mathbf{u}, F)$. The index $\mathbf{u}$ is determined so as to represent a directional rank associated with $\mathbf{x}$ and thus to define a rank function $\mathbf{R}(\mathbf{x}, F)$, with the proviso that $\|\mathbf{R}(\mathbf{x}, F)\| = \|\mathbf{u}\| = O(\mathbf{x}, F)$, showing that “central” and “extreme” quantile levels correspond to $\mathbf{Q}(\mathbf{u}, F)$ with $\|\mathbf{u}\|$ close to 0 and 1, respectively. Also, via the representation $\mathbf{x} = \mathbf{Q}(\mathbf{u}, F) = \mathbf{Q}(\mathbf{R}(\mathbf{x}, F), F)$, it is seen that the inverse of the quantile function $\mathbf{Q}(\mathbf{u}, F)$ is not the distribution function F but rather the rank function $\mathbf{R}(\mathbf{x}, F)$. The median corresponds to $\mathbf{u} = \mathbf{0}$ and is denoted by $\mathbf{Q}(\mathbf{0}, F)$. Relative to a sample $\mathbb{X}_n$ from F , sample versions $D(\mathbf{x}, \mathbb{X}_n)$, etc., are defined analogously. Note that the *sample* spatial median is defined as the point maximizing $D(\mathbf{x}, \mathbb{X}_n)$, which is not necessarily the (computationally quicker) maximum depth *data point*. As discussed in Serfling (2010), one can start with any one of these D , O , Q , and R functions and generate the others.

2.2. The spatial depth approach for functional data

Consider functional data curves in a Hilbert space $\mathcal{H}$ and their discretization to points in $\mathbb{R}^d$. We treat the spatial approach first in $\mathbb{R}^d$ and then in infinite-dimensional spaces.

2.2.1. The spatial approach in $\mathbb{R}^d$

Following Chaudhuri (1996), for a given distribution F on $\mathbb{R}^d$, the associated *spatial* (or “geometric”) *quantile function* in $\mathbb{R}^d$, indexed by $\mathbf{u}$ in $\mathbb{B}^d$, is defined as the minimizer $\mathbf{Q}(\mathbf{u}, F)$ of $E(\Phi(\mathbf{u}, \mathbf{X} - \mathbf{x}) - \Phi(\mathbf{u}, \mathbf{X}))$ with respect to $\mathbf{x}$, where $\Phi(\mathbf{u}, \mathbf{z}) = \|\mathbf{z}\| + \langle \mathbf{u}, \mathbf{z} \rangle$ for $\mathbf{z}$ in $\mathbb{R}^d$ and $\langle \cdot, \cdot \rangle$ denotes Euclidean inner product. A rationale for indexing $\mathbf{Q}(\mathbf{u}, F)$ by $\mathbf{u}$ in $\mathbb{B}^d$ is discussed in Section 2.2.2.

This generalizes, of course, the well-known characterization of the p th quantile $F^{-1}(p)$, of univariate Z with $E|Z| < \infty$ as any value θ minimizing $E(|Z - \theta| + u(Z - \theta))$, where $u = 2p - 1$ (Ferguson, 1967, p. 51). The resulting $Q(u, F)$, $-1 < u < +1$, simply re-indexes the usual quantile function $F^{-1}(p)$, $0 < p < 1$, via the transformation $u = 2p - 1$. Note that, for $x = F^{-1}(p) = Q(u, F)$, the magnitude of the new index $|u| = |2p - 1|$ represents a measure of the *outlyingness* of the point x .

Equivalently, $\mathbf{Q}(\mathbf{u}, F)$ is obtained by solving the equation $E(\mathbf{S}(\mathbf{X} - \mathbf{x})) + \mathbf{u} = \mathbf{0}$ with respect to $\mathbf{x}$ for given $\mathbf{u}$, where $\mathbf{S}(\mathbf{y})$ denotes the spatial sign function

$$\mathbf{S}(\mathbf{y}) = \begin{cases} \frac{\mathbf{y}}{\|\mathbf{y}\|}, & \mathbf{y} \neq \mathbf{0}, \\ \mathbf{0}, & \mathbf{y} = \mathbf{0}. \end{cases}$$

The *inverse* of $\mathbf{Q}(\mathbf{u}, F)$ is obtained by solving the same equation with respect to $\mathbf{u}$ for given $\mathbf{x}$, yielding the *spatial rank function* $\mathbf{R}(\mathbf{x}, F) = E(\mathbf{S}(\mathbf{x} - \mathbf{X}))$, *spatial outlyingness function* $O(\mathbf{x}, F) = \|\mathbf{R}(\mathbf{x}, F)\|$, and *spatial depth function* $D(\mathbf{x}, F) = 1 - O(\mathbf{x}, F)$. Sample versions $\mathbf{Q}(\mathbf{u}, \mathbb{X}_n)$, etc., are obtained by substituting $\|n^{-1} \sum_{k=1}^n \mathbf{S}(\mathbf{x} - \mathbf{X}_k)\|$ for $E(\mathbf{S}(\mathbf{x} - \mathbf{X}))$ in these population expressions.

The spatial depth and quantile functions are invariant and equivariant, respectively, only under orthogonal and homogeneous scale transformations, which suffices for functional data applications, as noted in Remark 1(ii). In comparison, for the spatial approach in the standard multivariate setting, to attain full affine invariance or equivariance that is typically desired one must first standardize the data by the inverse square root of a suitable scatter matrix (see Serfling, 2010, for discussion), introducing additional computational complexity and additional robustness issues. It is fortuitous that such standardization can be avoided for use of the spatial approach in the functional data setting.

For detailed background technical discussion on spatial quantiles, see Chaudhuri (1996) and Serfling (2004). Also, see Oja (2010) for a comprehensive nonparametric multivariate data analysis methodology centered on the spatial rank function.

2.2.2. Extension of the spatial approach to the functional data setting

For X a random function in a Hilbert space $\mathcal{H}$, the (population) spatial quantile function $Q(u)$, indexed by elements u in the unit ball $\mathcal{B}$ in $\mathcal{H}$, minimizes $E[\phi(u, X-x) - \phi(u, X)]$ with respect to x , where $\phi(u, z) = |u| + \langle u, z \rangle$, $|\cdot|$ is the L^2 -norm, and $\langle \cdot, \cdot \rangle$ is the inner product on $\mathcal{H}$ (see Chaudhuri, 1996, Gervini, 2008, and Sguera, Galeano, and Lillo, 2014). Equivalently, via Gâteaux differentiation, $Q(u)$ may be obtained as the solution of $E[S(X-x)] + u = 0$ with respect to x for fixed u . The solution for $u = 0$ represents the “deepest” curve (median). The spatial outlyingness function is $|E[S(x-X)]|$ and the spatial depth function is $1 - |E[S(x-X)]|$. Relative to a sample $\mathbb{X}_n$, sample versions $Q_n(u)$, etc., are obtained by substituting $|n^{-1} \sum_{k=1}^n S(x - X_k)|$ for $E(S(x-X))$ in the population expressions.

Indexing of $Q(u)$ by u in the unit ball of $\mathcal{H}$ reflects the view of ϕ as a linear functional in a Banach space, in which case spatial quantiles may be indexed by the open unit ball in the dual Banach space of real-valued linear functionals. In the case of a Hilbert space, the dual is isomorphic to the space itself and hence to the unit ball $\mathcal{B}$ in $\mathcal{H}$. See Chaudhuri (1996) for detailed discussion. However, for plotting of particular quantiles $Q(u)$ in the functional data setting, only a select subset of values u are of interest, as discussed in Sections 3.1.1-3.1.2.

For fast computation of the functional spatial median in a Hilbert space, an averaged stochastic gradient algorithm is provided by Cardot, Cenac, and Zitt (2013). This median is applied by Chaouch and Goya (2012) as a robust measure of center for a data set of electricity loading curves. The functional spatial depth is applied by Squera, Galeano, and Lillo (2014) for classification of functional data. However, the functional spatial depth has not been explored previously for nonparametric description of a functional data set as per our four goals stated above, nor has it been used in a functional boxplot. We pursue these uses of spatial depth here, and, for increased computational efficiency as well as for technical convenience, we make heavy use of reduction by the discretization (1) to the spatial approach in a Euclidean space. That is, the curves are treated computationally as vectors in $\mathbb{R}^d$ for an appropriate dimension d (in some cases they have arisen initially as such vectors).

2.3. Four leading current depth-based procedures for functional data

Four methods in current use that we describe below are taken from the R packages *fda.usc* (Manuel Febrero Bande, Manuel Oviedo de la Fuente, Pedro Galeano, Alicia Nieto and Eduardo Garcia-Portugues, 2011), *depthTools* (Sara López-Pintado and Aurora Torrente, 2011), and *rainbow* (Han Lin Shang and Rob J. Hyndman, 2012).

2.3.1. Mahalanobis distance approach

The method “Foutliers” in *rainbow* includes several functional depths, in particular by default (method = “robMah”) the one based on robust Mahalanobis distance using the minimum covariance determinant estimators of location and scatter (Rousseeuw and Leroy, 1987). We choose this option here.

2.3.2. Integrated depth approach

The integrated depth (Fraiman and Muniz, 2001), or so-called FM depth, of the curve x_i in a sample of curves $\mathbb{X}_n = \{x_1(t), x_2(t), \dots, x_n(t), t \in [a, b]\}$ in $\mathcal{H}$ is defined by

$$\text{FMD}(x_i, \mathbb{X}_n) = \int_a^b \left[1 - \left| \frac{1}{2} - F_{n,t}(x_i(t)) \right| \right] dt,$$

where $F_{n,t}(x_i(t)) = n^{-1} \sum_{k=1}^n I(x_k(t) \leq x_i(t))$ and $I(\cdot)$ is the indicator function. For the discretization (1), we use

$$\text{FMD}(x_i, \mathbb{X}_n) = \sum_{j=2}^d (t_j - t_{j-1}) \left[1 - \left| \frac{1}{2} - n^{-1} \sum_{k=1}^n I(x_k(t_j) \leq x_i(t_j)) \right| \right]$$

for $1 \leq i \leq n$. (See also Febrero, Galeano, and Gozalez-Manteiga, 2008.) We include here two versions of FM depth, denoted by “FM(trim)” and “FM(pond)”, given by the commands “outlier.depth.trim” and “outlier.depth.pond” in *fda.usc*.

Remark 2. VULNERABILITY OF FM DEPTH TO “SPIKE” TYPE SHAPE OUTLIERS. *The computation of $FMD(x_i, \mathbb{X}_n)$, the empirical distribution function $n^{-1} \sum_{k=1}^n I(x_k(t_j) \leq x_i(t_j))$ is employed at each of the grid of points $t_1 < \dots < t_d$. Because the indicator function $I(\alpha \leq \beta)$ is insensitive to the magnitude of the difference $\beta - \alpha$, a spike outlier curve with very high $x_i(t_j)$ relative to values $x_i(t_{j-1})$ and $x_i(t_{j+1})$ at nearby points t_{j-1} and t_{j+1} , for some i and j , can have FM depth as high as that of curves entirely within the middle of the data set. In this case, such an outlier is classified among the deepest curves, is not identified as an actual outlier, and distorts the shape of the 50% middle region.* $\square$

2.3.3. Modified band depth

Again consider a sample of curves $\mathbb{X}_n = \{x_1(t), x_2(t), \dots, x_n(t), t \in [a, b]\}$ in $\mathcal{H}$. For given choice of $j \in \{2, \dots, n\}$, the modified band depth $MBD_n^{(j)}$ of Lopez-Pintado and Romo (2009) starts with bands in $\mathbb{R}^2$ formed by subsets $\{x_{i_1}, x_{i_2}, \dots, x_{i_j}\}$ of curves, i.e.,

$$B(x_{i_1}, x_{i_2}, \dots, x_{i_j}) = \left\{ (t, y) : t \in [a, b], \min_{i=1, \dots, j} x_i(t) \leq y \leq \max_{i=1, \dots, j} x_i(t) \right\}.$$

For any x in $\{x_1, x_2, \dots, x_n\}$, the set of t in $[a, b]$ with $x(t)$ in the band $B(x_{i_1}, x_{i_2}, \dots, x_{i_j})$ is given by

$$A(x; x_{i_1}, x_{i_2}, \dots, x_{i_j}) = \left\{ t \in [a, b] : \min_{r=i_1, \dots, i_j} x_r(t) \leq x(t) \leq \max_{r=i_1, \dots, i_j} x_r(t) \right\}.$$

Then, with λ the Lebesgue measure on $[a, b]$,

$$\lambda_r(A(x; x_{i_1}, x_{i_2}, \dots, x_{i_j})) = \lambda(A(x; x_{i_1}, x_{i_2}, \dots, x_{i_j})) / (b - a),$$

gives the “proportion of time” that x is in $B(x_{i_1}, x_{i_2}, \dots, x_{i_j})$. Accordingly, the average of these proportions over all choices of $i_1, \dots, i_j$, i.e.,

$$MBD_n^{(j)}(x) = \binom{n}{j}^{-1} \sum_{1 \leq i_1 < \dots < i_j \leq n} \lambda_r(A(x; x_{i_1}, \dots, x_{i_j}))$$

defines a certain “band depth” of the curve x in the set $\{x_1, x_2, \dots, x_n\}$. Then, with J the limits of tolerance for the computational burden, which increases with j , an overall (“modified”) band depth of x is given by the sum $\sum_{j=2}^J MBD_n^{(j)}(x)$. Here we use the case $J = 2$ (known as “MBD”), as found in *depthTools*.

Remark 3. VULNERABILITY OF MBD TO SPIKE OUTLIERS. *Because a spike outlier curve lies within many bands except for very short periods of “time”, its MBD can be high, thus classifying it among the deepest curves. In this case, it fails to become identified as an actual outlier and the middle 50% region becomes distorted. In practice, this is resolved by using MBD in conjunction with a further tool, the “modified epigram index” (MEI), as treated in Arribas-Gil and Romo (2014).* $\square$

3. Methodology for nonparametric description of functional data sets

Section 3.1 treats quantile-based description of functional data, which differs significantly from that for standard multivariate data. Section 3.2 provides our enhanced version of functional boxplot. Any functional depth approach may be used in these methods, except for the confidence band procedure of Section 3.1.3, which requires that the sample quantile function $\mathbf{Q}(\mathbf{u}, \mathbb{X}_n)$ be asymptotically normal, which indeed holds for the spatial approach.

3.1. Quantile-based description of functional data

3.1.1. Upper/lower orientation in plots of curves, and canonical directions for quantiles

First let us review quantile-based description of multivariate data in $\mathbb{R}^d$. Relative to the sample quantile function $\mathbf{Q}(\mathbf{u}, \mathbb{X}_n)$, the points $\mathbf{x}$ possessing a given fixed level of outlyingness, say $O(\mathbf{x}, \mathbb{X}_n) = c$, comprise a *contour* $\mathcal{C}(c, \mathbb{X}_n) = \{\mathbf{Q}(\mathbf{u}, \mathbb{X}_n) : \mathbf{u} \in \mathbb{B}^d, \|\mathbf{u}\| = c\}$. These nested contours defined for $0 < c < 1$ can be indexed and/or selected by specifying the outlyingness level c , or alternatively by specifying the probability weight or the volume of the enclosed central region, as may be desired. However, except for $d = 2$ and 3 , one cannot visualize $\mathcal{C}(c, \mathbb{X}_n)$, and there is no basis for emphasizing some of the underlying directions $\mathbf{u}$ over others. One simply identifies the set of points $\mathbf{x}$ belonging to $\mathcal{C}(c, \mathbb{X}_n)$.

On the other hand, with *functional data*, one may view the entire data set of curves in $\mathcal{H}$ in a single plot, and, consequently, the “contours” $\mathcal{C}_n(c) = \{Q_n(u) : u \in \mathcal{B}(\mathcal{H}), |u| = c\}$ at a given outlyingness level c may also be viewed. However, this advantage needs modification in order to become of practical use, since a straightforward rendering of $\mathcal{C}_n(c)$ will appear as a solid band extending above and below the median curve, defying useful visual interpretation. An effective solution is based on the feature that plots of curves naturally have an “upper” region and a “lower” region, as noted in Remark 1(iv). Thus we display just two special subsets of the contour $\mathcal{C}_n(c)$, those for which the quantile curves $Q_n(u)$ for u correspond to “upper” and “lower” directions, respectively. This yields, for given outlyingness level c , an *upper quantile band* located somewhat above the median curve and a *lower quantile band* below the median curve.

In terms of the discretized curves as vectors in $\mathbb{R}^d$, the upper band is given by curves in $\mathcal{H}$ corresponding to the discretized curves $\{\mathbf{Q}(\mathbf{u}, \mathbb{X}_n) : \mathbf{u} \in \mathbb{B}^d; u_i \geq 0, 1 \leq i \leq d; \|\mathbf{u}\| = c\}$, for which the directions $\mathbf{u}$ have *nonnegative* components, subject to $\|\mathbf{u}\| = c$. Likewise, the lower band is similarly based on $\{\mathbf{Q}(\mathbf{u}, \mathbb{X}_n) : \mathbf{u} \in \mathbb{B}^d \text{ with } u_i \leq 0, 1 \leq i \leq d; \|\mathbf{u}\| = c\}$, for which the directions $\mathbf{u}$ have *nonpositive* components, subject to $\|\mathbf{u}\| = c$.

Moreover, however, for simplicity and clarity, we replace the entire upper band for given $\|\mathbf{u}\| = c$ by a single representative upper quantile curve, and likewise the entire lower band by a single representative lower quantile curve. For this purpose, we introduce and use the (upper and lower) “*canonical directions*”

$$\mathbf{u} = \pm c \left(\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}, \dots, \frac{1}{\sqrt{d}} \right)', \quad (2)$$

which satisfy $\|\mathbf{u}\| = c$ and work very well in practice. For $d = 2$, these are the diagonal directions in the “positive” and “negative” quadrants, respectively. As d increases, these become nearly orthogonal to the coordinate directions (see Scott, 1992).

As noted in Remark 1(v), directions in $\mathbb{B}^d$ associated with discretized functional data vectors in $\mathbb{R}^d$ may be viewed and compared as curves in $\mathcal{H}$. **Figure 1** (upper) displays as curves in $\mathcal{H}$ a typical upper canonical direction and for comparison two other upper directions, randomly chosen. Also, with reference to a Canadian city data set (in the R package *fda*) consisting of monthly temperatures in 35 locations in Canada averaged over the years 1960 through 1994, **Figure 1** (lower) displays the sample quantile curves corresponding to these three directions. It is seen that the sample quantile curve based on the upper canonical direction follows especially well the general shape structure of the data set.

(Could put Figure 1 here)

3.1.2. Upper and lower sample p th canonical quantiles

For a given sample $\mathbb{X}_n$ and $p \in (0, 1)$, let c_p denote the sample p th quantile of the outlyingness values $O(\mathbf{X}_i, \mathbb{X}_n)$, $1 \leq i \leq n$, and let $\pm \mathbf{u}_p$ be given by (2) with $c = c_p$.

Then the upper and lower *sample p th canonical quantiles* are given by the curves $Q_n(\pm u_p)$ in $\mathcal{H}$ constructed from the vectors $\mathbf{Q}(\pm \mathbf{u}_p, \mathbb{X}_n)$ in $\mathbb{R}^d$. Geometrically, these curves are representative points on the equal-outlyingness contours enclosing the middle $100p\%$ of the data curves and have shapes following the general structure of the data curves.

Note that upper and lower p th quantile curves differ from boundary curves of the sample $100p\%$ central region enclosing the $\lceil pn \rceil$ least outlying curves in the sample. Such boundary curves also follow the general structure of the data curves but are more sensitive to anomalies and lack geometric quantile interpretations. They have different roles, as in the functional boxplot considered in Section 3.2.

For the data set of Canadian city temperature curves, **Figure 2** (upper row) displays the sample median and sample 95% upper and lower canonical quantiles. Each of the upper and lower 95% quantile curves has some data curves crossing it, as expected. However, it is noteworthy that the lower quantile curve has a data curve lying completely below it, whereas no data curve lies completely above the upper quantile curve. Note that displaying canonical quantile curves facilitates the perception and description of any asymmetry between the upper and lower parts of the data set, when present.

3.1.3. Confidence bands for population quantiles in Hilbert curve space

We provide a confidence band procedure for a population quantile curve $Q(u)$ in $\mathcal{H}$ which is applicable when our chosen depth function supports the following assumption:

The depth-based sample quantile function $\mathbf{Q}(\mathbf{u}, \mathbb{X}_n)$ is asymptotically normal, with mean $\mathbf{Q}(\mathbf{u}, F)$ and some covariance $n^{-1}\Sigma(\mathbf{u}, F)$.

More precisely, let $\mathbf{Q}(\mathbf{u}, F)$ be the quantile in $\mathbb{R}^d$ corresponding to $Q(u)$ via (1), and suppose that

$$\sqrt{n}\Sigma(\mathbf{u}, \mathbb{X}_n)^{-1/2}(\mathbf{Q}(\mathbf{u}, \mathbb{X}_n) - \mathbf{Q}(\mathbf{u}, F)) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_d), \quad (3)$$

with $\Sigma(\mathbf{u}, \mathbb{X}_n)$ a consistent sample estimator of $\Sigma(\mathbf{u}, F)$. In this case, we construct an ellipsoidal confidence region in $\mathbb{R}^d$ for $\mathbf{Q}(\mathbf{u}, F)$ and then convert its boundary points to curves in $\mathcal{H}$ which together represent a confidence band for $Q(u)$.

In particular, for the *spatial quantile function* in $\mathbb{R}^d$, it follows from Chaudhuri (1996, Theorem 3.1.2) that (3) holds with

$$\Sigma(\mathbf{u}, \mathbb{X}_n) = [\mathbf{D}_1(\mathbf{Q}(\mathbf{u}, \mathbb{X}_n))]^{-1} \mathbf{D}_2(\mathbf{Q}(\mathbf{u}, \mathbb{X}_n), \mathbf{u}) [\mathbf{D}_1(\mathbf{Q}(\mathbf{u}, \mathbb{X}_n))]^{-1}, \quad (4)$$

where

$$\mathbf{D}_1(\mathbf{y}) = n^{-1} \sum_{k=1}^n \frac{1}{\|\mathbf{X}_k - \mathbf{y}\|} \left(\mathbf{I}_d - \frac{(\mathbf{X}_k - \mathbf{y})(\mathbf{X}_k - \mathbf{y})'}{\|\mathbf{X}_k - \mathbf{y}\|^2} \right)$$

and

$$\mathbf{D}_2(\mathbf{y}, \mathbf{u}) = n^{-1} \sum_{k=1}^n \left(\frac{\mathbf{X}_k - \mathbf{y}}{\|\mathbf{X}_k - \mathbf{y}\|} + \mathbf{u} \right) \left(\frac{\mathbf{X}_k - \mathbf{y}}{\|\mathbf{X}_k - \mathbf{y}\|} + \mathbf{u} \right)'.$$

Remark 4. Note that since (3) is based on $n \rightarrow \infty$, its use in practice for fixed d and n imposes that $d < n$. Further, since with discretized functional data the coordinates are ordered and have dependence structure similar to a time series, we anticipate a possibly substantial degree of collinearity among adjacent coordinates, thus adversely affecting the required nonsingularity of $\Sigma(\mathbf{u}, \mathbb{X}_n)$, especially for higher choices of discretization dimension d . From empirical exploration (discussed below), we find that discretization dimension **20** works well. Accordingly, practical implementation of our confidence band procedure with a particular data set entails the following steps regarding d :

Step 1. For computation of the confidence ellipsoid, apply (3) with

$$d_0 = \min\{d, 20, n - 1\}.$$

Step 2. For display of quantile curves in $\mathcal{H}$ converted from points on the ellipsoidal surface in $\mathbb{R}^{d_0}$, discretize using the original choice of d as in (1) and interpolate to construct the final curve.

We now provide the details for construction of suitable confidence bands.

ELLIPSOIDAL CONFIDENCE REGION FOR $\mathbf{Q}(\mathbf{u}, F)$ IN $\mathbb{R}^{d_0}$. From (3) follows by standard techniques (for example, Serfling, 1980) that

$$n(\mathbf{Q}(\mathbf{u}, \mathbb{X}_n) - \mathbf{Q}(\mathbf{u}, F))' \boldsymbol{\Sigma}(\mathbf{u}, \mathbb{X}_n)^{-1} (\mathbf{Q}(\mathbf{u}, \mathbb{X}_n) - \mathbf{Q}(\mathbf{u}, F)) \xrightarrow{d} \chi_{d_0}^2, \quad n \rightarrow \infty,$$

where χ_m^2 denotes the chisquare distribution with m degrees of freedom. Letting $\chi_{m,\beta}^2$ denote the β quantile of χ_m^2 , for $0 < \beta < 1$, we thus have, by the usual inversion,

Theorem 5. *An asymptotic $100(1 - \alpha)\%$ confidence region for $\mathbf{Q}(\mathbf{u}, F)$ in $\mathbb{R}^{d_0}$ is given by*

$$\mathcal{E}_n = \{\boldsymbol{\theta} \in \mathbb{R}^{d_0} : n(\boldsymbol{\theta} - \mathbf{Q}(\mathbf{u}, \mathbb{X}_n))' \boldsymbol{\Sigma}(\mathbf{u}, \mathbb{X}_n)^{-1} (\boldsymbol{\theta} - \mathbf{Q}(\mathbf{u}, \mathbb{X}_n)) \leq \chi_{d_0, 1-\alpha}^2\}.$$

CONVERSION TO A CONFIDENCE BAND IN $\mathcal{H}$. In principle, we may convert $\mathcal{E}_n$ to an associated band in $\mathcal{H}$, by treating the surface points of $\mathcal{E}_n$ as discretized versions of curves in $\mathcal{H}$ and reconstructing and plotting them in the $(t, X(t))$ plane. We then attach to the band the confidence level associated with $\mathcal{E}_n$. That is, the envelope formed by the pointwise maxima and minima of these curves in $\mathcal{H}$ is treated as a $100(1 - \alpha)\%$ confidence band for the population quantile curve $Q(u)$.

However, the sampling variability in $\boldsymbol{\Sigma}(\mathbf{u}, \mathbb{X}_n)$ induces a significant deviation in the orientation of $\mathcal{E}_n$ from that of the ellipsoid using the true covariance $\boldsymbol{\Sigma}(\mathbf{u}, F)$, resulting in spurious curves and a widening of the envelope forming the associated confidence band, making it overly conservative. Consequently, we use just a strategic subset of points from the boundary of $\mathcal{E}_n$, thus also reducing the computational burden. In particular, we confine attention to the $2d_0$ surface points on $\mathcal{E}_n$ that intersect the plus/minus eigendirections of $\boldsymbol{\Sigma}(\mathbf{u}, \mathbb{X}_n)$ from the center $\mathbf{Q}(\mathbf{u}, \mathbb{X}_n)$ of $\mathcal{E}_n$. Moreover, we further reduce to just the points corresponding to the largest few eigenvalues.

More precisely, let us denote the eigenvalues of $\boldsymbol{\Sigma}(\mathbf{u}, \mathbb{X}_n)$ by $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{d_0}$ and the corresponding unit eigenvectors by $\mathbf{V}_i$, $1 \leq i \leq d_0$. Then, letting $\boldsymbol{\theta}_{i1}$ and $\boldsymbol{\theta}_{i2}$ be the surface points in the directions $\pm \mathbf{V}_i$ from $\mathbf{Q}(\mathbf{u}, \mathbb{X}_n)$, respectively, it is readily seen that $\boldsymbol{\theta}_{i1} = \mathbf{Q}(\mathbf{u}, \mathbb{X}_n) + \beta_i \mathbf{V}_i$ and $\boldsymbol{\theta}_{i2} = \mathbf{Q}(\mathbf{u}, \mathbb{X}_n) - \beta_i \mathbf{V}_i$, with $\beta_i = (\lambda_i \chi_{d_0, 1-\alpha}^2 / n)^{1/2}$, $1 \leq i \leq d_0$. We take these $2d_0$ points as candidates for conversion to curves in $\mathcal{H}$, as illustrated in **Figure 2** (middle row). This corresponds to choosing the extreme points on the ellipsoid along the orthogonal directions of the principal component coordinate space. In this spirit, we may consider using just the first few principal directions. Accordingly, for each $J = 1, \dots, d_0$, let $CB^{(J)}$ denote the confidence band in $\mathcal{H}$ generated from $\{\boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_{i2}, 1 \leq i \leq J\}$.

In a small simulation study described below, the approximate asymptotic validity of this confidence procedure is verified, and it is further seen that very low J is effective. On this basis, we arrive at the following practical rule of thumb. *Use $CB^{(J)}$ with $d = d_0$ and J selected according to sample size n as follows:*

1. For $n \leq 75$, use $CB^{(4)}$;
2. For $75 < n \leq 450$, use $CB^{(2)}$;
3. For $n > 450$, use $CB^{(1)}$.

The present study simply confirms the feasibility of our CB procedure as a sound heuristic when the requisite asymptotic normality of sample quantiles is available. As a companion to using the sample quantiles to emphasize key structural features of a functional data set, our CB procedure provides an indication of their reliability as estimators of the associated target quantiles (even if purely hypothetical) and accommodates comparative viewing of the reliabilities of the quantiles displayed in a given plot.

As an illustration of CB procedure, we use Spatial with the Canadian city temperature curves, for which $n = 35$ and $d = 12$ and thus $d_0 = 12$. **Figure 2** (lower row) exhibits the 95% CBs so obtained for the median, the upper 95% quantile, and the lower 95% quantile, respectively. (The agreement between our choice of confidence level and our choice of quantile level is completely coincidental.) The bands displayed are the $\text{CB}^{(12)}$ versions, but in fact the respective $\text{CB}^{(1)}$ versions are virtually identical to the $\text{CB}^{(12)}$ versions for this data set. To see how the CBs contribute useful information in a visual way, note the relatively wider CB covering the sample lower 95% quantile, compared to that covering the sample upper 95% quantile. This reflects and emphasizes the relatively higher spread of the lower curves within this relatively small data set. Although this variability is perceivable from viewing the data, the CB helps assess its direct impact and also provides a note of caution about the sample lower quantile for this data set.

(Could put Figure 2 here)

SMALL SIMULATION STUDY ON FEASIBILITY OF $\text{CB}^{(J)}$ AND CHOICES OF d AND J . To confirm empirically the feasibility of the confidence band procedure $\text{CB}^{(J)}$ and to explore effective choices of d and J , a small simulation study was carried out. The parent distribution on $\mathcal{H}$ was taken as the main component of Model I (see Section 5.1.1) used in our large simulation study of Section 5 that compares Spatial with four other depth-based approaches.

For the present study, the target parameters to receive confidence bands in $\mathcal{H}$ were the “population” quantile curves $Q(0)$ and $Q(u_{0.95})$, represented in discretized form as $\mathbf{Q}(\mathbf{0}, F)$ and $\mathbf{Q}(\mathbf{u}_{0.95}, F)$ in $\mathbb{R}^d$, with F the parent distribution on $\mathbb{R}^d$ induced by the model on $\mathcal{H}$. These *target curves* were determined by discretizing the model with $d = 50$ and estimating them by $\mathbf{Q}(\mathbf{0}, \mathbb{X}_{n_0})$ and $\mathbf{Q}(\mathbf{u}_{0.95}, \mathbb{X}_{n_0})$ using a very large sample of size $n_0 = 20,000$. The study determined *empirical coverage probabilities* by the 95% confidence bands $\text{CB}^{(J)}$ for each of $J = 1, 2$, and 4 , based on discretization of the parent model using $d = 20$, taken over 1900 replications each, for samples of sizes $n = 25, 50, 100, 200, 300, 400, 500, 1000$, and 2000 . The number 1900 of replications was chosen to obtain a standard deviation of approximately 0.005 for the empirical coverage probabilities of order 0.95, which then individually are within ± 0.01 of the true coverage probabilities with 95% confidence.

The findings are reported in **Table 1** below. For $\text{CB}^{(1)}$, the coverage probabilities for $\mathbf{Q}(\mathbf{0}, F)$ and $\mathbf{Q}(\mathbf{u}_{0.95}, F)$, respectively, increase from 0.38 and 0.31 for $n = 25$ to 0.94 and 0.96 for $n = 2000$. The low rates for $\text{CB}^{(1)}$ for $n = 25$ can be improved to 0.97 and 0.78 using $\text{CB}^{(4)}$, and likewise for $n = 50$. $\text{CB}^{(2)}$ succeeds well for $n = 100, 200$, and 300 , and $\text{CB}^{(1)}$ and $\text{CB}^{(2)}$ are about equally effective at $n = 400$. $\text{CB}^{(1)}$ is best for $n \geq 500$. Also, we see that as n increases the coverage probabilities appear to be converging to the target value 0.95. On this basis, we confirm the feasibility of this confidence band procedure and arrive at the above-stated rule of thumb.

It is of interest that the coverage probabilities for the confidence bands in $\mathcal{H}$ tend to be significantly better than those for the corresponding ellipsoids in $\mathbb{R}^d$. This is because when a relatively “thin” ellipsoid in $\mathbb{R}^d$ misses covering the true center $\mathbf{Q}(\mathbf{u}, F)$, the extreme points along its leading principal axis nevertheless still tend to convert to curves lying above and below the “true” quantile curve $Q(u)$ in $\mathcal{H}$. Indeed, as n increases, these two extreme points carry most of the relevant information.

n	CB ⁽¹⁾		CB ⁽²⁾		CB ⁽⁴⁾	
	$\mathbf{u} = \mathbf{0}$	$\mathbf{u} = \mathbf{u}_{0.95}$	$\mathbf{u} = \mathbf{0}$	$\mathbf{u} = \mathbf{u}_{0.95}$	$\mathbf{u} = \mathbf{0}$	$\mathbf{u} = \mathbf{u}_{0.95}$
25	0.38	0.31	0.87	0.65	0.97	0.78
50	0.59	0.61	0.94	0.83	0.98	0.88
100	0.75	0.80	0.97	0.91	—	—
200	0.87	0.91	0.98	0.94	—	—
300	0.88	0.93	0.98	0.95	—	—
400	0.91	0.96	0.98	0.97	—	—
500	0.92	0.96	0.98	0.97	—	—
1000	0.93	0.96	0.98	0.96	—	—
2000	0.94	0.96	0.97	0.96	—	—

Table 1: Empirical coverage probabilities by confidence band CB^(J) for population median based on $\mathbf{Q}(\mathbf{u}, F)$ with $\mathbf{u} = \mathbf{0}$ and population 95% upper quantile based on $\mathbf{Q}(\mathbf{u}, F)$ with $\mathbf{u} = \mathbf{u}_{0.95}$, with discretization dimension $d = 20$, for $J = 1, 2$, and 4, and sample sizes $n = 25, 50, 100, 200, 300, 400, 500, 1000$, and 2000, with 1900 replications for each sample size.

Extensive additional experimentation was carried out exploring $d = 30, 40$, and 50 . It was found that as d increases, regularization of $\Sigma(\mathbf{u}, \mathbb{X}_n)$ becomes necessary to avoid nonsingularity. However, regularization is not needed when using $d = 20$, a simple choice found to be either better than or competitive with the other options.

A more thorough study of the confidence band procedure CB^(J) would be of interest, exploring other types of parent population and further investigating optimal choices of d and J . However, this is beyond the scope of the present paper.

3.2. Double-fence functional boxplot

Functional boxplots have been developed by Sun and Genton (2011). Essentially, their boxplot displays the median curve, the 50% most central region of curves based on the MBD, and fences demarking regions of potential location outliers. The sample 50% central region is given by the envelope of the sample curves $X_{[k]}$, $1 \leq k \leq \lceil n/2 \rceil$, where $X_{[k]}$ denotes the curve with k th smallest outlyingness, $1 \leq k \leq n$. The fences are curves following a level 1.5 IQR above and below the 50% central region, where “IQR” is the (pointwise) interquartile range, i.e., the pointwise vertical range of the 50% central region. Because this functional boxplot does not do well with certain shape outliers, its usage is sometimes accompanied by *ad hoc* approaches introduced as extra steps to detect such outliers, as we have mentioned in discussing MBD in Section 2.3.3. In any case, there may still be distortion of the middle half and of the fences, as discussed in Section 1 and further in Section 6.

However, if we employ a depth that reliably does not classify a severe shape outlier into the middle half, we can enhance the boxplot in a way that visually flags these outliers as well as the location outliers, by adding inner fences. Specifically, we introduce a *double-fence version* of functional boxplot by adding fences at levels 0.5 IQR above and below the 50% central region, thereby enhancing the opportunity for identification of shape outliers which otherwise would not be detected as unusual and would escape special scrutiny. The multiplier 1.5 for the IQR in the classical univariate boxplot is such that for normal data the tail probability outside each fence is 0.0035, corresponding to an extreme threshold for outliers. Our multiplier 0.5 for defining the inner fence is found to work well empirically and corresponds to outer tail probabilities of 0.09 in the case of normal data. Thus, in using the double-fence functional boxplot,

- (i) *shape outliers* tend to become detected visually as curves that cross an inner fence,

- (ii) *extreme location outliers* are detected visually as curves that cross or lie beyond an outer fence, and
- (iii) *moderate location outliers* are identified visually as curves which lie entirely between the two upper fences or between the two lower fences.

Further, advantageously using this approach, curves which *deviate in both shape and location simultaneously* can become identified as such when they simultaneously satisfy both (i) and (ii). Of course, the depth for any functional boxplot must be selected *judiciously*. As seen systematically in Sections 4 and 5, the effectiveness of the functional boxplot, double-fence or otherwise, varies with the choice of depth function. Especially successful is the *spatial depth*.

Illustration. **Figure 3** illustrates the double-fence functional boxplot, using *Spatial*, with the Canadian city temperature data. Fences are shown in heavy dark boldface, the median in heavy gray boldface, and the 50% central region in light gray. Outliers, as given by curves with high spatial outlyingness, are shown with light lines. In this data set, there are no upper extreme location outliers, all upper outliers being pure shape outliers, crossing just the upper inner fence; these curves are similar to the majority in their central parts but differ in their tails, due to somewhat higher temperatures at the beginning and end of the year. In the lower part of the plot, there is a single pure extreme location outlier, detected by the lower outer fence being crossed. Several other outliers, which cross the lower inner fence as well as the lower outer fence, exhibit shape deviation (in the tails) combined with location deviation. As it should, the boxplot suppresses some detail and displays the general structure of the data set including any asymmetrical differences between the upper and lower regions.

(Could put Figure 3 here)

In conjunction with any boxplot, its general features draw attention to special points or groups of points. For example, which actually is the city with the “median” temperature curve? Also, as noted above, with this data set the most pronounced *upper* outliers are two *shape* outliers having less peakedness and with both tails relatively high, the most pronounced *lower* outlier is a *pure extreme location* outlier, and there are several lower outliers of *mixed location-shape* type. Which cities are these? For such details, we examine actual spatial outlyingness values and select particular cities, as shown in **Table 2**.

<i>City</i>	<i>Spatial outlyingness</i>
Sherbrooke	0.19 (lowest)
Thunder Bay	0.19 (lowest)
Inuvik	0.87
Vancouver	0.90
Victoria	0.90
Resolute	0.96 (highest)

Table 2: Canadian cities with temperature curves of smallest and largest spatial outlyingness.

4. Comparison of methods using actual data sets

Using seven well known actual functional data sets representing a range of types, we compare *Spatial*, *Foutliers*, *FM(trim)*, *FM(pond)*, and *MBD* with respect to performance in

outlier detection, both in the case of each data set taken in its initial form and in the case that each data set is contaminated with an additional curve strikingly different in shape from the curves in the initial data set. We use five different outlier shapes, each in five different locations. For uniformity of treatment, outlier detection for each method employs the double-fence boxplot based on its particular depth. Percentages of correct identifications are compared. We also compare methods with respect to their computation times.

4.1. Data sets, outlier types, and performance measures

4.1.1. The seven data sets

Our data sets are as follows.

- A: *Canadian City Temperatures data (CCT)*, from the *fda* package in R. Average monthly temperatures over the years 1960-1994 are shown as curves for each of $n = 35$ Canadian cities. We have used this data in Section 3 for illustrative purposes (**Figures 2-3**). For discretization, we use dimension $d = 12$ (corresponding to months).
- B: *Australian Male Mortality data (AMM)*, from the *fds* package in R. The Australian male log mortality at different ages is shown as curves for each of the $n = 103$ years 1901-2003. For discretization, we use dimension $d = 101$ (corresponding to years of age).
- C: *Hourly Electricity Loadings data (HEL)*, from <http://www.pjm.com/markets-and-operations/energy/real-time/loadhryr.aspx>. Hourly electricity loadings across a week ($d = 168$) are given for each of the $n = 732$ weeks over the years 1998-2013, for the East Pennsylvania/New Jersey/Maryland power pool. Each curve connects the 168 hourly loadings for a given week. For better viewing of structure, we use a sample of just $n = 50$ of the 732 curves.
- D: *Near-Infrared Reflectance data (NIR)*, from the *fds* package in R. Inverse log reflectance is measured in 2 nm intervals over wavelengths 1100-2500 nm, for $n = 100$ wheat samples. For discretization, we use dimension $d = 701$ (corresponding to measurements at 2 nm intervals).
- E: *Sea Surface Temperatures data (SST)*, from the *rainbow* package in R. Average monthly sea surface temperatures ($d = 12$) are given for each of the $n = 62$ years 1950-2011 in the "Nino region" of the Pacific.
- F: *Spanish City Temperatures data (SCT)*, from the *fda.usc* package in R. For $n = 73$ weather stations in Spain, daily temperatures ($d = 365$) are averaged over the years 1980-2009.
- G: *US Electricity Consumption data (USEC)*, from the *fds* package in R. The monthly electricity consumption ($d = 12$) in the U.S. residential and commercial sectors is given for each of the $n = 29$ years 1973-2001.

The data sets B-G are displayed in **Figure 4**.

(Could put Figure 4 here)

4.1.2. Outlier types

We use five types of contaminating outlier, each with five variations:

1. *Reflection of median curve about zero line*, with variations by shifting up and down.
2. *Zigzag curve alternately changing signs of coordinates of median curve*, with variations by shifting up and down.

3. *Straight line*, with variations by different slopes.
4. *Step function*, with variations by shifting up and down.
5. *Spike function*, similar to main body of curves except for a high jump over a short duration of time, with variations by shifting up and down.

These types are shown in **Figure 5**.

(Could put Figure 5 here)

4.1.3. Performance measures

Our *measures of performance* are developed as follows. First, for each method i ($1 \leq i \leq 5$), each outlier type j ($1 \leq j \leq 5$), and each data set k ($1 \leq k \leq 7$), we compute:

- $p_1(i, j, k)$ = fraction of the 5 variations of contaminating outlier type j that are detected by an *inner* fence, for data set k using method i ,
- $p_2(i, j, k)$ = fraction of the 5 variations of contaminating outlier type j that are detected by an *outer* fence, for data set k using method i ,
- $q_1(i, j, k)$ = fraction of curves in the original data that are detected as outliers by an *inner* fence, for data set k using method i in the presence of outlier j ,
- $q_2(i, j, k)$ = fraction of curves in the original data that are detected as outliers by an *outer* fence, for data set k using method i in the presence of outlier j .

From these, we produce overall measures by averaging over data sets, separately for each method and type of outlier:

- $\bar{p}_1(i, j)$ = average of $p_1(i, j, k)$ over the 7 data sets,
- $\bar{p}_2(i, j)$ = average of $p_2(i, j, k)$ over the 7 data sets,
- $\bar{q}_1(i, j)$ = average of $q_1(i, j, k)$ over the 7 data sets,
- $\bar{q}_2(i, j)$ = average of $q_2(i, j, k)$ over the 7 data sets.

The measures $\bar{p}_1(i, j)$ and $\bar{p}_2(i, j)$ concern detection of a highly unusual “true positive”. A good method i should exhibit, for *each outlier type* j , high values (near 1) of $\bar{p}_1(i, j)$ or $\bar{p}_2(i, j)$ or both.

On the other hand, the measures $\bar{q}_1(i, j)$ and $\bar{q}_2(i, j)$ concern identification of both “true positives” and “false positives”, respectively, in the original data, with the respective proportions being unknown and perhaps subjective. These measures should take relatively low values, although not necessarily equal to 0, since some low level of “positives”, both “true” and “false”, is expected in a typical data set. If these measures are too high, then favorable values of $\bar{p}_1(i, j)$ and $\bar{p}_2(i, j)$ might be due in part to permitting an excess of false positives. Although we cannot specify ideal target values for $\bar{q}_1(i, j)$ and $\bar{q}_2(i, j)$, we can compare methods for mutual consistency with respect to these measures.

In general, we view a method i as successful with respect to identification of the shape outlier of type j if the corresponding $\bar{p}_1$ is high while the corresponding $\bar{q}_2$ is relatively low. The measures $\bar{p}_2$ and $\bar{q}_1$ provide useful auxiliary information on performance.

4.2. Our findings

For each triplet (i, j, k) of (method, outlier type, data set), the associated performance measures (p_1, p_2, q_1, q_2) are given in **Table 3**. Although these measures differ across the seven data sets, this variation is not particularly substantive, and so for convenience we compare methods according to the averages $(\bar{p}_1, \bar{p}_2, \bar{q}_1, \bar{q}_2)$, also included in **Table 3**. The variation across different outliers is not consistent across methods, however, so we do not average over outlier types.

Our first observation from Table 3 is that the measures $(\bar{q}_1, \bar{q}_2)$ are nearly constant across methods, approximately equal to (25%, 5%). Thus the new method Spatial and the established methods Foutliers, FM(trim, pond), and MBD all agree on nonparametric description of these data sets when free of artificial contaminants. Here, all five methods generate very similar double-fence functional boxplots, as shown for Spatial in Figure 3 and for the others in Figure 6.

(Could put Figure 6 here)

Turning to comparison over data sets contaminated by a very unusual shape outlier, we consider the measures $(\bar{p}_1, \bar{p}_2)$, with $\bar{p}_1$ having the ideal value 100% and $\bar{p}_2$ ideally having a relatively high value but not necessarily 100%. For Outliers 1, 3, and 4, all methods perform well and quite similarly, across all data sets. On the other hand, Outlier 2 is especially challenging for Foutliers, whose performance is not competitive. Also, Outlier 5 is not only challenging for Foutliers, whose performance is not competitive, but extremely challenging for FM and MBD, which fail completely. However, Spatial (alone) performs well for all five outliers.

Of course, it is also important to take into account the computation time to obtain the measures (p_1, p_2, q_1, q_2) . These are provided in **Table 4** and show that MBD, Foutliers, and Spatial are closely competitive (in that order) with low computational times, whereas FM requires significantly more time. (Note that our computation for Spatial has been done simply using straightforward R code and is already competitive even without the benefit of a special package optimizing the computation.) For each method, most of the computational time is for ranking the curves according to the chosen outlyingness function.

O	D	Method															
		Spatial				Foutliers				FM(trim, pond)				MBD			
1	A	100	100	44	20	100	100	28	18	100	100	43	15	100	100	43	15
	B	100	100	34	14	100	100	31	11	100	100	34	17	100	100	34	17
	C	100	0	12	0	100	20	10	0	100	20	7	0	100	20	8	0
	D	100	100	23	1	100	100	20	0	100	100	25	1	100	100	24	1
	E	100	100	13	3	100	100	18	3	100	100	11	3	100	100	16	5
	F	100	100	16	1	100	100	34	1	100	100	15	1	100	100	16	1
	G	100	100	29	4	100	100	9	0	100	100	25	0	100	100	25	0
		100	86	24	6	100	89	21	5	100	89	23	5	100	89	24	6
2	A	100	100	43	11	100	100	28	18	100	100	45	17	100	100	43	15
	B	60	60	34	14	40	40	36	13	80	60	34	16	80	60	34	16
	C	100	0	12	0	0	0	10	0	80	0	7	0	80	0	8	0
	D	100	60	23	1	40	40	20	0	80	20	25	1	100	40	23	1
	E	80	20	13	4	0	0	16	3	100	40	11	3	100	40	15	4
	F	20	0	17	1	0	0	32	1	40	0	16	1	20	0	16	1
	G	100	80	29	4	80	60	12	0	100	80	25	0	100	60	25	0
		80	46	24	5	37	34	22	5	83	43	23	6	83	43	23	5
3	A	100	100	42	12	100	100	28	18	80	80	44	16	80	80	43	16
	B	100	80	34	14	80	60	30	10	100	100	34	17	100	100	34	17
	C	80	0	12	0	60	0	10	0	80	0	7	0	80	0	8	0
	D	100	100	23	1	100	80	20	0	100	100	25	1	100	100	23	1
	E	100	100	12	3	100	80	19	3	100	80	12	4	100	80	15	5
	F	100	100	16	1	100	60	34	1	100	60	16	1	100	60	15	1
	G	100	100	29	4	100	100	14	0	100	100	25	0	100	100	26	0
		97	97	24	5	91	69	22	5	94	74	23	6	94	74	23	6
4	A	100	100	43	12	100	100	29	16	100	100	44	16	100	100	43	15
	B	100	100	34	14	100	80	33	12	100	100	37	17	100	100	37	17
	C	100	80	12	0	100	80	10	0	100	60	7	0	100	80	08	0
	D	100	100	23	1	100	100	20	0	100	100	25	1	100	100	24	1
	E	100	60	13	4	100	60	16	3	100	60	13	3	100	80	16	5
	F	100	80	17	1	100	100	34	1	100	80	16	1	100	100	16	1
	G	100	100	29	2	100	100	13	0	100	100	25	0	100	100	23	0
		100	89	24	5	100	89	22	5	100	86	24	6	100	94	22	6
5	A	100	100	43	13	80	80	31	19	0	0	45	17	0	0	45	17
	B	100	80	34	14	60	60	30	10	0	0	39	15	0	0	34	17
	C	100	100	12	16	0	0	11	0	0	0	7	0	0	0	7	0
	D	100	100	23	1	0	0	20	0	0	0	25	1	0	0	25	1
	E	100	100	12	3	80	80	15	4	0	0	15	3	0	0	17	4
	F	100	100	18	1	20	20	36	1	0	0	16	1	0	0	16	1
	G	100	100	29	2	100	100	12	0	0	0	24	0	0	0	28	2
		100	97	25	6	63	49	22	5	0	0	24	5	0	0	24	6

Table 3: Performance measures (p_1 , p_2 , q_1 , q_2) as percentages, by method, outlier type ($\mathcal{O}$), and data set ($\mathcal{D}$). High p_1 and p_2 are desired. Relatively small (but not necessarily zero) q_1 and q_2 are desired. Averages ($\bar{p}_1$, $\bar{p}_2$, $\bar{q}_1$, $\bar{q}_2$) over data sets are shown in boldface.

$\mathcal{D}$	<i>Method</i>				
	<i>Spatial</i>	<i>Foutliers</i>	<i>FM(trim)</i>	<i>FM(pond)</i>	<i>MBD</i>
A	17	12	27	27	12
B	4	3	6	5	3
C	135	105	170	160	105
D	212	206	620	616	203
E	5	4	5	5	3
F	53	55	138	134	50
G	4	3	5	5	3

Table 4: Times (sec) for computing performance measures (p_1, p_2, q_1, q_2) , by method and data set ($\mathcal{D}$).

4.3. Summary discussion

Considering just the uncontaminated data sets, all methods perform effectively and very similarly. Also, for the contaminated data sets, and considering just Outliers 1, 3, and 4, all the methods again perform very effectively and comparably. On the other hand, regarding Outlier 2, Foutliers is significantly outperformed by the others, which perform well and comparably. When the spike type Outlier 5 is considered, only Spatial performs very well, with Foutliers weakly competitive, while FM and MBD fail completely. Considering all five outlier types together, Spatial alone is successful, followed by Foutliers. With only Outliers 1-4 considered, Spatial and MBD together are best, followed by Foutliers. Because FM is not competitive with respect to computation time and does not improve upon the other methods, it can be dropped from consideration.

In conclusion, for rather “ordinary” functional data sets, all five methods perform rather equivalently with respect to outlier detection. However, in the presence of rather severe contamination, only Spatial continues to perform reliably, whereas the other methods are subject to significant shortcomings, depending on the particular type of outlier.

The reason behind the failure of FM and MBD with respect to Outlier 5 is that FM and MBD include this outlier within the 50% central region, thereby distorting that region and in turn the fences for outlier identification. See **Figure 7**, which shows how a spike outlier is detected by Spatial and excluded from its 50% central region, but is undetected by MBD and included in its 50% central region.

(Could put Figure 7 here)

5. Comparison of methods using simulated data

Here the five methods are compared as in Section 4, but now using simulated data following four different contamination models representing a range of outlier types. Three models are from a study of Arribas-Gil and Romo (2014), the fourth model we introduce to cover the case of spike outliers. Each model comprises a mixture of a “main” component X of specified type and a “contamination” Y of specified type. As in Arribas-Gil and Romo (2014), we use $N = 200$ samples of size $n = 50$, with some curves in each sample from the main component and the others from the contamination. Here, for each model, we consider four contamination rates, $c = 0.05, 0.10, 0.15$ and 0.20 , with each sample consisting of $n - \lceil cn \rceil$ curves from the main component and $\lceil cn \rceil$ curves from the contamination. We adopt four performance measures similar to those of Section 4 and again represent each method by its double-fence functional boxplot as the tool for outlier detection.

5.1. Simulation models and performance measures

5.1.1. Four simulation models

Model I

Main component: $X(t) = 30t(1-t)^{3/2} + \varepsilon(t)$,

Contamination component: $Y(t) = 30t^{3/2}(1-t) + \varepsilon(t)$,

for $0 \leq t \leq 1$, where $\varepsilon(\cdot)$ is Gaussian with mean zero and covariance function $\gamma(s, t) = 0.3e^{-|s-t|/0.3}$, $0 \leq s, t \leq 1$. The contaminating curves *are of the main type with time reversed*: $Y(\cdot) \stackrel{d}{=} X(1 - \cdot)$.

Model II

Main component: $X(t) = 4t + \varepsilon(t)$,

Contamination component: $Y(t) = 4t + (-1)^u 1.8 + \frac{1}{\sqrt{2\pi 0.01}} e^{-(t-\mu)^2/0.02} + \varepsilon(t)$,

for $0 \leq t \leq 1$, where $\varepsilon(\cdot)$ is Gaussian with mean zero and covariance function $\gamma(s, t) = e^{-|s-t|}$, $0 \leq s, t \leq 1$, μ is Uniform[0.25, 0.75], and u is Bernoulli(0.5). The contaminating curves *add constants and a normal shape to the main type*.

Model III

Main component: $X(t) = 4t + \varepsilon(t)$,

Contamination: $Y(t) = 4t + 2 \sin(4(t + \mu)\pi) + \varepsilon(t)$,

for $0 \leq t \leq 1$, with $\varepsilon(\cdot)$ and μ as in Model II. The contaminating curves *are sinusoidal additions to the main type*.

Model IV

Main component: $X(t) = 4t + \varepsilon(t)$,

Contamination component: $Y(t) = 4t + \frac{1}{\sqrt{2\pi 0.0001}} e^{-(t-\mu)^2/0.0002} + 0.2\varepsilon(t)$,

for $0 \leq t \leq 1$, with $\varepsilon(\cdot)$ and μ as in Model II. The contaminating curves *add spikes to the main type*.

Samples from these models are illustrated in **Figure 8** for $n = 50$, $c = 0.05$.

(Could put Figure 8 here)

5.1.2. Performance measures

Measures of performance analogous to those of Section 4.1.3, are defined as follows. For each of $N = 200$ samples of size $n = 50$ from Model j ($j = I, \dots, IV$), at contamination rate c ($c = 0.05, 0.10, 0.15, 0.20$), double-fence boxplots are constructed according to each method i ($1 \leq i \leq 5$). Then we compute:

- $P_1(i, j, c)$ = fraction of the $N \times \lceil cn \rceil$ curves from the contamination component of Model j that are detected by *inner* fences of the boxplots according to method i ,
- $P_2(i, j, c)$ = fraction of the $N \times \lceil cn \rceil$ curves from the contamination component of Model j that are detected by *outer* fences of the boxplots according to method i ,
- $Q_1(i, j, c)$ = fraction of the $N \times (n - \lceil cn \rceil)$ curves from the main component of Model j that are detected by *inner* fences of the boxplots according to method i ,
- $Q_2(i, j, c)$ = fraction of the $N \times (n - \lceil cn \rceil)$ curves from the main component of Model j that are detected by *outer* fences of the boxplots according to method i .

The interpretations and roles of (P_1, P_2, Q_1, Q_2) are similar to those discussed for $(\bar{p}_1, \bar{p}_2, \bar{q}_1, \bar{q}_2)$ in Section 4.

5.2. Our findings

For each triplet (i, j, c) of (method, outlier model, contamination rate), the associated performance measures (P_1, P_2, Q_1, Q_2) are given in **Table 5**. Although these measures differ across the four contamination rates, the variation is sufficiently low that we may consider the averages $(\overline{P}_1, \overline{P}_2, \overline{Q}_1, \overline{Q}_2)$ over contamination rates, and we include these in **Table 5**. However, the variation across the different outlier models is not consistent across the methods, so we do not average across the models.

Regarding comparative performance for outlier detection with just the main component curves, we examine measures $(\overline{Q}_1, \overline{Q}_2)$ for each of the models. For Models I, II, and III, the values of these measures are quite consistent across methods. For Model IV, however, Spatial and Foutliers are similar but differ from FM and MBD, which agree with each other. This anomaly is explained by the fact that the presence of the spike contamination is sufficiently influential as to corrupt the performance even regarding the main component curves. Note that here we have not treated these curves as a separate sample. Except for this anomaly in the case of spike outliers, the five methods perform similarly with “ordinary data” from the main components of the models.

Regarding detection of the contaminant curves, for Models I and II all methods perform well and similarly, across all contamination rates. On the other hand, Model III is a little more challenging for Foutliers, while Model IV is extremely challenging for FM and MBD. Spatial (alone) performs well for all four contamination models.

Again we examine computation times. For computation of (P_1, P_2, Q_1, Q_2) , the times (in minutes) are 5 for MBD, 6 for Foutliers, 8 for Spatial, and 25 for FM (both trim and pond). Thus MBD, Foutliers, and Spatial are closely competitive (in that order) with low computation times, significantly outperforming FM.

$\mathcal{M}$	c	Method															
		Spatial				Foutliers				FM(trim, pond)				MBD			
I	0.05	100	61	20	0	100	66	18	0	100	54	17	0	100	57	18	0
	0.10	100	50	18	0	100	60	17	0	100	46	16	0	100	37	14	0
	0.15	100	38	15	0	100	56	14	0	100	36	13	0	99	32	13	0
	0.20	98	30	13	0	100	50	13	0	99	33	12	0	99	32	13	0
		100	45	17	0	100	58	16	0	100	42	15	0	100	40	15	0
II	0.05	98	56	23	0	90	49	19	0	97	56	23	0	98	57	32	1
	0.10	98	53	21	0	89	46	17	0	97	55	21	1	98	54	21	0
	0.15	97	52	18	0	87	43	14	0	94	51	17	0	96	52	19	0
	0.20	96	48	17	0	86	40	1	0	93	46	16	0	95	48	17	0
		98	52	20	0	88	45	13	0	95	52	19	0	97	53	22	0
III	0.05	98	13	22	0	65	8	16	0	96	12	20	0	98	13	22	0
	0.10	98	14	21	0	62	6	14	0	98	13	20	0	98	14	21	0
	0.15	96	10	19	0	64	5	12	0	95	11	17	0	96	11	18	0
	0.20	96	48	17	0	61	3	11	0	93	8	16	0	94	9	16	0
		97	21	20	0	63	5	13	0	96	11	18	0	97	12	19	0
IV	0.05	100	100	23	0	95	95	20	0	1	1	28	1	1	1	30	1
	0.10	100	100	21	0	94	94	19	0	1	1	31	1	0	0	32	1
	0.15	100	100	18	0	93	93	16	0	2	1	35	3	1	1	37	3
	0.20	100	100	17	0	91	90	17	0	3	2	39	4	2	2	40	4
		100	100	20	0	93	93	19	0	2	1	33	2	1	1	35	2

Table 5: Performance measures (P_1, P_2, Q_1, Q_2) as percentages, by method, model ($\mathcal{M}$) and contamination rate (c). High P_1 and P_2 are desired. Relatively small (but not necessarily zero) Q_1 and Q_2 are desired. Averages $(\overline{P}_1, \overline{P}_2, \overline{Q}_1, \overline{Q}_2)$ over contamination rates are shown in boldface.

5.3. Summary discussion

For detection of contaminating outliers under Model I, all methods perform very well and similarly, with Foutliers slightly ahead. For Model II, again all methods perform very well and similarly, but now with Foutliers slightly behind. For Model III, the Spatial, FM, and MBD methods perform very well and similarly, with Foutliers only weakly competitive. Finally, for Model IV, Spatial and Foutliers both succeed, with Foutliers slightly behind, while FM and MBD fail completely. Since FM has significantly poorer computation time and does not improve upon any other method, it may be dropped, leaving Foutliers, MBD, and Spatial for consideration. Then, if only Models I, II, and III are considered, Spatial and MBD perform equally well and slightly outclass Foutliers. However, regarding Model IV, MBD fails completely (as does FM), while Spatial succeeds perfectly and Foutliers is strongly competitive. Overall, therefore, on the basis of all four models together for outlier detection using double-fence functional boxplots, Spatial outperforms the others, with Foutliers second-best.

Of course, as noted earlier, if also the epigram is used with MBD, then it too is successful with spike outliers. However, we emphasize that the functional boxplot is a basic and desirable tool in exploratory description of functional data, and by design one of its roles is outlier detection. Thus the performance of the boxplot in this regard is of central importance and interest.

As with the actual data sets, we also compare the methods across the four models without presence of the artificial contamination curves. In this case, for just the “ordinary” data from the main components of Models I-IV, it appears that all the methods perform effectively and very similarly.

6. Concluding discussion

In treating depth-based approaches to nonparametric description of functional data, we have contributed methods for displaying the median and other quantile curves, for providing confidence bands for associated *population* quantile curves, and for enhancement of the functional boxplot to flag both spatial and location outliers. Also, we have examined the merits of the “Spatial” approach. In Section 6.1 we provide added perspective on the double fence functional boxplot, and in Section 6.2 we develop a general recommendation.

6.1. Perspectives on the double-fence functional boxplot

The classical boxplot is designed to accomplish identification and display of the “center” of the data, the “middle half” of the data, and “outliers” in the data. Specifically, the boxplot incorporates a) the median, b) the box corresponding to the 50% central region, and c) lower and upper “fences” situated above and below the box by a distance equal to 1.5 times the spread of the box as measured by the IQR. See Wang and Serfling (2015) for discussion of how the boxplot is an *ingenious hybrid*, using *quantiles* to determine the middle half but *scaled-deviation outlyingness* to set thresholds for identifying outliers. Because of its appeal and strength for such purposes, the boxplot has enjoyed wide application with univariate data and is equally desirable in the functional data setting.

Maintaining the feature that location of fences is based on spread of the middle half, Rousseeuw, Ruts, and Tukey (1999), using halfspace depth, extend the univariate boxplot to a “bagplot” for bivariate data, and this in turn is extended by Hyndman and Shang (2010) to functional data using the first two functional PCs. On the other hand, using MBD, Sun and Genton (2011) provide a very natural and visually appealing extension of the univariate boxplot to a “functional boxplot”, although they also recommend some alternatives for situations when the weakness of MBD for shape outliers of spike type might be of concern.

It is very convenient, however, if the functional boxplot can stand alone as a reliable tool, just as its classical antecedant. For this, it is essential that the chosen depth function *not* classify any shape outlier embedded in the main body of the data as belonging to the boxplot’s middle half. Otherwise, there are two serious consequences: (i) distortion and misrepresentation of the middle half, and (ii) accompanying distortion and mislocation of the fences intended for outlier identification. Thus, *in the functional data setting, the choice of depth function for the boxplot is of paramount importance*. With an effective choice of depth function, (i) can be avoided and hence also (ii). Specifically, the outlyingness value attached to any shape outlier must be sufficiently high as to classify it as *outside the middle half* (whether or not identified as an “outlier”). While this is accomplished by simply ranking the curves using the depth function, the boxplot is a visual device and employs fences to display the whereabouts of outliers. For this purpose, in order that the boxplot visually flag not only location outliers but also shape outliers, we have introduced the use of additional fences above and below the middle half, with factors 0.5 instead of 1.5.

As seen in our findings of Sections 4 and 5, this double-fence functional boxplot indeed works in practice and enables the boxplot to be used as a stand-alone device even with functional data. Of course, this resolution carries with it the usual proviso with outlier identification that points flagged as outliers must be individually screened for confirmation. As shown in Figure 7, using MBD the functional boxplot fails at (i) and hence also at (ii), and a similar result can be shown for FM. On the other hand, as we have demonstrated with a range of both actual and simulated data sets, *Spatial* presents no such concerns and, very conveniently, enables the double-fence boxplot to serve as a reliable stand-alone tool for nonparametric description of a functional data set.

6.2. Summary recommendations

- If inference on population quantiles is desired, *Spatial offers confidence bands*.
- For outlier detection,
 - If the data set is free of extremely anomalous shape outliers, then *all of Spatial, Foutliers, FM, and MBD perform similarly*.
 - If the data set includes contamination by strikingly different shape outliers, then
 - * If spike outliers are present, *only Spatial succeeds across diverse data sets*.
 - * In all other shape outlier scenarios, *MBD and Spatial are preferred*.
- The double-fence version of functional boxplot seems to have useful application, and we particularly recommend the *double-fence functional boxplot with Spatial*.

All in all, for quantile-based nonparametric description of functional data, *Spatial offers a very competitive and appealing combination of robustness, efficiency, computational ease, and versatility*.

Acknowledgments

Very constructive comments by two reviewers and the Associate Editor are appreciated greatly and have led to substantial improvements. Also, the first author thanks G. L. Thompson for valued encouragement. Finally, support under National Science Foundation Grant DMS-1106691 is sincerely acknowledged.

References

- [1] Arribas-Gil, A. and Romo, J. (2014). Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics* **15** 603–619.
- [2] Cardot, H., Cenac, P. and Zitt, P.-A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli* **19** 18–43.
- [3] Cardot, H., Chaouch, M., Goga, C. and Labruère, C. (2010). Properties of design-based functional principal components analysis. *Journal of Statistical Planning and Inference* **140** 75–91.
- [4] Cardot, H., Goga, C. and Lardin, P. (2013). Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data. *Electronic Journal of Statistics* **7** 562–596.
- [5] Cardot, H. and Josserand, E. (2011). Horvitz-Thompson estimators for functional data: asymptotic confidence bands and optimal allocation for stratified sampling. *Biometrika* **98** 107–118.
- [6] Chaouch, M. and Goga, C. (2012). Using complex surveys to estimate the L1-median of a functional variable: application to electricity load curves. *International Statistical Review* **80** 40–59.
- [7] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association* **91** 862–872.
- [8] Claeskens, G., Hubert, M., Slaets, L. and Vakili, K. (2014). Multivariate functional halfspace depth. *Journal of the American Statistical Association* **109** 411–423.
- [9] Cuesta-Albertos, J. A. and Nieto-Reyes, A. (2008). The random Tukey depth. *Computational Statistics and Data Analysis* **52** 4979–4988.
- [10] Cuevas, A., Febrero, M. and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth functions. *Computational Statistics* **22** 481–496.
- [11] Febrero, M., Galeano, P. and Gonzalez-Manteiga, W (2008). Outlier detection on functional data by depth measures, with application to identify abnormal NO_x levels. *Environmetrics* **19** 331–345.
- [12] Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York.
- [13] Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *Test* **10** 419–440.
- [14] Fraiman, R. and Zvarc, M. (2013). Resistant estimates for high dimensional and functional data based on random projections. *Computational Statistics and Data Analysis* **58** 326–338.
- [15] Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika* **95** 587–600.
- [16] Hyndman, R. J. and Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data (2010). *Journal of Computational and Graphical Statistics* **19** 29–45.

- [17] López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data (2009). *Journal of the American Statistical Association* **104** 718–734.
- [18] Nieto-Reyes, A. and Battey, H. (2015). A topologically valid definition of depth for functional data. Preprint.
- [19] Oja, H. (2010). *Multivariate Nonparametric Methods with R: An approach based on spatial signs and ranks*. Springer.
- [20] Rousseeuw, P. and Leroy, A. (1987). *Robust Regression and Outlier Detection*. Wiley.
- [21] Rousseeuw, P., Ruts, I. and Tukey, J. W. (1999). The bagplot: a bivariate boxplot. *The American Statistician* **53** 382–387.
- [22] Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, New York.
- [29] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.
- [24] Serfling, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles. *Journal of Statistical Planning and Inference* **123** 259–278.
- [25] Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization. *Journal of Nonparametric Statistics* **22** 915–936.
- [26] Sguera, C., Galeano, P. and Lillo, R. (2014). Spatial depth-based classification for functional data. *Test* **23** 725–750.
- [27] Sun Y. and Genton M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics* **20** 316–334.
- [28] Wang, S. and Serfling, R. (2015). On masking and swamping robustness of leading nonparametric outlier identifiers for univariate data. *Journal of Statistical Planning and Inference* 162 62–74.
- [29] Zhu, H. T., Fan, J. Q., and Kong, L. L. (2014). Spatially varying coefficient models with applications in neuroimaging data with jumping discontinuity. *Journal of the American Statistical Association* **109** 977–990.

Figure 1: Illustrating three upper directions with the same $\|\mathbf{u}\|$ and with $d = 12$, and, for a given data set of curves (shown in light gray), the corresponding sample upper quantiles. **Upper**: Upper canonical direction and two random upper directions. **Lower**: Corresponding sample upper quantiles.

Figure 2: Illustration of sample quantiles and confidence bands. **Upper left:** Annual temperature curves, averaged over 1960 to 1994, for 35 Canadian cities. **Upper right:** Sample median curve and sample 95% upper and lower quantile curves, Canadian city temperature curves. **Middle:** Generic illustration of conversion of points on ellipsoid in $\mathbb{R}^d$ to curves in $\mathcal{H}$. The center $Q(u, \mathbb{X}_n)$ of the 95% confidence ellipsoid is reconstructed in $\mathcal{H}$ as a curve estimating the target quantile curve $Q(u)$ in $\mathcal{H}$. It is enclosed by a band of curves in $\mathcal{H}$ generated by the boundary points of the ellipsoid. **Lower:** 95% confidence bands, respectively, for the “population” median, 95% upper quantile, and 95% lower quantile, Canadian city temperature curves.

Figure 3: Illustration of double-fence functional boxplot. **Upper:** Canadian city temperature data. **Lower:** Corresponding double-fence functional boxplot using Spatial. Several upper shape outliers are identified but an absence of upper location outliers is indicated. Several lower shape outliers, some of which are also location outliers, are identified, as well as a lower pure location outlier.

Figure 4: Data sets AMM (top left), HEL (top right), NIR (middle left), SST (middle right), SCT (bottom left), and USEC (bottom right).

Figure 5: Outlier types: 1) Reflection of median (top left), 2) Zigzag (top right), 3) Straight line (middle), 4) Step function (lower left), and 5) Spike (lower right).

Figure 6: Double-fence functional boxplots of Canadian city temperature data with no contamination added, using FM(trim) (upper left), FM(pond) (upper right), Foutliers (lower left), and MBD (lower right).

Figure 7: Double-fence functional boxplots of Canadian city temperature data with spike outlier added, using Spatial (top) and MBD (bottom). Note that MBD classifies the spike outlier within its 50% central region, distorting that region and consequently also the associated fences.

Figure 8: Illustration of the four simulation models. **Model I** (upper left): the contamination is of the main type with time reversed. **Model II** (upper right): the contamination adds a constant and a normal shape to the main type. **Model III** (lower left): the contamination is a sinusoidal addition to the main type. **Model IV** (lower right): the contamination adds a spike to the main type. Each plot displays a sample of size 50, with 47 curves (light) from the main component and 3 curves (dark) from the contamination component.