



*J. R. Statist. Soc. B* (2016)  
78, Part 1, pp. 77–97

## A tilting approach to ranking influence

Marc G. Genton

*King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*

and Peter Hall

*University of Melbourne, Australia, and University of California at Davis, USA*

[Received October 2013. Final revision October 2014]

**Summary.** We suggest a new approach, which is applicable for general statistics computed from random samples of univariate or vector-valued or functional data, to assessing the influence that individual data have on the value of a statistic, and to ranking the data in terms of that influence. Our method is based on, first, perturbing the value of the statistic by ‘tilting’, or reweighting, each data value, where the total amount of tilt is constrained to be the least possible, subject to achieving a given small perturbation of the statistic, and, then, taking the ranking of the influence of data values to be that which corresponds to ranking the changes in data weights. It is shown, both theoretically and numerically, that this ranking does not depend on the size of the perturbation, provided that the perturbation is sufficiently small. That simple result leads directly to an elegant geometric interpretation of the ranks; they are the ranks of the lengths of projections of the weights onto a ‘line’ determined by the first empirical principal component function in a generalized measure of covariance. To illustrate the generality of the method we introduce and explore it in the case of functional data, where (for example) it leads to generalized boxplots. The method has the advantage of providing an interpretable ranking that depends on the statistic under consideration. For example, the ranking of data, in terms of their influence on the value of a statistic, is different for a measure of location and for a measure of scale. This is as it should be; a ranking of data in terms of their influence should depend on the manner in which the data are used. Additionally, the ranking recognizes, rather than ignores, sign, and in particular can identify left- and right-hand ‘tails’ of the distribution of a random function or vector.

**Keywords:** Band depth; Data weights; Functional boxplot; Functional data; Image data; Outlier; Robustness

### 1. Introduction

The analysis of functional data has received sustained attention in recent years due to the increasing availability and collection of observations that arise as functions or images. Data of this type arise in many disciplines and include growth curves in biology, climate variables from networks of monitors or from climate model outputs in geosciences and profiles derived by monitoring manufacturing processes, to name a few; see Ramsay and Silverman (2005), Ferraty and Vieu (2006), Ramsay *et al.* (2009) and references therein.

Classical statistical methods based on ranks can be extended to the functional setting. For this, López-Pintado and Romo (2009) introduced the notion of band depth. It yields an ordering of a sample of functional data from the centre outwards and therefore defines a measure

*Address for correspondence:* Marc G. Genton, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.  
E-mail: marc.genton@kaust.edu.sa

of the centrality or outlyingness of an observation. For instance, the median function, or a trimmed mean function, can be defined for robust statistical analysis, and functional boxplots (Sun and Genton, 2011, 2012a) or surface boxplots (Genton *et al.*, 2014) can be constructed for visualization and outlier detection. Sun and Genton (2012b) proposed a functional median polish algorithm, based on using a band depth functional median, to fit robustly a functional analysis-of-variance model. Sun *et al.* (2012) derived a fast algorithm to compute band depth for functional data and López-Pintado *et al.* (2014) introduced simplicial band depth for multivariate functional data. Yu *et al.* (2012) proposed a test for functional outlier detection, founded on functional principal components. Hyndman and Shang (2010) applied functional principal component analysis to a robust covariance matrix to detect outliers. The last two approaches have the drawback of being basis dependent. Genton and Ruiz-Gazen (2010) introduced another approach to defining and visualizing influential observations in dependent data based on additive data perturbations.

In this paper we suggest a new, widely applicable approach to ranking influential data, based on data tilting. Relatively to previously considered methods, this technique has the advantages of being basis independent, of ranking data according to their influence on specific statistics (for example our rankings in the context of location estimation can be quite different from those for scale estimation), of allowing the identification of ‘tails’ in distributions of random functions or vectors and of being applicable in an exceptionally wide range of settings. For instance, the method can be employed to identify influential data in random samples, no matter whether those data are univariate or vector valued or function valued, for many types of estimator. In the case of functional data, the functions can have almost arbitrary arguments, of any dimension, and the statistics being considered can involve tuning parameters, in which case the rankings reflect those parameters. For brevity our notation and our examples will address the functional data case.

Data tilting involves replacing uniform weights on the data by more general weights. It can be employed to render parametric procedures more robust, or to produce a natural ordering of the data in terms of their contributions to the fit of a model (Choi *et al.*, 2000). The use of tilting for assessing robustness, or the influence of particular data values, has been investigated by Hall and Presnell (1999a), Critchley and Marriott (2004), Lazar (2005) and Camponovo and Otsu (2012), among others. Many applications of tilting and perturbation arguments to contemporary, non-parametric statistical problems have been developed, including those of Hall and Presnell (1999b), Critchley *et al.* (2001), Hall and Yao (2003), Bravo (2005), Hall *et al.* (2009), Xu and Phillips (2011) and the vast literature on empirical likelihood.

In this paper we use tilting to assign a rank to each observation, e.g. each data function. We show that this ranking admits a simple, general interpretation in terms of projections of functions of the data onto the space that is spanned by an empirical eigenfunction corresponding to the largest eigenvalue, where the linear transformation that determines the eigenvalue and eigenfunction is determined by the statistic under investigation. This interpretation is not asymptotic in terms of sample size and so is available, and relevant, even for small sample sizes.

The remainder of this paper is organized as follows. In Section 2 we give a general definition of ranking based on tilting, and we introduce our methodology. We also outline theory that is associated with this approach. The methodology and its numerical properties are illustrated through a real data example in Section 3. The proof of our main result is relegated to Appendix A. The data and code can be obtained from [stsda.kaust.edu.sa/Pages/Software.aspx](http://stsda.kaust.edu.sa/Pages/Software.aspx).

## 2. Methodology

### 2.1. Examples of tilting

Assume that we observe data  $X_1, \dots, X_n$ , supported in a region  $\mathcal{R}$ . For definiteness we take the data to be functions, but it will be appreciated that our methodology is available more generally. The data  $X_i$ s may be paired with other data  $Y_i$ , e.g. response variables in a regression problem, and in this case the pairs  $(X_i, Y_i)$  are assumed to be distributed as  $(X, Y)$ . In the present subsection we shall discuss tilted versions of four statistics  $\hat{\omega}$ , which are defined in equations (1), (2), (4) and (5). There,  $W$  is a kernel function,  $h$  a bandwidth,  $t \in \mathcal{R}$  in equations (1) and (4),  $t = (t_1, t_2)$  in equation (2) is a vector in  $\mathcal{R} \times \mathcal{R}$ ,  $v = (t, u) \in \mathcal{R} \times \mathbb{R}$  in equation (5) and the quantities  $\hat{\theta}_j$  and  $\hat{\phi}_j$  will be defined and discussed below equation (6):

$$\hat{\omega}(t) = \bar{X}(t) = \frac{1}{n} \sum_{i=1}^n X_i(t), \quad (1)$$

$$\hat{\omega}(t) = \hat{K}(t_1, t_2) \equiv \frac{1}{n} \sum_{i=1}^n \{X_i(t_1) - \bar{X}(t_1)\} \{X_i(t_2) - \bar{X}(t_2)\} \quad (2)$$

$$= \sum_{j=1}^{\infty} \hat{\theta}_j \hat{\phi}_j(t_1) \hat{\phi}_j(t_2), \quad (3)$$

$$\hat{\omega}(t) = \hat{\phi}_j(t), \quad (4)$$

$$\hat{\omega}(v) = \left[ \sum_{i=1}^n Y_i W \left\{ \frac{X_i(t) - u}{h} \right\} \right] / \left[ \sum_{i=1}^n W \left\{ \frac{X_i(t) - u}{h} \right\} \right]. \quad (5)$$

We consider  $\hat{\omega}$  to be an estimator of a quantity  $\omega$ . In equation (1),  $\hat{\omega}(t)$  is the empirical mean of the functions  $X_i$ , evaluated at  $t$ , and estimates  $\omega(t) = E\{X(t)\}$ , and  $t \in \mathcal{R}$ . In equation (2),  $\hat{\omega}(t)$  denotes an estimator of the covariance function,  $\omega = K$ , defined by

$$K(t_1, t_2) = \text{cov}\{X(t_1), X(t_2)\} = \sum_{j=1}^{\infty} \theta_j \phi_j(t_1) \phi_j(t_2). \quad (6)$$

The singular value decompositions at equations (3) and (6) are the versions, for functional data, of respectively empirical and theoretical principal component covariance expansions expressed in terms of (eigenvalue, eigenfunction) pairs  $(\hat{\theta}_j, \hat{\phi}_j)$  and  $(\theta_j, \phi_j)$ , and ordered so that  $\hat{\theta}_1 \geq \hat{\theta}_2 \geq \dots$  and  $\theta_1 \geq \theta_2 \geq \dots$ . In equation (4),  $\omega = \phi_j$ . In equation (5),  $\hat{\omega}(v)$  is an estimator of the conditional mean,  $\omega(v) = E\{Y|X(t) = u\}$ . Result (6) holds, with convergence in mean square, provided that  $\int_{\mathcal{R}} E\{X(t)^2\} dt < \infty$ , and similarly equation (3) holds, with convergence in the same sense, as long as each data function is square integrable.

Let  $p = (p_1, \dots, p_n)$  denote an  $n$ -vector whose components are non-negative and satisfy  $\sum_i p_i = 1$ . (Such a vector is often referred to in the literature as a ‘multinomial distribution’.) Let  $\hat{\omega}_p$  denote the version of  $\hat{\omega}$  when each data value  $X_i$ , or the pair  $(X_i, Y_i)$ , is weighted by the respective value of  $p_i$ , rather than by  $n^{-1}$ . We say that  $X_i$  has been tilted using  $p$ . In the examples at equations (1), (2), (4) and (5) we have respectively

$$\hat{\omega}_p(t) = \bar{X}_p(t) = \sum_{i=1}^n p_i X_i(t), \quad (7)$$

$$\begin{aligned} \hat{\omega}_p(t) &= \hat{K}_p(t_1, t_2) = \sum_{i=1}^n p_i \{X_i(t_1) - \bar{X}_p(t_1)\} \{X_i(t_2) - \bar{X}_p(t_2)\} \\ &= \sum_{j=1}^{\infty} \hat{\theta}_{pj} \hat{\phi}_{pj}(t_1) \hat{\phi}_{pj}(t_2), \end{aligned} \quad (8)$$

$$\hat{\omega}_p(t) = \hat{\phi}_{pj}(t),$$

$$\hat{\omega}_p(v) = \left[ \sum_{i=1}^n p_i Y_i W \left\{ \frac{X_i(t) - u}{h} \right\} \right] / \left[ \sum_{i=1}^n p_i W \left\{ \frac{X_i(t) - u}{h} \right\} \right]. \tag{9}$$

If we write

$$p_0 = (n^{-1}, \dots, n^{-1}) \tag{10}$$

for the uniform distribution on  $n$  points, then the quantities  $\hat{\omega}$ , defined at equations (1), (2), (4) and (5), are denoted equivalently by  $\hat{\omega}_{p_0}$ . The  $L_2$ -distance between the tilted and non-tilted forms of  $\hat{\omega}$  is

$$d_1(p) = \int_{\mathcal{J}} \{\hat{\omega}_p(v) - \hat{\omega}(v)\}^2 dv = \int_{\mathcal{J}} \{\hat{\omega}_p(v) - \hat{\omega}_{p_0}(v)\}^2 dv, \tag{11}$$

where  $\mathcal{J}$  is a suitable set; for example, it would be taken equal to  $\mathcal{R}$  in the first and third examples, to  $\mathcal{R} \times \mathcal{R}$  in the second example and to a subset of  $\mathcal{R} \times \mathbb{R}$  in the fourth.

### 2.2. Methodology based on tilting

Let  $p = (p_1, \dots, p_n)$  represent a general multinomial distribution on  $n$  points, introduced in Section 2.1; let  $p_0$  be the uniform probability distribution, defined at equation (10); let  $d_1(p)$ , at equation (11), denote the distance between the tilted and untilted forms of the statistic  $\hat{\omega}$ ; and measure the distance between  $p$  and  $p_0$  by using the criterion

$$d_2(p) = \sum_{i=1}^n p_i \log(n p_i). \tag{12}$$

Given  $\varepsilon > 0$ , choose  $p$  to minimize  $d_2(p)$  subject to

$$d_1(p) = \varepsilon^2. \tag{13}$$

Minimizing  $d_2(p)$  is equivalent to maximizing a measure of entropy, or equivalent to minimizing a measure of Kullback–Leibler divergence. The resulting value of  $p$ , which is denoted by  $(p_1(\varepsilon), \dots, p_n(\varepsilon))$ , is our tilted form of the discrete uniform distribution. We rank the values of  $p_i(\varepsilon)$ , obtaining

$$p_{\hat{i}_1}(\varepsilon) \leq \dots \leq p_{\hat{i}_n}(\varepsilon), \tag{14}$$

say. This induces an ordering of the data set, and we express that ordering as  $\hat{i}_1(\varepsilon) \geq \dots \geq \hat{i}_n(\varepsilon)$ .

The distance measure at equation (12) is one of the power divergence distances (see for example Cressie and Read (1984)) and has advantages over other contenders in that it allows one or more of the  $p_i$ s to equal 0 and leads automatically to non-negative  $p_i$ s without requiring an algorithm such as quadratic programming to ensure that property.

Our ranking of influence recognizes, rather than ignores, sign. In particular, if  $\hat{\omega}$  is a mean then at one end of the sequence  $\hat{i}_1, \dots, \hat{i}_n$  the indices identify data values that are having a relatively large positive influence, whereas at the other end they are having a relatively large negative influence. This issue motivates our preference for the ranking  $\hat{i}_1 \geq \dots \geq \hat{i}_n$  rather than  $\hat{i}_n \geq \dots \geq \hat{i}_1$ , because the former acknowledges the positive role that is played by  $X_{\hat{i}_1}$ , and the negative role of  $X_{\hat{i}_n}$ . This is transparent in the case of a scalar mean, and similar arguments can be given in the case of function-valued data and other statistics  $\hat{\omega}$ , where our approach allows identification of left- and right-hand ‘tails’ of general distributions. See, for example, the discussion in Section 3.2.

If the distribution of the data is continuous then, in the limit as  $\varepsilon \downarrow 0$ , the inequality signs in ranking (14) are strict with probability 1, i.e., for each strictly positive, deterministic choice of  $\varepsilon$ ,  $P\{p_{\hat{i}_1}(\varepsilon) < \dots < p_{\hat{i}_n}(\varepsilon)\} = 1$ . Therefore the ordering is strict; rather than  $\hat{i}_1(\varepsilon) \succeq \dots \succeq \hat{i}_n(\varepsilon)$ , we have

$$\hat{i}_1(\varepsilon) \succ \dots \succ \hat{i}_n(\varepsilon). \tag{15}$$

Moreover, as we shall argue in Section 2.3, the strict ranking in expression (15) remains valid in the limit as  $\varepsilon \downarrow 0$ .

Given that we have tilted  $p_0$  to  $p = p(\varepsilon) = (p_1(\varepsilon), \dots, p_n(\varepsilon))$ , we rank the values  $p_1(\varepsilon), \dots, p_n(\varepsilon)$ , obtaining the inequalities at ranking (14). We assert that  $X_{\hat{i}_1}, \dots, X_{\hat{i}_n}$  is an ordering of  $X_1, \dots, X_n$  in decreasing order of their immediate influence on the value of  $\hat{\omega}$ , respecting the sign of influence, and we express the corresponding ordering of the indices  $1, \dots, n$  as  $\hat{i}_1(\varepsilon) \succeq \dots \succeq \hat{i}_n(\varepsilon)$ .

Potentially, the application of these ideas is hindered by the fact that  $\hat{i}_1(\varepsilon), \dots, \hat{i}_n(\varepsilon)$  depend on  $\varepsilon$ , since an appropriate choice of  $\varepsilon$  may require delicate empirical methods. However, under mild assumptions, the quantities  $\hat{i}_j(\varepsilon)$  have proper, well-defined limits as  $\varepsilon \downarrow 0$ . Moreover, the process of convergence to the limit here has several statistically attractive properties.

- (a) The entire sequence  $\hat{i}_1(\varepsilon), \dots, \hat{i}_n(\varepsilon)$  has a well-defined joint limit as  $\varepsilon \downarrow 0$ .
- (b) The ordering of the limiting values  $\hat{i}_1(0), \dots, \hat{i}_n(0)$  has attractive, statistically interpretable features, expressible in terms of generalized principal components and which we discuss in the next section.
- (c) If the distribution of the random function  $X$  is continuous then, with probability 1, there are no ties in the limiting ordering.
- (d) In the setting of (c), the ordering is identical to  $\hat{i}_1(\varepsilon) \succeq \dots \succeq \hat{i}_n(\varepsilon)$  for all sufficiently small  $\varepsilon$ , including  $\varepsilon = 0$ , and so is straightforward to identify simply by considering successively smaller values of  $\varepsilon$ .

Using the ideas in this paper as a foundation, various directions might be pursued in the future. First, breakdown points could be considered, although they are arguably beyond the scope of the present paper, not least because we are developing a general approach to ranking for general statistics for general types of data. In comparison, the study of breakdown points usually demands a much narrower focus. For example, in the case of location estimation it is often necessary to assume that the sampled distribution is symmetric about its centre, whereas quite different models are needed for estimators of other quantities. Secondly, one might try to measure and compare the ‘outlyingness’ of outlying data. For that problem we can offer at present no better suggestion than to compute  $\hat{\omega}$  with the outlying value removed, and to determine in that way just how much leverage the value removed has. Thirdly, motivated by the theoretical results that we shall discuss in Section 2.3, one could estimate higher order principal component functions that correspond to a general statistic  $\hat{\omega}$  and use that analysis to gain insight into higher order aspects of rankings of influence. However, this approach is likely to be challenged by the sheer complexity of those high order features.

### 2.3. Limit of the ranking as $\varepsilon \downarrow 0$

First we state our assumptions. We assume that, as  $\varepsilon \downarrow 0$ , we can write

$$\hat{\omega}_p - \hat{\omega} = \sum_{i=1}^n (p_i - n^{-1}) \Delta_i + O(\varepsilon^2) \tag{16}$$

and

$$\frac{1}{2} \frac{\partial}{\partial p_j} \int_{\mathcal{R}} (\hat{\omega}_p - \hat{\omega})^2 = \sum_{i=1}^n (p_i - n^{-1}) \int_{\mathcal{R}} \Delta_i \Delta_j + A_j + O(\varepsilon^3), \tag{17}$$

where the integral is over the support  $\mathcal{R}$  of  $\hat{\omega}_p$  and  $\hat{\omega}$ , which is assumed to be compact, and  $\Delta_i$  denotes an approximation to the first derivative of  $\hat{\omega}_p$  with respect to  $p_i$  (the approximation is accurate up to a remainder of order  $\varepsilon^2$ ). We ask also that, as  $\varepsilon \downarrow 0$ ,

$$\sum_{i=1}^n \Delta_i = \sum_{i=1}^n A_i = 0, \quad \Delta_i = O(1) \text{ and } A_i = O(\varepsilon^2) \text{ for each } i. \quad (18)$$

The quantities  $\Delta_i$  may depend on  $p$ , although only negligibly so; they will satisfy

$$\Delta_i = \Delta_i^0 + O(\varepsilon) \quad \text{for each } i \quad (19)$$

as  $\varepsilon \downarrow 0$ , where  $\Delta_i^0$  depends only on the data.

To convey intuition about these assumptions we mention that result (16) asserts that, to first order,  $\hat{\omega}_p - \hat{\omega}$  is linear in perturbations  $p_i - n^{-1}$  of the data weights, with a quadratic remainder, and equation (17) essentially follows from equation (16) on squaring both sides and integrating; see the discussion three paragraphs below. The first part of expression (18) reflects the fact that the  $p_i$ s are related by the formula  $\sum_i p_i = 1$  and in particular are not independent variables; and the second and third parts of expression (18) stipulate that the first, second and third terms on the right-hand side of equation (17) are of orders  $\varepsilon$ ,  $\varepsilon^2$  and  $\varepsilon^3$  respectively. Condition (19) asserts only that  $\Delta_i$ , which as equation (18) states is of order 1, does not depend on  $\varepsilon$  to first order. These assumptions are satisfied widely, as will be shown through examples later in this subsection.

Define

$$\hat{M}(t_1, t_2) = \frac{1}{n} \sum_{i=1}^n \Delta_i(t_1) \Delta_i(t_2), \quad (20)$$

for  $t_1, t_2 \in \mathcal{R}$ . We also write  $\hat{M}$  for the operator itself: if  $\chi$  is a function then

$$(\hat{M}\chi)(t_1) = \int_{\mathcal{R}} \hat{M}(t_1, t_2) \chi(t_2) dt_2.$$

Although  $\hat{M}$  may depend on  $p$ , through the quantities  $\Delta_i$ , in view of condition (19) this dependence becomes negligible as  $\varepsilon \downarrow 0$ . Let  $\hat{M}^0$  denote the limit of  $\hat{M}$  as  $\varepsilon$  decreases. Then we can write

$$\hat{M}^0(t_1, t_2) = \frac{1}{n} \sum_{i=1}^n \Delta_i^0(t_1) \Delta_i^0(t_2) = \sum_{j=1}^{\infty} \hat{\alpha}_j \hat{\beta}_j(t_1) \hat{\beta}_j(t_2), \quad (21)$$

representing a conventional singular value decomposition of the positive definite operator with kernel  $\hat{M}^0$ , having (eigenvalue, eigenfunction) pairs denoted by  $(\hat{\alpha}_j, \hat{\beta}_j)$ . It is assumed that the terms in the second series in definition (20) are ordered so that  $\hat{\alpha}_1 \geq \hat{\alpha}_2 \geq \dots$

There is, of course, a simple relationship between equations (16) and (17). Indeed, since  $\Delta_i$  equals the first derivative of  $\hat{\omega}_p - \hat{\omega}$  with respect to  $p_i$ , up to a remainder of order  $\varepsilon^2$ , and since, by result (16),  $\hat{\omega}_p - \hat{\omega}$  equals  $\sum_i (p_i - n^{-1}) \Delta_i$ , up to a remainder of the same order, then, differentiating under the integral sign, it is clear that, up to a remainder of order  $\varepsilon^2$ , half the value of the first derivative of  $\int (\hat{\omega}_p - \hat{\omega})^2$ , with respect to  $p_j$ , should equal

$$\int \left\{ \sum_{i=1}^n (p_i - n^{-1}) \Delta_i \right\} \Delta_j.$$

This is just the first term on the right-hand side of equation (17). The second term there, i.e.  $A_j$ , represents the dominant contributions of order  $\varepsilon^2$ . Therefore equation (17) simply confirms

the influence of equation (16) on the derivative of  $\int (\hat{\omega}_p - \hat{\omega})^2$ . The term  $A_j$  vanishes in many important examples.

The orders of magnitude of remainder terms, above and in the work below, and also in the proof of theorem 1 in Appendix A, are to be interpreted as holding as  $\varepsilon \downarrow 0$  for fixed  $n$ , in fact for a fixed sample. They are not asymptotic in  $n$ .

To illustrate these assumptions we consider instances where  $\hat{\omega}$  is defined by equations (1) or (2). In the first of these cases,  $\Delta_i = X_i - \bar{X}$  and  $A_i \equiv 0$ , where these random functions are from  $\mathcal{R}$  to the real line. Here  $\Delta_i^0$ , in condition (19), is identical to  $\Delta_i$ . In the case of equation (2),

$$\Delta_i(t_1, t_2) = \delta_{2i}(t_1, t_2) - \delta_{1i}(t_1)\xi(t_2) - \delta_{1i}(t_2)\xi(t_1), \tag{22}$$

and again  $A_i \equiv 0$ , where  $t_1, t_2 \in \mathcal{R}$ ,  $\delta_{1i} = X_i - \bar{X}$ ,  $\delta_{2i} = Z_i - \hat{K}$ ,  $\xi = \bar{X}_p - \bar{X}$  and

$$Z_i(t_1, t_2) = \{X_i(t_1) - \bar{X}(t_1)\}\{X_i(t_2) - \bar{X}(t_2)\}.$$

Although the quantities  $\Delta_i$ , in equation (22), depend on  $p$ , they enjoy property (19) with  $\Delta_i^0 = \delta_{2i}$ .

Likewise, the non-parametric regression example at equation (5) satisfies conditions (16)–(18). This indicates that the usefulness of our approach to ranking the influence of data extends well beyond the context of functional data analysis, where inference is typically semiparametric. Indeed, the example at equation (5) can be viewed as one of non-parametric regression, particularly if we fix  $t$  and take the region  $\mathcal{R}$  to be a subset of the real line, reflecting only the values that are taken by  $u$  in equation (5), rather than to be a subset of  $\mathcal{I} \times \mathbb{R}$  (if the random functions are defined on  $\mathcal{I}$ ). Theorem 1 below is valid in settings where the data are not just functions  $X_1, \dots, X_n$ , but functions with other quantities adjoined, as in the paired data  $(X_1, Y_1), \dots, (X_n, Y_n)$ , and also more complex settings where the  $i$ th data ‘point’ is actually a sample. The essential ingredient is that there should be an analogue of integration, so that the operator  $\hat{M}$  at equation (20) can be defined.

Recall that the ordering  $\hat{t}_1(\varepsilon) \geq \dots \geq \hat{t}_n(\varepsilon)$  was introduced in Section 2.1, that the perturbation  $\Delta_i^0$  was introduced at condition (19), that  $\mathcal{R}$  denotes the support of the random functions and that the eigenvalues  $\hat{\alpha}_j$  and eigenvectors  $\hat{\beta}_j$  are as at equation (21). When interpreting theorem 1 it helps if we note that  $\hat{\beta}_1$  and  $-\hat{\beta}_1$  are both eigenfunctions corresponding to the largest eigenvalue  $\hat{\alpha}_1$  and are indistinguishable in terms of that definition. Intuition behind expressions (16)–(19) is given in the paragraph immediately below condition (19).

*Theorem 1.* If conditions (16)–(19) hold, and if  $\hat{\alpha}_1 > \hat{\alpha}_2$ , then, with probability 1, the limit as  $\varepsilon \downarrow 0$  of the ranking  $\hat{t}_1(\varepsilon) \geq \dots \geq \hat{t}_n(\varepsilon)$  is well defined and is in fact the ordering  $\hat{t}_1 \geq \dots \geq \hat{t}_n$  given by

$$\int_{\mathcal{R}} \hat{\beta}_1 \Delta_{i_1}^0 \geq \dots \geq \int_{\mathcal{R}} \hat{\beta}_1 \Delta_{i_n}^0, \tag{23}$$

for one of the choices  $\pm \hat{\beta}_1$ .

In summary, theorem 1 asserts that the limit, as  $\varepsilon \downarrow 0$ , of the ordering of  $X_1, \dots, X_n$  in terms of their influence on  $\hat{\omega}$ , is identical to the ordering of the projections of the perturbations  $\Delta_1^0, \dots, \Delta_n^0$  onto the univariate space that is spanned by the first empirical principal component function for  $\hat{M}^0$ . Since that function explains the greatest amount possible, for a single function, of the variability of the empirical distribution of perturbations corresponding to respective data functions  $X_i$  (or data pairs  $(X_i, Y_i)$ ), then this result enhances the intuitive appeal of assessing influence in terms of tilting.

To appreciate the implications of theorem 1 it is helpful to consider in detail the case  $\hat{\omega} = \bar{X}$ : the sample mean for functional data. There theorem 1 asserts that the limit, as  $\varepsilon \downarrow 0$ , of the

ordering of  $X_1, \dots, X_n$  in terms of their influence, is identical to the ordering of the projections of  $X_1, \dots, X_n$  onto the univariate space that is spanned by the first empirical principal component function for  $\hat{K}$ . In this case,  $\Delta_i = X_i - \bar{X}$  and  $\hat{\beta}_1 = \hat{\phi}_1$ , the latter introduced at equation (3), and therefore the ranking at expression (23) is identical to

$$\int_{\mathcal{R}} \hat{\phi}_1 X_{i_1} > \dots > \int_{\mathcal{R}} \hat{\phi}_1 X_{i_n}. \tag{24}$$

Since  $\hat{\phi}_1$  explains the greatest amount possible, for a single function, of the variability of the empirical distribution of the data set  $X_1, \dots, X_n$ , then the ranking at expression (24) is particularly reasonable.

For a general statistic  $\hat{\omega}$  the role of  $X_i$  in the ranking, e.g. in expression (24), is replaced by  $\Delta_i$ , but the index  $i$  still relates directly to the  $i$ th observation, e.g. the  $i$ th function  $X_i$  or the  $i$ th data pair  $(X_i, Y_i)$ . Perturbations of data weights result in eigenvalue estimators changing and can lead to ties between those quantities where they did not exist previously. That can result in difficulty identifying principal component functions. However, under the continuity assumption that was discussed in Section 2.2, the probability that this occurs in the case  $\varepsilon = 0$  is 0, and so it is not a problem at least in theory. We have not noticed it in numerical work.

### 3. Numerical properties

#### 3.1. Computational aspects

Assume that we observe functional data  $X_1(t_k), \dots, X_n(t_k)$  at  $m$  points  $t_1, \dots, t_m$  on the interval  $\mathcal{I}$ , which here plays the role of  $\mathcal{R}$ . At first sight, the tilting procedure that was described in Section 2 may seem to require solving an optimization problem of dimension  $n$  to obtain the weights  $p_1, \dots, p_n$ . However, it needs only the solution of a system of non-linear equations, the dimension of which depends on  $m$ . The value of  $m$  generally would be determined by the amount of computing power at our disposal, not by the number of points at which the functions were recorded. Therefore it is difficult to be prescriptive about  $m$ , although in numerical experiments we have found that the rankings are not very sensitive to the choice of that quantity.

For instance, tilting the functional mean leads to solving a system of  $m + 2$  non-linear equations in  $C(t_1), \dots, C(t_m), \lambda_1, \lambda_2$ :

$$\begin{aligned} \sum_{i=1}^n \exp\left\{ \lambda_1 - 2\lambda_2 \sum_{k=1}^m C(t_k) X_i(t_k) \right\} X_i(t_k) - \bar{X}(t_k) &= C(t_k), & k = 1, \dots, m, \\ \sum_{k=1}^m C(t_k)^2 &= \varepsilon^2, \\ \sum_{i=1}^n \exp\left\{ \lambda_1 - 2\lambda_2 \sum_{k=1}^m C(t_k) X_i(t_k) \right\} &= 1, \end{aligned}$$

where  $\lambda_1$  and  $\lambda_2$  are Lagrange multipliers,  $C(t) = \bar{X}_p(t) - \bar{X}(t)$  and

$$p_i = \exp\left\{ \lambda_1 - 2\lambda_2 \sum_{k=1}^m C(t_k) X_i(t_k) \right\}.$$

When  $m = 1$  the problem reduces to ranking univariate observations.

Similarly, tilting the functional covariance leads to solving a system of  $\frac{1}{2}m(m + 3) + 2$  non-linear equations in  $C_1(t_k)$  for  $k = 1, \dots, m$ , and in  $C_2(t_k, t_l)$  for  $k, l = 1, \dots, m$ , where, assuming that  $C_2(t_k, t_l) = C_2(t_l, t_k)$ , we impose the constraints

**Table 1.** Computing time for ranking functional data based on tilting the functional mean for various sample sizes  $n$  and time point numbers  $m$

$n$	$m$	Tilting time (s)
10000	12	0.5
	25	0.8
	50	3.1
100000	12	4.2
	25	12.7
	50	40.5
1000000	12	52.7
	25	163.6
	50	543.2

$$\sum_{i=1}^n p_i X_i(t_k) = C_1(t_k), \quad k = 1, \dots, m,$$

$$\sum_{i=1}^n p_i X_i(t_k) X_i(t_l) = C_2(t_k, t_l), \quad k, l = 1, \dots, m,$$

$$\sum_{k=1}^m \sum_{l=1}^m \{C_2(t_k, t_l) - C_1(t_k) C_1(t_l) - \hat{K}(t_k, t_l)\}^2 = \varepsilon^2, \quad \sum_{i=1}^n p_i = 1.$$

Here the constraints are again imposed via Lagrange multipliers,  $C_1(s) = \bar{X}_p(s)$ ,  $C_2(s, t) - C_1(s) C_1(t) = \hat{K}_p(s, t)$  and

$$p_i = \exp \left[ \lambda_1 - 2\lambda_2 \sum_{k=1}^m \sum_{l=1}^m \{C_2(t_k, t_l) - C_1(t_k) C_1(t_l) - \hat{K}(t_k, t_l)\} \{X_i(t_k) - C_1(t_k)\} \{X_i(t_l) - C_1(t_l)\} \right].$$

In the same manner, tilting the  $j$ th eigenfunction requires solving a non-linear system of  $m + 2$  equations, and tilting the conditional mean requires solving a non-linear system of  $mq + 2$  equations, where  $q$  is the number of values that  $u$  can take.

As an illustration, we consider tilting the functional mean for various sample sizes  $n$  and time point numbers  $m$ . The original sample curves are generated from an outlier model (Sun and Genton, 2011):  $X_i(t) = 4t + e_i(t) + \eta_i S_i L$ ,  $i = 1, \dots, n$ ,  $t \in [0, 1]$ , where  $\eta_i = 1$  with probability 0.1 and equals 0 with probability 0.9,  $L = 6$  is the size of the contamination,  $S_i$  is a sequence of random variables independent of  $\eta_i$  taking values  $\pm 1$  with probability  $\frac{1}{2}$ , and  $e_i(t)$  is a Gaussian stochastic process with zero mean and exponential covariance function  $K(t_1, t_2) = \exp(-|t_2 - t_1|)$ . All the computations are done in MATLAB on a 2.80-GHz Intel Xeon X5560 chip with 48 Gbytes of random-access memory. The computing times (in seconds) are reported in Table 1. It can be seen that the tilting approach is feasible even for very large sample sizes  $n$ .

### 3.2. Sea surface temperature data

The sea surface temperature data set consists of monthly temperatures measured in degrees Celsius over the east-central tropical Pacific Ocean. In this case, each curve represents 1 year of observed sea surface temperatures from January 1951 to December 2007, and there are  $n = 57$  such curves.

We apply our tilting method with  $\varepsilon = 0.2$  to this data set, for tilting and ranking based on the functional mean and the functional covariance. The value 0.2 was determined simply by reducing  $\varepsilon$  a little more, for security, beyond the point where there was no further apparent change in the ranking. Choosing  $\varepsilon$  generally is simple and unsophisticated; it is unrelated to the problem of selecting a smoothing parameter, for example.

The results are presented in Fig. 1, where the red curve represents the median, the broken curve represents the mean, and the blue and green curves represent the lower 25% and upper 25% quantiles respectively. The shading of the data curves corresponds to a ranking based on the weights  $p_i$ , from the highest ranked curve (i.e. the curve with index  $\hat{t}_1(\varepsilon)$  in equation (15), indicated by light grey in Fig. 1) to the lowest ranked curve (indicated by dark grey).

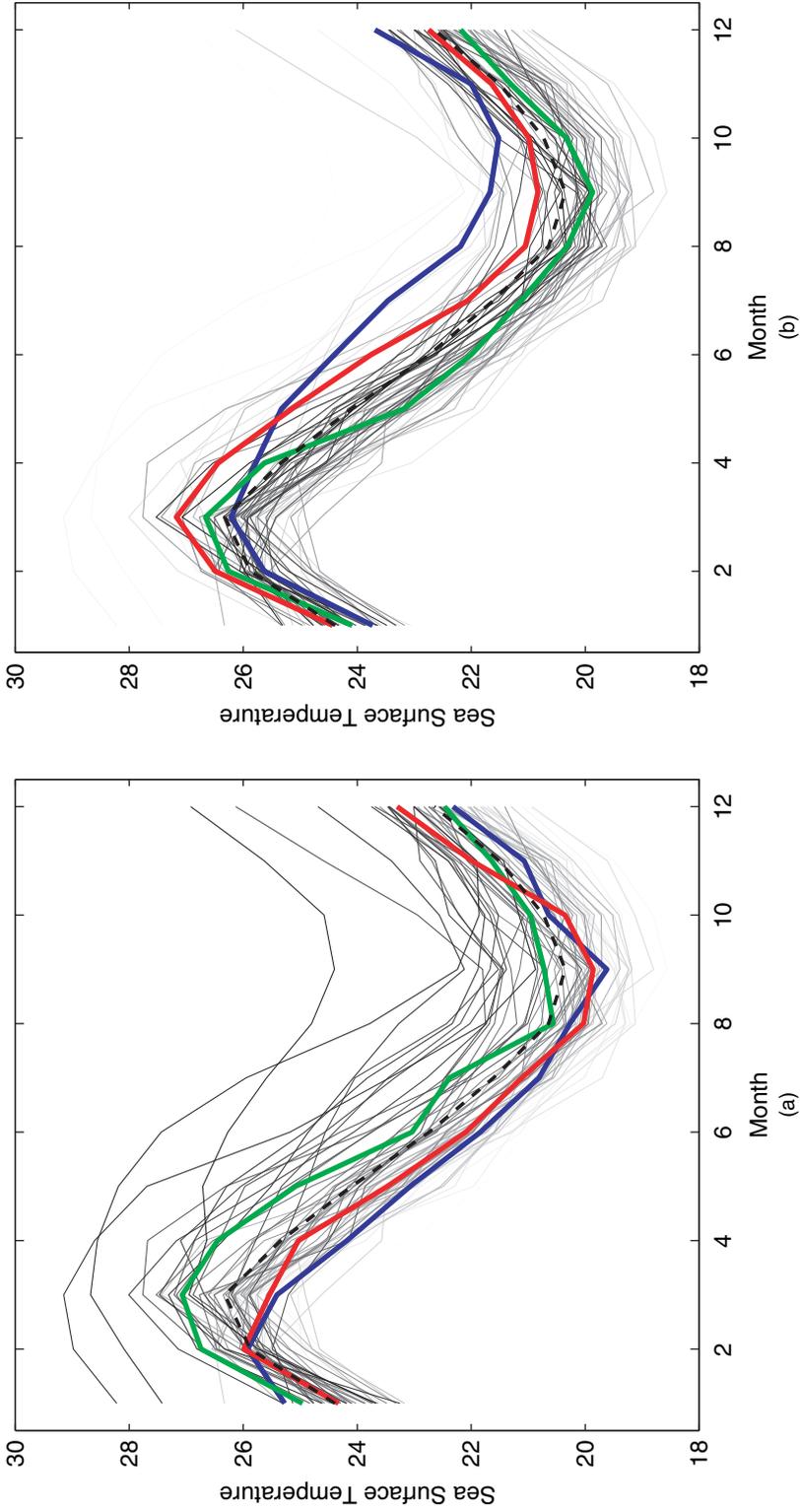
The plots in the case of the sample mean, which are shown in Fig. 1(a), reveal that curves that have greatest influence (i.e. which correspond to small or large values of  $p_i$ ) are those that are positioned relatively high up or low down the temperature axis. Moreover, the distribution of influential curves is quite asymmetric, with the least influential curves being more tightly bunched than the most influential. In particular, the lower 25% quantile curve is, on average, significantly closer to the median curve than is the upper 25% quantile curve. If these data were real valued then the distribution would be quite asymmetric, with one tail (say, the right-hand tail) significantly longer than the other. In this case, moving from left to right on the real line, the lower 25% quantile, the median, the mean and the upper 25% quantile would be arranged in that order. This is also the order of those quantities in Fig. 1(a), between months 2 and 9, and in this sense the influence of data curves on the mean exhibits a pattern that is familiar for real data drawn from a skewed distribution.

A ranking based on the functional mean is similar to that for real-valued data in other respects, also. For example, the two darkest curves, i.e. the two high up most influential functions, correspond to the years 1983 and 1997 during which there was a so-called El Niño effect: the unusual warming of sea surface temperature in the region where the data were gathered.

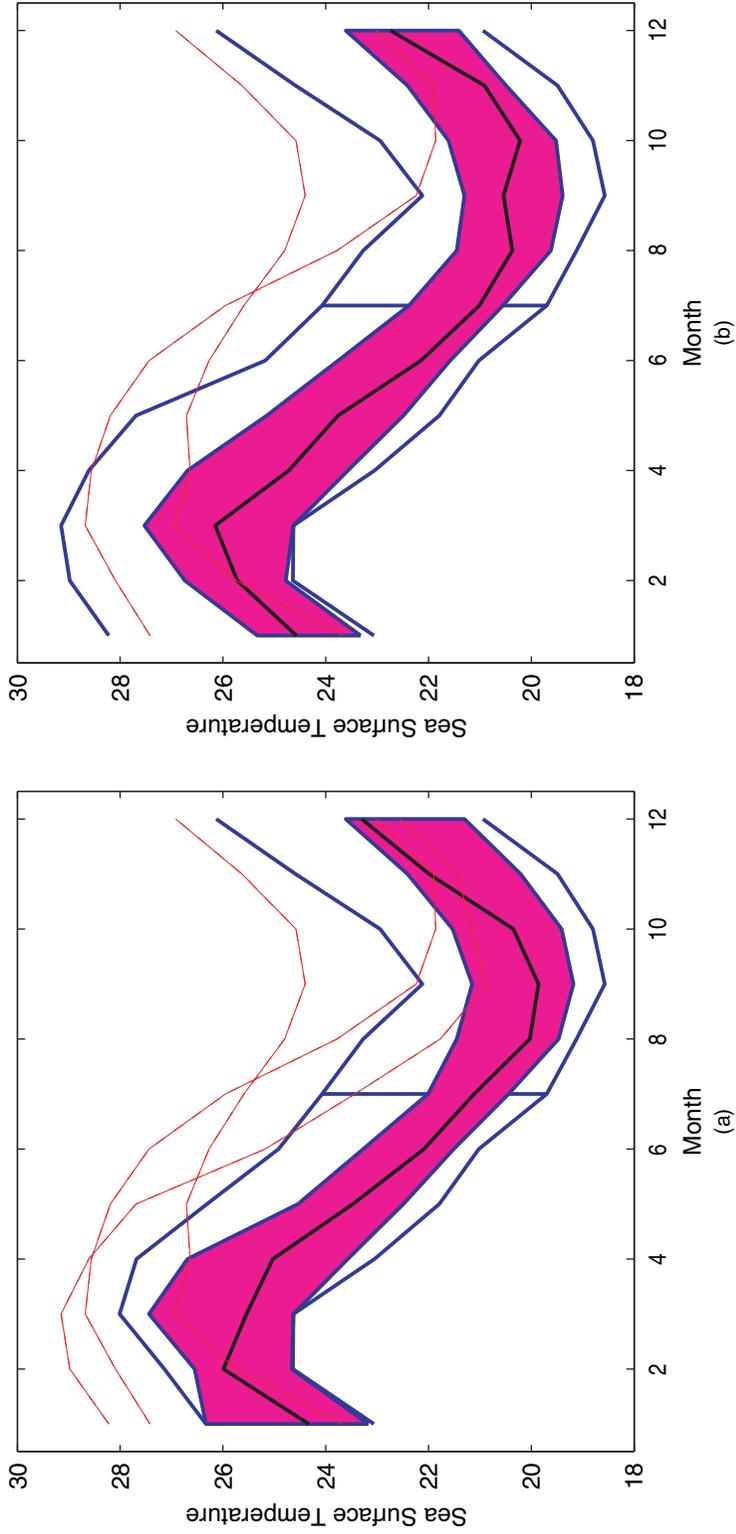
Of course, the data depictions in either panel of Fig. 1 represent projections, shown in a two-dimensional plot, of higher dimensional quantities, and so we should not expect them to reflect the familiar ‘ordering’ for real data always. Fig. 1(b), which depicts the data functions in terms of their relative influence on functional covariance, is a case in point. There the functions representing upper and lower 25% quantiles lie on the same side, not opposite sides, of the median function throughout the first 5 months of the year, indicating that the above-mentioned projections do not, in the case of covariance, produce a separation that usefully grades the data from more to less influential. Variability through much of the year is seen to be relatively haphazard, in terms of its depiction in Fig. 1(b), with both low and high variability, relative to median variability, tending to be caused by functions which are relatively high on the temperature scale, whereas the ‘more average’ curves (again in the sense of the median) are generally positioned towards relatively low temperatures.

The approach to functional ranking that is proposed in this paper can be used to construct a functional boxplot as in Sun and Genton (2011, 2012a). Fig. 2 depicts functional boxplots based on the functional mean ranking (Fig. 2(a)) and on modified band depth ranking (Fig. 2(b)). The general shape of the functional boxplot is similar for the two rankings, but the ranking based on the method in this paper detects one additional outlier corresponding to the year 1998, when the El Niño effect continued from 1997.

To provide further guidance in interpreting measures of influence for functional data we give, in Fig. 3, results for scalar data. Specifically, Fig. 3(a) shows a plot of tilted weights in the case of the mean,  $\bar{X}_p(t_0) = \sum_i p_i X_i(t_0)$ , where  $t = t_0 \equiv$  month 7. Here the results are easily interpreted; they are similar to those for Fig. 1(a), and in particular the median dot lies between the dots



**Fig. 1.** Tilting and ranking of sea surface temperature functional data (in degrees Celsius) with  $\varepsilon = 0.2$  based on (a) the functional mean and (b) the functional covariance: —, upper 25% quantile; —, 50% quantile; —, lower 25% quantile; - - -, mean



**Fig. 2.** Functional boxplots of sea surface temperature functional data (in degrees Celsius) based on (a) tilting and ranking the functional mean with  $\varepsilon = 0.2$  and (b) the modified band depth

that represent lower and upper quantiles. (In both panels we use the colour and grey shade convention in Fig. 1.) Fig. 1(b) addresses the case of covariance,  $\hat{K}_p(t, t_0)$ , viewed as a function of  $t$ . Here the interpretation reflects that for Fig. 1(b), where the median does lie between the 25% and 75% quantiles for month 7. In particular, as was also found in the context of Fig. 1, the data projection that is represented by Fig. 3(b) does order the data in an informative way for month 7.

Next, in Fig. 4(a) we visualize the weights  $p_i$ , for  $i = 1, \dots, 57$ , based on tilting the functional mean, when  $d_1(p)$  is constrained as at expression (13). In particular, the values of  $p_i$  are now functions of  $\varepsilon > 0$ . The vertical black line corresponds to  $\varepsilon = 0.2$ . The green, red and blue curves correspond to the upper 25%, 50% and lower 25% quantiles of the weights. As anticipated by theorem 1 in Section 2.3, the ranking of the weights remains the same as  $\varepsilon \downarrow 0$ . To interpret the weights further we investigate the empirical influence of one outlying observation. We take that curve to be the median function, represented as the red curve in Fig. 1(a), and we perturb that observation,  $X_{28}(t)$ , at a single time point,  $t = 6$ , representing June, by an additive quantity  $\xi$ , i.e. we consider the function  $X_{28}(6) + \xi$ . In Fig. 4(b) we plot the weights  $p_i$ , for  $i = 1, \dots, 57$ , based on tilting the functional mean with  $\varepsilon = 0.2$  and associated with each sea surface temperature observation viewed as a function of  $X_{28}(6) + \xi$ . The vertical line corresponds to  $\xi = 0$  (no contamination). The broken curve depicts the change in the weight  $p_{28}$  when the functional observation  $X_{28}(t)$  is perturbed to  $X_{28}(6) + \xi$ . We see that  $p_i$  becomes larger or smaller as the positive or negative amount of the contamination  $\xi$  changes in the same way and eventually becomes the largest or smallest weight among those for the functional sample.

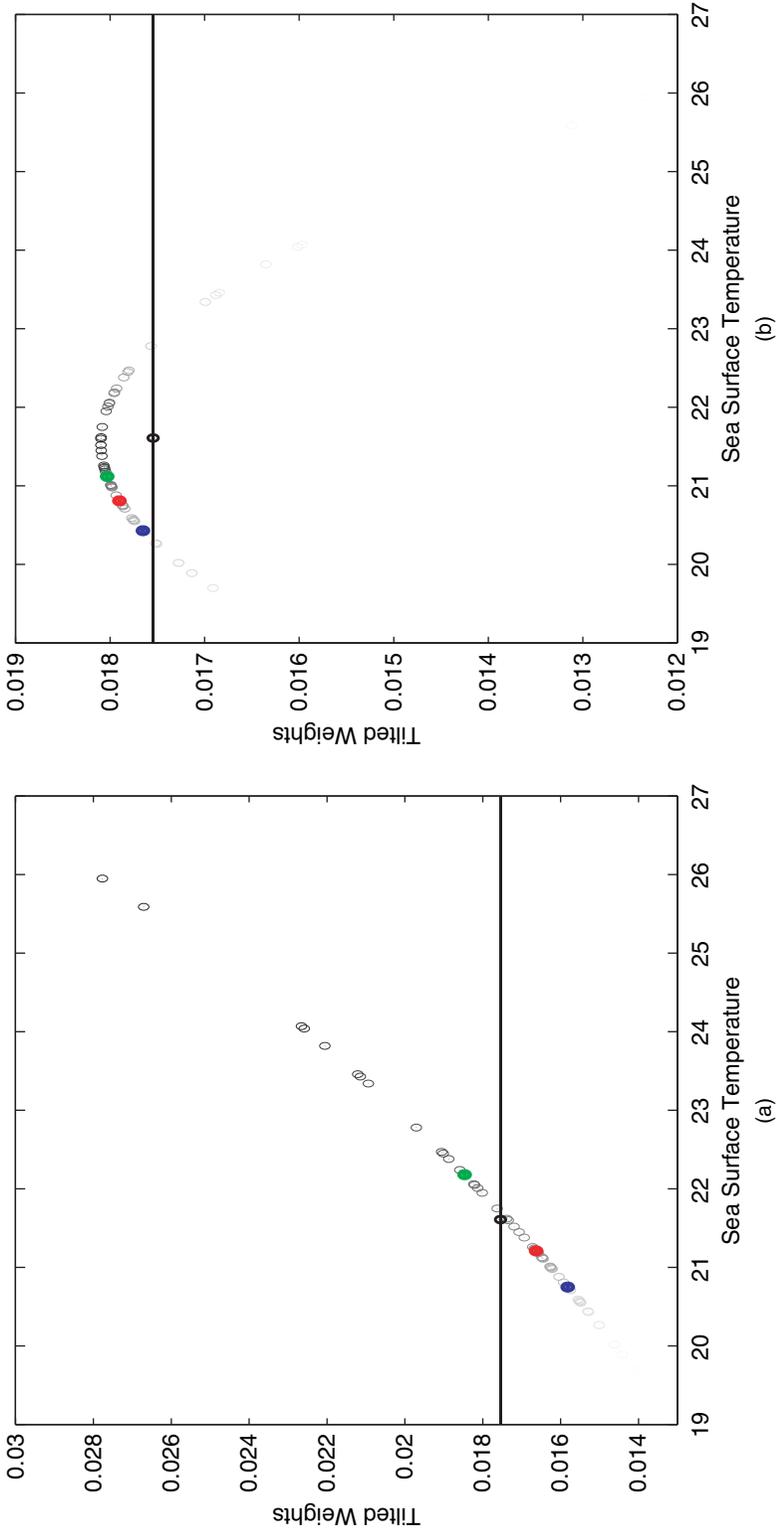
### 3.3. Simulation comparison

To compare the tilting approach to ranking with the band depth ranking, we perform the following simulation experiment. We simulate  $n = 100$  functional curves, each observed at 12 time points, with a mean 0 unit variance Gaussian distribution and with an exponential auto-covariance function  $\gamma(h) = \exp(-h/0.001)$ . We then find the median curve from this sample with the tilting approach and with the band depth approach. We repeat this experiment 1000 times. Hence, we have 1000 medians from tilting and 1000 medians from band depth. Fig. 5 summarizes this information by means of functional boxplots. It can be seen that the medians that are obtained by tilting and the medians that are obtained by band depth have approximately similar distribution patterns.

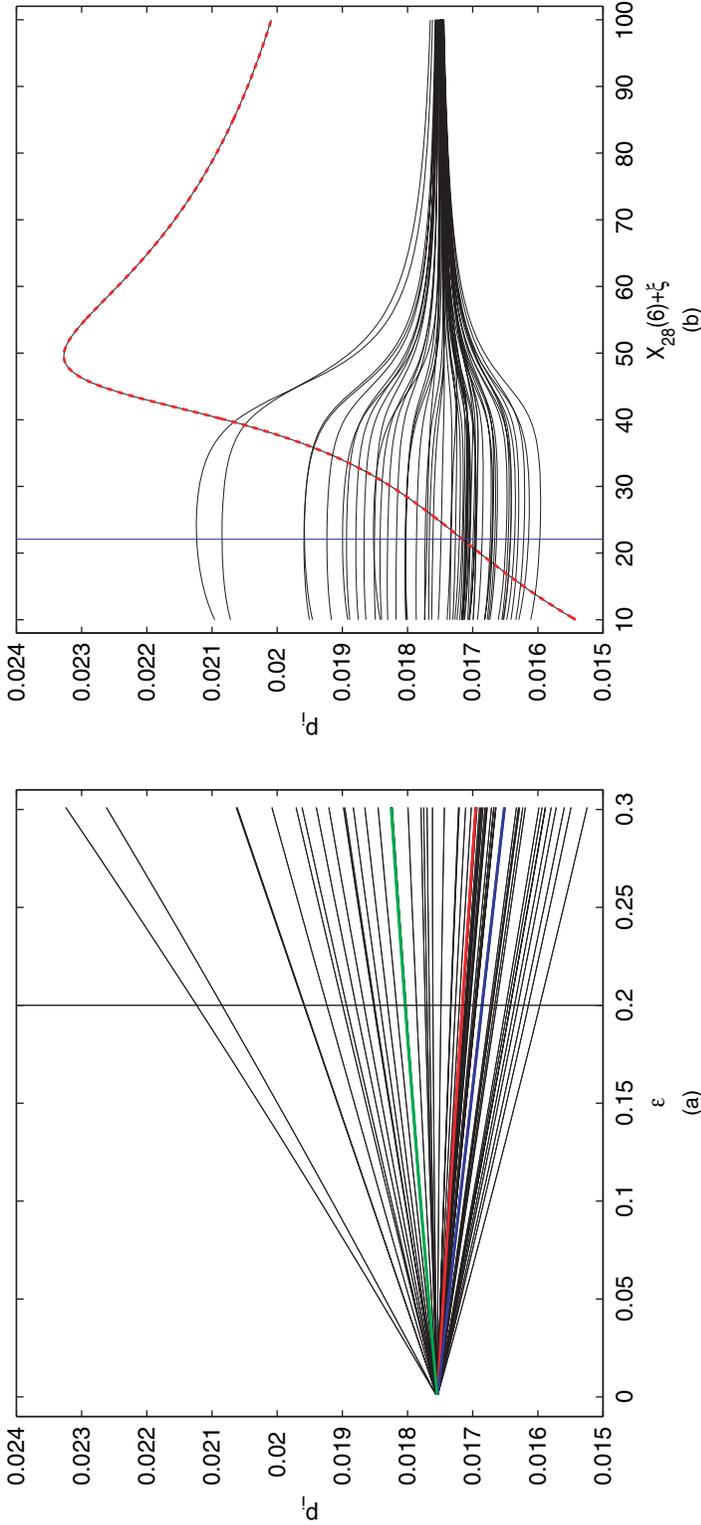
### 3.4. Child growth data

The child growth data set consists of the heights, in centimetres, of 39 boys and 54 girls, measured at 31 unequally spaced ages from 1 year to 18 years; see Ramsay and Silverman (2005) for details. We consider the boys and girls to come from separate populations. Within each population the growth curves have similar shape, and in particular are monotone increasing. Therefore it is not straightforward to identify unusual growth patterns. We shall first determine and discuss the rankings of these curves, using our tilting approach, and then we shall compare the level of information that can be extracted from the data by using functional boxplots, based on rankings provided by tilting or by modified band depth (López-Pintado and Romo, 2009) respectively.

First we apply our tilting method, using the functional data sample mean as the statistic of interest. The results are presented in Fig. 6, where the red curve represents the median, the broken curve depicts the mean, and the blue and green curves show the lower 25% and upper 25% quantiles respectively. The shading of the data curves corresponds to a ranking based on the weights  $p_i$ , from the highest ranked curve (i.e. the curve with index  $\hat{i}_1(\varepsilon)$  in



**Fig. 3.** Tiling and ranking of univariate data with  $\varepsilon = 0.2$  based on (a) the functional mean  $X_{\rho}(t_0)$ , where  $t_0$  is month 7, and (b) the functional covariance  $K_{\rho}(t, t_0)$ : ●, upper 25% quantile of the weights; ●, 50% quantile of the weights; ●, lower 25% quantile of the weights; ○, mean; —, constant weight 1/57



**Fig. 4.** (a) Weights  $p_i$ , for  $i = 1, \dots, 57$ , based on tilting the functional mean and associated with each sea surface temperature curve (the  $p_i$ :s are viewed as functions of  $\epsilon$ ),  $\epsilon = 0.2$ ; —, upper 25% quantile; —, 50% quantile; —, lower 25% quantile; (b) weights  $p_i$ , for  $i = 1, \dots, 57$ , based on tilting the functional mean with  $\epsilon = 0.2$  and associated with each sea surface temperature curve viewed as a function of  $X_{28}(t) + \xi$  ( $\xi = 0$  (no contamination); —, change in the weight  $p_{28}$  when the curve  $X_{28}(t)$  is perturbed by adding  $\xi$ )

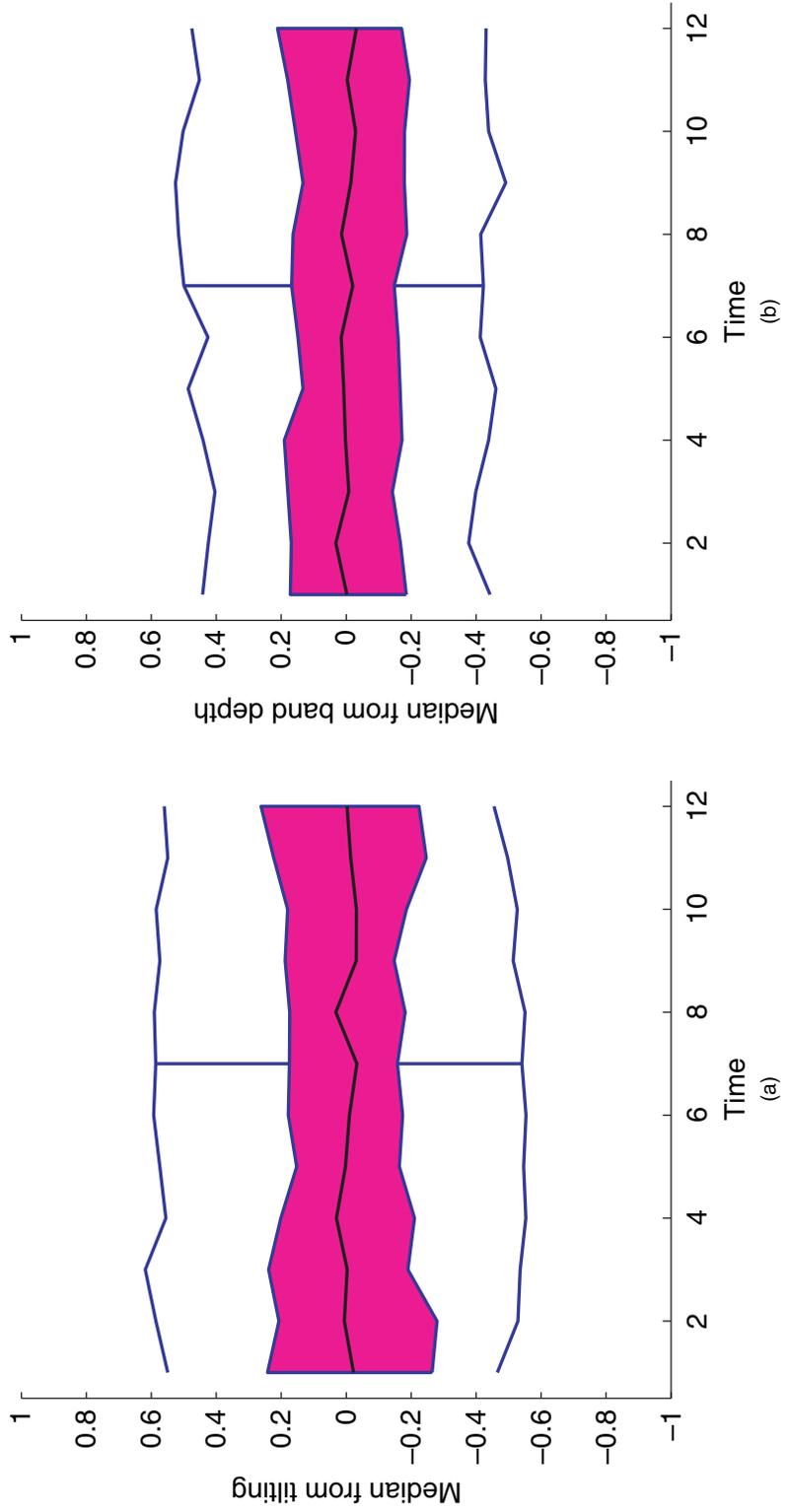
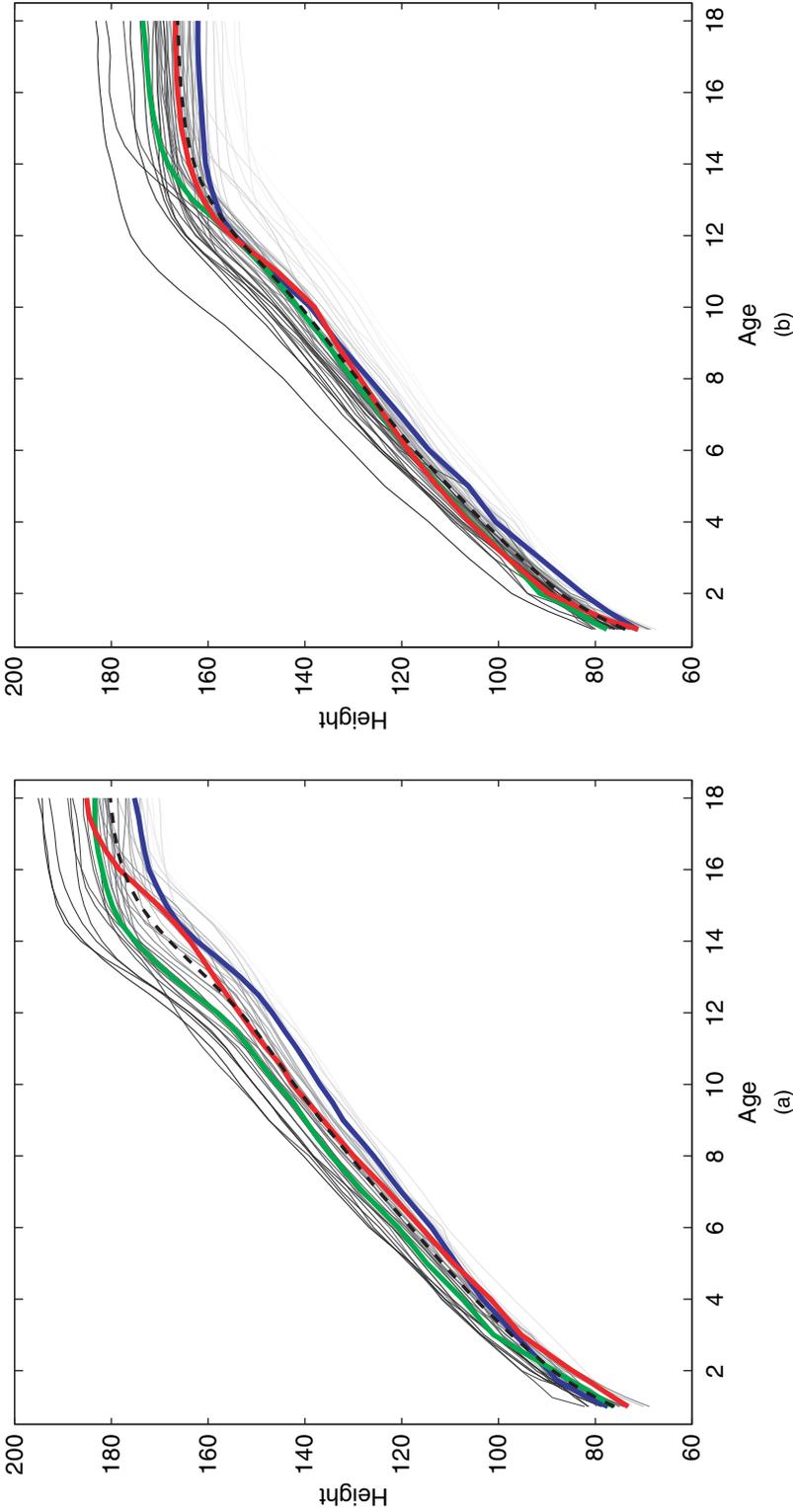
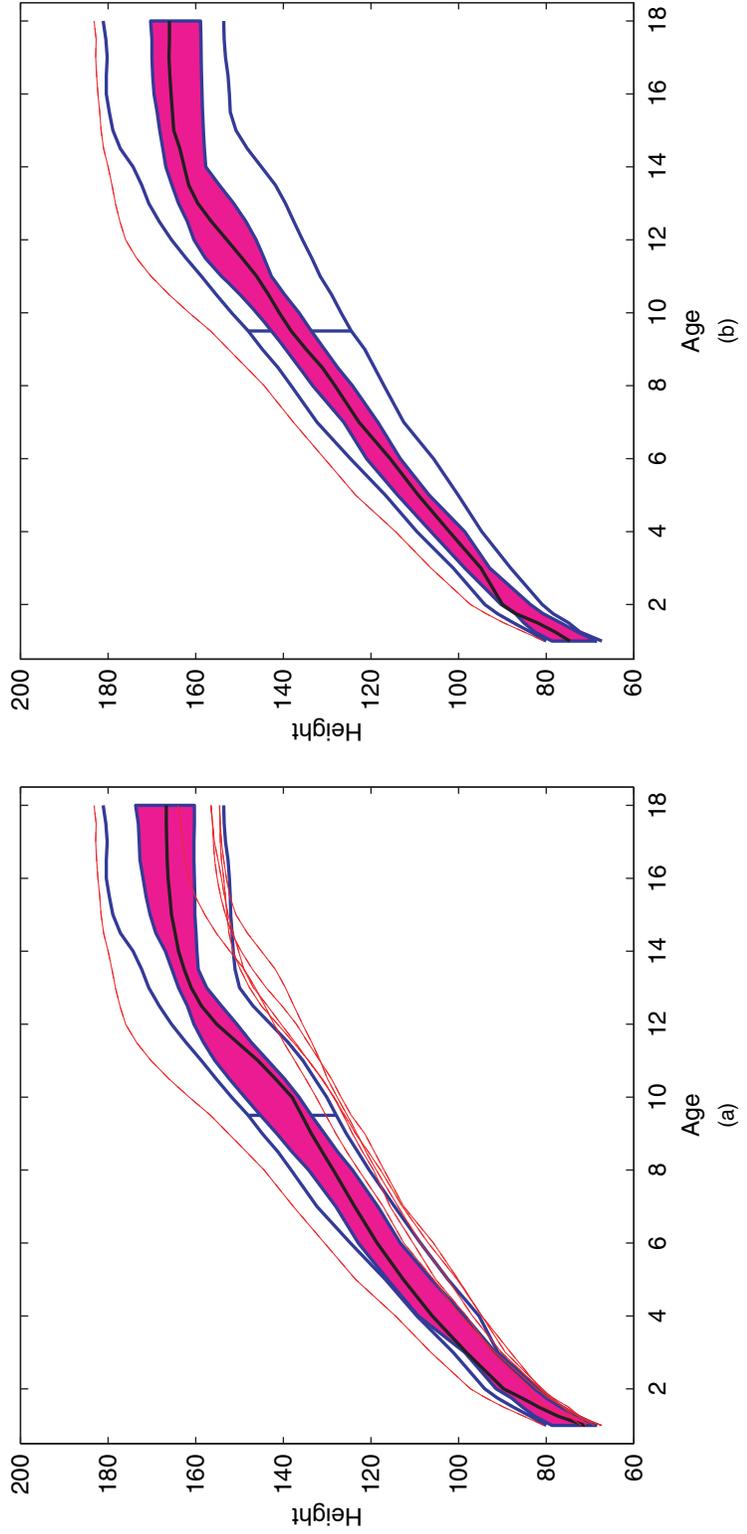


Fig. 5. Functional boxplots of 1000 medians obtained by (a) tilting and (b) band depth



**Fig. 6.** Tilting and ranking of child growth data, using the functional data sample mean as the statistic of interest and in the case of (a) boys and (b) girls: —, upper 25% quantile; —, lower 25% quantile; —, 50% quantile; — — —, mean



**Fig. 7.** Functional boxplots of the child growth data for the sample of girls, (a) based on tilting and using the functional data sample mean as the statistic of interest and (b) based on the modified band depth

equation (15), indicated by light grey in Fig. 6) to the lowest ranked curve (indicated by dark grey). The plots reveal that curves that have greatest influence (i.e. which correspond to small or large values of  $p_i$ ) are positioned relatively high up or low down the height axis. The distribution of influential curves is more symmetric in this example than in the example that was treated in Section 3.2.

Our approach to functional data ranking can also be used to construct functional boxplots, as in Sun and Genton (2011, 2012a). Fig. 7 depicts, for the girls, functional boxplots based on the functional mean ranking (Fig. 7(a)) and on modified band depth ranking (Fig. 7(b)). The general shape of the functional boxplot is similar for the two rankings, but that based on the method that is suggested in this paper detects five additional outliers, corresponding to girls with very low growth curves, especially in their teenage years. For the boys, the two functional boxplots are similar in shape and neither approach detects outlying data curves. This is a classic benchmark data set, and readers seeking further discussion of the data will find it in Ramsay and Silverman (2005); see particularly pages 1, 62, 88, 112 and 165.

**Acknowledgements**

The authors thank three referees, the Associate Editor and the Joint Editor for their helpful feedback, which has greatly improved the manuscript.

Genton’s research was partially supported by a grant from the University of Melbourne as an Honorary Fellow.

**Appendix A: Proof of theorem 1**

Assumption (17) implies that, if  $p$  is an extremum of

$$\int (\hat{\omega}_p - \hat{\omega})^2 + \lambda_1 \sum_{i=1}^n p_i \log(p_i) + \lambda_2 \left( \sum_{i=1}^n p_i - 1 \right),$$

where  $\lambda_1$  and  $\lambda_2$  denote Lagrange multipliers, then

$$\begin{aligned} p_j &= \exp \left\{ \lambda_3 + \lambda_4 \sum_{i=1}^n (p_i - n^{-1}) \int \Delta_i \Delta_j + \lambda_4 A_j + O(\varepsilon^3) \right\}, \\ &= \frac{1}{n} \left\{ \lambda_5 + \lambda_4 \int \psi \Delta_j + \lambda_4 A_j + \frac{1}{2} \left( \lambda_4 \int \psi \Delta_j \right)^2 + O(\varepsilon^3) \right\}. \end{aligned} \tag{25}$$

Here and below,  $\lambda_1, \dots, \lambda_5$  do not depend on  $p$ , and we have defined

$$\psi = \sum_{i=1}^n (p_i - n^{-1}) \Delta_i, \tag{26}$$

equal to the dominant term on the right-hand side of result (16). The property  $\Delta_i = O(\varepsilon)$  in equation (18) implies that  $\psi = O(\varepsilon)$ .

Using equation (18), and the fact that  $\sum_j p_j = 1$ , it can be deduced from equation (25) that, for each  $j$ ,

$$p_j = \frac{1}{n} \left\{ 1 + \lambda_4 \int \psi \Delta_j + \lambda_4 A_j + \frac{1}{2} \lambda_4^2 S_j + O(\varepsilon^3) \right\}, \tag{27}$$

where

$$S_j = \left( \int \psi \Delta_j \right)^2 - \frac{1}{n} \sum_{i=1}^n \left( \int \psi \Delta_i \right)^2.$$

Therefore, again using equation (18),

$$\begin{aligned} \sum_{j=1}^n p_j \log(np_j) &= \frac{1}{n} \sum_{j=1}^n \left( 1 + \lambda_4 \int \psi \Delta_j + \lambda_4 A_j + \frac{1}{2} \lambda_4^2 S_j \right) \\ &\quad \times \left\{ \lambda_4 \int \psi \Delta_j + \lambda_4 A_j + \frac{1}{2} \lambda_4^2 S_j - \frac{1}{2} \left( \lambda_4 \int \psi \Delta_j \right)^2 \right\} + O(\varepsilon^3) \\ &= \frac{1}{2} \sum_{j=1}^n \left( \lambda_4 \int \psi \Delta_j \right)^2 + O(\varepsilon^3). \end{aligned}$$

Furthermore, in view of results (16) and (26),

$$\int (\hat{\omega}_p - \hat{\omega})^2 = \int \psi^2 + O(\varepsilon^3).$$

Hence, the algorithm chooses  $p$  (and therefore  $\psi$ ) to minimize

$$\sum_{j=1}^n \left( \lambda_4 \int \psi \Delta_j \right)^2 + O(\varepsilon^3)$$

or, equivalently, to minimize

$$\lambda_4^2 \int \psi (\hat{M} \psi) + O(\varepsilon^3), \tag{28}$$

subject to

$$\int \psi^2 = \varepsilon^2 + O(\varepsilon^3). \tag{29}$$

Since  $p_j$  satisfies equation (27) then

$$\begin{aligned} \psi &= \sum_{i=1}^n (p_i - n^{-1}) \Delta_i = \frac{1}{n} \sum_{i=1}^n \left( \lambda_4 \int \psi \Delta_i + \lambda_4 A_i + \frac{1}{2} \lambda_4^2 S_i \right) \Delta_i + O(\varepsilon^3) \\ &= \lambda_4 \frac{1}{n} \sum_{i=1}^n \Delta_i \int \psi \Delta_i + O(\varepsilon^3) = \lambda_4 \hat{M} \psi + O(\varepsilon^3). \end{aligned}$$

Therefore, up to an error that is negligible as  $\varepsilon \downarrow 0$ ,  $\psi$  is an eigenfunction of  $\hat{M}$  with eigenvalue  $\lambda_4^{-1}$ . Hence, minimizing the quantity at expression (28) is equivalent to minimizing  $\lambda_4 \int \psi^2 + O(\varepsilon^3)$ , which, we know from condition (29), is equivalent to minimizing  $\lambda_4 \varepsilon^2 + O(\varepsilon^3)$ . Hence,  $\lambda_4$  should be as small as possible, and so  $\psi$  should equal the first-ranked eigenvector of  $\hat{M}$ , which in turn equals  $\hat{\beta}_1$ , plus a term of order  $\varepsilon$ , and  $\lambda_4 = \hat{\alpha}_1^{-1} + O(\varepsilon)$ . These results and result (27) together imply that the ordering of indices  $\hat{i}_1, \dots, \hat{i}_p$  corresponding to  $p_{\hat{i}_1}(\varepsilon) > \dots > p_{\hat{i}_p}(\varepsilon)$ , for all sufficiently small  $\varepsilon$ , is identical to the ordering (23).

## References

Bravo, F. (2005) Blockwise empirical entropy tests for time series regressions. *J. Time Ser. Anal.*, **26**, 185–210.  
 Camponovo, L. and Otsu, T. (2012) Breakdown point theory for implied probability bootstrap. *Econometr. J.*, **15**, 32–55.  
 Choi, E., Hall, P. and Presnell, B. (2000) Rendering parametric procedures more robust by empirically tilting the model. *Biometrika*, **87**, 453–465.  
 Cressie, N. and Read, T. R. C. (1984) Multinomial goodness-of-fit tests. *J. R. Statist. Soc. B*, **46**, 440–464.  
 Critchley, F., Atkinson, R. A., Lu, G. and Biazi, E. (2001) Influence analysis based on the case sensitivity function. *J. R. Statist. Soc. B*, **63**, 307–323.

- Critchley, F. and Marriott, P. (2004) Data-informed influence analysis. *Biometrika*, **91**, 124–140.
- Ferraty, F. and Vieu, P. (2006) *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer.
- Genton, M. G., Johnson, C., Potter, K., Stenchikov, G. and Sun, Y. (2014) Surface boxplots. *Stat*, **3**, 1–11.
- Genton, M. G. and Ruiz-Gazen, A. (2010) Visualizing influential observations in dependent data. *J. Computnl Graph. Statist.*, **19**, 808–825.
- Hall, P. and Presnell, B. (1999a) Biased bootstrap methods for reducing the effects of contamination. *J. R. Statist. Soc. B*, **61**, 661–680.
- Hall, P. and Presnell, B. (1999b) Intentionally biased bootstrap methods. *J. R. Statist. Soc. B*, **61**, 143–158.
- Hall, P., Titterton, D. M. and Xue, J.-H. (2009) Tilting methods for assessing the influence of components in a classifier. *J. R. Statist. Soc. B*, **71**, 783–803.
- Hall, P. and Yao, Q. (2003) Data tilting for time series. *J. R. Statist. Soc. B*, **65**, 425–442.
- Hyndman, R. J. and Shang, H. L. (2010) Rainbow plots, bagplots, and boxplots for functional data. *J. Computnl Graph. Statist.*, **19**, 29–45.
- Lazar, N. A. (2005) Assessing the effect of individual data points on inference from empirical likelihood. *J. Computnl Graph. Statist.*, **14**, 626–642.
- López-Pintado, S. and Romo, J. (2009) On the concept of depth for functional data. *J. Am. Statist. Ass.*, **104**, 718–734.
- López-Pintado, S., Sun, Y., Lin, J. K. and Genton, M. G. (2014) Simplicial band depth for multivariate functional data. *Adv. Data. Anal. Classific.*, **8**, 321–338.
- Ramsay, J. O., Hooker, G. and Graves, S. (2009) *Functional Data Analysis with R and MATLAB*. New York: Springer.
- Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*, 2nd edn. New York: Springer.
- Sun, Y. and Genton, M. G. (2011) Functional boxplots. *J. Computnl Graph. Statist.*, **20**, 313–334.
- Sun, Y. and Genton, M. G. (2012a) Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics*, **23**, 54–64.
- Sun, Y. and Genton, M. G. (2012b) Functional median polish. *J. Agric. Biol. Environ. Statist.*, **17**, 354–376.
- Sun, Y., Genton, M. G. and Nychka, D. W. (2012) Exact fast computation of band depth for large functional datasets: how quickly can one million curves be ranked? *Stat*, **1**, 68–74.
- Xu, K. L. and Phillips, P. C. B. (2011) Tilted nonparametric estimation of volatility functions with empirical applications. *J. Bus. Econ. Statist.*, **29**, 518–528.
- Yu, G., Zou, C. and Wang, Z. (2012) Outlier detection in functional observations with applications to profile monitoring. *Technometrics*, **54**, 308–318.