Generalized Linear Models for Geometrical Current predictors. An application to predict garment fit

Sonia Barahona [1], Pablo Centella [1], Ximo Gual-Arnau [2], M. Victoria Ibáñez [3], and Amelia Simó [4]

[1] Department of Mathematics, Universitat Jaume I, Catellón, Spain

[2] Department of Mathematics-INIT, Universitat Jaume I, Catellón, Spain

[3] Department of Mathematics-IMAC, Universitat Jaume I, Catellón, Spain

---

Address for correspondence: M. Victoria Ibáñez, Department Mathematics- IMAC, Universitat Jaume I 12071, Castellón Spain.

E-mail: mibanez@uji.es.

Phone: (+34) 964 728 393.

Fax: (+34) 964 728 429.

Abstract: The aim of this paper is to model an ordinal response variable in terms of vector-valued functional data included on a vector-valued RKHS. In particular, we focus on the vector-valued RKHS obtained when a geometrical object (body) is characterized by a *current* and on the ordinal regression model. A common way to solve this problem in functional data analysis is to express the data in the orthonormal basis given by decomposition of the covariance operator. But our data present very

important differences with respect to the usual functional data setting. On the one hand, they are vector-valued functions, and on the other, they are functions in an RKHS with a previously defined norm. We propose to use three different bases: the orthonormal basis given by the kernel that defines the RKHS, a basis obtained from decomposition of the integral operator defined using the covariance function, and a third basis that combines the previous two. The three approaches are compared and applied to an interesting problem: building a model to predict the fit of children's garment sizes, based on a 3D database of the Spanish child population. Our proposal has been compared with alternative methods that explore the performance of other classifiers (Suppport Vector Machine and k-NN), and with the result of applying the classification method proposed in this work, from different characterizations of the objects (landmarks and multivariate anthropometric measurements instead of currents), obtaining in all these cases worst results.

Key words: Statistical Shape and Size Analysis; Vector-valued Reproducing Kernel Hilbert Space; Functional Data Analysis; Ordinal Regression

1 Introduction

In many scientific fields, such as Biology, Medicine and Anthropometry, we can find a great number of applications where it is necessary to predict a categorical variable as a function of a geometrical object predictor. These geometrical objects can be mathematically characterized in different ways, the most popular being as a set of landmarks (Bookstein, 1978; Kendall, 1984; Dryden and Mardia, 2016), compact

sets (Serra, 1982; Baddeley and Molchanov, 1998; Simó et al., 2004; Molchanov, 2006) or functions (Loncaric, 1998; Kindratenko, 2003; Gual-Arnau et al., 2013). In this paper, the contour of each geometrical object (surface in $\mathbb{R}^3$) is represented by a mathematical structure named current. This framework was introduced by Vaillant and Glaunès (2005) and Glaunes and Joshi (2006) and it provides a unifying framework in which to process any sets of points, curves and surfaces or a mixture of these. No hypothesis on the topology of the shapes is assumed. Moreover, it is weakly sensitive to the sampling of shapes and it does not depend on the choice of parameterizations.

Currents are mathematically complex objects but, fortunately, it is possible to associate a subspace of currents with a vector-valued Reproducing Kernel Hilbert Space (RKHS) by duality and, as a result, we can represent each geometrical object with a function in an RKHS (Durrleman, 2010; Barahona et al., 2017b). A vector-valued RKHS is a Hilbert vector space of functions with useful properties.

This work is motivated by an experimental study carried out by the Biomechanics Institute of Valencia, whose ultimate objective was to implement a web application for online shopping for children's wear. In particular, that application should make it possible to select the right size of children's clothing without requiring the child to try on the clothes. Selecting the proper size of any garment for a child without trying it on constitutes a problem when buying these items both in a physical store and, especially, online. Seventy-eight randomly selected children between the ages of 3 and 12 years participated in this study. Firstly, the children were scanned using a 3D body scanner. Next, garments were tried on in different sizes and an expert classified the fit of each garment as "too small", "correct fit" or "too large".

Authors have been previously working with a part of this data set in Barahona et al.

(2017b) and Barahona et al. (2017a). In Barahona et al. (2017b) we reviewed the theory necessary to represent geometrical objects as elements of an RKHS through currents. There, we saw how to calculate the sample mean in a vector-valued RKHS (Hsing and Eubank, 2015), and checked that the distance between two surfaces (or curves) is defined as the distance between the corresponding elements in the RKHS, with the particularity that the inner product of the RKHS will differ from the usual L^2 inner product. Then, we adapted the classic *k-means* partitioning algorithm to this space and applied it to synthetic and real data sets.

In Barahona et al. (2017a) we explored supervised classification methods of geometrical objects when these objects are characterized by functions that "live" on a vector-valued RKHS. We focused on supervised classification methods that are not just based on distances between objects (as Functional Discriminant Analysis (Fisher, 1936; Shin, 2008) or Classification Trees), where a reduction of the dimension of the data is compulsory.

Now, we are interested in a "new" ordinal variable "fit of each garment", that can be seen as a categorized version of the latent unmeasured continuous variable "width of the mismatch", and our main aim is to model it, in terms of vector-valued functional data included on a vector-valued RKHS.

Different approaches to the problem of predicting the correct garment sizes, can be found in the literature. Most of them are based on taking the user's anthropometric measurements and their relationship with the dimensions of the garment. In this paper, from the 3D scanning of the child's body and using the currents approach, a child will be represented by a function in a vector-valued RKHS and Functional Data Analysis (FDA) will be used on this space. Unlike the methods based on landmarks, curves or parameterized surfaces, the great advantage of working with functions is

that the shift from two to three dimensions does not increase the complexity of the expressions or the calculations.

RKHSs (usually, scalar-valued) have a long history in the statistical and machine learning literature. RKHSs have been largely used to facilitate statistical modeling and estimation. For example, in the 1940s, probabilists had already begun to employ Hilbert space methods to clarify the structure of time series (Parzen, 1961). Preda (2007); Yuan and Cai (2010) and Cai and Yuan (2012) use the RKHS framework in prediction problems under which the unknown slope function is assumed to reside in a reproducing kernel Hilbert space $H(K)$ but originally data are in L^2 with the usual inner product. In the literature of Support Vector Machines, RKHSs are used to map original data in a higher dimensional space (Cristianini and Shawe-Taylor, 2000; Steinwart and Christmann, 2008). RKHSs provide a convenient framework for efficient computation.

An important and crucial difference between previous applications of RKHSs and ours is that, in our case, original data are functions in a given vector-valued Reproducing Kernel Hilbert Space with a previously defined inner product, that is generally different from the L^2 inner product. For example, polynomial functions can never be included in an RKHS with a Gaussian kernel (Steinwart and Christmann, 2008).

Nowadays, the theory of statistics with functional data is an important field of research in statistics. It is used when data are in an infinite-dimensional function space. Although this theory is often a generalization of classic parametric or multivariate statistics, the infinite-dimensional nature of the sample space poses particular problems. Key references in the FDA literature are the books by Ramsay and Silverman (2005) and Ferraty and Vieu (2006). A more theoretical treatment is taken in Hsing

and Eubank (2015).

With respect to the particular problem of regression with a scalar response and a functional predictor, the first papers focused on the continuous version of the multiple linear model, the functional linear model. In these cases, a direct estimation of the parameter function of the proposed functional regression models through the use of least squares methods is not possible. The most commonly used approximated solution for this estimation problem is to consider that functional observations belong to a space generated by a basis of functions and to perform a multiple treatment based on this approach. Different bases have been used in the literature, such as spline functions, trigonometric functions or wavelet functions (see Ramsay and Silverman, 2005, and the references therein).

A slightly different approach, and one of the most popular in the literature, is to use principal component functional regression. This approach uses the orthonormal basis of eigenfunctions of the covariance function (Cardot et al., 1999). Unlike the previous ones, this is a data-driven basis. As in the multivariate case, this technique makes use of the data covariance function to determine the subspace where the data are projected. This subspace is spanned by the data covariance eigenfunctions and it is always an RKHS. This approach solves a typical problem in functional regression: the great dependence between coefficients, that causes that the estimation of the model is not very accurate.

Nevertheless, as Morris (2015) advises, care must be taken when using principal component functional regression for very complex, high-dimensional functional data for which the decay rate in the eigenvalues is slow, especially when the number of functions is small. In certain high-dimensional, low-sample-size settings, PCs have

been shown to be inconsistent. These problems can be at least partially mitigated using functional principal component analysis with regularization (Yuan and Cai, 2010).

In the same spirit as in the multivariate setting, functional generalized linear models (Escabias et al., 2004), which are the functional version of generalized linear models (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989), were introduced more recently in the literature. Such models are based on similar ideas to the linear case. In James (2002) the predictors are modeled as cubic splines, and in Cardot and Sarda (2005) the functional coefficient of the functional generalized linear model is estimated via penalized likelihood with spline approximation.

As our data (functions in a vector-valued RKHS), are not expressed in the standard form of functional data, in Barahona et al. (2017a) we investigated how to obtain an appropriate orthonormal basis of the vector-valued RKHS in which to project our data. The proposed basis was the orthonormal basis given by the integral operator defined by its reproducing kernel (Quang et al., 2010).

On the other side, Dou et al. (2012) use the functional principal component analysis (FPCA) approach. However, when we use this approach in logistic or multinomial functional generalized linear models, we face an additional problem to the one mentioned previously. Logistic or, in general, multinomial generalized linear models, are used to solve classification problems and, as explained in Jolliffe (2002, chap. 9), when PCA is used in classification problems in order to reduce the dimensionality of the analysis, we have to be aware that there is no guarantee that the separation between groups will be in the direction of the high-variance PCs; the separation between groups may be in the directions of the last few PCs.

On the other hand, as González (2010, chap. 3) notes, an improvement in the classification results can be achieved by using other kernels that capture nonlinear dependencies between the data, because the covariance function only deals with linear ones.

We can also explore this approach and consider the integral operator from the covariance function regarding our data as a realization of a stochastic process. The relationship between L^2, our original RKHS and the RKHS defined from the covariance function provides another non-orthonormal base. Moreover, we will prove a result on simultaneous diagonalization that provides an alternative basis system that combines the properties of both.

Therefore, the main aim of this work arises in the analysis of the ordinal variable "fit of each garment" (classified as "too small", "correct fit" or "too large"). It turns into a regression problem with an ordinal response and a functional predictor, that can be considered as an intermediate problem between regression and classification. To estimate the coefficients of a regression model with functional predictors, it is usual to consider that functional observations belong to a space generated by a basis of functions and to perform the analysis based on this approach. The last novelty of our work will consist of comparing the performance of three different bases in which to project our data. Our implementations have been written in MATLAB (2015) and R (R Core Team, 2018).

The article is organized as follows: Section 2 provides a detailed explanation of the practical case that motivated this work. Section 3 introduces the concepts of currents and Reproducing Kernel Hilbert Spaces. Section 4 reviews the basis of functional generalized linear models and Section 5 gives the different bases of functions in the

RKHS. The application for predicting children's garment fit is detailed in Section 6. Finally, conclusions are discussed in Section 7.

2 Motivating example

In the current process of buying children's clothing online, consumers base the size selection on their previous experience or on the size chart that is normally included in the online store. The consumer's previous experience is not usually very reliable, because each brand uses its own sizing system that usually evolves over time according to the needs of each company. Size tables indicate the ranges of the main anthropometric measures covered by each size. This method is also unreliable, because taking measurements at home is subject to significant errors, and ambiguous, because users can fit into different sizes according to the measurements used. The final result is a high percentage of returns on children's clothing sold online, meaning that many consumers are reluctant to buy through this channel, thus increasing the cost of sales.

Selection based on the child's anthropometry seems the most appropriate approach to predict garment size and fit in the child population. However, the 3D anthropometry acquisition systems currently available have several drawbacks. 3D body scanners are too expensive for home use. Parametric avatars configured from manual measurements by the user present three important sources of inaccuracy: they are not based on statistics of real populations but on models and proportions that prevail in aesthetics, the number of measurements entered does not depend on the type of garment selected or the critical measurements for the associated adjustment, and the measurements are taken by an untrained user, with the error that this may entail.

Low-cost systems that use domestic technology to capture body measurements have not yet achieved sufficient precision for size allocation or prediction of fit.

Ergonomic childrenswear design and size definition processes have several differences with regard to those of adult apparel. Firstly, childrenswear size designation is usually labeled in ages, which is not a body measurement, so it is usually related to a specific body height per age, which may not necessarily be close to a child of that age, due to the high variability of height by age in children. According to the European standard UNE-EN 13402-3, the 3 to 12 years age range has 10 different sizes associated with it (950-1010 mm, 1010-1070 mm, 1070-1130 mm, 1130-1190 mm, 1190-1250 mm, 1250-1310 mm, 1310-1370 mm, 1370-1430 mm, 1430-1490 mm and 1490-1550 mm).

As online clothes shopping is a problem for both the customer and the apparel industry, in recent years both national administrations and industrial groups from the clothing sector have been developing national anthropometric surveys in different countries. Emerging technology for body scanning has also promoted these new sizing surveys.

In order to help solve all these problems, the Biomechanics Institute of Valencia (IBV) started an ambitious research project in 2004, of which this work is a part. This project has two objectives: first, to develop a system for capturing the child body's 3D morphometry that is precise, easy to use and can be done at home. Second, to build a model to predict how a given garment size of garment fits a child based on the aforementioned 3D reconstruction. Our work in this article focuses on the latter.

With respect to the first objective of the project addressed by the IBV group, an application has already been developed. The system reconstructs the body of the

child in 3D from two or three photographs taken with domestic technology (smartphone, tablet or digital camera) using models representative of the European infant morphometry as a base for reconstruction (Ballester et al., 2016).

To achieve both objectives, a 3D anthropometric study of the child population in Spain was conducted in 2004. In this study, a randomly selected sample of Spanish children between the ages of 3 and 12 years was scanned using a Vitus Smart 3D body scanner from Human Solutions, a non-intrusive laser system which performs a sweep of the body. Several cameras capture images and associated software provided by the scanner manufacturer provides information about the 3D spatial location of up to 200000 points on the body surface. 3D scan data was processed for the creation of posture harmonized homologous models to obtain a database of individual 3D homologous avatars with anatomical one-to-one vertex correspondence among them. Next all the scans were rigidly aligned (Ballester et al., 2014).

Seventy-eight of these children of different ages performed an additional fit test, where they tested up to three different consecutive sizes of the same shirt model: the supposedly correct size, the size above and the size below. Then, an expert in clothing and design evaluated each fit qualitatively (as small, correct fit or large). There were 7 possible shirt sizes available, nominally corresponding to ages 3, 4, 5, 6, 8, 10 and 12. In 24 cases, only two sizes were evaluated. In 18 of these cases, the children tested either had a correct shirt size corresponding to ages 3 or 12. The 6 remaining cases with only two sizes evaluated were due to lack of cooperation by the children. Additionally, 9 children tested just one shirt size because the age 12 size was too small for them or the age 3 size was too large for them.

So, our data set contains the 3D body scans of a total of 78 children (37 boys and

41 girls, between 3 and 12 years old). It also includes the expert's opinion on the goodness of the fit of different (consecutive) sizes of the same shirt model on the children, codifying the goodness of fit as -1 (if the shirt is too small), 0 (for a good fit) or 1 (if the shirt is too big). The total number of expert observations is 192 (3 evaluations of 45 children, 2 evaluations of 24 children and 1 evaluation of 9 children).

As the children's head, hands, legs and feet do not come into play in shirt size selection, these parts were discarded from the scans, and a total of 1423 points representing the remaining surface per child were considered. This amount of detail is enough to characterize a child for our purposes, while keeping the time and memory requirements to perform the calculations reasonable. These points were grouped into 2766 triangles forming a mesh. The body contour from each child in our data set was therefore represented by an oriented triangulated smooth surface, S_k (see Fig. 2).

3 3D geometrical objects as elements in a Reproducing Kernel Hilbert Space

Currents are mathematical elements that can be used to model general geometrical objects (Durrleman et al., 2009; Barahona et al., 2017b). In this paper, they will be used to model the bodies of the children in our data set.

Let D be a compact set in $\mathbb{R}^3$ and $K : D \times D \longrightarrow \mathbb{R}^{3\times 3}$, the matrix-valued kernel associated with a vector-valued RKHS $H_K(D, \mathbb{R}^3)$.

The current representation of a surface $S \subset D$ is defined by the integral of K along the surface:

$$C_S(y) = \int_S K(x, y)(\tau(x))\, dx.$$

Where $\tau(x)$ is the normal vector to the surface S at point x (Durrleman et al., 2009; Barahona et al., 2017b).

In the discrete setting, the vectors $\tau(x)$ are constant over each mesh cell. Then, if x_j is located at the center of mass of mesh cell j, and τ_j is $\tau(x_j)$ scaled by the size of the mesh cell,

$$S \longrightarrow C_S(\cdot) \cong \int_S K(x, \cdot)(\tau(x))\, dx \approx \varphi = \sum_j K(x_j, \cdot)(\tau_j), \tag{3.1}$$

So, the vector field φ_k associated with each surface S_k will be defined on a different set of points $\{x_k\}$ given by the centers of mass of the respective mesh cells. Using the "Representer Theorem" (Cucker and Smale, 2001), given $\{a_i\}_{i=1}^N$ a sample grid in D, we can find a smooth function $\overline{\varphi_k}$, defined as:

$$\overline{\varphi_k} = \sum_{i=1}^{N} K(a_i, \cdot)(\beta_i^k), \tag{3.2}$$

where $\overline{\varphi_k}(a_i)$ is closest to $\varphi_k(a_i)$ (Barahona et al., 2017a). $\beta_i^k \in \mathbb{R}^3$ for $i = 1, \cdots, N$.

The inner product of two geometrical objects represented as currents is given by the inner product of the corresponding elements in $H_K(D, \mathbb{R}^3)$.; that is, if $\overline{\varphi_1} = \sum_j K(a_j^1, \cdot)(\beta_j^1)$ and $\overline{\varphi}_2 = \sum_j K(a_j^2, \cdot)(\beta_j^2)$ are two elements in the RKHS, associated with two surfaces S_1 and S_2, then

$$\langle \overline{\varphi_1}, \overline{\varphi_2} \rangle_{H_K} = \sum_j \sum_l \beta_j^1 \cdot K(a_j^2, a_l^1)(\beta_j^2),$$

where $\cdot$ is the inner product in $\mathbb{R}^3$.

From now on, the vector fields in $H_K(D, \mathbb{R}^3)$ will be named functions and $H_K(D, \mathbb{R}^3)$ will be denoted H_K for the sake of simplicity. The space of quadratic integrable vector fields from D to $\mathbb{R}^3$ with the Lebesgue measure will simply be denoted by L^2.

Moreover, a Gaussian kernel, $K(x, y) = k(x, y)I_{3\times3} := \exp(\frac{-\|x-y\|^2_{\mathbb{R}^3}}{\lambda^2})I_{3\times3}$ will be used in the definition of the operator-valued reproducing kernels K, where $I_{3\times3}$ denotes the identity matrix (Barahona et al., 2017a).

4 The functional generalized linear model

As mentioned in the introduction, the aim of this paper is to predict the fit of a particular garment on a child as "too small" ($Y = -1$), "good fit" ($Y = 0$) or "too big" ($Y = 1$), given different predictors, including the child's surface, which, as shown in the previous section, is modeled as a current.

Generalized linear models (GLMs) (McCullagh and Nelder, 1989) provide a natural generalization of classical linear models. Given a dependent variable Y, with $E(Y) = \mu$, they assume that it is distributed following a probability function in the exponential family (not necessarily a Gaussian distribution), and the relationship between predictors $\varphi_1, \ldots, \varphi_m$ and response Y is modeled by means of a link function g as:

$$g(\mu) = \alpha + \sum_{i=1}^{m} \sigma_i \varphi_i \tag{4.1}$$

where g can be any monotonic differentiable function, and α and $\sigma = (\sigma_1, \ldots, \sigma_m)$ are the parameters to estimate.

When the response variable is a score, representing an ordered category, i.e. when

Y may take one of several discrete ordered values indexed as $1, \ldots, J$ with probabilities $\pi_1, \ldots, \pi_J$: $\sum_{j=1}^{J} \pi_j = 1$, these probabilities can be modeled using cumulative logits (Agresti, 2010) as:

$$\text{logit}[P(Y \leq j)] = \alpha_j + \sum_{i=1}^{m} \sigma_i \varphi_i, \qquad \forall j \in \{1, \ldots, J-1\}. \tag{4.2}$$

GLMs provide a very flexible class of procedures. However, they assume that the predictors have a finite dimension. That is why James (2002) extended GLMs to functional generalized linear models (FGLMs), which directly model the relationship between a single response from any member of the exponential family of distributions and a functional predictor. Then, when the predictor $\varphi(\cdot)$ is functional, as in our case, the link given by Eqs. (4.1) and (4.2) cannot be applied directly, but a natural generalization is to replace the summation over the finite-dimensional space with an integral over the infinite-dimensional one. Then, if we were working with functional predictors in L^2, the cumulative logit model (Eq. 4.2) would become:

$$\text{logit}[P(Y \leq j)] = \alpha_j + \int \sigma(x)\varphi(x)\, dx, \forall j \in \{1, \ldots, J-1\},$$

where $\sigma(\cdot)$ is the functional analogue of σ in Eq. (4.2). But, in our case, we have functional predictors in H_K, so:

$$\text{logit}[P(Y \leq j)] = \alpha_j + \langle \sigma, \varphi \rangle_{H_K}, \forall j \in \{1, \ldots, J-1\}. \tag{4.3}$$

The most widely used approach to estimate these models considers that the functions $\sigma(\cdot)$ and $\varphi(\cdot)$ belong to spaces generated by bases of functions (see Ramsay and Silverman, 2005). In our case, $\sigma(\cdot)$ and $\varphi(\cdot)$ belong to a common space (a vector-valued RKHS), so if $\{\phi_l(\cdot)\}_{l=1}^{\infty}$ is a basis of this space, there will exist coefficients

$\{c_l\}_{l=1}^{\infty}$, $\{b_l\}_{l=1}^{\infty}$ such that:

$$\varphi(x) = \sum_{l=1}^{\infty} c_l \phi_l(x); \qquad \sigma(x) = \sum_{l=1}^{\infty} b_l \phi_l(x).$$

In practice, these developments are truncated, and $\varphi(x)$ and $\sigma(x)$ are usually approached by a summation of a finite number of terms, as:

$$\varphi(x) \cong \sum_{l=1}^{r} c_l \phi_l(x); \qquad \sigma(x) \cong \sum_{l=1}^{r} b_l \phi_l(x). \tag{4.4}$$

Ideally, these basis functions should have similar features to the functions being estimated. Different bases have been used in the literature, such as trigonometric functions, spline functions (Aguilera et al., 1996), wavelet functions (Ocaña et al., 1998) or the orthonormal basis of eigenfunctions of the covariance function (Cardot et al., 1999).

Then from Eqs. (4.3) and (4.4), the link function of the cumulative logit models can be written as:

$$\text{logit}[P(Y \leq j)] = \alpha_j + \sum_{p=1}^{r}\sum_{l=1}^{r} b_p c_l \langle \phi_p, \phi_l \rangle_{H_K}, \qquad \forall j \in \{1, \ldots, J-1\}. \tag{4.5}$$

And then:

$$P(Y \leq j) = \frac{\exp\left(\alpha_j + \sum_{p=1}^{r}\sum_{l=1}^{r} b_p c_l \langle \phi_p, \phi_l \rangle_{H_K}\right)}{1 + \exp\left(\alpha_j + \sum_{p=1}^{r}\sum_{l=1}^{r} b_p c_l \langle \phi_p, \phi_l \rangle_{H_K}\right)}, \ \forall j \in \{1, \ldots, J-1\}, \tag{4.6}$$

where $\{\alpha_j\}_{j=1}^{J-1}$ and $\{b_l\}_{l=1}^{r}$ are the parameters to estimate.

If the basis is orthonormal for H_K, then Eq. (4.6) becomes:

$$P(Y \leq j) = \frac{\exp\left(\alpha_j + \sum_{l=1}^{r} b_l c_l\right)}{1 + \exp\left(\alpha_j + \sum_{l=1}^{r} b_l c_l\right)}, \qquad \forall j \in \{1, \ldots, J-1\}.$$

5 Bases of functions in H_K

This section focuses on the methodological novelties of our work. It is interesting to remember at this point that we are working in a vector-valued RKHS, i.e. our data are vectorial fields. Theoretical properties about bases in scalar RKHSs are well known and scalar RKHSs and their bases have been largely used in the statistical literature. However, nowadays theoretical properties of vector-valued RKHSs in general and their bases in particular are a research field in the functional analysis literature and, as far as we know, they have never been used in classical statistical applications. Vector-valued RKHSs were used in Image colorization problems, a particular case of a mathematical extension problem (Quang et al., 2010).

It should be noted that in conventional functional data analysis applications, the original data are functions in the L^2 Hilbert space and RKHSs are used to facilitate statistical modeling and estimation. For instance, Preda (2007); Yuan and Cai (2010); Cai and Yuan (2012) use the RKHS framework in prediction problems under which the unknown slope function is assumed to reside in a reproducing kernel Hilbert space $H(K)$ with a reproducing kernel K, but originally the data are in L^2 with the usual inner product. As mentioned previously, in our case the original data are functions in a given vector-valued Reproducing Kernel Hilbert Space with the previously defined inner product, which is different from the L^2 inner product, and this is an important difference between conventional functional data analysis applications and ours.

It is known that the eigen structures of integral operators in an RKHS provide bases of functions; this idea will be explored in the following section. Moreover, when working with Hilbert spaces, compact operator can be approximated by finite-dimensional

operators (matrix) and, as a result, exhibits similar properties. This make easier the implementation of the models.

5.1 Basis from the operator integral of the kernel

Let $L_K \colon L^2 \longrightarrow L^2$ be the integral operator of the kernel K in the space H_K, defined by

$$L_K f(x) := \int_D K(x,y)(f(y))\, dy.$$

Since L_K is a compact, continuous, self-adjoint, positive operator, there are eigenvalues $\{\lambda_l\}_{l=1}^{\infty}$ and the corresponding eigenfunctions $\{\psi_l\}_{l=1}^{\infty}$ of L_K, with $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ and $\lim_{l\to\infty} \lambda_l = 0$ (Hsing and Eubank, 2015).

Moreover,

$$\langle \psi_i, \psi_j \rangle_{L^2} = \delta_{ij}, \qquad \text{and} \qquad \langle \psi_i, \psi_j \rangle_{H_K} = \delta_{ij}/\lambda_i,$$

where δ_{ij} is the Kronecker delta.

Therefore, if we denote $\rho_l = \sqrt{\lambda_l}\psi_l$, $\{\rho_l\}_{l=1}^{\infty}$ is an orthonormal basis for H_K, and the hypersurfaces S_k in Eq. (3.2), as elements of H_K, can be represented as

$$\overline{\varphi_k} = \sum_{l=1}^{\infty} \langle \overline{\varphi_k}, \rho_l \rangle_{H_K}\, \rho_l = \sum_{l=1}^{\infty} \mu_l^k\, \rho_l. \tag{5.1}$$

The basis $\{\rho_l\}_{l=1}^{\infty}$ is determined by the RKHS where our functions are included and a generalization to our vector-valued case of Theorems 4.4.7 and 4.6.8 in Hsing and Eubank (2015) can be proved. Because our functional data are of the form $\overline{\varphi_k}(x) = \sum_{i=1}^{N} K(a_i, x)(\beta_i^k)$, these results ensure that the truncated eigenvalue-eigenvector decomposition provides the best approximation to K and, as a result, the truncation of Eq. 5.1 reduces the dimension in an optimal way.

This basis was previously used in Barahona et al. (2017a) in a Supervised Classification problem. As usual in practice, the coefficients are estimated using the matrix approach to the kernel function (Barahona et al., 2017a).

In particular, we consider $K|_a$ the matrix defined as $(K|_a)(i,j) = k(a_i, a_j)$, $i,j = 1, \cdots, N$. $v_q \in \mathbb{R}^N$ are the eigenvectors of $K|_a$ and ℓ_q are the eigenvalues of $K|_a$. In Barahona et al. (2017a) it was proved that if

$$\overline{\varphi}_k(\cdot) = \sum_{i=1}^{N} K(a_i, \cdot)(\beta_i^k) = \sum_{q=1}^{\infty} \sum_{j=1}^{3} \mu^k_{3(q-1)+j} \left(\sqrt{\lambda_q} \psi_q^j(\cdot) \right), \tag{5.2}$$

where $l = 3(q-1)+j$ and $\rho_l = \left(\sqrt{\lambda_q}\psi_q^j(\cdot)\right)$ in Eq.5.1, the first coefficients μ_l^k can be approximated by

$$\widehat{\mu_l^k} = \sqrt{\ell_q}(v_q \cdot \beta^{k,j}), \text{ where } l = 3(q-1)+j$$

for $j = 1,2,3$, $q = 1, \cdots, d$ ($d = rank(K|_a)$) and where $\beta^{k,j}$ is the j-th column of the $N \times 3$-matrix β^k whose rows are the vectors β_i^k of Eq. 3.2

If we truncate the first summation in Eq. 5.2 to a low number r of terms, $r \leq d = rank(K|_a)$, each hypersurface S_k for $k = 1, \ldots, m$, is given by the coefficients $\mu^k_{q,j}$ for $j = 1,2,3$ and $q = 1, \ldots, r$ (estimated by $\widehat{\mu^k_{q,j}}$).

As a result, it can be represented as a $(3 \cdot r)$-dimensional vector.

5.2 Basis from the operator integral of the covariance function

Our functional data are realizations of random variables that take values in a vector-valued RKHS. Classical L^2 FPCA is based on the eigenvalue-eigenvector decomposition of the integral operator of the covariance function. We will show that this

operator is also well defined in the case of random elements in vector-valued RKHS and the Karhunen-Lóeve Theorem is fulfilled (Hsing and Eubank, 2015). The eigenvectors of the decomposition of this operator are different from those of the covariance operator (see Hsing and Eubank (2015) page 197) and, as a result, they do not form a base of H_K. We will prove that they are included in $H(K)$ and then we will calculate the inner products in Eq. 4.6 with respect to the H_K norm.

It is known that a random element of H_K is a stochastic process (Hsing and Eubank, 2015). Therefore, we could consider the covariance functions $\gamma_{ij}(x,y) := Cov(\Phi_i(x), \Phi_j(y))$, $\forall i,j = 1,2,3$. Let $\gamma(x,y)$ be the (3×3)-matrix whose elements are $\gamma_{ij}(x,y)$. Then $\Gamma(x,y)(\alpha) := \gamma(x,y)\alpha$ is a symmetric and nonnegative-definite vector-valued function. We consider the integral operator L_Γ:

$$L_\Gamma f(x) := \int_D \Gamma(x,y)(f(y))\, dy.$$

By again using the eigenvalue-eigenvector decomposition for a self-adjoint compact operator (Quang et al., 2010; Hsing and Eubank, 2015), the eigenfunctions of the operator L_Γ, $\{v_l\}_{l=1}^{\infty}$ form an orthonormal basis for L^2, that is, $\langle v_i, v_j \rangle_{L^2} = \delta_{ij}$.

Because $\overline{\varphi_k} \in H_K \subset L^2$:

$$\overline{\varphi_k} = \sum_{l=1}^{\infty} \langle \overline{\varphi_k}, v_l \rangle_{L^2} v_l = \sum_{l=1}^{\infty} \varsigma_l^k v_l. \tag{5.3}$$

As Γ is a symmetric and non-negative-definite function, we can consider the vector-valued RKHS associated with the kernel Γ, H_Γ (Aronszajn, 1950). Assuming that Γ is continuous, it is known that $H_\Gamma \subset H_K$ (Lukić and Beder, 2001), and then $v_j \in H_K$.

The inner products satisfy the following relationship with respect to the products in

L^2:

$$\langle v_i, v_j \rangle_{H_K} = \sum_{k=1}^{\infty} \lambda_k^{-1} \langle v_i, \psi_k \rangle_{L^2} \langle v_j, \psi_k \rangle_{L^2}. \tag{5.4}$$

The basis given by the integral operator defined from the covariance functions $\{v_l\}_{l=1}^{\infty}$ depends on the random sample, and the Karhunen-Lóeve Theorem guarantees that there are random variables $I_\Phi(v_l)$ with mean zero, decreasing variances and uncorrelated such that:

$$\lim_{n\to\infty} \sup_{t\in D} E[\|\Phi(t) - \sum_{l=1}^{n} I_\Phi(v_l) v_l(t)\|] = 0,$$

and as a result, the truncated development of Eq. 5.3 is optimal in this regard.

As usual in practice, the coefficients ς_l^k are estimated using the matrix approach to the covariance function (Barahona et al., 2017a).

In particular, the coefficients have been estimated by applying PCA method to a $m \times 3N$-matrix X. This matrix contains in each row the evaluations in the grid $\{a_i\}_{i=1}^{N}$ of a vector field of the database. To calculate the inner products of Eq. 5.4 we use the corresponding eigenvectors.

It is important to note that although it could seem more natural to consider the eigenfunctions of the operator $L_\Gamma : H_K \longrightarrow H_K$ we work with have to work with the operators defined from L^2 to L^2 for several reasons. Firstly, because if we define the operator from H_K to H_K we can not estimate, as far as we know, the coefficients with respect to the base. In a discrete setting the product of L^2 is equivalent to the product of matrices or vectors and it is important to emphasize that despite the sets inclusion $H_K \subset L^2$, we are considering in H_K the inner-product of H_K and in L^2 the one of L^2. Another reason is related to the good theoretical properties of the integral operators defined in L^2.

This basis has some drawbacks, as mentioned in the introduction. When we use the PCA approach in logistic or multinomial functional generalized linear models, there is no guarantee that the separation between groups will be in the direction of the high-variance PCs. Moreover, following the covariance procedure, only linear relations are captured (González, 2010).

5.3 Mixed basis

In the preceding subsections we have seen that we can express a function $\overline{\varphi_k}$ representing a hypersurface S_k as an infinite linear combination with respect to two different bases: $\{v_l\}_{l=1}^{\infty}$, which depends on the random sample and is related to the covariance operator, and $\{\rho_l\}_{l=1}^{\infty}$, which is determined by the kernel that defines the RKHS. Both of them have different optimality properties. The aim of this section is to obtain a new basis $\{u_l\}_{l=1}^{\infty}$, from a relationship between both operators L_K and L_Γ. This expression will represent a compromise between the optimality given by the sample information and the one given by the RKHS in which our functions are included.

Define the linear operator $G := L_K^{\frac{1}{2}} \circ L_\Gamma \circ L_K^{\frac{1}{2}}$ and let $\{\eta_j, w_j\}_{j=1}^{\infty}$ be the eigenvalue-eigenvector pairs of G.

As a consequence of Mercer's theorem, we have $\mathrm{Im}(L_K^{\frac{1}{2}}) = H_K$ (Quang et al., 2010). Then, since $H_\Gamma \subset H_K$, we have $\mathrm{Im}(G) \subseteq H_K$.

Note that if we assume that the operators L_K and L_Γ are perfectly aligned, that is, they share the same ordered set of eigenfunctions $\{\psi_j\}_{j=1}^{\infty}$, then

$$G(\psi_k) = l_k \lambda_k \psi_k.$$

In this case, $\eta_j = l_k \lambda_k$ and the three bases $\{\psi_j\}_{j=1}^{\infty}$, $\{v_j\}_{j=1}^{\infty}$ and $\{w_j\}_{j=1}^{\infty}$ used in the application coincide.

From the basis $\{w_l\}_{l=1}^{\infty}$ we will define in Theorem 5.1, a new set of functions in H_K, $\{u_l\}_{l=1}^{\infty}$, which generate all the functions in H_K, and from which we will obtain a relationship ("simultaneous diagonalization") between the operators L_K and L_Γ.

Theorem 5.1 *Let $\{\eta_j, w_j\}_{j=1}^{\infty}$ be the eigenvalue-eigenvector pairs of G.*
Define $u_j := L_K^{\frac{1}{2}}(w_j)$, $\forall j$.

$$\textit{Then, } \forall f \in H_K, \quad f = \sum_{j=1}^{\infty} \xi_j \, u_j, \quad \textit{where} \quad \xi_j = \langle f \ , \ L_K^{-1} \, u_j \rangle_{L^2},$$

$$L_\Gamma \, u_j = \eta_j \, L_K^{-1} \, u_j$$

and

$$\langle u_i \ , \ L_\Gamma \, u_j \rangle_{L^2} = \eta_j \, \delta_{ij}.$$

Proof. Details regarding operators and proof of the theorem can be found in the appendix. □

As in the preceding subsection, the products of vectors u_j, which are not necessarily orthonormal with respect to the L^2-metric, satisfy a similar equation to Eq. 5.4.

Let denote K to a $3N \times 3N$-matrix, which is the discrete estimation of the integral operator L_K. It is defined by N blocks 3×3. The block i, j of the matrix is $K(a_i, a_j) = k(a_i, a_j) \times I_{3\times 3}$, with $i, j = 1, \cdots, N$ and $\{a_i\}_{i=1}^{N}$ the grid. If X is the $m \times 3N$-matrix that contains in each row the sampling of a vector field of the database in

the grid $\{a_i\}_{i=1}^N$ and $G = K^{1/2} \times cov(X) \times K^{1/2}$, the implementation of this basis is based on obtaining the eigenvectors ϖ_j, $j = 1, \cdots, 3N$ of the matrix G, which estimate w_j functions. Then, it is possible to estimate the basis functions u_j by $K^{1/2} \times \varpi_j$, $j = 1, \cdots, 3N$ and to calculate the coefficients by solving the matrix system $X = \varepsilon \times U$. U is a $3N \times 3N$-matrix containing each $K^{1/2} \times \varpi_j \in \mathbb{R}^{3N}$ in a row, and ε is the $m \times 3N$-matrix whose rows are the coefficients of each vector field in the database in relation to the mixed basis.

Similar bases to those used in Subsections 5.2 and 5.3 have been considered in Yuan and Cai (2010) and Cai and Yuan (2012), but in a different framework. In these papers the authors consider a prediction problem in FDA where the functional predictor is a real function defined over a domain in $\mathbb{R}$, and the slope function is also a real function which is assumed to reside in a real-valued RKHS. In our case, both the functional predictor and the slope are vector-valued functions in a vector-valued RKHS. Then, a metric defined in the vector-valued RKHS is used instead of the L^2-metric.

In Section 1 we noted that PC basis can fail in capturing directions that are best suited for a classification problem. As with this criticism, we can never be sure which basis performs better in each particular supervised problem. Our advise is to try the different alternatives and choose the one that provides better cross validation results. In section 6.3 a experimental study is done to illustrate that.

Given a 3d-triangulated surface, the source code to compute its coefficients in the three bases of functions described in the paper can be found as supplementary material through the link: `http://www.statmod.org/smij/archive.html`.

6 Application

We revisit the motivating example presented in Section 2 and apply the proposed modeling approach.

As stated in Section 2, the points representing the surface of each child were grouped into 2766 triangles, forming a mesh. If a_j, b_j, c_j denote the vertices of the j-th oriented triangle for a child k, the center of this triangle was defined as $x_j^k = (a_j + b_j + c_j)/3$ and its area vector (that is, its unit normal vector, scaled by its area) was $\tau_j^k = (b_j - a_j) \times (c_j - a_j)$, $\forall j = 1, \ldots, 2766$ (see Fig. 1).

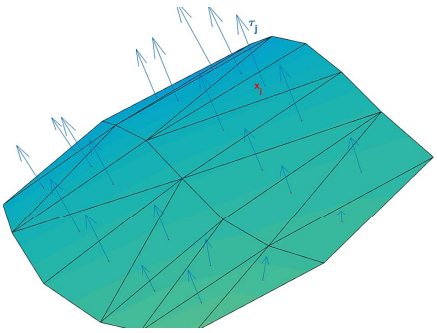

Figure 1: Section of the triangulated surface in $\mathbb{R}^3$, centers of the triangles and area vectors.

Then, each child's body surface was associated with a function $\varphi_k = \sum_{j=1}^{2766} K(x_j^k, \cdot)(\tau_j^k)$ in H_K, where the points x_j^k differ from one hypersurface to another. All these vector fields were represented in the same sample grid of points $\{a_i\}_{i=1}^N$ chosen in the compact subset $D \subset \mathbb{R}^3$, so that each φ_k was approximated by a smooth function $\overline{\varphi_k} = \sum_{j=1}^N K(a_j, \cdot)(\beta_j^k)$ evaluated on this common grid (Sec.3). In this case we considered $D = [-472.73, 487.27] \times [-824.72, 735.28] \times [-156.70, 203.30]$ and the grid was defined on this domain considering a set of points separated by a fixed gap $\Delta = 200$

in the three dimensions. So a grid with $N = 90$ points was obtained (see Fig. 2).

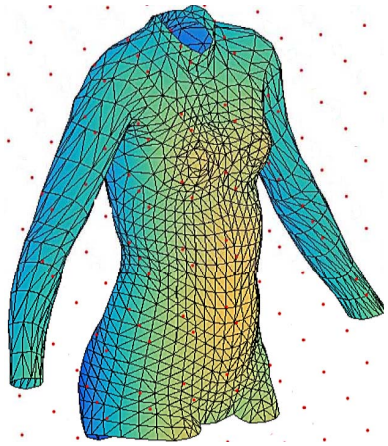

Figure 2: Triangulated surface that represents the body contour from the upper body of a child in the sample. Red points represent the common grid $\{a_i\}_{i=1}^{N}$.

A Gaussian kernel, $K(x, y) := \exp(\frac{-\|x-y\|^2_{\mathbb{R}^3}}{\lambda^2})I_{3\times 3}$ was used in the definition of the operator-valued reproducing kernels K (Eq. 3.1) and the value of the parameter λ was chosen by cross-validation.

The robustness of this representation regarding the pre-processing parameters and the kernel assumption has been discussed in (Barahona et al., 2017b) and Barahona et al. (2017a), using a data base with much simpler geometrical objects and so with much less computational cost.

In Section 5, it has been seen that we can express a function $\overline{\varphi_k}$ by representing a surface S_k as an infinite linear combination with respect to a basis of H_K that is determined by the RKHS (Sec. 5.1), with respect to a basis of H_K that depends on the random sample (Sec. 5.2), and with respect to a basis of H_K found from both non-negative-definite operators (Sec. 5.3).

As our aim is to predict the goodness of fit of a given shirt size for the k-th child as

small ($Y_k = -1$), good fit ($Y_k = 0$) or large ($Y_k = 1$) as a function of the garment size, the age of the child, his/her body shape and his/her sex, we will get three different models depending on the basis used to express the children's body surface in the RKHS. Following the notation used in Section 4, let us denote the basis used each time with $\{\phi_l\}_{l=1}^{\infty}$, and the corresponding coefficients for the k-th surface with $\{c_l^k\}_{l=1}^{\infty}$. $\forall l = 1, \ldots, \infty$ we will consider *'case 1'*, where $\phi_l = \rho_l$ and $c_l^k = \langle \overline{\varphi_k}, \rho_l \rangle_{H_K}$ (as in Eq. 5.1), *'case 2'*, where $\phi_l = v_l$ and $c_l^k = \langle \overline{\varphi_k}, v_l \rangle_{L^2}$ (as in Eq. 5.3), and *'case 3'*, where $\phi_l = u_l$ and $c_l^k = \langle \overline{\varphi_k},\ L_K^{-1}\ u_l \rangle_{L^2}$ (as in Theorem 5.1). Since much of the information inherent in the original data is captured by the first few functional components and their associated coefficients, these bases are truncated with a low number r of terms. In *'case 1'*, the approach $\overline{\varphi(x)} \cong \sum_{l=1}^{r} c_l \phi_l(x)$ is truncated considering $r = 7$ elements in the basis. In *'case 2'*, $r = 8$ elements are considered in the basis and in *'case 3'* $r = 7$ elements are also considered. Then, Eq. (4.5) to include functional and non-functional predictors in the model become:

$$\begin{aligned} \text{logit}[P(Y_k \leq j)] =& \alpha_j + \beta_1 shirt.size + \beta_2 sex_k + \beta_3 age_k + \\ &+ \sum_{p=1}^{r} \sum_{l=1}^{r} b_p c_l^k \langle \phi_p, \phi_l \rangle_{H_K}, \forall j \in \{-1, 0\}, \forall k \end{aligned}$$

with $\{\alpha_j\}_{j \in -1,0}$, β_1, β_2, β_3 and $\{b_p\}_{p=1}^{r}$, parameters to estimate.

This model assumes independent observations, but in our case several measurements are taken on each child, so this model is modified to overcome this fact into the following cumulative link mixed model (Agresti, 2010):

$$\begin{aligned} \text{logit}[P(Y_{k,i} \leq j)] =& \alpha_j + \beta_1 shirt.size_i + \beta_2 sex_k + \beta_3 age_k + \\ &+ \sum_{p=1}^{r} \sum_{l=1}^{r} b_p c_l^k \langle \phi_p, \phi_l \rangle_{H_K} + u(k), \\ & \forall i = 1, \cdots, n_k; j \in \{-1, 0\}; k = 1, \cdots, 78, \end{aligned} \tag{6.1}$$

where $n_k \in \{1, 2, 3\}$ is the number of observations taken on the k-th child, and the child's effects are assumed to be random, independent and identically distributed following a Gaussian distribution, i.e. $u(k) \sim N(0, \sigma_k^2)$.

The *clmm* function of the R-package *ordinal* (Christensen, 2015) is used to fit the model in all three cases.

Parameters $\beta = (\beta_1, \beta_2, \beta_3)$ in Eq. (6.1) are clearly interpretable unlike the parameters $\{b_p\}_{p=1,\cdots,r}$ associated to the different basis, that are not.
The estimation procedure provides very similar estimates for the three interpretable parameters with the three different representations (the three different basis): $\hat{\beta}_{case1} = (2.576, -0.741, -0.603)$; $\hat{\beta}_{case2} = (2.249, -0.207, -0.512)$ $\hat{\beta}_{case3} = (2.253, -0.211, -0.523)$. As $\beta_1 > 0$, each cumulative logit increases as the evaluated shirt's size increases and equivalently, as $\beta_3 < 0$, each cumulative logit decreases as the child's age increases, as was expected. Moreover, given the difference in the signs of the parameters, fixed a sex, $\beta_1 shirt.size + \beta_3 age_k$ becomes somehow a measure distance between the age of the k-th child and the size of the garment tested.

Additionally, a leave-one-out cross-validation (CV) analysis is performed to check the predictive power of the model for the different bases. For each basis, the model is iteratively estimated taking into account all the data except for the observations available for each child in turn and a prediction is made for the different observations of this child (i.e. the model selection is performed within the CV and the predictions are completely out-of-sample). A percentage of agreement between predictions and real observations is computed and used to evaluate the different cases (see table 1).

As can be seen, although in the functional data analysis literature, the basis of eigen-

		Case 1			Case 2			Case 3		
		Prediction			Prediction			Prediction		
		-1	0	1	-1	0	1	-1	0	1
Expert decision	-1	47	16	1	51	12	1	51	12	1
	0	10	41	10	11	38	12	10	39	12
	1	1	12	54	0	13	54	1	11	55
% of agreement		73.95%			74.48%			75.52%		

Table 1: Results of the cross-validation procedure for the mixed ordinal regression models estimated with the different bases.

functions of the covariance function is one of the most commonly used to approach functional observations, and to work with them, we have found two additional bases, and we have obtained a similar predictive power for this particular application with all of them. Although the differences are quite small, the results are slightly better with the mixed basis, as initially expected.

Other link functions (probit, cloglog and loglog) have been tested using the function *pom* of the R-package OCAPIS (Heredia-Gómez et al., 2018a), getting similar results.

6.1 Robustness against the number of points in the grid

To evaluate the influence of the gap Δ that determines the number of points in the grid, we have repeated the previous procedure for two additional values: $\Delta = 100$ and $\Delta = 250$. The experimental results shown in Table 4 reflect that the procedure is robust enough and the fact that better results are obtained from a balance between small and large values of Δ.

Δ	Points in the grid	% of agreement		
		Case 1	Case 2	Case 3
100	450	71.88%	69.79%	71.35%
200	90	73.96%	74.48%	75.52%
250	60	70.31%	69.79%	70.83%

Table 2: Results of the cross-validation procedure for different values of Δ.

6.2 Comparison with other methods

Functions *svmo* and *wknnor*, implemented in the R-package OCAPIS (Heredia-Gómez et al. (2018b)), allow us to compare our results with those obtained with other classifiers such as the Support Vector Machine with Ordered Partitions (SVMOP) (Waegeman and Boullart, 2009), and the weighted k-Nearest Neighbors classifier for ordinal data (Hechenbichler and Schliep, 2004). We have checked the performance of these algorithms, from the estimated coefficients (corresponding to the three bases) of the current-based representation of the geometrical objects, separately for boys and girls. Table 3 shows the percentages of correct classifications obtained in each case. The best results for the k-NN algorithm have been achieved for $k = 5$. As can be seen, in all cases, the percentages of agreement is lower than the obtained with the method proposed in this work (Table 1).

The European Standard CEN - EN 13402-3, establishes tables for body measurements and intervals to be used for compiling standard garment sizes for men, women, boys, girls and infants. Additionally, each brand usually has its own sizing chart that relates consecutive ranges of the main anthropometric measurements with size assignations. So, several new technologies and online services have been developed in

			Case 1	Case 2	Case 3
% of agreement	SVMOP classifier	Boys	29.35%	35.87%	34.78%
		Girls	25.27%	31.87%	31.87%
	k-NN classifier (k=5)	Boys	46.74%	53.26%	47.82%
		Girls	48.35%	50.55%	49.45%

Table 3: Results of the SVMOP classifier and of the k-NN classifier ($k = 5$), with the different bases.

recent years to address the selection of proper garment sizes or models based on the user's anthropometric measurements (see, for instance, `www.fits.me`). In the case of T-shirts, shirts and/or blouses, the main anthropometric dimensions established by the European standard are height, chest and neck for boys, and height and bust for girls.

In Pierola et al. (2016), the authors used ordered logistic regression and random forest methodologies to predict a garment's goodness of fit from the differences between the measurements of the reference mannequin for the evaluated size and the child's anthropometric measurements.

Following this line, we could also have used different children's anthropometric measurements to fit a mixed proportional odds model, as in McCullagh (1980). So if $Y_{k,i}$ denotes the response as (small (-1)/good (0)/large (1)) fit of a shirt size i for the $k-th$ child, and X_k denotes a vector of explicative variables formed by the sex of the $k-th$ child, his/her age and the values of the 27 anthropometric measurements considered by Pierola et al. (2016), we could have fitted:

$$\text{logit}[P(Y_{k,i} \leq j)] = \alpha_j + \beta_1 shirt.size_i + \beta X_k + u(k), \tag{6.2}$$
$$\forall i = 1, \cdots, n_k; j \in \{-1, 0\}; k = 1, \cdots, 78,$$

where once again $n_k \in \{1, 2, 3\}$ is the number of observations taken on the k-th child, and the child's effects are assumed to be random, independent and identically distributed following a Gaussian distribution, i.e. $u(k) \sim N(0, \sigma_k^2)$.

Performing a leave-one-out cross-validation study, choosing the model on each step by a forward stepwise model selection based on likelihood ratio tests (Christensen, 2015), we obtain worse results than those obtained with our methodology. The percentage of correct classifications is now 68.27% (see table 4).

As in our case, many problems in medical imaging analysis and computer vision involve the classification of bodies (geometrical objects with bounded boundaries), based on their size and shape. Several mathematical frameworks have been proposed in the literature to deal with such objects, three of these being the most widely used. Firstly, functions can be used to represent closed contours of the objects (curves in 2D and surfaces in 3D). Secondly, geometrical objects can be treated as compact subsets of $\mathbb{R}^m$ and, finally, these geometrical objects can be characterized by sequences of points with certain geometrical or anatomical properties (landmarks).

In our application, we are working with rigidly aligned 3D homologous avatars with anatomical one-to-one vertex correspondence among them (Ballester et al., 2014), so we can consider these 1423 points as landmarks, and define $X_k \in M_{1423\times 3}$ as the configuration matrix of the k-th child. As the shape of an object is all the geometric information that remains invariant with translations, rotations and changes of scale,

the shape space and size-and-shape space are not flat Euclidean spaces, so classical statistical methods cannot be directly applied to the manifold-valued data (Pennec, 2006). However, if the sample has little variability, the problem can be transferred to a tangent space (at the Procrustes mean of these shapes or size-and-shapes, for example) and then standard multivariate procedures can be performed in this space (Dryden and Mardia, 2016), such as Principal Component Analysis (PCA), where the first p PC scores, which summarize most of the variability in the tangent plane data, are usually chosen in order to reduce the dimensionality of the data set.

The tangent space is defined from a point called pole, so the distance from the shape to the pole is preserved. As one moves away from the pole, the Euclidean distances between some pairs of points in the tangent space are smaller than their corresponding shape distances. This distortion becomes larger as one considers points further from it. For this reason, the pole should be taken close to all of the points and the mean of the observed shapes is the best choice (Dryden and Mardia, 2016).

So, given the configuration matrices $X_k \in M_{1423\times 3}$, the size s_k of each child is obtained and the full Procrustes mean shape is computed. Then, the coordinates of the projection of $X_k \in M_{1423\times 3}$ onto the tangent plane defined at its corresponding mean shape is obtained. The first PC scores of these coordinates are calculated and they will be used as covariates in our predictive model. The first PC components that explain 98% of the variability are considered.

So, given the response variable Y with 3 ordered categories and given the garment size to evaluate and a vector X with the child's size, his/her sex, his/her age and the first PC scores of his/her coordinates in the tangent space, we can fit the model given by Eq. 6.2.

Once again, performing a leave-one-out cross-validation study using this model, we obtain worse results than those obtained with our methodology. The percentage of correct classification is now 66.67 (see Table 4).

		Multivariate			Landmarks		
		Prediction			Prediction		
		-1	0	1	-1	0	1
Expert decision	-1	47	12	3	48	14	2
	0	13	31	15	14	34	13
	1	1	15	49	0	21	46
% of agreement		68.27%			66.67%		

Table 4: Results of the cross validation procedure for the two alternative methods tested

The outperformance of our method compared with Procrustes landmark methods might be due to the parametrization dependence of these methods. In this sense, other methods which need a far greater computational complexity, as Large Deformation Diffeomorphic Metric Mapping (LDDMM) or Diffeomorphic Demon techniques (Beg et al., 2005; Vercauteren et al., 2007) could give more accurate results. However, we consider that this complexity is not necessary for the purpose of this work.

6.3 Experimental study of basis performance

Two types of experimental studies are done in this section to obtain more information on the behavior of the different bases for classification. In order to simplify the experiments, a problem of binary classification is considered instead of an ordinal

one.

Firstly, we want to see that, as it was said in the introduction, the PCA basis may fail in a classification problem because there is no guarantee that the separation between groups will be in the direction of the high-variance PCs; the separation between groups may be in the directions of the last few PCs.

In order to show this behavior, two samples of size 200 of currents were simulated. All them were simulated from a multivariate gaussian distribution with the sample covariance function of our data set (child's data set), and two different means. The mean of the samples of each group is taken as the sample mean, plus and minus respectively the p-th vector of the basis from the operator integral of the covariance function (Sec. 5.2). The values of p were $p = 1, 5, 40, 80$.

Thereafter, we obtain the expression of all these simulated currents with respect to the tree different bases stated in Sec. 5 of H_K. Finally the classification of each current of each group is predicted using as explicative variables the first 10 and the fist 20 coefficients in the bases. In Table 5 the goodness of the classification is given. As can be seen, if we define the groups using the the first's vectors of the basis from the covariance kernel ($p = 1$ and $p = 5$) all the bases provide very good classification rates. When the direction is defined by the 40-th vector, i.e. the direction that the defines the groups is further away from the first vectors, the basis from the reproducing kernel provides better results. When we take the last vectors all the bases provide bad results.

Unfortunately, the currents simulated in this first experiment could not correspond to a random surface, because the application that corresponds to each surface a current

		first 10 coefficients	first 20 coefficients
basis 1	$p = 1$	0.9700	0.9650
	$p = 5$	0.9125	0.9350
	$p = 40$	0.6000	0.7375
	$p = 80$	0.4425	0.4650
basis 2	$p = 1$	0.9700	0.9675
	$p = 5$	0.9575	0.9625
	$p = 40$	0.4700	0.4725
	$p = 80$	0.4825	0.4675
basis 3	$p = 1$	0.9700	0.9675
	$p = 5$	0.9575	0.9625
	$p = 40$	0.4775	0.4800
	$p = 80$	0.4825	0.4750

Table 5: Results of the first experimental study

is not surjective. As a result, no illustration of the representatives of each group can be shown. For this reason, we conduct another similar experimental study using the landmark-based shape methodology of Section 6.2.

In this experiment, two hundred shapes are generated from a multivariate Gaussian distribution in the tangent space of the mean shape of Section 6.2, all of them using the covariance matrix of the tangent coordinates. In a similar way to the previous experiment, the mean is taken using the tangent coordinates of the mean shape, plus and minus respectively the p-th PC vector multiplied by twice its standard deviance (Dryden and Mardia, 2016). The values of p were $p = 1, 5, 40, 50$. In figure 3 it can be seen the total sample mean shape and, in Figure 4 the mean shape of each simulated group. We analyze this synthetic sample of surfaces using the methodology of

Figure 3: Total sample mean shape

Section 6.2. As the Tangent Space is a finite Euclidean space it is easier to analyze the results. In Table 6 we can find the proportion of agreement between actual and predicted classification, when we use as information the coefficients of the surfaces expressed on the PC basis on the Tangent Space. In the first case we use the first 15 coefficients and in the second case the coefficients from 16 to 30. We can also

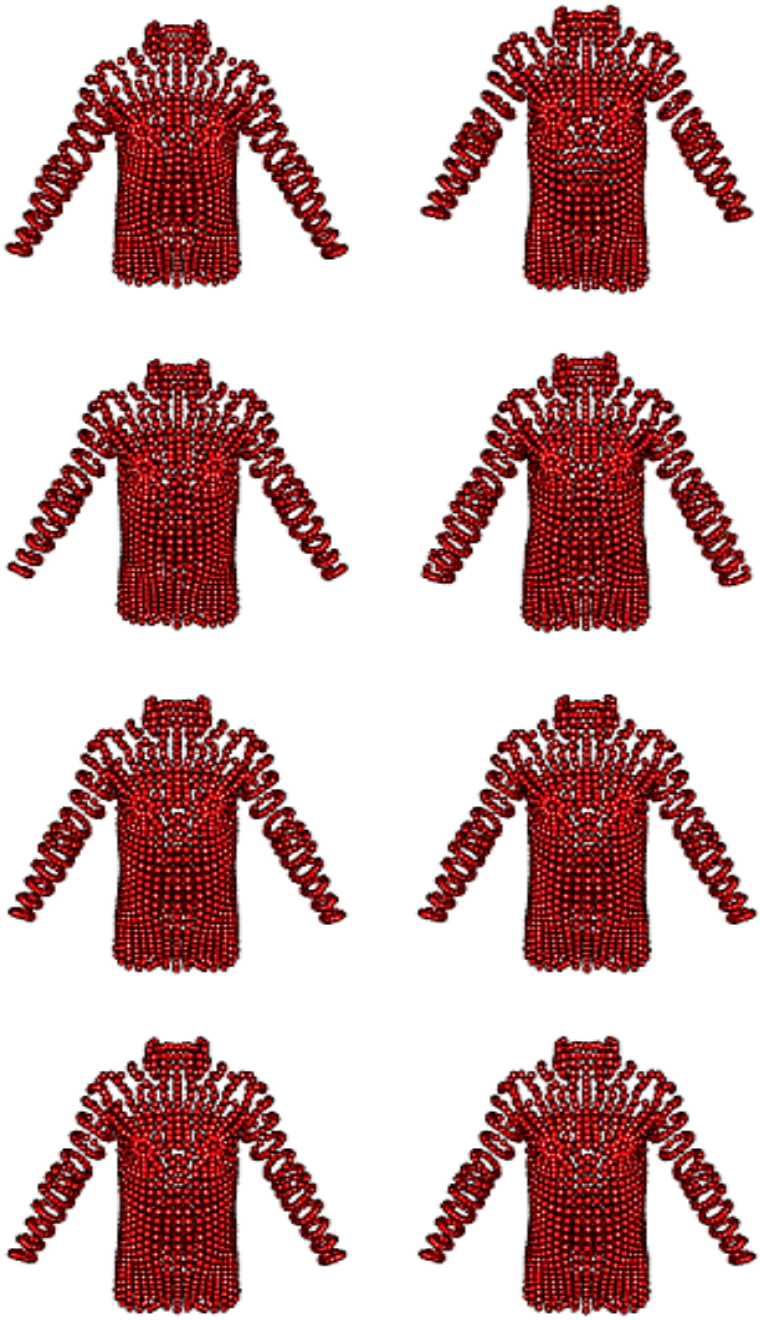

Figure 4: Means of each simulated group of experiment two. Each row corresponds to the p-th eigen vector with $p = 1, 5, 40, 50$. Left column shows the mean of the first group and right column the mean of the second

see the percentage of variability explained using these coefficients. We can see, again

	first 15 coefficients		coefficients from 16 to 30	
	agreement	variablity	agreement	variablity
$p = 1$	0.986	95.7	0.320	2.9
$p = 5$	0.970	92.1	0.354	5.4
$p = 40$	0.568	90.3	0.930	6.7
$p = 50$	0.550	90.3	0.926	6.6

Table 6: Results of the second experimental study.

and more clearly, that there is no guarantee of success when we use the first PC's directions in classification problems. With lower values of p the first PC's directions provide excellent results but with greater values of p, the directions of high-variance don't provide a good separation between groups but good results are provided by subsequent PC's directions.

Finally a third experiment has been performed, quite different from the previous ones.

A sample of currents of size 400 is simulated taking the 10 first vectors of each of the three bases of Section 6, and with random coefficients extracted from a standard gaussian distribution. We have to note that in this way, our simulated data are in a subspace of H_K of dimension 10.

For each sample, taking these simulated currents φ_k as explicative variables, a sample of 400 binary variables Y_k was simulated with mean:

$$P(Y_k = 1) = \frac{1}{1 + \exp -\langle \sigma, \varphi_k \rangle_{H_K}},$$

being σ a functional parameter previously set. Different values of σ have been proved

with similar classification results, we show the results with $\sigma = (1, ..., 1)$.

The results of this experiment don't show clear differences between bases and no clear conclusions can be derived from them. We guess that this could be due to the low dimensionality of the data. The means of 10 repetitions are shown in Table 7.

	first 2 coefficients of			first 6 coefficients of		
	basis 1	basis 2	basis 3	basis 1	basis 2	basis 3
vectors basis 1	0.805	0.800	0.812	0.975	0.927	0.935
vectors basis 2	0.795	0.712	0.720	0.827	0.757	0.797
vectors basis 3	0.490	0.510	0.522	0.520	0.535	0.535

Table 7: Results of the third experimental study

7 Discussion

In this paper we have proposed a new methodology for modeling an ordinal response variable in terms of 3D geometrical objects. It is based on their characterization by means of currents and the expression of each geometrical object in terms of three different bases of functions that generate the corresponding vector-valued RKHS.

Firstly, the predictors were expressed in the orthonormal basis given by the kernel of the RKHS. Secondly, we used a basis obtained in a similar way as the usual basis given by the covariance kernel in the scalar setting. Thirdly, a basis of functions that connect the benefits of the two previous expressions was sought by "simultaneous diagonalization" of operators. The coefficients of each geometrical predictor in relation to the three bases of functions were estimated. The goodness of the method was

checked by leave-one-out cross-validation.

Then, it was applied to predict whether the size fits a customer or is too large or small for him/her, which is useful for an application to online clothing sales. This was done using a 3D training database obtained from an anthropometric survey of the Spanish child population. The results were quite promising, taking into account the difficulty of the application. Although the results obtained are quite similar with the three bases tested, as expected they are slightly better with the mixed basis. In near future this methodology could be incorporated into the mobile application (kidsize) recently developed by the Biomechanical Institute of Valencia.

We compare our methodology with alternative methods that explore the performance of other classifiers (Suppport Vector Machine and k-NN), and with another two methods traditionally used in biometric size determination. The first of these traditional methods is based on considering children's anthropometric measurements and classical ordinal regression. The second is based on landmark configuration and its projection in the tangent space where classical multivariate statistical methods can be applied. In all cases, the classification results obtained are slightly worse than those obtained with the methodology developed in this paper.

It is important to note that the success of the three bases proposed in our methodology depends on the data and the application. We illustrate with an experimental study that we can never be sure which basis performs better in each particular supervised problem. Our suggestion to practitioners is to check the performance of all of them for each particular problem.

Acknowledgements

This paper has been partially supported by the Spanish Ministry of Science and Innovation Project DPI2013-47279-C2-1-R with FEDER funds. We would also like to thank the Biomechanics Institute of Valencia for providing us with the data set.

A Theorem 5.1

Before starting with the proof of the Theorem 5.1, let us recall some definitions regarding operators. If an operator $\mathcal{F}$ is compact, self-adjoint and nonnegative, there is a unique nonnegative operator $\mathcal{F}^{\frac{1}{2}}$, called the square-root of the operator, such that $(\mathcal{F}^{\frac{1}{2}})^2 = \mathcal{F}^{\frac{1}{2}} \circ \mathcal{F}^{\frac{1}{2}} = \mathcal{F}$, and it commutes with any operator that commutes with $\mathcal{F}$. When an operator $\mathcal{F}$ is bijective, there is an operator $\mathcal{F}^{-1}$, called the inverse operator, such that $\mathcal{F}^{-1} \circ \mathcal{F}$ and $\mathcal{F} \circ \mathcal{F}^{-1}$ are the identity operator.

Due the fact that L_K is compact, positive-definite and self-adjoint, the operator $L_S^{\frac{1}{2}} \colon L^2 \longrightarrow L^2$ can be defined, and it works as follows:

$$L_K^{\frac{1}{2}}(f) := \sum_{j=1}^{\infty} \sqrt{\lambda_j} \langle f, \psi_j \rangle_{H_K} \psi_j.$$

$L_K^{\frac{1}{2}}$ is also compact, positive-definite and self-adjoint.

Note that L_K and $L_K^{\frac{1}{2}}$ are injective because they are positive-definite (their eigenvalues are strictly positive). Hence, restricting their arrival spaces to their respective images converts them into bijective operators on pre-Hilbert spaces. These restricted

operators will also be denoted by L_K and $L_K^{\frac{1}{2}}$; then, there are the inverse operators o L_K^{-1}, $L_K^{-\frac{1}{2}}$, and we have $L_K^{-\frac{1}{2}} \circ L_K^{-\frac{1}{2}}\ f = L_K^{-1}\ f,\ \forall f \in H_K$.

Proof of Theorem 5.1.

Since $u_j := L_K^{\frac{1}{2}}(w_j)$ and $G\, w_j = \eta_j\, w_j$, we prove that, for all j.

$$\begin{aligned} L_\Gamma\ u_j = \ L_\Gamma\ L_K^{\frac{1}{2}}\ w_j = L_K^{-\frac{1}{2}}\ L_K^{\frac{1}{2}} L_\Gamma\ L_K^{\frac{1}{2}}\ w_j = \ L_K^{-\frac{1}{2}}\ G\ w_j \\ = L_K^{-\frac{1}{2}}\ \eta_j\ w_j = \eta_j\ L_K^{-\frac{1}{2}}\ L_K^{-\frac{1}{2}}\ u_j = \eta_j\ L_K^{-1}\ u_j. \end{aligned}$$

Moreover, $L_K^{\frac{1}{2}}$ is self-adjoint and $\{w_j\}_{j=1}^{\infty}$ is an orthogonal system. Thus,

$$\begin{aligned} \langle u_i\ ,\ L_\Gamma\ u_j\rangle_{L^2} = \ \langle L_K^{\frac{1}{2}}\ w_i\ ,\ L_\Gamma\ L_K^{\frac{1}{2}}\ w_j\rangle_{L^2} = \langle w_i\ ,\ L_K^{\frac{1}{2}}\ L_\Gamma\ L_K^{\frac{1}{2}}\ w_j\rangle_{L^2} \\ = \langle w_i\ ,\ \eta_j\ w_j\rangle_{L^2} = \eta_i \delta_{ij}. \end{aligned}$$

Finally, if $f \in H_K$,

$$\begin{aligned} f = L_K^{\frac{1}{2}}\ \left(L_K^{-\frac{1}{2}}\ f\right) = L_K^{\frac{1}{2}} \left(\sum_{j=1}^{\infty} \langle L_K^{-\frac{1}{2}}\ f\ ,\ w_j\rangle_{L^2}\ w_j\right) \\ = \sum_{j=1}^{\infty} \langle L_K^{-\frac{1}{2}}\ f\ ,\ w_j\rangle_{L^2}\ L_K^{\frac{1}{2}}\ (w_j) \\ = \sum_{j=1}^{\infty} \langle L_K^{-\frac{1}{2}}\ f\ ,\ L_K^{-\frac{1}{2}}\ u_j\rangle_{L^2}\ L_K^{\frac{1}{2}}\ \left(L_K^{-\frac{1}{2}}\ u_j\right) \\ = \sum_{j=1}^{\infty} \langle L_K^{-\frac{1}{2}}\ f\ ,\ L_K^{-\frac{1}{2}}\ u_j\rangle_{L^2}\ u_j = \sum_{j=1}^{\infty} \langle f\ ,\ L_K^{-\frac{1}{2}}\ L_K^{-\frac{1}{2}}\ u_j\rangle_{L^2}\ u_j \\ = \sum_{j=1}^{\infty}\ \langle f\ ,\ L_K^{-1}\ u_j\rangle_{L^2}\ u_j = \sum_{j=1}^{\infty} \xi_j\ u_j, \end{aligned}$$

$\{w_j\}_{j=1}^{\infty}$ being an orthogonal basis of L^2 with respect to the L^2-metric and using the definition of $\{u_j\}_{j=1}^{\infty}$. □

References

Agresti, A. (2010). *Analysis of ordinal categorical data*, volume 656. John Wiley & Sons.

Aguilera, A. M., Gutiérrez, R., and Valderrama, M. J. (1996). Approximation of estimators in the pca of a stochastic process using b-splines. *Communications in Statistics-Simulation and Computation*, **25**(3), 671–690.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, pages 337–404.

Baddeley, A. and Molchanov, I. (1998). Averaging of random sets based on their distance functions. *Journal of Mathematical Imaging and Vision*, **8**, 79–92.

Ballester, A., Parrilla, E., Uriel, J., Pierola, A., Alemany, S., Nacher, B., Gonzalez, J., and Gonzalez, J. C. (2014). 3d-based resources fostering the analysis, use, and exploitation of available body anthropometric data. In *5th international conference on 3D body scanning technologies*.

Ballester, A., Parrilla, E., Piérola, A., Uriel, J., Pérez, C., Piqueras, P., Nácher, B., Vivas, J. A., and Alemany, S. (2016). Data-driven three-dimensional reconstruction of human bodies using a mobile phone app. *International Journal of the Digital Human*, **1**(4), 361–388.

Barahona, S., Centella, P., Gual-Arnau, X., Ibáñez, M. V., and Simó, A. (2017a). Classification of geometrical objects by integrating currents and functional data analysis. An application to a 3D database of Spanish child population. *ArXiv e-prints 1707.02147*. URL http://arxiv.org/abs/1707.02147.

Barahona, S., Gual-Arnau, X., Ibáñez, M. V., and Simó, A. (2017b). Unsupervised classification of children's bodies using currents. *Advances in Data Analysis and Classification*, pages 1–33.

Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, **61**(2), 139–157.

Bookstein, F. (1978). Lecture notes in biomathematics. In *The measurement of biological shape and shape change*. Springer-Verlag.

Cai, T. T. and Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association*, **107**(499), 1201–1216.

Cardot, H. and Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *Journal of Multivariate Analysis*, **92**(1), 24–41.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, **45**(1), 11–22.

Christensen, R. H. B. (2015). ordinal—regression models for ordinal data. R package version 2015.6-28. http://cran.r-project.org/package=ordinal/.

Cristianini, N. and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Cucker, F. and Smale, S. (2001). On the mathematical foundations of learning. *American Mathematical Society*, **39**(1), 1–49.

Dou, W. W., Pollard, D., Zhou, H. H., et al. (2012). Estimation in functional regression for general exponential families. *The Annals of Statistics*, **40**(5), 2421–2451.

Dryden, I. L. and Mardia, K. V. (2016). *Statistical Shape Analysis: With Applications in R*. John Wiley & Sons.

Durrleman, S. (2010). *Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution*. PhD thesis, Université Nice Sophia Antipolis.

Durrleman, S., Pennec, X., Trouvé, A., and Ayache, N. (2009). Statistical models of sets of curves and surfaces based on currents. *Medical image analysis*, **13**(5), 793–808.

Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2004). Principal component estimation of functional logistic regression: discussion of two different approaches. *Journal of Nonparametric Statistics*, **16**(3-4), 365–384.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, **7**(2), 179–188.

Glaunes, J. A. and Joshi, S. (2006). Template estimation form unlabeled point set data and surfaces for computational anatomy. In *1st MICCAI Workshop on Mathematical Foundations of Computational Anatomy: Geometrical, Statistical and Registration Methods for Modeling Biological Shape Variability*.

González, J. (2010). *Representing functional data in reproducing Kernel Hilbert spaces with applications to clustering, classification and time series problems*. PhD thesis, Universidad Carlos III de Madrid.

Gual-Arnau, X., Herold-García, S., and Simó, A. (2013). Shape description from generalized support functions. *Pattern Recognition Letters*, **34**(6), 619–626.

Hechenbichler, K. and Schliep, K. (2004). Weighted k-nearest-neighbor techniques and ordinal classification. collaborative research center 386. In *University of Munich discussion paper*, volume 399, page 16.

Heredia-Gómez, M. C., García, S., Gutiérrez, P. A., and Herrera, F. (2018a). Ocapis: R package for ordinal classification and preprocessing in scala. *Progress in Artificial Intelligence*, pages 1–6.

Heredia-Gómez, M. C., García, S., Gutiérrez, P. A., and Herrera, F. (2018b). Ocapis: R package for ordinal classification and preprocessing in scala. *arXiv preprint arXiv:1810.09733*.

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons.

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(3), 411–432.

Jolliffe, I. (2002). *Principal component analysis*. Wiley Online Library.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, **16**(2), 81–121.

Kindratenko, V. (2003). On using functions to describe the shapes. *J. Math. Imaging Vision*, **18**, 225–245.

Loncaric, S. (1998). A survey of shape analysis techniques. *Pattern recognition*, **31** (8), 983–1001.

Lukić, M. and Beder, J. (2001). Stochastic processes with sample paths in reproducing kernel hilbert spaces. *Transactions of the American Mathematical Society*, **353**(10), 3945–3969.

MATLAB (2015). *version 8.6.0 (R2015b)*. The MathWorks Inc., Natick, Massachusetts.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)*, pages 109–142.

McCullagh, P. and Nelder, J. (1989). *Generalised linear models. Second edition.* London: Chapman and Hall.

Molchanov, I. (2006). *Theory of random sets*. Springer Science & Business Media.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application*, **2**, 321–359.

Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A*, pages 370–384.

Ocaña, F. A., Valenzuela, O., and Aguilera, A. M. (1998). A wavelet approach to functional principal component analysis. In *COMPSTAT*, pages 413–418. Springer.

Parzen, E. (1961). An approach to time series analysis. *The Annals of Mathematical Statistics*, pages 951–989.

Pennec, X. (2006). Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vis.*, **25**, 127–154.

Pierola, A., Epifanio, I., and Alemany, S. (2016). An ensemble of ordered logistic regression and random forest for child garment size matching. *Computers & Industrial Engineering*, **101**, 455–465.

Preda, C. (2007). Regression models for functional data by reproducing kernel hilbert spaces methods. *Journal of statistical planning and inference*, **137**(3), 829–840.

Quang, M. H., Kang, S. H., and Le, T. M. (2010). Image and video colorization using vector-valued reproducing kernel hilbert spaces. *Journal of Mathematical Imaging and Vision*, **37**(1), 49–65.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis. Second edition.* Springer.

Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press.

Shin, H. (2008). An extension of fisher's discriminant analysis for stochastic processes. *Journal of Multivariate Analysis*, **99**(6), 1191–1216.

Simó, A., De Ves, E., and Ayala, G. (2004). Resuming shapes with applications. *Journal of Mathematical Imaging and Vision*, **20**(3), 209–222.

Steinwart, I. and Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media.

Vaillant, M. and Glaunès, J. (2005). Surface matching via currents. In *Information Processing in Medical Imaging*, pages 381–392. Springer.

Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2007). Non-parametric diffeomorphic image registration with the demons algorithm. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 319–326. Springer.

Waegeman, W. and Boullart, L. (2009). An ensemble of weighted support vector machines for ordinal regression. *International Journal of Computer Systems Science and Engineering*, **3**(1), 47–51.

Yuan, M. and Cai, T. T. (2010). A reproducing kernel hilbert space approach to functional linear regression. *The Annals of Statistics*, **38**(6), 3412–3444.