

Recent advances in functional data analysis and high-dimensional statistics

Germán Aneiros^a, Ricardo Cao^a, Ricardo Fraiman^b, Christian Genest^c, Philippe Vieu^{d,*}

^aMODES, Departamento de Matemáticas, Facultad de Informática, CITIC,
Universidad da Coruña, Campus de Elviña, 15071 A Coruña, Spain and ITMATI

^bCentro de Matemática, Universidad de la República, Uruguay

^cDepartment of Mathematics and Statistics, McGill University, 805, rue Sherbrooke ouest, Montréal (Québec) Canada H3A 0B9

^dInstitut de Mathématiques de Toulouse, Université Paul-Sabatier, 118, route de Narbonne, F-31062 Toulouse Cedex 9, France

Abstract

This paper provides a structured overview of the contents of this Special Issue of the *Journal of Multivariate Analysis* devoted to Functional Data Analysis and Related Topics, along with a brief survey of the field.

Keywords: Functional data analysis, High-dimensional statistics.

1. Introduction

Functional data analysis (FDA) is a branch of statistics concerned with the analysis of infinite-dimensional variables such as curves, sets, and images. It has undergone a phenomenal growth over the past 20 years, stimulated in part by major advances in data collection technology that have brought about the “Big Data” revolution. Often perceived as a somewhat arcane area of research at the turn of the century, FDA is now one of the most active and relevant fields of investigation in data science. Today, nearly all methodological problems traditionally associated with classical multivariate analysis have a functional analog. The *Journal of Multivariate Analysis* (JMVA) gradually established itself as the main vehicle for the dissemination of research in FDA and this Special Issue reinforces the journal’s leadership position in this regard.

The purpose of this introductory article is to give a structured overview of this Special Issue on Functional Data Analysis and Related Topics, which is the third to appear in JMVA. The previous two were published in 2011 and 2016. This Special Issue comprises 19 research articles describing new methodological developments in the field. A companion Special Issue in *Computational Statistics* is devoted to numerical and/or applied issues in FDA [6]. Some papers included here were first presented at the Fourth International Workshop on Functional and Operatorial Statistics (IWFOS) held in A Coruña, Spain, June 15–17, 2017. Other publications related to these regular meetings include [5, 25].

In the overview that follows, the various contributions to this Special Issue have been divided into two broad categories. Papers promoting new methodology for data varying over a continuum are described in Section 2, and papers concerned with very high- but finite-dimensional problems arising, e.g., in the analysis of “Big Data” are discussed in Section 3. Some articles actually qualify on both counts and had to be classified somewhat arbitrarily. Connections between FDA and high-dimensional statistics have already been highlighted, e.g., in [2]. A recent Special Issue of *Statistics & Probability Letters* has also helped to bring the two fields closer together; see [107].

For background material about FDA, the reader can refer to textbooks such as [26, 44, 59, 63, 74, 103, 108, 113] and broad surveys such as [35, 51, 73, 91, 112]. Current developments on various aspects of FDA are also reported in numerous review papers, notably regression analysis [43, 56, 104], functional principal component analysis [43], clustering [65], depth analysis [92, 95], spatial functional data [40, 87, 89], dependent functional data [72], nonparametric modeling [49, 82], semiparametric modeling [50, 111], and testing [34, 61]. The statistical literature

*Corresponding author

URL: ganeiros@udc.es (Philippe Vieu), rcao@udc.es (Philippe Vieu), fraimanricardo@gmail.com (Philippe Vieu), christian.genest@mcgill.ca (Philippe Vieu), vieu@math.univ-toulouse.fr (Philippe Vieu)

on complex data also includes contributions to the analysis of multi-functional objects [55], statistics on manifolds [52, 100], and object-oriented data analysis [83, 84, 89].

2. Contributions to functional data analysis

This Special Issue assembles contributions covering a wide range of statistical problems involving functional data that can be roughly broken down as follows.

2.1. Mathematical foundations

As emphasized, e.g., in [63, 96], methodological research in FDA is rooted in probabilistic and analytical methods for infinite-dimensional spaces. Operator theory plays a key role and this Special Issue includes a paper by Boudou and Viguier-Pla [30] which studies the algebraic properties (more specifically commutativity) of operators in Hilbert spaces. It extends previous work from the same group of researchers [28, 29], who are leaders in the area.

2.2. Covariance operator estimation

Since the origins of FDA [103], the covariance operator has played a central, multifaceted role in Functional Principal Component Analysis (FPCA), as illustrated in [43], but also in functional classification and clustering [65], as well as in linear [57], semi-linear [78] and nonparametric [44] functional regression. The covariance operator has a prominent role in contributions [22] and [77] to this Special Issue. In a third contribution [1], it arises as a by-product of more general methodology.

In [22], Boente, Rodríguez, and Sued introduce a modification of the covariance operator inspired by the signed covariance matrix from classical multivariate analysis [42]. They use this signed covariance operator to design a new type of FPCA, state the asymptotic properties of its empirical analog, and show its value for testing in two-sample problems. In [77], Li, Huang, and Härdle rely on the standard covariance operator to derive new estimators leading to empirical eigen-elements (and then to FPCA) that are appropriate for dealing with spatial FDA. Covariance operator estimation is also studied in the paper by Aaron, Cholaquidis, Fraiman, and Ghattas [1] which describes a robust procedure for situations in which it is impossible to deal with the whole sample because of its exceedingly large size. This general fusion method will be further discussed in Section 3.

2.3. Functional depth

The notion of depth was originally proposed as an exploratory tool for analyzing multivariate data; see [116] for the state of the art. The concept was recently extended and adapted in various ways to functional data contexts; see [47] for early results, [36] for a general approach, and [92] for a recent survey. Two contributions to this Special Issue are concerned with functional depth. In [93], Nagy and Ferraty investigate the frequently encountered case where functional variables are observed with noise. Depth also plays a role in the fusion method proposed by Aaron, Cholaquidis, Fraiman, and Ghattas [1] discussed in Section 3. Very recently, Fraiman et al. [45] also published in JMVA a paper investigating the notion of depth in the broader framework of data taking values in Riemannian manifolds. This paves the way, inter alia, to FDA for data that cannot be modeled as elements of a Euclidean space.

2.4. Functional autoregressive processes

There is now a great deal of literature on modeling functional time series by means of autoregressive processes; see [26] for early work and [85] for a survey. This Special Issue includes two contributions related to this theme. One important question when dealing with such models is the estimation of the covariance operator. While early writings focused on Hilbert-valued autoregressive processes, extensions to Banach-valued processes have now been developed; see [27]. In their contribution to this Special Issue, Ruiz-Medina, and Álvarez-Liébaná [106] derive new consistency results for the autocorrelation operator of Banach-valued autoregressive processes and discuss the impact of these results for prediction. The second contribution, by Rossini and Canale [105], was motivated by a study of daily demand-and-supply curves from the Italian gas market. The specific constraints that such curves must satisfy in this context led them to propose a new functional autoregressive process with shape restrictions. A recent issue of JMVA also featured a paper by Allam and Mourid [3] on the theoretical properties of the empirical covariance and giving the optimal root- n rate of convergence when the coefficients involved in the autoregressive model are random.

2.5. Linear regression

As one might expect, the earliest extension of regression analysis to FDA settings was in the linear case; see [43] for a recent discussion. This Special Issue includes one contribution to functional linear modeling in which Berrendero, Bueno-Larraz, and Cuevas [18] consider the estimation of impact points. As their work involves a sparse functional linear regression model, it is discussed in greater detail below with other papers dealing with sparsity issues.

2.6. Semi-parametric regression

Semi-parametric and other regression models have been widely studied over the past few years as intermediate approaches between linear and nonparametric solutions to the dimensionality reduction problem; see [50, 111] for surveys. Partial linear modeling ideas were introduced in [8]; see [81] and references therein for recent advances. This Special Issue includes one contribution to this topic in which Zhu, Zhang, Yu, Lian, and Liu [115] study a partial functional linear model in the general case where the covariates are subject to error measurement. These authors develop statistical methodology for estimating the model parameters and for testing purposes.

2.7. Nonparametric regression

Nonparametric modeling in FDA has been popularized by Ferraty and Vieu [44] and the area has been widely developed in the last decade; see [49, 82] for surveys. Two contributions to nonparametric FDA are featured in this Special Issue. Most of the literature on functional regression concerns kernel or k NN smoothers; see [68, 69, 75] for recent advances. In contrast, the contribution by Benhenni, Hajj Hassan, and Su [15] concerns local linear type estimators of the regression operator for dependent functional samples. In their paper, short- and long-term dependence scenarios are considered and broad asymptotic results are stated for local linear functional estimators. The second contribution, by Chaouch [33], also pertains to nonparametric functional regression modeling. Its aim is to estimate the conditional variance operator (rather than the regression itself), which can be interpreted in econometrics as volatility. Papers on nonparametric functional conditional variance estimation are relatively rare but include [7, 64]. One of the main novelties in Chaouch's paper is that his asymptotic results allow for highly dependent errors.

2.8. Spatial functional statistics

Developing statistical methodology for spatial functional samples is a challenging topic both from a mathematical perspective and because of the wide scope of applied scientific fields that face such issues; see [40, 87, 89] for surveys. This Special Issue includes two contributions to this topic, one on spatial regression and another on spatial FPCA. The first paper, [13], is from recent contributors to the spatial FDA literature; see, e.g., [80] and references therein. Specifically, Arnone, Azzimonti, Nobile, and Sangalli [13] propose a functional version of the spatial regression model [101] involving a time dependent partial differential equation structure. Coincidentally, their work is motivated by a medical image analysis problem, as is the second contribution by Li, Huang, and Härdle [77]. The latter paper develops a new way of performing FPCA. A smoothing step (based on B-splines) is proposed to construct a new covariance operator estimator for which asymptotic theory is developed. This smooth estimator leads to new estimators of the eigen-elements of the operator and, in turn, to a new kind of FPCA that can capture the spatial feature of the data.

2.9. Robust functional data analysis

This Special Issue welcomes a contribution by Alvarez, Boente, and Kudraszow [4] on robust statistics. For earlier contributions to the area, including in the FDA setting, see, e.g., [14, 23]. The present paper revisits canonical correlation analysis in a functional framework by combining sieves and robustness ideas; this is especially relevant to the detection of influential observations. Asymptotics are stated and finite-sample studies are presented.

2.10. Sparsity in FDA

Sparsity effects are related to the dimension of the variables in the sample, and for this reason they appear under many guises in FDA; see [12] for a broad discussion on sparsity in FDA settings. In this Special Issue, sparsity issues arise within each functional observation in contribution [79] and through the model in contribution [18].

In FDA, sparsity is often a concern in relation to the discretized grid on which each curve is observed; see, e.g., [114] for recent advances. This is precisely how it occurs in the paper by Liebl [79], where a double-asymptotic study

is performed. In his work, the rates of convergence are expressed not only as a function of the sample size n , but also as a function of the size m of the discretization of each functional observation. Liebl explains how the asymptotics behave in terms of the ratio m/n and highlights the need to develop finite-sample bias and variance corrections.

Sparsity is also a concern in FDA when constructing dimensionality reduction models. Much progress has recently been made, mainly in a regression setting for linear [9, 70, 88], semi-parametric [10] or nonparametric [11] models, but also in classification or clustering [38], and more generally in shape analysis [46]. Reproducing kernel Hilbert spaces are of great interest [17]. In their contribution to this Special Issue, Berrendero, Bueno-Larraz, and Cuevas [18] use this tool to construct an alternative to the L_2 -criterion minimization for variable selection in sparse functional regression in the linear framework; the consistency and finite-sample behavior of their procedure are studied.

2.11. Tests for functional data

As in classical multivariate analysis, the question of developing testing procedures arises naturally in FDA. This Special Issue includes four papers which address this topic. Two of them pertain to a functional analog of the usual two-sample problem; see, e.g., [58, 60] for recent advances on this question, and [61] for a wide set of references. The contribution by Jiang, Hušková, Meintanis, and Zhu [67] proposes a new statistic, based on an integrated L_2 distance between the empirical characteristic functions of each group. The explicit terms involved in the asymptotic distribution are rather intricate, and resampling techniques are investigated, making the proposed statistic easy to implement and competitive with earlier methods for functional two-sample problems. The second contribution along these lines is by Boente, Rodríguez, and Sued [22]. Although their main purpose is covariance operator estimation, as mentioned in Section 2.2, their derivations lead to a new statistic for two-sample testing as a by-product.

The third contribution, by Pini, Spreafico, Vantini and Vietti [99], considers more than two functional samples. Motivated by a medical image analysis problem, these authors show how nonparametric inferential techniques lead to interval-wise testing (IWT) procedures; their approach can be used to test a very wide variety of hypotheses. General asymptotic results are provided which extend earlier work on functional IWT techniques by the same group of authors [97, 98], and the methodology is used to test mean and variance in more-than-two-sample problems.

The fourth contribution is by Zhu, Zhang, Yu, Lian, and Liu [115]. As mentioned in Section 2.6, partial functional linear regression is the focus of their work but the question of testing the scalar linear part of the model is investigated, and they propose a new statistic based on the comparison of the residuals under the null hypothesis and the alternative.

2.12. Analysis of estimated functional data

While FDA was originally conceived for the analysis of samples of observed curves, extensions have been considered to handle situations where the functional data to be analyzed are estimated probabilistic elements. An archetypical example concerns samples of densities; see [71] for early work and [39, 62] for recent advances. However, this idea was also developed for samples of distribution functions [94], samples of level sets [41], etc. In this Special Issue, Bongiorno and Goia [24] revisit the theory of FPCA for dealing with samples of Lorenz curves. These curves are of interest in econometrics but some of their characteristics pose mathematical challenges, e.g., in connection with the construction of a suitable vector space structure. In their paper, Bongiorno and Goia overcome these difficulties by means of embedding techniques leading to a new form of FPCA suitably adapted to Lorenz curves.

2.13. Dependent FDA

Many contributions discussed above account for possible dependence between functional samples; see [72] for a wide scope discussion. In some instances, the dependence structure is fully part of the model. This is the case in contributions [105, 106] to autoregressive functional time series analysis, but also in the paper by Gao, Shang, and Yang [48], which will be discussed in Section 3. The latter contribution also considers the broader setting where not just one but a large number of functional variables are observed over time. In its own way, the partial differential equation method featured in the contribution by Arnone, Azzimonti, Nobile, and Sangalli [13], already discussed in Section 2.8, is also geared to deal with dependent data.

In some other situations, a dependence structure is introduced in order to extend existing iid methodology to time series contexts. This is, in particular, the case of non/semi-parametric models. Early papers in this direction have considered mixing samples [44, 86] but extensions to more general dependence structures have now been considered, e.g., to ergodic samples [76] and long memory processes [16]. The contribution by Benhenni, Hajj Hassan, and Su [15] is part of this literature on highly dependent FDA.

3. Contributions to high-dimensional data analysis

As a way to foster links between FDA and high-dimensional statistics, this Special Issue includes some contributions on the latter theme, most notably the paper by Bodnar, Okhrin, and Parolya [20]. However, other papers are at the cross-roads between the two fields and are therefore highly representative of the wide variety of connections between FDA and “Big Data”. In the papers by Gao, Shang, and Yang [48] and Berrendero, Bueno-Larraz, and Cuevas [18], high-dimensional methods are adapted to FDA. In Aaron, Cholaquidis, Fraiman, and Ghattas [1], the connection comes from a general modeling approach that provides a unified view of the two fields.

3.1. High-dimensional mean vector estimation

In [20], Bodnar, Okhrin, and Parolya consider situations in which the number, p , of variables is (asymptotically) as high as the sample size, n , itself. This is not these authors’ first contribution to high-dimensional statistical methodology; see, e.g., [19, 21]. The present paper uses random matrix theory to derive optimality properties for a new estimator based on shrinkage ideas. These optimality properties go together with appealing finite-sample behavior, as illustrated through simulation studies.

3.2. High-dimensional functional time series analysis

In the contribution by Gao, Shang, and Yang [48], the high-dimensional aspect of the work lies in the fact that the statistical samples are composed of a large number of variables, each of which is possibly infinite-dimensional. The proposed methodology borrows both from the FDA and the high-dimensional statistical literature. FPCA is used to reduce the infinite-dimensional variables to vectors, and the latter are then treated with high-dimensional factor analysis techniques. Asymptotics are stated for this two-stage procedure and simulation studies compare its behavior in finite samples with alternative multi-functional times series analysis methods.

3.3. Fusion methods

When the sample size is large, it may prove infeasible to perform estimation over the whole sample but only over various subsamples. It is then natural to wonder what would be the best way of combining the subsample estimates. Such methodologies are known as fusion methods, and this Special Issue includes one such contribution by Aaron, Cholaquidis, Fraiman, and Ghattas [1]. Based on the notion of depth, these authors construct a robust fusion method for which both asymptotic properties and computational issues are studied.

3.4. Model selection

In high-dimensional statistics, the question of variable selection (or more generally model selection) has been one major topic in the last twenty years; see [31, 90] for broader discussions. In functional regression, the variable selection question arises, e.g., through the need to select the most informative discretized points of a curve, also called impact points. Contributions to sparse modeling presented in Section 2.10 could thus have been discussed here. In particular, the paper by Berrendero, Bueno-Larraz, and Cuevas [18] on sparse functional linear regression is both a contribution to FDA and to “Big Data” analysis.

4. Applications and numerical issues

The primary focus of this Special Issue is methodological. Nevertheless, many papers included herein provide concrete data illustrations. Applications of FDA and numerical issues surrounding its use are described, e.g., in the book by Ramsay and Silverman [102], the survey paper by González-Manteiga and Vieu [54], as well as Special Issues of applied journals such as [53, 110]. Today, FDA plays a key role in many subject-matter areas such as chemometrics [32], e-commerce [66], econometrics [73], management science [37], and medicine [109]. Many novel applications of FDA methodology can also be found in a companion Special Issue of *Computational Statistics* [6].

Acknowledgments

The first two authors acknowledge partial support from MINECO grants MTM2014-52876-R and MTM2017-82724-R (EU ERDF support included). They received additional financial support from the Xunta de Galicia (Centro Singular de Investigación de Galicia accreditation ED431G/01 2016-2019 and Grupos de Referencia Competitiva ED431C2016-015), as well as from the European Union (European Regional Development Fund - ERDF). The fourth author is grateful for continuing financial support from the Canada Research Chairs Program and the Natural Sciences and Engineering Research Council of Canada.

References

- [1] C. Aaron, A. Cholaquidis, R. Fraiman, B. Ghattas, Multivariate and functional robust fusion methods for structured Big Data, *J. Multivariate Anal.* **170** (2019) [in press](#).
- [2] S.E. Ahmed, *Big and Complex Data Analysis: Methodologies and Applications*, Springer International Publishing, Cham, Switzerland, 2017.
- [3] A. Allam, T. Mourid, Optimal rate for covariance operator estimators of functional autoregressive processes with random coefficients, *J. Multivariate Anal.* **169** (2019) 130–137.
- [4] A. Alvarez, G. Boente, N.L. Kudraszow, Robust sieve estimators for functional canonical correlation analysis, *J. Multivariate Anal.* **170** (2019) [in press](#).
- [5] G. Aneiros, E.G. Bongiorno, R. Cao, P. Vieu, An introduction to the 4th edition of the International Workshop on Functional and Operatorial Statistics, In: G. Aneiros, E.G. Bongiorno, R. Cao, P. Vieu (Eds.), *Functional Statistics and Related Fields: Contributions to Statistics*, Springer, Cham, Switzerland, 2017, pp. 1–5.
- [6] G. Aneiros, R. Cao, P. Vieu, Editorial on functional data analysis and related topics, *Comput. Statist.* 2019, [in press](#).
- [7] G. Aneiros, N. Ling, P. Vieu, Error variance estimation in semi-functional partially linear regression models, *J. Nonparametr. Stat.* **27** (2015) 316–330.
- [8] G. Aneiros, P. Vieu, Semi-functional partial linear regression, *Statist. Probab. Lett.* **76** (2006) 1102–1110.
- [9] G. Aneiros, P. Vieu, Variable selection in infinite-dimensional problems, *Statist. Probab. Lett.* **94** (2014) 12–20.
- [10] G. Aneiros, P. Vieu, Partial linear modelling with multi-functional covariates, *Comput. Statist.* **30** (2015) 647–671.
- [11] G. Aneiros, P. Vieu, Sparse nonparametric model for regression with functional covariate, *J. Nonparametr. Stat.* **28** (2016) 839–859.
- [12] G. Aneiros, P. Vieu, Comments on: Probability enhanced effective dimension reduction for classifying sparse functional data, *TEST* **25** (2016) 27–32.
- [13] E. Arnone, L. Azzimonti, F. Nobile, L.M. Sangalli, Modeling spatially dependent functional data via regression with differential regularization, *J. Multivariate Anal.* **170** (2019) [in press](#).
- [14] J.L. Bali, G. Boente, Robust estimators under a functional common principal components model, *Comput. Statist. Data Anal.* **113** (2017) 424–440.
- [15] K. Benhenni, A. Hajj Hassan, Y. Su, Local polynomial estimation of regression operators from functional data with correlated errors, *J. Multivariate Anal.* **170** (2019) [in press](#).
- [16] K. Benhenni, S. Hedli-Griche, M. Rachdi, P. Vieu, Consistency of the regression estimator with functional data under long memory conditions, *Statist. Probab. Lett.* **78** (2008) 1043–1049.
- [17] A. Berlinet, C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer Academic, Boston, MA, 2004.
- [18] J. Berrendero, B. Bueno-Larraz, A. Cuevas, An RKHS model for variable selection in functional linear regression, *J. Multivariate Anal.* **170** (2019) [in press](#).
- [19] T. Bodnar, H. Dette, N. Parolya, Spectral analysis of the Moore–Penrose inverse of a large dimensional sample covariance matrix, *J. Multivariate Anal.* **148** (2016) 160–172.
- [20] T. Bodnar, O. Okhrin, N. Parolya, Optimal shrinkage estimator for high-dimensional mean vector, *J. Multivariate Anal.* **170** (2019) [in press](#).
- [21] T. Bodnar, M. Reiß, Exact and asymptotic tests on a factor model in low and large dimensions with applications, *J. Multivariate Anal.* **150** (2016) 125–151.
- [22] G. Boente, D. Rodríguez, M. Sued, The spatial sign covariance operator: Asymptotic results and applications, *J. Multivariate Anal.* **170** (2019) [in press](#).
- [23] G. Boente, A. Vahnovan, Robust estimators in semi-functional partial linear regression models, *J. Multivariate Anal.* **154** (2017) 59–84.
- [24] E.G. Bongiorno, A. Goia, Describing the concentration of income populations by functional principal component analysis on Lorenz curves, *J. Multivariate Anal.* **170** (2019) [in press](#).
- [25] E.G. Bongiorno, A. Goia, E. Salinelli, P. Vieu, An overview of IWFOS’2014, In: *Contributions in Infinite-dimensional Statistics and Related Topics*, E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu (Eds.), Società Editrice Esculapio, Bologna, Italy, 2014, pp. 1–5.
- [26] D. Bosq, *Linear Processes in Function Spaces: Theory and Applications*, Springer-Verlag, New York, 2000.
- [27] D. Bosq, Estimation of mean and covariance operator of autoregressive processes in Banach spaces, *Stat. Inference Stoch. Process.* **5** (2002) 287–306.
- [28] A. Boudou, Y. Romain, On spectral and random measures associated to discrete and continuous-time processes, *Statist. Probab. Lett.* **59** (2002) 145–157.
- [29] A. Boudou, S. Viguier-Pla, Gap between orthogonal projectors: Application to stationary processes, *J. Multivariate Anal.* **146** (2016) 282–300.
- [30] A. Boudou, S. Viguier-Pla, Commuter of operators in a Hilbert space, *J. Multivariate Anal.* **170** (2019) [in press](#).
- [31] P. Bühlmann, S. van de Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer, Heidelberg, Germany, 2011.

- [32] R. Burfield, C. Neumann, C. Saunders, Review and application of functional data analysis to chemical data: The example of the comparison, classification, and database search of forensic ink chromatograms, *Chemom. Intell. Lab. Syst.* 149 (2015) 97–106.
- [33] M. Chaouch, Volatility estimation in a nonlinear heteroscedastic functional regression model with martingale difference errors, *J. Multivariate Anal.* 170 (2019) [in press](#).
- [34] J.A. Cuesta-Albertos, E. García-Portugués, M. Febrero-Bande, W. González-Manteiga, Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes, *Ann. Statist.* (2018) [in press](#), [arXiv:1701.08363v3](#).
- [35] A. Cuevas, A partial overview of the theory of statistics with functional data, *J. Statist. Plann. Inference* 147 (2014) 1–23.
- [36] A. Cuevas, R. Fraiman, On depth measures and dual statistics: A methodology for dealing with general data, *J. Multivariate Anal.* 100 (2009) 753–766.
- [37] M. Dass, C. Shropshire, Introducing functional data analysis to managerial science, *Organ. Res. Methods* 15 (2012) 693–721.
- [38] A. Delaigle, P. Hall, N. Bathia, Componentwise classification and clustering of functional data, *Biometrika* 99 (2012) 299–313.
- [39] P. Delicado, Functional k -sample problem when data are density functions, *Comput. Statist.* 22 (2007) 391–410.
- [40] P. Delicado, R. Giraldo, C. Comas, J. Mateu, Statistics for spatial functional data: Some recent contributions, *Environmetrics* 21 (2010) 224–239.
- [41] P. Delicado, P. Vieu, Optimal level sets for bivariate density representation, *J. Multivariate Anal.* 140 (2015) 1–18.
- [42] A. Dürre, D. Vogel, D.E. Tyler, The spatial sign covariance matrix with unknown location, *J. Multivariate Anal.* 130 (2014) 107–117.
- [43] M. Febrero-Bande, P. Galeano, W. González-Manteiga, Functional principal component regression and functional partial least-squares regression: An overview and a comparative study, *Int. Stat. Rev.* 85 (2017) 61–83.
- [44] F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, New York, 2006.
- [45] R. Fraiman, F. Gamboa, L. Moreno, Connecting pairwise geodesic spheres by depth: DCOPS, *J. Multivariate Anal.* 169 (2019) 81–94.
- [46] R. Fraiman, Y. Gimenez, M. Svarc, Feature selection for functional data, *J. Multivariate Anal.* 146 (2016) 191–208.
- [47] R. Fraiman, G. Muniz, Trimmed means for functional data, *TEST* 10 (2011) 419–440.
- [48] Y. Gao, H.L. Shang, Y. Yang, High-dimensional functional time series forecasting: An application to age-specific mortality rates, *J. Multivariate Anal.* 170 (2019) [in press](#).
- [49] G. Geenens, Curse of dimensionality and related issues in nonparametric functional regression, *Stat. Surv.* 5 (2011) 30–43.
- [50] A. Goia, P. Vieu, Some advances on semi-parametric functional data modelling, In: E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu (Eds.), *Contributions in Infinite-Dimensional Statistics and Related Topics*, Società Editrice Esculapio, Bologna, 2014, pp. 135–140.
- [51] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, *J. Multivariate Anal.* 146 (2016) 1–6.
- [52] W. González-Manteiga, G. Henry, D. Rodríguez, Partly linear models on Riemannian manifolds, *J. Appl. Stat.* 39 (2012) 1797–1809.
- [53] W. González-Manteiga, P. Vieu, Statistics for functional data, *Comput. Statist. Data Anal.* 51 (2007) 4788–4792.
- [54] W. González-Manteiga, P. Vieu, Methodological richness of functional data analysis, In: *Statistical Learning and Data Science*, Chapman & Hall/CRC Press Comput. Sci. Data Anal. Ser., Boca Raton, FL, 2012, pp. 197–203.
- [55] T. Górecki, M. Krzyśko, Ł. Waszak, W. Wołyński, Selected statistical methods of data analysis for multivariate functional data, *Statist. Papers* 59 (2018) 153–182.
- [56] S. Greven, F. Scheipl, A general framework for functional regression modelling, *Stat. Model.* 17 (2017) 1–35.
- [57] P. Hall, J.L. Horowitz, Methodology and convergence rates for functional linear regression, *Ann. Statist.* 35 (2007) 70–91.
- [58] P. Hall, I. Van Keilegom, Two-sample tests in functional data analysis starting from discrete data, *Statist. Sinica* 17 (2007) 1511–1531.
- [59] L. Horváth, P. Kokoszka, *Inference for Functional Data with Applications*, Springer, New York, 2012.
- [60] L. Horváth, G. Rice, Testing equality of means when the observations are from functional time series, *J. Time Series Anal.* 36 (2015) 84–108.
- [61] L. Horváth, G. Rice, An introduction to functional data analysis and a principal component approach for testing the equality of mean curves, *Rev. Mat. Complut.* 28 (2018) 505–548.
- [62] K. Hron, A. Menafoglio, M. Templ, K. Hřůzová, P. Filzmoser, Simplicial principal component analysis for density functions in Bayes spaces, *Comput. Statist. Data Anal.* 94 (2016) 330–350.
- [63] T. Hsing, R. Eubank, *Theoretical Foundations of Functional Data Analysis with an Introduction to Linear Operators*, Wiley, Chichester, 2015.
- [64] Y. Hu, Nonparametric estimation of variance function for functional data under mixing conditions, *Comm. Statist. Theory Methods* 42 (2013) 1774–1786.
- [65] J. Jacques, C. Preda, Functional data clustering: A survey, *Adv. Data Anal. Classif.* 8 (2014) 231–255.
- [66] W. Jank, G. Shmueli, Functional data analysis in electronic commerce research, *Statist. Sci.* 21 (2006) 155–166.
- [67] Q. Jiang, M. Hušková, S.G. Meintanis, L. Zhu, Asymptotics, finite-sample comparisons and applications for two-sample tests with functional data, *J. Multivariate Anal.* 170 (2019) [in press](#).
- [68] L. Kara-Zaitri, A. Laksaci, M. Rachdi, P. Vieu, Data-driven k NN estimation in nonparametric functional data analysis, *J. Multivariate Anal.* 153 (2017) 176–188.
- [69] L. Kara-Zaitri, A. Laksaci, M. Rachdi, P. Vieu, Uniform in bandwidth consistency for various kernel estimators involving functional data, *J. Nonparametr. Stat.* 29 (2017) 85–107.
- [70] A. Kneip, D. Poss, P. Sarda, Functional linear regression with points of impact, *Ann. Statist.* 44 (2016) 1–30.
- [71] A. Kneip, K.J. Utikal, Inference for density families using functional principal component analysis, *J. Amer. Statist. Assoc.* 96 (2001) 519–542.
- [72] P. Kokoszka, Dependent functional data, *Int. Scholar. Res. Notices* (2012) 1–30.
- [73] P. Kokoszka, H. Oja, B. Park, L.M. Sangalli, Special issue on functional data analysis, *Econometr. Statist.* 1 (2017) 99–100.
- [74] P. Kokoszka, M. Reimherr, *Introduction to Functional Data Analysis*, CRC Press, Boca Raton, FL, 2017.
- [75] N.L. Kudraszow, P. Vieu, Uniform consistency of k NN regressors for functional variables, *Statist. Probab. Lett.* 83 (2013) 1863–1870.
- [76] N. Laïb, D. Louani, Nonparametric kernel regression estimation for functional stationary ergodic data: Asymptotic properties, *J. Multivariate Anal.* 101 (2010) 2266–2281.
- [77] Y. Li, C. Huang, W.K. Härdle, Spatial functional principal component analysis with applications to brain image data, *J. Multivariate Anal.* 170 (2019) [in press](#).

- [78] H. Lian, Functional partial linear model, *J. Nonparametr. Stat.* 23 (2011) 115–128.
- [79] D. Liebl, Inference for sparse and dense functional data with covariate adjustments, *J. Multivariate Anal.* 170 (2019) in press.
- [80] E. Lila, J.A.D. Aston, L.M. Sangalli, Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging, *Ann. Appl. Stat.* 10 (2016) 1854–1879.
- [81] N. Ling, G. Aneiros, P. Vieu, *k*NN estimation in functional partial linear modeling, *Statist. Papers* (2017) <https://doi.org/10.1007/s00362-017-0946-0>.
- [82] N. Ling, P. Vieu, Nonparametric modelling for functional data: Selected survey and tracks for future, *Statistics* 52 (2018) 934–949.
- [83] J.S. Marron, Object oriented data analysis: Open problems regarding manifolds, In: E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu (Eds.), *Contributions in Infinite-Dimensional Statistics and Related Topics*, Società Editrice Esculapio, Bologna, 2014, pp. 185–190.
- [84] J.S. Marron, A.M. Alonso, Overview of object oriented data analysis, *Biom. J.* 56 (2014) 732–753.
- [85] A. Mas, B. Pumo, Linear processes for functional data, In: F. Ferraty, Y. Romain (Eds.), *The Oxford Handbook of Functional Data*, Oxford University Press, Oxford, UK, 2010, pp. 1–27.
- [86] E. Masry, Nonparametric regression estimation for dependent functional data: Asymptotic normality, *Stochastic Process. Appl.* 115 (2005) 155–177.
- [87] J. Mateu, E. Romano, Advances in spatial functional statistics, *Stoch. Environ. Res. Risk Assess.* 31 (2017) 1–6.
- [88] I.W. McKeague, B. Sen, Fractals with point impact in functional linear regression, *Ann. Statist.* 38 (2010) 2559–2586.
- [89] A. Menafoglio, P. Secchi, Statistical analysis of complex and spatially dependent data: A review of object oriented spatial statistics, *Europ. J. Oper. Res.* 258 (2017) 401–410.
- [90] A. Miller, *Subset Selection in Regression*, 2nd Ed., Chapman & Hall/CRC, Boca Raton, FL, 2002.
- [91] H.-G. Müller, Peter Hall, functional data analysis and random objects, *Ann. Statist.* 44 (2016) 1867–1887.
- [92] S. Nagy, An overview of consistency results for depth functionals, In: G. Aneiros, E.G. Bongiorno, R. Cao, P. Vieu P. (Eds.), *Functional Statistics and Related Fields: Contributions to Statistics*, Springer, Cham, Switzerland, 2017, pp. 189–196.
- [93] S. Nagy, F. Ferraty, Data depth for measurable noisy random functions, *J. Multivariate Anal.* 170 (2019) in press.
- [94] D. Nerini, B. Ghattas, Classifying densities using functional regression trees: Applications in oceanology, *Comput. Statist. Data Anal.* 51 (2007) 4984–4993.
- [95] A. Nieto-Reyes, H. Battey, Statistical functional depth, In: G. Aneiros, E.G. Bongiorno, R. Cao, P. Vieu (Eds.), *Functional Statistics and Related Fields: Contributions to Statistics*, Springer, Cham, Switzerland, 2017, pp. 197–202.
- [96] A. Petersen, Theoretical foundations to functional data analysis with an introduction to linear operators (book review), *J. Amer. Statist. Assoc.* 112 (2017) 463–464.
- [97] A. Pini, S. Vantini, The interval testing procedure: A general framework for inference in functional data analysis, *Biometrics* 72 (2016) 835–845.
- [98] A. Pini, S. Vantini, Interval-wise testing for functional data, *J. Nonparametr. Stat.* 29 (2017) 404–424.
- [99] A. Pini, L. Spreafico, S. Vantini, A. Vietti, Multi-aspect local inference for functional data: Analysis of ultrasound tongue profiles, *J. Multivariate Anal.* 170 (2019) in press.
- [100] S.M. Pizer, J.S. Marron, Object statistics on curved manifolds, In: *Statistical Shape and Deformation Analysis: G. Zheng, S. Li, G. Székely (Eds.), Methods, Implementation and Applications*, Elsevier, 2017, pp. 137–164.
- [101] L.M. Sangalli, J.O. Ramsay, T.O. Ramsay, Spatial spline regression models, *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 75 (2013) 681–703.
- [102] J.O. Ramsay, B.W. Silverman, *Applied Functional Data Analysis*, Springer-Verlag, New York, 2002.
- [103] J.O. Ramsay, B.W. Silverman, *Functional Data Analysis*, 2nd Ed., Springer, New York, 2005.
- [104] P.T. Reiss, J. Goldsmith, H.L. Shang, R. Ogden, Methods for scalar-on-function regression, *Int. Stat. Rev.* 85 (2017) 228–249.
- [105] J. Rossini, A. Canale, Quantifying prediction uncertainty for functional-and-scalar to functional autoregressive models under shape constraints, *J. Multivariate Anal.* 170 (2019) in press.
- [106] M.D. Ruiz-Medina, J. Álvarez-Liébana, Strongly consistent autoregressive predictors in abstract Banach spaces, *J. Multivariate Anal.* 170 (2019) in press.
- [107] L.M. Sangalli, The role of Statistics in the era of Big Data, *Statist. Probab. Lett.* 136 (2018) 1–3.
- [108] J.Q. Shi, T. Choi, *Gaussian Process Regression Analysis for Functional Data*, CRC Press, Boca Raton, FL, 2011.
- [109] S. Ullah, C.F. Flinch, Applications of functional data analysis: A systematic review, *BMC Med. Res. Methodol.* 13 (2013) 13–43.
- [110] M.J. Valderrama, An overview to modelling functional data, *Comput. Statist.* 22 (2007) 331–334.
- [111] P. Vieu, On dimension reduction models for functional data, *Statist. Probab. Lett.* 136 (2018) 134–138.
- [112] J.-L. Wang, J.-M. Chiou, H.-G. Müller, Review of functional data analysis, *Annu. Rev. Stat. Appl.* 3 (2016) 257–295.
- [113] J.-T. Zhang, *Analysis of Variance for Functional Data*, CRC Press, Boca Raton, FL, 2014.
- [114] X. Zhang, J.-L. Wang, From sparse to dense functional data and beyond, *Ann. Statist.* 44 (2016) 2281–2321.
- [115] H. Zhu, R. Zhang, Z. Yu, H. Lian, Y. Liu, Estimation and testing for partially functional linear errors-in-variables models, *J. Multivariate Anal.* 170 (2019) in press.
- [116] Y. Zuo, R.J. Serfling, General notions of statistical depth function, *Ann. Statist.* 28 (2000) 461–482.