1. ARTÍCULOS DE ESTADÍSTICA

A PRESENT OVERVIEW ON FUNCTIONAL DATA ANALYSIS

Manuel Febrero Bande*
Dpto. de Estadística e Investigación Operativa
Universidad de Santiago de Compostela

Abstract

In this work, a general overview on the state of the art in functional data analysis (FDA) is given. After an essentially historical introduction, the basic concepts in FDA are defined in the Section 2. The third section is devoted to some comments on several recent and interesting works for exploratory functional data analysis, regression with functional variables and other techniques.

Keywords: Functional Data Analysis, Functional Linear Regression, Functional Space.

1. Introducción

En los últimos tiempos la computación aplicada a diversas áreas ha provocado un cambio tecnológico muy importante. Este cambio tecnológico viene de incorporar equipos de medición más rápidos y precisos que son capaces de proporcionar información más fiable y más rápidamente. Esta evolución tecnológica cambia o cambiará algunos de los paradigmas en los que se ha basado la estadística clásica, por ejemplo, el paradigma de que en un conjunto de datos siempre el número de datos es mayor que el número de variables. En muchas áreas se ha empezado a trabajar con grandes bases de datos, que cada vez con más frecuencia, corresponden a observaciones de una variable aleatoria tomadas a lo largo de un intervalo continuo (o en discretizaciones cada vez más extensas de este intervalo continuo). Así, por ejemplo, en campos como la espectrometría, el resultado de la medición es una curva que representa a la muestra concreta que al menos se ha evaluado en una centena de puntos. Este tipo de datos, que llamaremos datos funcionales, surgen de manera natural en muchas disciplinas. En Economía podríamos hablar de curvas intra-día de cotizaciones en bolsa, en Ingeniería podríamos hablar de curvas minutales de producción o demanda eléctrica, en Medio Ambiente se dispone de mediciones continuas de redes de vigilancia atmosférica, fluvial o meteorológica y es bien conocido el auge del reconocimiento de imágenes o de la información espacial. Ante estos nuevos retos surge como respuesta la estadística de datos funcionales que originalmente identificaba dato funcional con función en un intervalo continuo.

Básicamente, los problemas a los que se debe enfrentar la estadística con datos funcionales responde a las mismas necesidades que la estadística clásica. Estos se podrían categorizar de la siguiente manera:

- Explorar y describir el conjunto de datos funcionales resaltando sus características más importantes.
- Explicar y modelar la relación entre una variable dependiente y una independiente (modelos de regresión)
- Métodos de Clasificación Supervisada o no Supervisada de un conjunto de datos respecto a alguna característica.
- Contraste, validación y predicción.

Sin duda alguna, el libro que más ha contribuido a popularizar las técnicas estadísticas para datos funcionales es el de Ramsay y Silverman [18] (en adelante RS2002) cuya primera edición de 1997 trata muchos de los problemas básicos de la estadística funcional con un lenguaje muy asequible dirigido tanto a investigadores del área de estadística como de otras áreas. A este libro le sigue otro de carácter muy aplicado [19] donde se explora el uso de las técnicas en conjuntos de datos interesantes. En ambos casos todas las técnicas incluidas están

*Corresponding Author. E-mail: mfebrero@usc.es

restringidas al espacio de funciones L^2 que como veremos más adelante es un espacio con características específicas que lo hacen especialmente tratable. La página web http://www.functionaldata.org/ mantenida por Jim Ramsay puede ser un buen comienzo para familiarizarse con su trabajo. El otro hito bibliográfico relevante es el reciente libro de Ferraty y Vieu [15] (en adelante FV2006) donde se tratan los datos funcionales desde un punto de vista no paramétrico y se establecen marcos teóricos apropiados para su tratamiento. En este caso, el planteamiento es más general, considerando espacios funcionales normados o semi-normados que en algunos casos pueden ser más apropiados para describir la realidad. Estos autores forman parte del grupo francés STAPH que mantienen la página http://www.lsp.ups-tlse.fr/staph/ donde se puede encontrar más información sobre sus actividades y trabajo. Por supuesto, el número de artículos dedicados al tema en los últimos años es muy importante. Una búsqueda en scholar.google.com con la frase exacta "functional data analysis"proporciona más de 1800 resultados y si nos restringimos a entradas posteriores al 2003 aparecen más de 900. El libro de Ramsay y Silverman ha sido citado según esta herramienta más de 800 veces.

2. ¿Qué es un dato funcional?

Seguiremos la definición de FV2006 por su generalidad y facilidad de comprensión.

Definición 2.1. *Una variable aleatoria $\mathcal{X}$ se dice que es una variable funcional si toma valores en un espacio funcional $\mathcal{E}$ (Espacio normado o semi-normado completo)*

Definición 2.2. *Un conjunto de datos funcionales $\{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$ es la observación de n variables funcionales $\mathcal{X}_1, \ldots, \mathcal{X}_n$ identicamente distribuidas.*

Estas definiciones se pueden aplicar a muchos tipos de espacios. En particular, $\mathbb{R}^p$ con las métricas usuales es un espacio funcional y por tanto puede deducirse que toda técnica que se desarrolle para datos funcionales puede ser aplicada con ciertas garantías en el entorno multivariante. El espacio más comunmente usado cuando se habla de datos funcionales es el espacio $L^2[\mathcal{S}]$, esto es, las funciones de cuadrado integrable en el intervalo $\mathcal{S} = [a, b] \subset \mathbb{R}$. Desde un punto de vista más general podemos tener datos funcionales en la familia: $L^p[\mathcal{S}, \mu] = \{f : \mathcal{S} \to \mathbb{R}$ tal que $\int |f(t)|^p d\mu < \infty\}$, donde $(\mathcal{S}, \mu)$ es un espacio de medida y $1 < p < \infty$. Estos espacios son semi-normados salvo el caso $p = 2$ que es el único de esta familia que es un espacio de Hilbert separable. Cuando se desarrolla una nueva técnica para datos funcionales la primera preocupación es siempre determinar en qué espacio funcional vamos a trabajar. Esto determinará decisivamente el conjunto de herramientas que podremos usar. Una preocupación similar la tendremos al aplicar una técnica de datos funcionales a un conjunto de datos. La métrica del espacio funcional que se elija para encuadrar estos datos debe ser coherente con la interpretación física del fenómeno que describan. Por ejemplo, en el conjunto de datos Tecator (http://lib.stat.cmu.edu/datasets/tecator) que se usa extensamente en FV2006 como hilo conductor, las curvas de absorbancia registradas en el intervalo $[850, 1050]$ para el análisis de trozos de carne presentan formas similares aunque se aprecia un cambio de escala entre ellas.

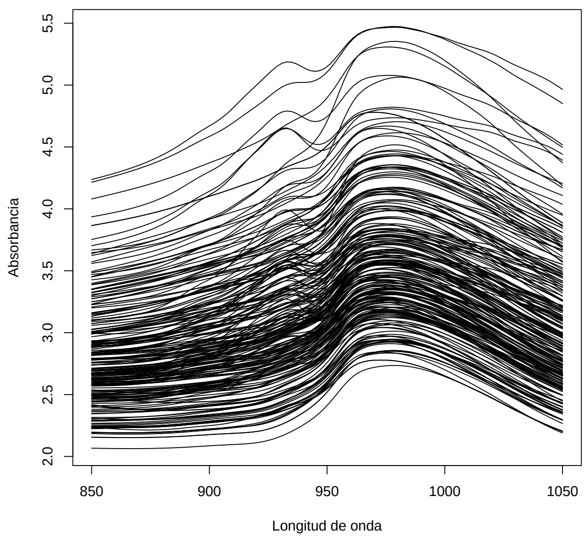

Figura 1: Ejemplo curvas espectrométricas

La primera dificultad que siempre tendremos al analizar datos funcionales, es encontrar una representación adecuada para los datos. Habitualmente, como se refleja también en la Figura 1, la representación de las curvas en el clásico eje X-Y podría esconder las características interesantes. Así, es difícil apreciar el cambio de escala que se mencionaba si no limpiamos un poco la muestra. Si entendemos que este cambio de escala es informativo, la elección L^2 como espacio de referencia sería la más aconsejable. Si por el contrario este cambio de escala es irrelevante y la información está en la forma, una semi-norma como la de las derivadas $\left(d(f,g) = \sqrt{\int_{\mathcal{S}}(f'(t) - g'(t))^2 dt}\right)$ sería una elección más adecuada. Este conjunto de datos presenta como posible variable respuesta el porcentaje de grasa en la muestra. En este caso, atendiendo a esta variable y como se puede ver en la figura siguiente, parece que el cambio de escala es menos importante y el análisis se debe focalizar en el cambio de forma.

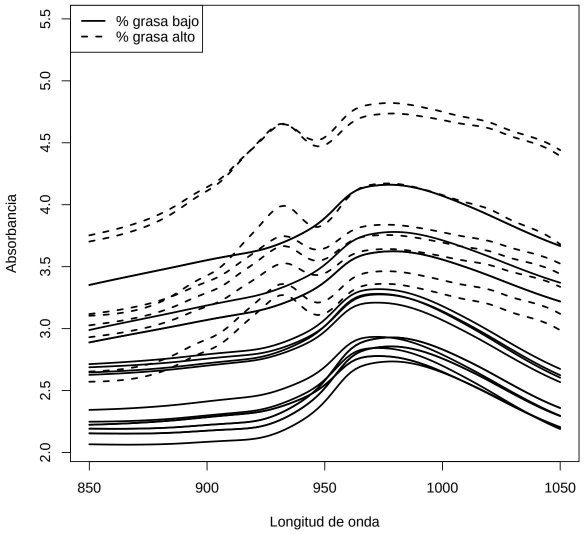

Figura 2: Curvas más extremas por su contenido en grasa

Sin embargo, ambas figuras están diseñadas para apreciar distancias en L^2 entre los datos y no distancias con otro tipo de norma o semi-norma como sería el de la primeras derivadas. Dependiendo del espacio elegido el conjunto de herramientas disponibles cambia notablemente. El caso del espacio más utilizado $L^2[\mathcal{S}]$ es el más favorable. Éste, por ser separable, dispone de bases ortonormales que dan mucho juego a la hora de diseñar procedimientos. En general, la representación de un dato funcional en una base ortonormal proporcionará ventajas tanto desde el punto de vista teórico como práctico sirviendo de puente entre la inevitable discretización del dato funcional y su verdadera forma funcional.

Definición 2.3. *Una base es un conjunto de funciones conocidas e independientes* $\{\phi_k\}_{k\in\mathbb{N}}$ *tales que cualquier función puede ser aproximada, tan bien como se quiera, mediante una combinación lineal de* K *de ellas con* K *suficientemente grande. De esta forma, la observación funcional puede aproximarse como* $\mathcal{X}(t) \approx \sum_{k=1}^{K} c_k\phi_k(t)$.

Básicamente, la idea clave cuando se pueden usar bases ortonormales es representar cada dato funcional en la base usando aquellas coordenadas que son más significativas. Debido a la alta dimensión de los datos funcionales, se elige en general un número K para representar los datos en el subespacio, convirtiendo el problema de dimensión infinita en un problema multidimensional. La elección del parámetro K y de la base más adecuada para los datos observados se antoja crucial y, en principio, no hay ninguna regla que permita hacer una selección óptima de forma universal. El parámetro K es, en cierto modo, un parámetro de suavización de los datos funcionales. Si K es bajo tendremos un modelo muy manejable pero posiblemente habremos perdido información relevante. Si K es alto representaremos muy bien los datos pero el problema de la dimensión cobra importancia. Si atendemos a la elección de la base, para datos periódicos se suele emplear la base de Fourier y para datos no periódicos la base B-spline o la wavelet. Una base muy popular está basada en la expansión de Karhunen-Loève que no es más que la extensión del análisis de componentes principales multivariante a procesos estocásticos y por añadidura a datos funcionales. Calculando a partir del operador momento de segundo orden muestral las correspondientes autofunciones y autovalores es posible construir específicamente una base ortonormal adaptada para cada conjunto de datos. Esta técnica se denomina Componentes Principales Funcionales (FPCA) y ha dado lugar a muchas técnicas interesantes para datos funcionales. Sin embargo, y por incluir alguna sombra, esta técnica puede ser muy sensible a la aparición de datos atípicos y la representación del dato funcional puede no ser relevante para el objetivo del estudio como podría ser la relación con otra variable funcional o no. La decisión sobre qué base elegir debe tomarse en función del objetivo del estudio y los datos y aprovechando las ventajas e inconvenientes que presenta cada tipo de base. Si se trunca cualquiera de estas bases en un número determinado de elementos obtendremos una semimétrica que también podremos usar para manejar los datos funcionales. En este caso, cualquier métrica o semi-métrica en el espacio no es más que una forma de determinar qué elementos del espacio están cercanos y cuáles lejanos.

La estadística con datos funcionales tiene frontera con otros campos relevantes de la estadística como el análisis multivariante, el análisis de datos longitudinales o las series temporales. Como se comentó anteriormente, una técnica de datos funcionales puede aplicarse con ciertas garantías a datos multivariantes. El reverso, en general, no es cierto. Para la mayoría de las técnicas multivariantes que basan mucho de su trabajo en propiedades del álgebra matricial puede ser un problema casi insalvable tratar datos funcionales de alta frecuencia con seguramente, muy fuerte colinealidad. Según aumenta el grado de resolución con el que somos capaces de ver una curva, más difícil resulta para las técnicas multivariantes obtener un resultado convirtiendo el aumento de resolución en una dificultad más que en una oportunidad de obtener mejor información. Algo similar podría decirse del análisis de datos longitudinales. En este campo se obtienen medidas repetidas a lo largo del tiempo para el mismo sujeto, pero en general, éste es un número pequeño y las técnicas multivariantes pueden adaptarse para trabajar con ellas. La principal dificultad para tratar datos longitudinales como datos funcionales suele ser precisamente la baja calidad de representación de las curvas. La relación con el campo de las series temporales es totalmente diferente. Así, ejemplos clásicos de datos funcionales se han construido a base de cortar una serie temporal en ciclos homogéneos. Por ejemplo, en RS2002 se usan los datos de un índice bursátil estadounidense troceados por años (como unidad funcional) para deducir a partir de la forma de cada curva anual la tipología de los distintos años (de expansión, de crisis, ...). Considerados los datos como una serie temporal, el objetivo es predecir alguno de los periodos del próximo año. Como conjunto de datos funcionales, el objetivo es resumir la información y el resultado será siempre un dato funcional, esto es, un ciclo anual completo. Por supuesto, se pueden mezclar ambos mundos para obtener herramientas para series de tiempo funcionales (véase por ejemplo, [14]). Por tanto, la relación entre estos dos campos es peculiar. Muchas veces trabajan sobre la misma información pero desde ópticas completamente diferentes.

3. Estado de la cuestión

3.1. Técnicas exploratorias para Datos Funcionales

Aunque probablemente cualquier estudio estadístico de un conjunto de datos debiera empezar por un análisis descriptivo, sin duda alguna este apartado no ha merecido demasiada atención hasta el momento. En RS2002 en el Capítulo 2 sólo se recogen como herramientas para resumir los datos: la media funcional, la varianza funcional y la función de covarianza. En un capítulo posterior se emplean las componentes principales funcionales como herramientas del análisis descriptivo. Básicamente esto era todo el análisis descriptivo de un conjunto de datos funcionales. Sin embargo, el análisis descriptivo se revela decisivo para el tratamiento de datos funcionales. Como decíamos, un vistazo rápido a la Figura 1 demuestra que el gráfico por defecto de un conjunto de datos funcionales puede ser enormemente no informativo. Esto no ocurre con los típicos gráficos de nube de puntos en $\mathbb{R}^2$ donde una mirada entrenada puede encontrar características relevantes de la población. Para datos funcionales la cuestión se complica si pensamos que nuestros datos pueden estar sujetos a métricas no usuales y por tanto, las representaciones usuales engañarían nuestra mirada. En este campo se echan en falta herramientas descriptivas que en otros ámbitos como el multivariante se han desarrollado expresamente. Esta falta de atención está cambiando en los últimos años donde han aparecido varios trabajos que hacen más hincapié en este apartado. Así en el trabajo de Dabo-Niang et al [10] se definen extensiones de la moda a datos funcionales. Manejando diferentes conceptos sobre profundidad estadística también se han definido extensiones de medidas robustas para datos funcionales como las referidas en los trabajos [17] y [8], incluyendo en este último el bootstrap para datos funcionales como herramienta para analizar la variabilidad de los distintos estimadores. También se han hecho avances en la detección de outliers (véase por ejemplo [12] y [13]).

3.2. Regresión

El apartado de regresión, es el que ha recibido más atención por parte de la comunidad científica. Se pueden establecer distintos modelos bajo las diferentes condiciones que deben cumplir la variable respuesta y las variables regresoras. En sus distintas variantes el problema es inicialmente estudiado en RS2002 al que dedica varios capítulos. En su análisis cobra mucha importancia la necesidad de penalizar la falta de suavidad de los estimadores y con esta visión revisa varios problemas de regresión sin adentrarse en el marco teórico. El caso más estudiado es posiblemente el de variable respuesta escalar y variable regresora funcional. Bajo diseño aleatorio, los trabajos de Cardot et al ([2],[3]) se han centrado en el modelo lineal funcional, esto es, $y = \alpha_0 + \int_{\mathcal{C}} \alpha(t)\mathcal{X}(t)dt + e(t)$. En este caso, la métrica del espacio funcional siempre es L^2. Una extensión natural de este modelo viene dada por considerar $y = r(\mathcal{X}(t)) + e(t)$ donde r es una función suave general y no está restringida a un operador lineal. Este modelo se conoce como regresión funcional no paramétrica y ha sido extensamente estudiado en FV2006 y trabajos anteriores de los mismos autores (véase por ejemplo [14]). La idea subyacente es usar un estimador similar al de regresión no paramétrica pero adaptado a la norma o semi-norma de los datos funcionales. Para ello se deben definir funciones núcleo apropiadas así como determinar condiciones teóricas del espacio funcional para la convergencia del estimador. Una característica importante de este modelo es que no sufre el desastre de la dimensionalidad al usar las funciones núcleo sobre la métrica (unidimensional).

Para el modelo con respuesta funcional y variables regresoras funcionales no se dispone en general de tantas herramientas fuera de L^2. Este modelo que viene dado por $Y(t) = \alpha(t) + \int_{\mathcal{S}} \mathcal{X}(s)\beta(s,t)ds + e(t)$ está resuelto en RS2002 representando en bases restringidas tanto la respuesta como la variable regresora. En este modelo todavía parece un problema abierto las condiciones teóricas necesarias para la convergencia de los estimadores. En el contexto de diseño fijo algunas condiciones sobre convergencia de los estimadores se puede encontrar en el trabajo de Cuevas et al [6]. Otra referencia reciente de interés es [23].

Las técnicas de análisis de la varianza pueden considerarse como un modelo de regresión con respuesta funcional y variable regresora discreta. Este es el punto de vista que se usa en RS2002 para la obtención de estimadores. Siguiendo la estela marcada por este libro, han aparecido varios trabajos que, básicamente, transforman los datos funcionales en un conjunto de datos multivariante (mediante una base ortonormal truncada) y resuelven un MANOVA. Desde otro punto de vista, el trabajo de Cuevas et al [7] está enfocado al estudio tanto teórico como aplicado de un contraste ANOVA general para datos funcionales con bootstrap. Usando la idea de los modelos de regresión en otros contextos podemos destacar el trabajo de Ferraty y Vieu [14] donde se usa el estimador de la regresión funcional no paramétrico en el contexto de discriminación de curvas o de predicción de series de tiempo. En series de tiempo es destacable el trabajo de Aguilera et al [1] como un primer precedente de aplicación a series de tiempo funcionales. Sería muy largo y prolijo comentar extensiones de los modelos de regresión en datos funcionales siguiendo las ideas del libro de Ramsay y Silverman y sólo a modo de ejemplo citaré aquí el trabajo de Escabias et al [11] como aplicación al campo de la regresión logística.

3.3. Otros métodos

Fundamentalmente, en la literatura abundan trabajos que se dedican de una u otra manera al problema de la clasificación bien supervisada (discriminación) o no supervisada (cluster). Como uno de los trabajos más clásicos en el tema se puede citar el de James y Sugar [16] que está basado en los B-splines y que posteriormente fue seguido por el trabajo de Yao et al [22] aplicando componentes principales funcionales al problema de discriminación. Otro clásico es el trabajo de Tarpey y Kinateder [21]. Más recientemente en FV2006 el tercer capítulo está integramente dedicado a estas cuestiones. En este libro, el problema de discriminación se resuelve, como se ha comentado, como un problema de regresión calculando la esperanza de una variable indicadora. El problema de la clasificación no supervisada es abordado mediante la elaboración de un índice de heterogeneidad que sirve para ir particionando la muestra mediante un método jerárquico. El trabajo de Cuesta-Albertos y Fraiman [4] proporciona una versión robusta del algoritmo de k-medias para datos funcionales mientras que el trabajo de Cuevas et al [9] presenta varios métodos para discriminación basados en conceptos de profundidad. El análisis de datos funcionales está llegando a todos los ámbitos de la estadística y como ejemplo se cita aquí el trabajo de Rossi y Conan-Guez [20] dedicado a redes neuronales con entradas funcionales. En mayor o menor medida los trabajos citados se han dedicado más a la estimación que al contraste. Un trabajo reciente que posiblemente cambiará este planteamiento es el trabajo de Cuesta-Albertos et al [5] que, en pocas palabras, permite caracterizar poblaciones funcionales mediante el uso de proyecciones aleatorias. Esto abre la puerta a poder utilizar un amplio abanico de herramientas uni o multi-dimensionales para el contraste de características en datos funcionales con técnicas muy simples.

4. Conclusiones

El análisis de datos funcionales es una disciplina emergente en la estadística actual. Las causas de esta eclosión hay que buscarlas en la necesidad creciente de tratar realidades cada vez más complejas que evolucionan rápidamente gracias a las nuevas tecnologías de medición. De este interés creciente da fe el que revistas del máximo nivel hayan dedicado números especiales al tratamiento de datos funcionales como por ejemplo *Statistica Sinica* (Vol. 14, nº3, 2004), *Computational Statistics & Data Analysis* (Vol. 51, 10, 2007) o *Computational Statistics* (Vol. 22, nº 3, 2007). Otra iniciativa que demuestra este interés es la creación de un grupo de trabajo especializado en estadística funcional dentro del grupo de trabajo de Computing & Statistics en el ERCIM (European Research Consortium for Informatics and Mathematics). Puede consultarse la información disponible en el siguiente vínculo http://www.dcs.bbk.ac.uk/ercim/TrackSFD.html. Se ha recorrido mucho camino en poco tiempo fruto de este interés creciente pero todavía queda mucho más por recorrer. A medida que estas herramientas de datos funcionales se vayan popularizando surgirán nuevos tipos de datos funcionales que necesitarán de desarrollos específicos o adaptación de los ya existentes para su correcto tratamiento. Esto augura un inmenso campo de trabajo para los próximos años.

Referencias

[1] Aguilera A.M., Ocaña F.A. y Valderrama M.J. (1999). Forecasting time series by functional PCA. Discussion of several weighted approaches. *Computational Statistics*, **14**, 442-467.

[2] Cardot H., Ferraty F. y Sarda P. (1999). Functional linear model. *Statistics and Probability Letters*, **45**, nº 1, 11-22.

[3] Cardot H., Ferraty F., Mas A. y Sarda P. (2003). Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics*, **30**, nº 1, 241-255.

[4] Cuesta-Albertos J. y Fraiman, R. (2007). Impartial Trimmed k-means for Functional Data. *Computational Statistics and Data Analysis*, **51**, 4864-4877.

[5] Cuesta-Albertos J., Fraiman, R. y Ransford, T. (2007). A sharp form of the Cramer-Wold theorem. *Journal of Theoretical Probability*, **20**, 201-209.

[6] Cuevas A., Febrero M. y Fraiman, R. (2002). Linear functional regression: the case of fixed design and functional response. *The Canadian Journal of Statistics*, **30**,2 285-300.

[7] Cuevas A., Febrero M. y Fraiman, R. (2004). An ANOVA test for functional. *Computational Statistics and Data Analysis*, **47**, 111-222.

[8] Cuevas A., Febrero M. y Fraiman, R. (2007). On the use of bootstrap for estimating functions with functional data. *Computational Statistics and Data Analysis*, **51**, 1063-1074.

[9] Cuevas A., Febrero M. y Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, **22**, 1481-496.

[10] Dabo-Niang S., Ferraty F. y Vieu, Ph. (2007). On the using of modal curves for radar waveforms classification. *Computational Statistics and Data Analysis*, **51**, 4957-4968.

[11] Escabias M., Aguilera A.M. y Valderrama, M.J. (2005). Modelling environmental data by functional component logistic regression. *Environmetrics*, **16**, 1, 95-107.

[12] Febrero M., Galeano, P. y González-Manteiga, W. (2007). Outlier detection in functional data by depth measures with application to identify abnormal NOx levels. *Environmetrics*, DOI:10.1002/env.878.

[13] Febrero M., Galeano, P. y González-Manteiga, W. (2007). A functional analysis of NOx levels:location and scale estimation and outlier detection. *Computational Statistics*, **22**, 411-427.

[14] Ferraty F. y Vieu, Ph. (2004). Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *Nonparametric Statistics*, **16**, 111-125.

[15] Ferraty F. y Vieu Ph. (2006). *Nonparametric Functional Data Analysis*, Springer, New York.

[16] James, G.M. y Sugar, C.A.(2003). Clustering for Sparsely Sampled Functional Data. *Journal of the American Statistical Association*, **98**, 397-408.

[17] López-Pintado S. y Romo, J (2007). Depth-based inference for functional data. *Computational statistics and data analysis*, **51**, 4878-4890.

[18] Ramsay J.O. y Silverman, B.W. (2002). *Functional Data Analysis* Second Edition, Springer, New York.

[19] Ramsay J.O. y Silverman, B.W. (2002). *Applied Functional Data Analysis Methods Case and Studies*, Springer, New York.

[20] Rossi, F. y Conan-Guez, B. (2006). Theoretical Properties of Projection Based Multilayer Perceptrons with Functional Inputs. *Neural Processing Letters*, **23**, 55-70.

[21] Tarpey, T. y Kinateder, K. (2003). Clustering Functional Data. *Journal of Classification*, **20**, 93-114.

[22] Yao, F., Müller, H.G. y Wang J.L. (2005). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*, **100**, nº 470, 577-590.

[23] Zhang J.T. y Chen J. (2007). Statistical inferences for functional data. *Annals of Statistics*, **35**, nº 3, 1052-1079.

Manuel Febrero Bande es profesor del Departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela. Entre sus líneas de investigación se incluye el análisis de series temporales, la estadística espacial, el bootstrap y la estadística con datos funcionales con aplicaciones fundamentalmente en el campo medioambiental, industrial o económico. Es autor/coautor de una treintena de artículos en revistas del JCR y actualmente es co-chair del grupo especializado de Estadística para Datos Funcionales en el grupo de trabajo Computing & Statistics dentro del ERCIM (European Research Consortium for Informatics and Mathematics). Más detalles se pueden encontrar en su página web http://eio.usc.es/pub/febrero/.