The Annals of Statistics
2016, Vol. 44, No. 5, 1867–1887
DOI: 10.1214/16-AOS1492

PETER HALL, FUNCTIONAL DATA ANALYSIS AND RANDOM OBJECTS

BY HANS-GEORG MÜLLER[1,2]

University of California, Davis

Functional data analysis has become a major branch of nonparametric statistics and is a fast evolving field. Peter Hall has made fundamental contributions to this area and its theoretical underpinnings. He wrote more than 25 papers in functional data analysis between 1998 and 2016 and from 2005 on was a tenured faculty member with a 25% appointment in the Department of Statistics at the University of California, Davis. This article describes aspects of his appointment and academic life in Davis and also some of his major results in functional data analysis, along with a brief history of this area. It concludes with an outlook on new types of functional data and an emerging field of "random objects" that subsumes functional data analysis as it deals with more complex data structures.

1. Introduction: Peter Hall in Davis. This article highlights Peter Hall's contributions to functional data analysis and elucidates their scientific context. This introductory section is devoted to Peter's time as a tenured faculty member at the University of California, Davis.

Peter Hall visited the then Division of Statistics at the University of California, Davis, a number of times before he accepted a faculty position. Several faculty became friends with him over the years, which might have contributed to his eventual decision to join the Davis faculty. One of these visits in Davis, perhaps the first longer one, took place in June 1989. It was part of an early recruitment effort by then Division of Statistics chair and Associate Dean George Roussas. Peter's visit coincided with a lecture series on Stochastic Curve Estimation, which was given by Murray Rosenblatt (UC San Diego) in the framework of a NSF-CBMS Regional Conference in Davis, with lecture notes published two years later [Rosenblatt (1991)]. Peter was not ready to consider an appointment at Davis at that time, and even though he was being recruited by several prominent Statistics departments in the U.S., he decided to stay at the Australian National University.

More extensive visits of Peter were arranged for Fall 2003 and Spring 2004 (see Figure 1), upon the initiative of department chair Jane-Ling Wang and after the Division of Statistics had been converted into a Department of Statistics

Received June 2016.

[1]Supported in part by National Science Foundation Grant DMS-14-07852.

[2]Additional sources of material in this memorial article are listed in the Acknowledgments.

MSC2010 subject classifications. Primary 62G05; secondary 62G07, 62G20.

Key words and phrases. Functional principal component analysis, functional regression, functional classification, modes and densities for functional data.

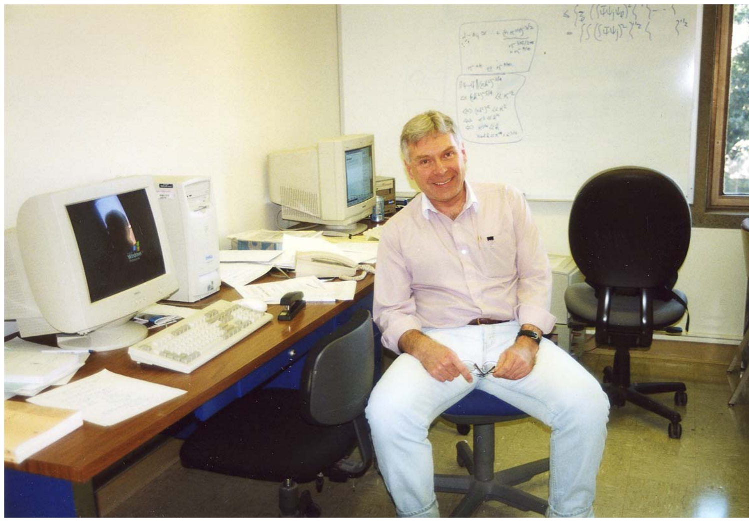

FIG. 1. *Peter Hall in his office at UC Davis, while visiting in the Fall of* 2003. *He joined the Department of Statistics as a tenured faculty member with a* 25% *appointment in* 2005. *Photo taken by Jane-Ling Wang.*

in 2000. This focused recruitment effort was continued by the next department chair, Rudy Beran, and eventually proved successful: In 2005, Peter joined the department as a tenured faculty member at the rank of Distinguished Professor. The arrangement was for a 25% level appointment, which suited Peter well. He was not ready to move away from his native Australia to which he remained strongly attached [Delaigle and Wand (2016)] throughout his life, both culturally and through his family, especially his wife Jeannie Hall, who had high profile posts in the Australian federal government that included Deputy Official Secretary to the Governor-General of Australia; Parliamentary Liaison Officer for the House of Representatives of the Parliament of Australia; Senior Adviser of the Cabinet Secretariat in the Department of the Prime Minister and Cabinet.

The 25% appointment had the advantage for Peter to give him a steady, even if periodic, presence in the U.S.; he found it beneficial to have exposure to the U.S. statistics enterprise. A downside was that because of the part-time appointment he had to forgo many of the regular faculty benefits. He liked the weather in Davis, which he felt matched well with Canberra, and also the Davis faculty colleagues and the location on the West Coast. As UC Davis operates on the quarter system, Peter was in Davis every Spring quarter from April to June. He would teach two quarter courses every second year, an upper division introductory probability course with a large enrollment and an advanced graduate course on the bootstrap, which was always well attended.

Peter typically would complete all course preparations such as syllabus, homework assignments, exams and lecture notes in a single weekend before teaching started so that he would only need to lecture and hold his office hours for the rest of the quarter. This would be impossible to do for most faculty, but Peter pulled it off due to his extremely fast mental processing and intense focus on his work. His unique focus and ability also led to his extraordinary productivity reflected in more than 600 publications, with most of them in the top journals of statistics and many of them highly cited.

Peter was well liked by his students. The uneven preparedness of some of the undergraduate students never fazed him. Occasionally, over lunch he would bring up teaching topics, such as how to answer nonroutine questions in the office hours for which one does not know the answer off-hand. We agreed a good method was to check online on the spot and thereby teach students how to effectively find answers on their own. Lunch was always a time for Peter to relax, which nevertheless included discussing research. It was always pleasant and stimulating for me to converse with him in the student cafeteria (the UC Davis Silo) and in hindsight it was a great privilege. The topics that interested Peter included of course statistics but ranged far beyond. Lunch conversations that I remember were on topics such as why a density does not exist in function space, or what drives research progress in statistics—Peter attributed the latter to technological changes outside of the control of statistics that lead to new types of data and then to demand for appropriate data analytic and statistical tools.

Favorite topics for Peter that were unrelated to statistics were aircraft, piloting and airline travel, a vast field, in which he was extremely knowledgable; economics and the economy, in which he was very interested; political developments, especially in the English-speaking world; photography, trains, especially nonelectric locomotives and how to best take pictures of them. Our lunch table conversations were facilitated by our common interests, including Peter's professed disinterest in sports events [Delaigle and Wand (2016)], while he would enjoy discussing functional data analysis challenges arising from longitudinal sports statistics.

In addition to teaching, Peter also took his other faculty obligations very seriously and was an excellent departmental citizen. When he was around in Davis and a faculty meeting was scheduled, he would always attend it for the whole duration and his advice and input at the meetings were invaluable. During 2012–2015 when I was department chair, I became aware of how beneficial his presence was to attenuate divisions in the faculty and to arrive at better decisions. He also had very good relations with the Dean and helped to promote our department. Peter was very kind and gentle to those around him and was much liked not only by the faculty but also by our department staff. He would always had research support from NSF while he was in Davis, which he primarily used to support graduate students.

As Peter was keenly aware that the developments in theory and methodology of statistics are primarily driven by new types of data, to keep fully informed about new developments made it particularly important for him to travel frequently and also enticed him to collaborate with many other researchers. In his earlier years in

Davis, he used to take one day off for an excursion to one of the railroad museums in the neighborhood, Napa Valley, Muir Woods or Point Reyes, which would be joined by his wife, Jeannie, but more recently he was less inclined to go on such trips, and seemed to intensify his working habit, if that was even possible. Peter's focus on his work was legendary and he was the hardest working scientist many of us will ever encounter. Equally legendary was his mind-boggling speed in working out complex arguments and expansions, which enabled him to produce complex results in a short time and in printer-ready form, sometimes even while he was juggling back and forth between several demanding tasks that he completed simultaneously.

Peter was honored through many awards and prizes; he received the highest honors our field bestows. Two local Davis events stand out: One was a conference in 2012 that was held on the occasion of his 60th birthday. This conference featured an excellent scientific program with leading statisticians as speakers and was well attended. It showcased the vibrancy and continuing relevance of the topics Peter worked on, including high-dimensional statistics, errors-in-variables and functional data analysis. Another major event during Peter's tenure at Davis was his election in 2013 as Foreign Associate of the National Academy of Sciences (USA). This is of course a rarified and very high honor for a statistician, especially a non-U.S. citizen. We had a small celebration in the department (see Figure 2), which was attended by the Dean. Indeed, Peter had been the first faculty in the

FIG. 2. *Peter cutting the cake at the celebration for his election as a Foreign Associate of the National Academy of Sciences* (*USA*) *in April* 2013 *in the lounge of the Statistics Department in Davis*. *Photo taken by Pete Scully.*

FIG. 3. *Peter's orphaned bicycle in Davis.*

Division of Mathematical and Physical Sciences at UC Davis who was elected to the Academy while being on the faculty.

A feature of Davis that Peter particularly enjoyed was that he could get around town and campus efficiently with his bicycle and there was no need for a car. As Jeannie Hall reports, "Peter had his bike light and lock in his backpack all the time, not just when he went to Davis, because he was afraid of forgetting to take them when he did go to Davis. And he always kept the bike inside the apartment and kept it well maintained while he was in Davis" (Figure 3). When he was not in Davis, it was stored in our garage. While Peter found bicycling relaxing, his main hobby was photography, and he was well known for the artistic quality of his photos of landscapes, cityscapes and especially of trains. When he was once uncharacteristically stymied with a difficult proof, he came to my office with the request to show him a location where he could photograph trains. So we went with our bicycles along dirt roads criss-crossing the vast UC Davis experimental fields until we came to a spot right next to the railroad tracks near Putah Creek outside of Davis that met Peter's scrutiny. Peter would then spend hours there waiting for trains to pass by, where he was most interested in the huge transcontinental freight trains and their various Diesel locomotives (he was slightly irritated that the trains were not following an advertised schedule).

Peter's contributions to the Statistics department at UC Davis are immeasurable, and he contributed in many ways to make the UC Davis Statistics department a better department, not just by lending his prestige and name to the department and through his active contributions, but also simply through his regular presence and his unmatched generosity toward colleagues and students. When Peter passed away far too early at the age of 64 on 9 January 2016 in Melbourne, Australia, we lost an eminent scientist, foremost leader of nonparametric statistics, beloved colleague and outstanding departmental citizen. We will miss Peter. In his honor, the department has named the positions offered through its visiting Assistant Professor program as Hall Assistant Professors and the UC Davis Statistics conference series that is organized by the department annually will be dedicated to Peter's memory and will be known as the Peter Hall Conference.

2. A brief history of functional data analysis. This section contains a very brief and incomplete history of functional data analysis (FDA) to set the stage for Peter's entry into the field in 1998. The data atoms of FDA are random functions, and FDA deals with samples of such random functions. A distinctive feature of functional data is that they are infinite dimensional from the outset and often considered as Hilbert space valued data, in contrast to other data types encountered in statistics (even in the large p, small n frameworks the data are still finite-dimensional for each n). This means that tools for dimension reduction are essential. A common approach are expansions of the random functions into basis functions, which then can be truncated at a suitably large number of included terms. An example is the eigenbasis expansion and associated functional principal component analysis, which has gained a prominent place in the field, due to its theoretical attractiveness and good practical performance. Current challenges and potential future directions of FDA are briefly surveyed in Section 4.

The origins of FDA can be found in early papers on the decomposition of square integrable stochastic processes into series expansions to obtain a representation in the Hilbert space L^2. Specifically, in their respective Ph.D. theses, Karhunen and Grenander [Grenander (1950), Karhunen (1946)] proposed the concept of expanding a square integrable stochastic process X into its eigencomponents. With mean functions $\mu(t) = EX(t)$, consider centered processes and variables $X^c(t) = X(t) - \mu(t)$, the covariance function $C(s,t) = \operatorname{cov}(X(s), X(t))$ and the auto-covariance operator of the process X,

$$A(g) = \int_T C(s,t)g(s)\,ds \qquad \text{for } g \in L^2(T). \tag{2.1}$$

Under mild conditions, this linear operator is a trace class and, therefore, compact Hilbert–Schmidt operator. By Mercer's theorem, $C(s,t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s)\phi_j(t)$, where (λ_j, ϕ_j) are the eigenvalues and eigenfunctions of the operator A, and the ϕ_j form an orthonormal basis, the eigenbasis.

Grenander emphasized Gaussian (and also Poisson) processes and in a side remark introduced the functional linear model in the Gaussian context in an implicit form, which for $Y \in \mathcal{R}$ and $X \in L^2(T)$ postulates that there is a function $\beta \in L^2(T)$ such that $E(Y^c|X) = \int_T X^c(t)\beta(t)\,dt$. Grenander framed this as an out of domain prediction problem, and in his analysis made use of the representation of X in the eigenbasis,

$$X^c(t) = \sum_{j=1}^{\infty} \xi_j \phi_j(t), \tag{2.2}$$

known as the Karhunen–Loève representation of X.

If X and Y are jointly Gaussian random processes on an interval $[0, 1]$, there exists a square integrable function β on $[0, 1] \times [0, 1]$ such that

$$E\big(Y^c(t)|X\big) = \int X^c(s)\beta(s,t)\,ds, \tag{2.3}$$

a model that can be viewed as an extension to the functional case of the simple linear regression model that one has for bivariate normal data, with $X, Y \in \mathcal{R}$, where $E(Y^c|X) = \beta X^c$ [Ramsay and Dalzell (1991)]. A simpler version of this model with scalar response Y is the above Grenander regression,

$$E\big(Y^c|X\big) = \int X^c(s)\beta(s)\,ds. \tag{2.4}$$

The idea to expand random curves in a suitable basis with random coefficients in more practical statistical settings seems to have appeared for the first time in Rao (1958) and Tucker (1958), in a somewhat rudimentary form. More rigorous ideas about functional principal component analysis and the asymptotic distribution of the eigenvalues appeared in Kleffe (1973), which is, for example, cited in the comprehensive Ph.D. thesis by Dauxois and Pousse (1976) with subsequent publication Dauxois, Pousse and Romain (1982). This thesis laid the groundwork for theoretical analysis of estimators for the case where the random trajectories $X_1, \dots, X_n$ that form a sample of realized trajectories of the underlying process X are assumed to be fully observed, and it introduced perturbation theory as an important tool to study corresponding estimators [Kato (1995)].

It is not surprising that functional data as observed in practice hardly ever consist of completely observed functions and at best are observed on a dense grid, often with measurement errors, and at worst with sparse measurements, which introduces additional challenges. There were also early hints of limitations of linear methods due to the presence of time warping, that is, random distortions of the time axis, which led to the curve registration or alignment problem.

There are two "classical" challenges for theoretical analysis in FDA. The first challenge is that the compactness of the operator A in (2.1) implies that A is not invertible. This corresponds to an inverse problem, which affects, for example the estimation of the regression function β in model (2.3), where one would like to be able to obtain the least squares solution, which however is not feasible. For example, considering the regression model (2.3), define $C_{XY}(s,t) = \operatorname{cov}(X(s), Y(t))$ and the linear operator $R : L^2 \times L^2 \to L^2 \times L^2$ by $R\beta(s,t) = \int C_{XY}(s,w)\beta(w,t)\,dw$. Then a "functional normal equation" takes the form [He, Müller and Wang (2000)] $C_{XY} = R\beta$ for $\beta \in L^2(I_X \times I_X)$. Since R as a compact operator is not invertible, this cannot be directly solved. Popular approaches to this inverse problem are to use truncated expansions of X and β for which the equation is solvable, or to regularize β, adopting a version of Tikhonov regularization, which is most often implemented by penalized splines.

The second challenge is a consequence of the fact that the integral operator A in (2.1) is actually not known and must be estimated through an estimate $\hat{C}$ of the covariance kernel C. This is the realm of perturbation theory, which provides tools to use bounds between $\hat{C}$ and C to infer bounds between estimated and true eigenvalues/eigenfunctions. So establishing asymptotic limit theory for the eigencomponents typically requires to derive the convergence of estimated mean and covariance functions in a desirable metric, and then to apply perturbation arguments to obtain convergence of the eigencomponents.

The most commonly used perturbation argument is encapsulated in Lemma 4.3 of Bosq (2000), which is a direct extension of corresponding multivariate results to the functional case [Hsing and Eubank (2015)]. If one has two covariance kernels C and $\tilde{C}$, with respective eigenvalues and eigenfunctions (λ_k, ϕ_k), $(\tilde{\lambda}_k, \tilde{\phi}_k)$, $k \geq 1$, then this perturbation result yields the bounds

$$(2.5) \qquad |\lambda_k - \tilde{\lambda}_k| \leq \|C - \tilde{C}\|, \qquad \|\phi_k - \tilde{\phi}_k\| \leq 2\sqrt{2}\delta_k^{-1}\|C - \tilde{C}\|,$$

where the norms correspond to the respective L^2 norms and δ_k is defined by $\delta_1 = \lambda_1 - \lambda_2$, $\delta_k = \min_{j \leq k}(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$, $k \geq 2$. Peter worked on both of the above challenges, deriving his own perturbation results and using these to make seminal contributions to functional linear regression and functional principal component analysis.

The work by Dauxois and Pousse marked the beginning of the French school of FDA or statistics in Hilbert space, with key researchers located in Toulouse. These pioneers were very productive over the years [Bosq (2000), Ferraty and Vieu (2006)] and developed basic theory, then became increasingly interested in applied methods and data analysis and hosted visitors that included Peter Hall more recently and Jim Ramsay earlier. Jim Ramsey coined the name "Functional Data Analysis" and has been a tireless promoter of this area [Ramsay and Silverman (2005)].

A smaller group of early researchers can be characterized as the Zürich–Heidelberg school which was applications and computing oriented at its inception. Its leader was Theo Gasser with inspirations from Peter Huber, prominent members included Wolfgang Härdle and Alois Kneip and it started out with smoothing methods in the late 1970s and early 1980s, with a peak of activity in the 1980s and 1990s. Early focus areas included derivatives, nonparametric estimation of growth curves [Gasser et al. (1984)] and time warping, with an emphasis on the landmark method and shape-invariant modeling [Gasser et al. (1984), Kneip and Gasser (1992)].

In addition, there were some other early pioneers with high impact in this area, including Bernard Silverman and John Rice [Rice and Silverman (1991), Rice and Wu (2001)], with several others making seminal contributions [Castro, Lawton and Sylvestre (1986), Staniswalis and Lee (1998)]. In addition to work that laid the foundations of linear modeling in FDA through functional principal components and the functional linear model and then nonlinear time warping, another line of early work addressed issues of functional correlation [Leurgans, Moyeed and Silverman (1993)]. Additional review of developments in FDA can be found in Wang, Chiou and Müller (2016).

When Peter Hall became interested in this area, his first paper was actually co-authored with Theo Gasser and Brett Presnell [Gasser, Hall and Presnell (1998)], while some of his last work in this area was joint with prominent representatives of the French-Toulouse school, Fréderic Ferraty and Philippe Vieu [Ferraty, Hall and Vieu (2010)]. Peter's collaborations with key researchers in FDA exposed him to a large variety of challenges that he enjoyed to address. Peter was a problem solver par excellence.

3. Peter Hall's contributions to functional data analysis. Peter worked on an incredible number of different problems in diverse areas including point processes, time series, extreme values, quantification of the roughness of surfaces and small area estimation, among others, with an emphasis on nonparametric methods. Apart from his early work on martingales and convergence of sums of random variables [Hall and Heyde (1980)] that was directly related to his Master's thesis at the Australian National University (advisor: Chris Heyde) and Ph.D. thesis at the University of Oxford (advisor: John Kingman), the areas of his earlier work that were particularly relevant for his contributions to functional data analysis (FDA) included: (1) The Bootstrap, where Peter made several early pioneering contributions to the theory; see Chen (2016). This was the area where he established his reputation as a top researcher. He also wrote one of the first papers to apply bootstrap to functional data. (2) Smoothing methods, where Peter's contributions were very broad, with particularly influential papers on the asymptotics of error measures, boundary problems, error variance estimation, and bandwidth choice for kernel type smoothers in nonparametric density estimation and nonparametric regression; see Cheng and Fan (2016). He employed various smoothing methods and utilized their properties in his work in FDA. (3) High-dimensional statistics, where Peter published an influential early article on the near-normality of random projections; see Samworth (2016). (4) Deconvolution and errors in variables, an area on which Peter worked throughout his career, including some of his very last papers; see Delaigle (2016).

His previous work in areas related to FDA provided Peter with a toolbox that allowed him to address some of the toughest theory problems in FDA. His primary motivation to work in this area was that this new type of data posed theoretical challenges that he found intriguing [Delaigle and Wand (2016)]. Peter worked primarily on problems that were within the realm of his formidable toolkit, and would often approach those aspects of a problem for which his tools were directly applicable. The theory challenges of FDA fitted him particularly well, as they often require complex expansions, in which Peter was an unsurpassed master. FDA also draws on tools from smoothing, functional analysis and stochastic processes, all areas in which he had accumulated vast expertise in the course of his previous research.

Peter's work in FDA was carried out jointly in collaboration with various co-authors, where Peter typically was the driving force of the mathematical developments and proofs. He pioneered a deeper theoretical understanding of existing techniques such as functional principal component analysis, functional regression and classification and also introduced seminal new concepts such as *perfect classification* that arise for functional data. His work in FDA alone would have been sufficient for a distinguished career. Peter's specific contributions to FDA can be clustered into three areas, as follows.

3.1. *Estimation of densities and modes in function space.* The nonexistence of a density w.r.t. Lebesgue measure in the function space L^2 had been previously attributed to the insufficient size of the small ball probabilities, which prevents the existence of a Radon–Nikodym derivative [Li and Linde (1999)]. Trying to estimate such a nonexisting density would seem like a fool's errand. Not so for Peter and his co-authors, who boldly went ahead in Gasser, Hall and Presnell (1998) and simply redefined the problem to find density and modes for a finite-dimensional approximation of the functional data,

$$X(t) \approx \sum_{j=1}^{p} \zeta_j \psi_j(t) \tag{3.1}$$

for a suitable basis ψ_j. Once one replaces the functions X by their finite-dimensional approximations, the density problem is essentially reduced to finding the density of the finite-dimensional vector of coefficients ζ_j, $j = 1, \dots, p$.

Of course, if p is larger than 3 there would still be a curse of dimensionality for nonparametric density estimation, so this method would still not be straightforward to deploy. The main purpose of Peter's work was to find the mode of the random distribution with a suitable algorithm, with a view toward finding a "typical" value for the distribution and toward clustering of functional data. Clustering of functional data has since become a popular area of research in FDA [Chiou and Li (2008)] that Peter revisited 14 years after this initial contribution [Delaigle, Hall and Bathia (2012)].

The mode finding problem was revisited in Hall and Heckman (2002), where the emphasis was on a mode climbing algorithm similar to a mean update algorithm for functional data. The authors also introduced a version of a kernel density estimator based on replacing differences in the kernel arguments by a general distance in function space, $\hat{f}(x) = \alpha_n \sum_{i=1}^{n} K(d(x, X_i)/h)$ for a multiplier α_n, a distance d in L^2 and a bandwidth h and derived the consistency of the mode climbing algorithm.

The density problem made yet another appearance in Peter's work. In Delaigle and Hall (2010), complex arguments are used to show that the finite approximation approach in Gasser, Hall and Presnell (1998) can be refined by replacing the fixed basis functions ψ_j by eigenfunctions ϕ_j of the underlying processes X, more precisely by estimated eigenfunctions. Then a finite approximation to the density (the "surrogate density") can be expressed in terms of the estimated functional principal component scores. The crux of this argument is that the densities must be estimated from estimated scores, and the critical step is to show that when substituting the functional principal component scores by their estimates one still obtains consistent density estimates. To prove this warranted a separate paper [Delaigle and Hall (2011)].

Overall, the construction of densities for functional data still remains a challenge. Various possible approaches invite further investigation. Since the full L^2 space does not admit the notion of a usual density, one approach to this problem is to restrict consideration to statistically sensible subspaces of L^2 that keep the infinite dimensionality intact and at the same time provide interpretable and meaningful representations of the functional elements.

3.2. *Theory of functional principal components*. The year 2006 saw a burst of activity from Peter in regard to FDA, with an emphasis on functional principal component analysis; it was his *annus mirabilis* in FDA. He published four papers on this topic, two in the *Journal of the Royal Statistical Society Series B* and two in the *The Annals of Statistics*. The properties of estimates of the eigenvalues and eigenfunctions of the operator A in (2.1) are derived by perturbation theory, which relies on expansions of the estimated eigencomponents around their targets. Two main expansions have been used: One that is grounded in expansions of resolvent operators [Dauxois and Pousse (1976)] and a second more popular approach that makes use of the more direct bounds in equation (2.5) [Bosq (2000)]. Given Peter's ingenuity to develop complex higher order expansions, it is not surprising that he took up the challenge to develop more advanced expansions. He chose the direct approach with a parallel development to (2.5) A key paper is Hall and Hosseini-Nasab (2006), which is highly cited and has a focus on higher order expansions for perturbations. In this paper, it is shown that the spacings of the eigenvalues have a first-order effect on eigenfunction estimation and a second-order effect on eigenvector estimation. These results are then used to derive inference for eigenvalues and eigenfunctions by bootstrapping. This is likely the first systematic study of using bootstrap methods for functional data.

These developments were instrumental for a follow-up paper [Hall, Müller and Wang (2006)], where the results of Hall and Hosseini-Nasab (2006) and Yao, Müller and Wang (2005) were extended to systematically compare the estimation of eigenvalues and eigenfunctions for two sampling scenarios: a classical scenario with fully observed functional data and a second longitudinal scenario, where one only has a sparse number of measurements per curve that are taken at random locations and are contaminated by noise. It is shown that eigenvalue estimation is not as sensitive to the design as eigenfunction estimation is, where the latter has a nonparametric convergence rate in the sparse and a parametric convergence rate in the fully observed case. The transition from the sparse to the dense scenario gives rise to a change in the asymptotic behavior. When the order of the number of sampled measurements per random function is larger than $n^{-1/4}$, where the sample size of the functional data is n, this behavior is first-order equivalent to that of completely observed functional data. This was the first time a *phase transition* was observed to occur in the sampling schemes of functional data, with the discovery of additional phase transitions to follow later [Cai and Yuan (2011)]. The influence of the sparsity of designs on functional principal component analysis also motivated other recent work [Li and Hsing (2010), Zhang and Wang (2016)].

It is sometimes of interest to assume that functional data are of finite dimensionality and then the question arises how to determine the correct dimension. This question was studied by Peter for the difficult situation where one has additional noise and discrete measurements of the functional data in Hall and Vial (2006a), where some asymptotically motivated criteria were proposed. Another question that has consequences for the interpretation of FPCA is to determine the structure of the eigenfunctions, and especially their shape properties. Here especially, extrema in the eigenfunctions that indicate points where increased variability occurs are of interest and Peter studied the properties of corresponding empirical estimates in Hall and Vial (2006b), including bootstrap methods to assess their strength. These results also make use of the perturbation results of Hall and Hosseini-Nasab (2006).

Another version of functional principal component analysis for a different data type, where one has repeatedly observed data of generalized non-Gaussian type for each subject, was developed in Hall, Müller and Yao (2008). The main feature of this approach was the combination of a latent Gaussian process that generates the observed data with an independent random mechanism that, given the trajectory value, generates an actual observation. For example, if an observation at time t is Bernoulli(p), the value of $p = p(t) = \exp(X(t))/[1 + \exp(X(t))]$ corresponds to the expit of the value of the smooth random trajectory $X(t)$ at t, and a Bernoulli response sampled at time t would be obtained by an independent Bernoulli(p) experiment with $p = p(t)$. This approach was developed in the framework of sparse data with few observations per subject and shown to produce estimates for the principal component scores that can serve as random effects in further statistical analysis. An additional feature was a result on projecting covariance surfaces obtained by smoothing onto L^2 projections that only contain positive eigenvalues, yielding positive definite symmetric surfaces.

3.3. *Functional regression, classification and related topics.* The methodology used by Peter in these core areas of FDA was primarily based on functional principal component analysis, where he skillfully deployed the expertise he had gained by working on the projects described in the previous subsection. His first two papers on functional regression models concerned model (2.4) with scalar response and pioneered the study of exact convergence rates in the framework of this model. In the high-impact paper Cai and Hall (2006), the emphasis was not on properties of estimates of the regression parameter function as in previous work [Cardot, Ferraty and Sarda (1999)], but rather on estimates of the predictor $\eta = \int \beta(t)X(t)\,dt$, which turns out to be an easier problem in the sense that one can obtain faster rates of convergence, due to the smoothing effect of the integral. In the paper, it was shown that undersmoothing of estimators of the parameter function β can lead to rates of convergence for estimates of η as fast as $n^{-1/2}$, that is, parametric rates, under certain assumptions. In many situations, the rate was shown to be nonparametric, with exact rates determined by the interplay between the eigenvalue decay rate of the auto-covariance operator (2.1) of the predictor process X, the smoothness of β, as measured by the decay rate of its coefficients in the eigenbasis and the smoothness of X, as measured by the decay rate of its functional principal component scores.

Convergence rates of estimates of the regression function β itself were derived in Hall and Horowitz (2007). The techniques used are similar to those in Cai and Hall (2006), invoking similar smoothness conditions on the auto-covariance operators and the function β to obtain optimal rates. Peter considered a weighted least squares version of fitting model (2.4) in subsequent work [Delaigle, Hall and Apanasovich (2009)] and studied the number of principal components one should include for principal component based implementations of model (2.4) [Hall and Yang (2010)]. He analyzed an extension of the functional linear model to a more general single index model with possibly multiple indices in Chen, Hall and Müller (2011). In one of his last papers in FDA, Peter revisited the functional linear model (2.4) and derived asymptotic results for the domain selection problem and also studied its identifiability [Hall and Hooker (2016)]. In this important but difficult problem, one seeks to determine the best subinterval of the domain of X, in the sense that linear regression with the selected subinterval as predictor domain provides the best linear predictor η in dependence on the selected domain. This problem is related to the historical functional linear model [Malfait and Ramsay (2003)] and extends history versions of functional varying coefficient models [Şentürk and Müller (2010)].

Sometimes it is of interest to quantify the change of a response when a predictor changes, where the rate of change corresponds to the derivative of the response with respect to the predictor. This problem is nonstandard when predictors are random functions; it was addressed in Hall, Müller and Yao (2009), where asymptotic properties for a kernel-based method were studied. Since this method involves functional predictors in a nonparametric model, it is subject to a serious curse of dimensionality. An additive version that breaks the problem down into a series of one-dimensional additive predictors was devised later [Müller and Yao (2010)]. This methodology provides a practical illustration of Gâteaux derivatives and can be used for gradient based optimization if one aims to determine the shape of predictor trajectories for which a response is maximized.

While many of Peter's papers in FDA were based on fully observed functional data, there are some notable exceptions. The effect of smoothing on discretely sampled functional data when the goal is a two sample test was studied in Hall and Van Keilegom (2007). In this paper, it is shown that it is best to use the same smoothing parameter across all random curves for maximizing power, and asymptotic characterizations of situations where the smoothing step is negligible are provided. A goodness-of-fit test for the null hypothesis that the observed random trajectories follow a parametric model was proposed in Bugni et al. (2009). The alternatives are nonparametric and this testing problem was illustrated with interesting applications in econometrics. Peter revisited the smoothing problem for discretely sampled data later [Carroll, Delaigle and Hall (2013)]. The simple smoothing rules advocated in Hall and Van Keilegom (2007) were found not to apply to functional classification. Instead, fairly complex relations between the degree of smoothing and classification results were discovered.

When one has completely observed functional predictors, it is often of interest to reduce the burden of recording the entire function, as is required in model (2.4). Instead, one would like to find critical points within the predictor domain so that a resulting multivariate linear model that only uses the observations at the selected predictor points provides a good approximation to the fitting of the entire functional linear model. This challenge is motivated by applications in chemometrics and led to a promising approach proposed in Ferraty, Hall and Vieu (2010). This work employs boosting methods that had been developed in prior work for functional data [Ferraty and Vieu (2009)] and has many possible ramifications [Ji and Müller (2016), Kneip, Poss and Sarda (2016)].

Another popular method in chemometrics is partial least squares for regression, where instead of maximizing the correlation between the responses and a linear combination of the predictors one aims at maximizing the covariance. This approach is particularly attractive for functional data as it avoids the inverse problem that is at the root of the problems one encounters in functional regression and canonical correlation modeling [Yang, Müller and Stadtmüller (2011)]. The full partial least squares algorithm is iterative, which makes its theoretical analysis very challenging, especially in the case of functional predictors. Peter braved these challenges in a tour de force and was able to derive rates of convergence [Delaigle and Hall (2012a)].

Peter also left a substantial legacy of works on functional classification, which recently has become a very active line of research [Cuevas (2014)]. The functional classification problem was on his radar screen already in 2001, as reflected in Hall, Poskitt and Presnell (2001). In this early approach, the random functions are projected onto their first K principal components and these random vectors are then used in vector-based discriminant analysis. An extension where additionally some nonlinear functionals are considered, but with the same idea of projecting the functional data onto finite-dimensional random vectors, appeared in Hall, Lee and Park (2007).

Peter's main and seminal contributions to functional classification appeared in two papers in 2012 and 2013, co-authored with Aurore Delaigle. In the 2012 paper [Delaigle and Hall (2012b)], a functional version of Fisher's linear discriminant analysis was considered. The main discovery in the paper is that even with this simple method that is known to be inefficient in many classical classification problems where the predictors are random vectors, one can obtain consistency in classification. This consistency is studied for both Gaussian and non-Gaussian situations and under specific conditions for the functional Mahalanobis distances between the two groups to be classified. Here, consistency means that under suitable regularity conditions the misclassification error converges to zero. This phenomenon is referred to in the paper as perfect classification.

The method that is used to demonstrate this property is based on projecting on a linear discriminant function. This leads to the convergence of the misclassification error to zero in the Gaussian case. This is plausible, since in the infinite-dimensional case any differences between the means of the two groups will eventually be reflected in a functional projection that targets directions for which the projections are small. Then the linear classifier will eventually detect the difference in the means of the predictors, leading to perfect classification. The non-Gaussian case is also studied in the paper.

This approach was further extended in Delaigle and Hall (2013), which introduces the additional challenge to the functional classification problem that the predictor functions are only observed on random subintervals that are not too short. A heuristic procedure is introduced to extend the incompletely observed curves. The theoretical part of the paper introduces a quadratic discriminant method that extends the results of Delaigle and Hall (2012b) to establish perfect classification in this more general setting. Undoubtedly, these works and in general Peter's overall contributions to functional data analysis have moved the field substantially forward and will have lasting impact.

4. From functional data to random objects. This section contains some speculative and subjective thoughts of where functional data analysis might be headed in the next couple of years. Peter's contributions were instrumental in building the foundations of what can be viewed as "first generation" functional data analysis, characterized by predominantly linear methods, such as functional principal component analysis, functional linear regression or linear functional discriminants.

As data of functional type are becoming more complex, the current trend is toward "next generation" functional data, to use a term coined by Jane-Ling Wang at the London Workshop on the Future of the Statistical Sciences; see page 23 of the report, available at http://www.worldofstatistics.org/wos/pdfs/Statistics&Science-TheLondonWorkshopReport.pdf. An incomplete list of such next generation data is as follows:

- Functional data that are irregularly and sometimes extremely sparsely observed, with maybe two measurements per subject that are made at random locations; Peter's work included major contributions toward a better understanding of such functional data of "longitudinal type" [Delaigle and Hall (2013), Hall, Müller and Wang (2006), Hall, Müller and Yao (2008)], but many open problems remain.
- The interface between high-dimensional and functional data, which may occur simultaneously among the predictors in regression models [Kneip and Sarda (2011), Kong et al. (2016)] or models where high-dimensional data are represented as functional data [Chen et al. (2011)], and more recently, functional data as part of "big data" [Chen et al. (2015)], where Peter's key contributions in this area were in the context of predictor selection for functional linear models [Ferraty, Hall and Vieu (2010)] and in classification [Delaigle and Hall (2012b)].
- Nonlinear models for functional data, where models that address time-warping have been well studied and nonlinear versions of functional principal component analysis are of interest [Chen and Müller (2012), Kneip and Ramsay (2008)], in addition to nonlinear regression models, to which Peter contributed in Chen, Hall and Müller (2011).
- Functional data in classical biostatistical models, such as survival analysis, where the Cox regression model plays a prominent role, with functional versions in Qu, Wang and Wang (2016), and longitudinal or repeated measurements designs where the repeated measurements are functional or where functions have a multivariate time domain [Chen, Delicado and Müller (2016)] and other forms of dependent functional data, such as vectors of functional data.
- Functional time series, where one has only one sequence of repeatedly observed functional data that are sampled over a regular time grid and stationarity is a common assumption; while a complete theory of functional auto-regressive models was already presented in Bosq (2000), this area recently has met with increasing interest, with models both in the spectral as well as time domains and also spatio-temporal extensions [Horváth and Kokoszka (2012), Panaretos and Tavakoli (2013)].

A promising approach that has emerged is to make the assumption that there exists a smooth underlying stochastic process from which a trajectory is sampled by a random mechanism and then a second random mechanism generates the actual observed data, conditional on the unobserved random trajectory. Functional dependencies are then modeled through dependency structures of the underlying stochastic processes. This idea is exemplified in Hall, Müller and Yao (2008).

While functional data analysis differs in key aspects from multivariate data analysis, mainly due to the infinite dimensionality of functional data and the topological aspects of proximity of predictors and continuity of the functional objects, both fields deal with data that lie in a vector space. For the case of functional data, this is commonly assumed to be the Hilbert space L^2 of square integrable functions or alternatively a reproducing kernel Hilbert space [Hsing and Eubank (2015)]. Such data are *Hilbertian*. The inner (scalar) product enables projections and linear methods, and makes it possible to extend the notions of mean, regression and principal component analysis, which are central for practical data analysis, from the multivariate to the functional case.

There are various situations where functional data lie in a subset or submanifold of a vector space that itself is not a vector space: Two prominent examples are random samples of density functions [Kneip and Utikal (2001)] and functional data with time warping components [Gasser, Sroka and Jennen-Steinmetz (1986)]. Other examples are provided by shape-constrained functional data such as samples of monotone [Ramsay and Silverman (2005)], unimodal or convex random functions. Since one cannot perform linear operations or projections and stay within the relevant subspace, the usual methods of functional data analysis will not work well for these types of inherently nonlinear data.

For such data, one can sometimes find a suitable transformation that maps the subset in which they live to a suitable vector space where there are no constraints. Such an approach can, for example, be implemented for functional data that are densities [Petersen and Müller (2016)]; similarly, for time-warped functional data one can apply a manifold embedding, for instance with Isomap [Tenenbaum, de Silva and Langford (2000)] that will then map the subset where the functional data live to a typically low-dimensional vector space [Chen and Müller (2012)] under certain geometric assumptions. Therefore, this type of functional objects can be characterized as *quasi-Hilbertian* in the sense that there exists a generally "smooth" invertible transformation of such objects into a Hilbert space, where the regular tools of functional or multivariate analysis can be applied.

However, this transformation approach will not always be applicable and statisticians increasingly encounter situations of functional or nonfunctional data that are more decidedly *non-Hilbertian*. An example are functional data that lie on a Hilbert sphere or other closed Hilbert manifold or nonfunctional data such as covariance and correlation matrices, trees and networks. Such data typically reside in a metric space for a suitably chosen metric, and in some special cases may lie on a manifold.

A key feature of functional data analysis is that one has available a sample of i.i.d. realizations of the underlying processes that are assumed to generate the observed data. There is currently a shortage of principled statistical methods to handle i.i.d. samples of non-Hilbertian data that lie in a metric space that is not a vector space and might not be a manifold. One may refer to i.i.d. samples of data in metric as well as linear spaces as *random objects*, which include functional data as a special case. Methods that have been successful for FDA will inspire the future development of methods for such random objects. Challenges include the choice of a suitable metric and the construction of means, modes of variation, regression and inference.

Acknowledgments. I wish to thank Aurore Delaigle, Jeannie Hall, Runze Li, Richard Samworth and Jane-Ling Wang for valuable feedback on earlier drafts of this paper, Ray Carroll for making accessible a memorial talk he gave at Texas A&M University, Aurore Delaigle for providing access to a preprint of Delaigle and Wand (2016), and Steve Marron for providing valuable information about Peter Hall on his personal web page at the University of North Carolina, http://marron.web.unc.edu/sample-page/peter-hall-memorial-page/. Some of the memorial material about Peter that is presented in the Introduction has appeared in similar form in an obituary I wrote for Bernoulli News, Vol. 23, No. 1 (2016), pages 9–11.

REFERENCES

BOSQ, D. (2000). *Linear Processes in Function Spaces*: *Theory and Applications*. *Lecture Notes in Statist*. **149**. Springer, New York. MR1783138

BUGNI, F. A., HALL, P., HOROWITZ, J. L. and NEUMANN, G. R. (2009). Goodness-of-fit tests for functional data. *Econom*. *J*. **12** S1–S18.

CAI, T. T. and HALL, P. (2006). Prediction in functional linear regression. *Ann*. *Statist*. **34** 2159–2179. MR2291496

CAI, T. T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann*. *Statist*. **39** 2330–2355. MR2906870

CARDOT, H., FERRATY, F. and SARDA, P. (1999). Functional linear model. *Statist*. *Probab*. *Lett*. **45** 11–22. MR1718346

CARROLL, R. J., DELAIGLE, A. and HALL, P. (2013). Unexpected properties of bandwidth choice when smoothing discrete data for constructing a functional data classifier. *Ann*. *Statist*. **41** 2739–2767. MR3161446

CASTRO, P. E., LAWTON, W. H. and SYLVESTRE, E. A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* **28** 329–337.

CHEN, X. (2016). Peter Hall's contribution to the bootstrap. *Ann*. *Statist*. **44** 1821–1836.

CHEN, K., DELICADO, P. and MÜLLER, H.-G. (2016). Modeling function-valued stochastic processes, with applications to fertility dynamics. *J*. *Roy*. *Statist*. *Soc*. *Ser*. *B*. To appear.

CHEN, D., HALL, P. and MÜLLER, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann*. *Statist*. **39** 1720–1747. MR2850218

CHEN, D. and MÜLLER, H.-G. (2012). Nonlinear manifold representations for functional data. *Ann*. *Statist*. **40** 1–29. MR3013177

CHEN, K., CHEN, K., MÜLLER, H.-G. and WANG, J.-L. (2011). Stringing high-dimensional data for functional analysis. *J*. *Amer*. *Statist*. *Assoc*. **106** 275–284. MR2816720

CHEN, K., ZHANG, X., PETERSEN, A. and MÜLLER, H.-G. (2015). Quantifying infinite-dimensional data: Functional data analysis in action. *Statistics in Biosciences* 1–23.

CHENG, M. Y. and FAN, J. (2016). Peter Hall's contributions to nonparametric function estimation and modeling. *Ann. Statist.* **44** 1837–1853.

CHIOU, J.-M. and LI, P.-L. (2008). Correlation-based functional clustering via subspace projection. *J. Amer. Statist. Assoc.* **103** 1684–1692. MR2724685

CUEVAS, A. (2014). A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference* **147** 1–23. MR3151843

DAUXOIS, J. and POUSSE, A. (1976). Les analyses factorielles en calcul des probabilités et statistique: Essai d'étude synthétique. Ph.D. thesis, Toulouse.

DAUXOIS, J., POUSSE, A. and ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** 136–154. MR0650934

DELAIGLE, A. (2016). Peter Hall's main contributions to deconvolution. *Ann. Statist.* **44** 1854–1866.

DELAIGLE, A. and HALL, P. (2010). Defining probability density for a distribution of random functions. *Ann. Statist.* **38** 1171–1193.

DELAIGLE, A. and HALL, P. (2011). Theoretical properties of principal component score density estimators in functional data analysis. *Sankt-Peterburgskii Universitet. Vestnik. Seriya* 1. *Matematika, Mekhanika, Astronomiya* **2** 55–69.

DELAIGLE, A. and HALL, P. (2012a). Methodology and theory for partial least squares applied to functional data. *Ann. Statist.* **40** 322–352. MR3014309

DELAIGLE, A. and HALL, P. (2012b). Achieving near perfect classification for functional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 267–286. MR2899863

DELAIGLE, A. and HALL, P. (2013). Classification using censored functional data. *J. Amer. Statist. Assoc.* **108** 1269–1283. MR3174707

DELAIGLE, A., HALL, P. and APANASOVICH, T. V. (2009). Weighted least squares methods for prediction in the functional data linear model. *Electron. J. Stat.* **3** 865–885. MR2534205

DELAIGLE, A., HALL, P. and BATHIA, N. (2012). Componentwise classification and clustering of functional data. *Biometrika* **99** 299–313. MR2931255

DELAIGLE, A. and WAND, M. (2016). A conversation with Peter Hall. *Statist. Sci.* **31** 275–304.

FERRATY, F., HALL, P. and VIEU, P. (2010). Most-predictive design points for functional data predictors. *Biometrika* **97** 807–824. MR2746153

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York. MR2229687

FERRATY, F. and VIEU, P. (2009). Additive prediction and boosting for functional data. *Comput. Statist. Data Anal.* **53** 1400–1413. MR2657100

GASSER, T., HALL, P. and PRESNELL, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 681–691. MR1649539

GASSER, T., SROKA, L. and JENNEN-STEINMETZ, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633. MR0897854

GASSER, T., MÜLLER, H.-G., KÖHLER, W., MOLINARI, L. and PRADER, A. (1984). Nonparametric regression analysis of growth curves. *Ann. Statist.* **12** 210–229. MR0733509

GRENANDER, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.* **1** 195–277. MR0039202

HALL, P. and HECKMAN, N. E. (2002). Estimating and depicting the structure of a distribution of random functions. *Biometrika* **89** 145–158. MR1888371

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. MR0624435

HALL, P. and HOOKER, G. (2016). Truncated linear models for functional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* To appear.

HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 109–126. MR2212577

HALL, P., LEE, Y. K. and PARK, B. U. (2007). A method for projecting functional data onto a low-dimensional space. *J. Comput. Graph. Statist.* **16** 799–812. MR2412483

HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. MR2278365

HALL, P., MÜLLER, H.-G. and YAO, F. (2008). Modelling sparse generalized longitudinal observations with latent Gaussian processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 703–723. MR2523900

HALL, P., MÜLLER, H.-G. and YAO, F. (2009). Estimation of functional derivatives. *Ann. Statist.* **37** 3307–3329. MR2549561

HALL, P., POSKITT, D. S. and PRESNELL, B. (2001). A functional data-analytic approach to signal discrimination. *Technometrics* **43** 1–9. MR1847775

HALL, P. and VAN KEILEGOM, I. (2007). Two-sample tests in functional data analysis starting from discrete data. *Statist. Sinica* **17** 1511–1531. MR2413533

HALL, P. and VIAL, C. (2006a). Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 689–705. MR2301015

HALL, P. and VIAL, C. (2006b). Assessing extrema of empirical principal component functions. *Ann. Statist.* **34** 1518–1544. MR2278366

HALL, P. and YANG, Y.-J. (2010). Ordering and selecting components in multivariate or functional data linear prediction. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 93–110. MR2751245

HE, G., MÜLLER, H.-G. and WANG, J.-L. (2000). Extending correlation and regression from multivariate to functional data. In *Asymptotics in Statistics and Probability* (M. L. Puri, ed.) 301–315. VSP International Science Publishers, Leiden.

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer, New York. MR2920735

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, Chichester. MR3379106

JI, H. and MÜLLER, H.-G. (2016). Optimal designs for longitudinal studies via functional data analysis. *J. Roy. Statist. Soc. Ser. B*. To appear.

KARHUNEN, K. (1946). Zur Spektraltheorie stochastischer Prozesse. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.* **1946** 7. MR0023012

KATO, T. (1995). *Perturbation Theory for Linear Operators*. Springer, Berlin. MR1335452

KLEFFE, J. (1973). Principal components of random variables with values in a separable Hilbert space. *Math. Operationsforsch. Statist.* **4** 391–406. MR0391402

KNEIP, A. and GASSER, T. (1992). Statistical tools to analyze data representing a sample of curves. *Ann. Statist.* **20** 1266–1305. MR1186250

KNEIP, A., POSS, D. and SARDA, P. (2016). Functional linear regression with points of impact. *Ann. Statist.* **44** 1–30. MR3449760

KNEIP, A. and RAMSAY, J. O. (2008). Combining registration and fitting for functional models. *J. Amer. Statist. Assoc.* **103** 1155–1165. MR2528838

KNEIP, A. and SARDA, P. (2011). Factor models and variable selection in high-dimensional regression analysis. *Ann. Statist.* **39** 2410–2447. MR2906873

KNEIP, A. and UTIKAL, K. J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96** 519–542. MR1946423

KONG, D., XUE, K., YAO, F. and ZHANG, H. H. (2016). Partially functional linear regression in high dimensions. *Biometrika* **103** 147–159. MR3465827

LEURGANS, S. E., MOYEED, R. A. and SILVERMAN, B. W. (1993). Canonical correlation analysis when the data are curves. *J. Roy. Statist. Soc. Ser. B* **55** 725–740. MR1223939

LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38** 3321–3351. MR2766854

LI, W. V. and LINDE, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.* **27** 1556–1578. MR1733160

MALFAIT, N. and RAMSAY, J. O. (2003). The historical functional linear model. *Canad. J. Statist.* **31** 115–128. MR2016223

MÜLLER, H.-G. and YAO, F. (2010). Additive modelling of functional gradients. *Biometrika* **97** 791–805. MR2746152

PANARETOS, V. M. and TAVAKOLI, S. (2013). Fourier analysis of stationary time series in function space. *Ann. Statist.* **41** 568–603. MR3099114

PETERSEN, A. and MÜLLER, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.* **44** 183–218. MR3449766

QU, S., WANG, J.-L. and WANG, X. (2016). Optimal estimation for the functional Cox model. *Ann. Statist.* **44** 1708–1738.

RAMSAY, J. O. and DALZELL, C. J. (1991). Some tools for functional data analysis. *J. Roy. Statist. Soc. Ser. B* **53** 539–572. MR1125714

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

RAO, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics* **14** 1–17.

RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243. MR1094283

RICE, J. A. and WU, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57** 253–259. MR1833314

ROSENBLATT, M. (1991). Stochastic curve estimation. IMS Lecture Notes, IMS.

SAMWORTH, R. J. (2016). Peter Hall's work on high-dimensional data and classification. *Ann. Statist.* **44** 1888–1895.

ŞENTÜRK, D. and MÜLLER, H.-G. (2010). Functional varying coefficient models for longitudinal data. *J. Amer. Statist. Assoc.* **105** 1256–1264. MR2752619

STANISWALIS, J. G. and LEE, J. J. (1998). Nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **93** 1403–1418. MR1666636

TENENBAUM, J. B., DE SILVA, V. and LANGFORD, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** 2319–2323.

TUCKER, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika* **23** 19–23.

WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional data analysis. *Annual Review of Statistics* **3** 257–295.

YANG, W., MÜLLER, H.-G. and STADTMÜLLER, U. (2011). Functional singular component analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 303–324. MR2815778

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44** 2281–2321.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, DAVIS
ONE SHIELDS AVENUE
DAVIS, CALIFORNIA 95616
USA
E-MAIL: hgmueller@ucdavis.edu