Contributions to Statistics

Germán Aneiros
Enea G. Bongiorno
Ricardo Cao
Philippe Vieu *Editors*

Functional Statistics and Related Fields

Springer

Contributions to Statistics

More information about this series at http://www.springer.com/series/2912

Germán Aneiros · Enea G. Bongiorno
Ricardo Cao · Philippe Vieu
Editors

Functional Statistics and Related Fields

Springer

Editors
Germán Aneiros
Department of Mathematics
University of A Coruña
A Coruña
Spain

Ricardo Cao
Department of Mathematics
University of A Coruña
A Coruña
Spain

Enea G. Bongiorno
Department of Studies in Economics and Business
University of Eastern Piedmont "Amedeo Avogadro"
Novara
Italy

Philippe Vieu
Toulouse Mathematics Institute (IMT)
Paul Sabatier University
Toulouse
France

ISSN 1431-1968
Contributions to Statistics
ISBN 978-3-319-55845-5 ISBN 978-3-319-55846-2 (eBook)
DOI 10.1007/978-3-319-55846-2

Library of Congress Control Number: 2017936648

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To the memory of Peter Gavin Hall

Photograph of Peter Hall taken by his wife, Jeannie, at the University of California, Davis, May 2012 (by kind permission of Jeannie Hall)

Preface

Along the past ten years the IWFOS meetings went hand-in-hand with the developments in Functional Data Analysis. After the success of the first editions (Toulouse 2008, Santander 2011 and Stresa 2014), this fourth one has been organized in A Coruña in June 2017. This volume presents a wide scope of works presented during this event. By following the orientation started along the last edition in Stresa, this meeting has been opened to contributions in other related fields in modern statistics such as Big Data Analysis and High Dimensional Statistical Modeling.

Diversity has always been an important tradition in IWFOS meetings, and this is appearing at different levels along this volume. First of all, the scope of contributions covers most of the challenging questions arising in Infinite and/or High Dimensional Statistical problems. Secondly, diversity appears through the nature of the contributions going from the most theoretical developments to the most applied ones. Finally, diversity is also in the contributors themselves, who include both high level senior statisticians and promising young active researchers coming from different parts of the world.

We would like to thank very much all the authors that presented their work in IWFOS 2017 -not only those appearing in this book- and also those who attended the workshop. This list would be very long, so we just mention them collectively. Our special gratitude is for John Aston (Cambridge, UK), Alexander Aue (Davis, USA), Antonio Cuevas (Madrid, Spain), Aurore Delaigle (Melbourne, Australia), Jeff Goldsmith (Columbia, USA), Stephan Huckemann (Göttingen, Germany), Alicia Nieto-Reyes (Santander, Spain), Victor Panaretos (Lausanne, Switzerland), Paula Raña (A Coruña, Spain) and Laura Sangalli (Milano, Italy) for having kindly accepted to take part in our conference as Invited Speakers. Also, the efforts of all the members of the Scientific Committee, both for promoting the meeting and for their scientific expertise along the reviewing process of the abstracts, are particularly appreciated. We wish to address special thanks to Ana Aguilera (Granada, Spain), Graciela Boente (Buenos Aires, Argentina), Pedro Delicado (Barcelona, Spain), Aldo Goia (Novara, Italy), Wenceslao González-Manteiga (Santiago de Compostela, Spain), Sonja Greven (München, Germany), Siegfried Hörmann (Bruxelles, Belgium), Marie Hušková (Prague, Czech Republic), Steve Marron (North Carolina, USA), Juhyun

Park (Lancaster, UK), Dimitris Politis (San Diego, USA), Juan Romo (Madrid, Spain), Piercesare Secchi (Milano, Italy), Han Lin Shang (Camberra, Australia) and Sara van de Geer (Zürich, Switzerland). Our gratitude is also extended to the current Editors-in-Chief of the Journal of Multivariate Analysis and Computational Statistics, who kindly accepted our proposal for special issues on Functional Data Analysis and related topics.

We also would like to thank all the institutions that have supported IWFOS 2017: the University of A Coruña (UDC), the Foundation Pedro Barrié de la Maza, Springer International Publishing, the International Society for Non Parametric Statistics (ISNPS), the Spanish Society for Statistics and Operations Research (SEIO), the Galician Society for Promotion of Statistics and Operations Research (SGAPEIO), the Université Franco-Italienne / Università Italo-Francese, the Galician Network on Cloud and Big Data Technologies -funded by Xunta de Galicia-, the Research Center for Information and Communication Technologies (CITIC), the Institut de Mathématiques of Toulouse, and the Research Group on Modeling, Optimization and Statistical Inference (MODES).

Nothing would have been possible without the hard work of many people. Our last thanks go to Rubén Fernández-Casal, Mario Francisco-Fernández, Salvador Naya, Paula Raña, Javier Tarrío-Saavedra, José Antonio Vilar and Juan Manuel Vilar from A Coruña, to Ernesto Salinelli from Novara and to Sylvie Viguier-Pla from Toulouse for their active participation in the Organizing Committee, and to Veronika Rosteck and Alice Blanck from the Springer's team.

A Coruña
June 2017

Germán Aneiros
Enea G. Bongiorno
Ricardo Cao
Philippe Vieu

Contents

List of Contributors

Catherine Aaron
Université Clermont Auvergne, Campus Universitaire des Cézeaux, France
e-mail: catherine.aaron@math.univ-bpclermont.fr

Nihan Acar-Denizli
Mimar Sinan Güzel Sanatlar Üniversitesi, Istanbul, Turkey
e-mail: nihan.acar@msgsu.edu.tr

Ana M. Aguilera
Univeristy of Granada, Department of Statistics and O.R. and IEMath-GR, Spain
e-mail: aaguiler@ugr.es

Javier Álvarez-Liébana
University of Granada, Campus Fuentenueva, 18071 Granada, Spain
e-mail: javialvaliebana@ugr.es

Germán Aneiros
Departamento de Matemáticas, Universidade da Coruña, A Coruña, Spain
e-mail: ganeiros@udc.es

Eleonora Arnone
MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
e-mail: eleonora.arnone@polimi.it

John A. D. Aston
Statistical Laboratory, Department of Pure Maths and Mathematical Statistics, University of Cambridge, Cambridge, UK
e-mail: j.aston@statslab.cam.ac.uk

Jean-Baptiste Aubin
INSA-Lyon, ICJ, 20, Rue Albert Einstein, 69621 Villeurbanne Cedex, France
e-mail: jean-baptiste.aubin@insa-lyon.fr

Alexander Aue
Department of Statistics, University of California, One Shields Avenue, Davis, CA 95616, USA
e-mail: aaue@ucdavis.edu

Laura Azzimonti
IDSIA - Department of Innovative Technologies, Università della Svizzera Italiana Galleria 1, via Cantonale, Switzerland
e-mail: laura.azzimonti@supsi.ch

Gülay Başarır
Mimar Sinan Güzel Sanatlar Üniversitesi, Istanbul, Turkey
e-mail: gulay.basarir@msgsu.edu.tr

Heather Battey
Department of Mathematics, Imperial College London, UK
e-mail: h.battey@imperial.ac.uk

Enea G. Bongiorno
Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, Via Perrone, 18, 28100, Novara, Italy
e-mail: enea.bongiorno@uniupo.it

Alain Boudou
Equipe de Stat. et Proba., Institut de Mathématiques, UMR5219, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France
e-mail: boudou@math.univ-toulouse.fr

Isabel Caballero
ICMAN-CSIC, Cádiz, Spain
e-mail: isabel.caballero@icman.csic.es

Alejandra Cabaña
Universidad Autónoma de Barcelona, Barcelona, Spain
e-mail: acabana@mat.uab.cat

Rebeca Campos-Sánchez
Centro de Investigacin en Biologa Celular y Molecular, Universidad de Costa Rica, Costa Rica
e-mail: rcamposs@cariari.ucr.ac.cr

Ricardo Cao
Research group MODES, INIBIC, CITIC, Departamento de Matemáticas, Facultade de Informática, Universidade da Coruña, 15071 A Coruña, Spain
e-mail: rcao@udc.es

Francesca Chiaromonte
Center for Medical Genomics and Department of Statistics, Pennsylvania State University, USA, and Sant'Anna School of Advanced Studies, Pisa, Italy
e-mail: chiaro@stat.psu.edu

Alejandro Cholaquidis
Universidad de la República, Facultad de Ciencias, Uruguay
e-mail: acholaquidis@cmat.edu.uy

Marzia A. Cremona
Department of Statistics, Pennsylvania State University, USA
e-mail: mac78@psu.edu

Paul Cristini
Aix-Marseille Univ., CNRS, Centrale Marseille, L.M.A., France
e-mail: cristini@lma.cnrs-mrs.fr

Pedro Delicado
Universitat Politècnica de Catalunya, Barcelona, Spain
e-mail: pedro.delicado@upc.edu

Benjamin Eltzner
Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, University of Göttingen, Goldschmidstr. 7, 37077 Göttingen, Germany
e-mail: beltzne@uni-goettingen.de

Manuel Escabias
University of Granada, Department of Statistics and O.R. Spain
e-mail: escabias@ugr.es

Ana M. Estrada
Universidad de Los Andes, Bogotá, Colombia
e-mail: am.estrada213@uniandes.edu.co

Nathalie Favretto-Cristini
Aix-Marseille Univ., CNRS, Centrale Marseille, L.M.A., France
e-mail: favretto@lma.cnrs-mrs.fr

Manuel Febrero-Bande
Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Spain
e-mail: manuel.febrero@usc.es

Rubén Fernández-Casal
Department of Mathematics, Faculty of Computer Science, Universidade da Coruña, Campus de Elviñ, A Coruña-España
e-mail: ruben.fcasal@udc.es

Miguel Flores
Departamento de Matemáticas, Facultad de Ciencias, Escuela Politécnica Nacional, Ladrón de Guevara E11-253, Quito-Ecuador
e-mail: miguel.flores@epn.edu.ec

Ricardo Fraiman
Universidad de la República, Facultad de Ciencias, Uruguay
e-mail: rfraiman@cmat.edu.uy

Mario Francisco-Fernández
Research group MODES, INIBIC, CITIC, Departamento de Matemáticas, Facultade de Informática, Universidade da Coruña, 15071 A Coruña, Spain
e-mail: mariofr@udc.es

Pedro Galeano
Department of Statistics and UC3M-BS Institute of Financial Big Data, Universidad Carlos III de Madrid, Spain
e-mail: pedro.galeano@uc3m.es

Yuan Gao
College of Business and Economics, Australian National University
26C Kingsley St. ACTON 2601, Australia
e-mail: u5758483@anu.edu.au

Badih Gathas
Aix Marseille Université, CNRS, Centrale Marseille, I2M UMR 7373, 13453, Marseille, France
e-mail: badihghattas@gmail.com

Aldo Goia
Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, Via Perrone, 18, 28100, Novara, Italy
e-mail: aldo.goia@uniupo.it

Jeff Goldsmith
Columbia University Department of Biostatistics, New York, USA
e-mail: ajg2202@cumc.columbia.edu

Wenceslao González-Manteiga
Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Spain
e-mail: wenceslao.gonzalez@usc.es

Jose M. Heredia-Jiménez
Univeristy of Granada, Department of Physical Education and Sport, Spain
e-mail: herediaj@ugr.es

Stephan F. Huckemann
Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, University of Göttingen, Goldschmidstr. 7, 37077 Göttingen, Germany
e-mail: huckeman@math.uni-goettingen.de

Qing Jiang
School of Statistics, Beijing Normal University
e-mail: 201631011006@mail.bnu.edu.cn

Zoulikha Kaid
Laboratoire de Statistique et Processus Stochastiques, Université Djillali Liabès, Sidi Bel-Abbès, BP. 89, Sidi Bel-Abbès 22000, Algeria
e-mail: kaedzoulekha@yahoo.com

Lydia Kara-Zaïtri
Université Djillali Liabès, LSPS, Sidi Bel Abbès, Algeria
e-mail: kara_zaitri@hotmail.fr

Johannes Klepsch
Center for Mathematical Sciences, Technische Universität München, 85748 Garching, Boltzmannstraße 3, Germany
e-mail: j.klepsch@tum.de

Ali Laksaci
Laboratoire de Statistique et Processus Stochastiques, Université Djillali Liabès, Sidi Bel-Abbès, BP. 89, Sidi Bel-Abbès 22000, Algeria
e-mail: alilak@yahoo.fr

Dominik Liebl
University of Bonn, Institute for Financial Economics and Statistics, Bonn, Germany
e-mail: dliebl@uni-bonn.d

Eardi Lila
Cambridge Centre For Analysis, University of Cambridge, Wilberforce Road, Cambridge Cb3 0Wb, UK
e-mail: e.lila@maths.cam.ac.uk

Nengxiang Ling
School of Mathematics, Hefei University of Technology, China
e-mail: hfut.lnx@163.com

Yang Liu
School of Mathematics, Hefei University of Technology, China
e-mail: liuyang0806@foxmail.com

Kateryna D. Makova
Center for Medical Genomics and Department of Biology, Pennsylvania State University, USA
e-mail: kdm16@psu.edu

Marcos Matabuena
Universidade de Vigo, Spain
e-mail: marcos.matabuena.tic@gmail.com

Simos G. Meintanis
Department of Economics, National and Kapodistrian University of Athens, Athens, Greece and Unit for Business Mathematics and Informatics, North–West University, Potchefstroom, South Africa
e-mail: simosmei@econ.uoa.gr

Stanislav Nagy
Department of Mathematics, KU Leuven, Belgium and Department of Probability and Mathematical Statistics, Charles University, Czech Republic
e-mail: stanislav.nagy@kuleuven.be

Salvador Naya
Department of Mathematics, Higher Polytechnic University College, Universidade da Coruña, Mendizábal s/n, Ferrol-España
e-mail: salva@udc.es

Alicia Nieto-Reyes
Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria. Avd/ Los Castros s/n. 39005 Santander, Spain
e-mail: alicia.nieto@unican.es

Fabio Nobile
MATHICSE-CSQI, École Polytechnique Fédérale de Lausanne, Route Cantonale, 1015 Lausanne, Switzerland
e-mail: fabio.nobile@epfl.ch

Eva Orantes-González
Univeristy of Granada, Department of Physical Education and Sport, Spain
e-mail: maevor@ugr.es

Manuel Oviedo de la Fuente
Technological Institute for Industrial Mathematics and Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Spain
e-mail: manuel.oviedo@usc.es

Laurent Pantera
CEA, DEN, DER/SRES, Cadarache,F13108 Saint-Paul-Lez-Durance, France
e-mail: laurent.pantera@cea.fr

Jairo Peña
Universidad de Los Andes, Bogotá, Colombia
e-mail: ji.pena45@uniandes.edu.co

Davide Pigoli
Statistical Laboratory, Department of Pure Maths and Mathematical Statistics, University of Cambridge, Cambridge, UK
e-mail: d.pigoli@statslab.cam.ac.uk

Alessia Pini
MOX - Department of Mathematics, Politecnico di Milano, Italy
e-mail: alessia.pini@polimi.it

Adolfo J. Quiroz
Universidad de Los Andes, Bogotá, Colombia
e-mail: aj.quiroz1079@uniandes.edu.co

Mustapha Rachdi
Université Grenoble Alpes, AGIM Team, AGEIS EA 7407, Grenoble, France
e-mail: mustapha.rachdi@univ-grenoble-alpes.fr

Paula Raña
Departamento de Matemáticas, Universidade da Coruña, A Coruña, Spain
e-mail: paula.rana@udc.es

M. Dolores Ruiz-Medina
University of Granada, Campus Fuentenueva, 18071 Granada, Spain
e-mail: mruiz@ugr.es

Laura M. Sangalli
MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
e-mail: laura.sangalli@polimi.it

Matthieu Saumard
Conservatoire National des Arts et Métiers, 292 Rue Saint-Martin, 75003 Paris, France
e-mail: matthieu.saumard@lecnam.net

Hanlin Shang
College of Business and Economics, Australian National University, Australia
e-mail: hanlin.shang@anu.edu.au

Lorenzo Spreafico
ALPs - Alpine Laboratory of Phonetics and Phonology Free University of Bozen-Bolzano, Bolzano, Italy
e-mail: lorenzo.spreafico@unibz.it

Aymeric Stamm
MOX - Department of Mathematics, Politecnico di Milano, Italy
e-mail: aymeric.stamm@polimi.it

Javier Tarrío-Saavedra
Department of Mathematics, Higher Polytechnic University College, Universidade da Coruña, Mendizábal s/n, Ferrol-España
e-mail: javier.tarrio@udc.es

Shahin Tavakoli
Statistical Laboratory, Department of Pure Maths and Mathematical Statistics, University of Cambridge, Cambridge, UK, and PRDA, Darwin College, Cambridge, UK
e-mail: s.tavakoli@statslab.cam.ac.uk

Oumar I. Traore
Aix-Marseille Univ., CNRS, Centrale Marseille, L.M.A., France
e-mail: toumarissiaka@gmail.com

Simone Vantini
MOX - Department of Mathematics, Politecnico di Milano, Italy
e-mail: simone.vantini@polimi.it

Alessandro Vietti
ALPs - Alpine Laboratory of Phonetics and Phonology Free University of Bozen-Bolzano, Bolzano, Italy
e-mail: alessandro.vietti@unibz.it

Philippe Vieu
Institut de Mathématiques, Université Paul Sabatier, Toulouse, France
e-mail: philippe.vieu@math.univ-toulouse.fr

Sylvie Viguier-Pla
Equipe de Stat. et Proba., Institut de Mathématiques, UMR5219, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France
Université de Perpignan via Domitia, LAMPS, 52 av. Paul Alduy, 66860 Perpignan Cedex 9, France
e-mail: viguier@univ-perp.fr

Juan Vilar
Departamento de Matemáticas, Universidade da Coruña, A Coruña, Spain
e-mail: juan.vilar@udc.es

Fabian Walders
University of Bonn, BGSE & Institute for Financial Economics and Statistics, Bonn, Germany
e-mail: fwalders@uni-bonn.de

Julia Wrobel
Columbia University Department of Biostatistics, New York, USA
e-mail: jw3134@cumc.columbia.edu

Yang Yang
Research School of Finance, Actuarial Studies and Statistics, Australian National University, Acton ACT 2601, Australia
e-mail: yang.yang@anu.edu.au

Yanrong Yang
College of Business and Economics, Australian National University, Australia
e-mail: yanrong.yang@anu.edu.au

Jiří Zelinka
Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Kotlářská 2, Brno, Czech Republic
e-mail: zelinka@emath.muni.cz

Lixing Zhu
School of Statistics, Beijing Normal University and Department of Mathematics, Hong Kong Baptist University, Hong Kong
e-mail: lzhu@hkbu.edu.hk

Chapter 1
An introduction to the 4th edition of the International Workshop on Functional and Operatorial Statistics

Germán Aneiros, Enea G. Bongiorno, Ricardo Cao and Philippe Vieu

Abstract The aim of this introductory chapter is to present the various contributions to the fourth edition of the International Workshop on Functional and Operatorial Statistics (IWFOS 2017) held in June 2017 in A Coruña, Spain. These contributions are put into the context of the recent developments on Functional Data Analysis and related fields.

1.1 Functional Data Analysis along the last decade

The idea of organizing international meetings in Europe concerning the various features of Infinite Dimensional Statistics grew a decade ago with the first edition of the International Workshop on Functional and Operatorial Statistics (IWFOS), which took place in Toulouse (June 2008). At that stage, this field of Statistics was not so much developed but its popularization was starting, mainly thanks to the book [7].

Less than ten years later, and after two more editions of the IWFOS conference (Santander, Spain, 2011 and Stresa, Italy, 2014), the situation turns out to be completely different. The number of scientific publications in the field has exponentially increased and most of international statistical conferences have now several contri-

Germán Aneiros (✉)
Research group MODES, CITIC, ITMATI, Departamento de Matemáticas, Facultade de Informática, Universidade da Coruña, 15071 A Coruña, Spain, e-mail: ganeiros@udc.es

Enea G. Bongiorno
DiSEI, Università del Piemonte Orientale, Italy e-mail: enea.bongiorno@uniupo.it

Ricardo Cao
Research group MODES, CITIC, ITMATI, Departamento de Matemáticas, Facultade de Informática, Universidade da Coruña, 15071 A Coruña, Spain, e-mail: rcao@udc.es

Philippe Vieu
Institut de Mathématiques, Université Paul Sabatier, Toulouse, France, e-mail: philippe.vieu@math.univ-toulouse.fr

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_1

butions and/or specialized sessions on this topic. A wide scope of general books on Functional Data Analysis (FDA) is now available, illustrating how most of the standard statistical ideas (standard in the sense of multivariate) can be efficiently adapted to this new kind of data. The books [15, 21, 33, 10, 16] present various methodological statistical advances for FDA, while [8] is more devoted to applied issues and [22] is concerned with the mathematical background necessary to the study of infinite dimensional statistical problems.

Undoubtedly, by bringing together statisticians coming from all parts of the world, the IWFOS conferences have been important events in this evolution, each of them being a moment for sharing experiences on FDA and for developing cooperations between researchers. While the first two editions (Toulouse 2008 and Santander 2011) were exclusively devoted to Infinite Dimensional Statistics, the third edition (Stresa 2014) intended to promote collaborations with people working on High Dimensional Statistics because of the existence of evident common features between both fields. The Special Issue of the Journal of Multivariate Analysis (see [17]) is illustrating the success of this opening strategy. For this fourth edition, it has again been decided to open to other scientific communities, including again High Dimensional Statistics but also new related fields like Big Data.

Of course, contributions on FDA take a major part in this volume and they are presented soon after a brief tribute to Professor Peter Hall in Section 1.2. Precisely, Section 1.3 is devoted to new methodological advances on FDA, while Section 1.4 is discussing contributions on mathematical background and Section 1.5 is devoted to applied issues. Contributions concerning related fields such as High Dimensional Statistics and Big Data are gathered all together in Section 1.6.

1.2 On Peter Hall's impact on IWFOS conferences

At the moment of writing this presentation it is worth being noted that the success of all the IWFOS conferences would have never been possible without the support of many people who took major role as well as participants and members of the various Scientific or Organizing Committees. Naming of all them would be a long and tedious task, but among all of them we wish to stress on the high role played by Professor Peter Hall in the success of these meetings. From a scientific point of view, his numerous contributions in Statistics (see [5]) meet naturally FDA (see [6]) and his mark is fully crossing many contributions in this volume. Moreover, by being an invited speaker of the first edition in 2008 at a moment when there were no evidence for organizing such events and by participating as Scientific Committee member of the second edition in 2011, Peter has highly contributed to the success of all IWFOS editions. At the moment of opening this 2017 edition our thanks to Peter are mixed with our sadness for his death. We would like to dedicate this IWFOS 2017 to the memory of Peter.

1.3 Methodological contribution for Statistics with Functional Data

This volume contains methodological contributions on many features of Statistics for Functional Data, including:

- Clustering (see Chapter **33** for a proposal based on the envelopes of the curves).
- Data Registration (see Chapter **35** for curves alignment based FPCA).
- Depth (see Chapter **25** for an overview on consistency results and Chapter **26** for a discussion on the adaptation of the notion of depth to the functional context).
- Dimension reduction (see Chapter **22** for PCA of functional data on manifolds, Chapter **31** for FPCA and FPC regression in grouped multivariate functional time series forecasting and Chapter **36** for random functional variables and fourier series).
- Functional conditional mode estimation (see Chapter **23** for a nonparametric proposal).
- Functional time series (see Chapters **4** and **30** for autoregressive models and Chapter **8** for invertible time series).
- Multifunctional variables (see Chapter **32** for tests on covariance of multifunctional variables and Chapter **19** for two–sample tests for multivariate functional data).
- Regression (see Chapter **14** for parametric regression, Chapters **21** and **20** for nonparametric regression and Chapters **29** and **34** for semiparametric regression).
- Robustness (see Chapter **2** for combination of estimators from different subsamples by robust fusion procedures).
- Small–ball probability (see Chapter **7** for small–ball probability factorization).
- Spatial data (see Chapter **6** for space–time regression with PDE penalization).
- Testing (see Chapter **11** for permutation tests, Chapter **13** for ANOVA methodology, Chapter **16** for outlier detection, Chapter **27** for differential interval–wise test and Chapter **28** for Hotelling statistic in Hilbert spaces).
- Variable selection (see Chapter **15** for variable selection in functional nonparametric additive regression models).

1.4 Contribution on Mathematical background for Infinite Dimensional Statistics

A few contributions are devoted to mathematical background for developing theory on Statistics with infinite dimensional variables. This includes:

- Analytical developments (see Chapter **18** for PCA in non–euclidean spaces).
- Operatorial developments (see Chapter **32** for testing separability of covariance operator and Chapter **10** for commutator of projectors and of unitary operators).

1.5 Applied Functional Data Analysis contributions

This volume also contains a few contributions fully devoted to applied case studies (see Chapter **3** for an analysis of the turbidity in a coastal zone from satellite data, Chapter **9** for detecting complexity of financial time series, Chapter **12** for a study on the effects of genomic landscape features on the integration and fixation of endogenous retroviruses and Chapter **24** for predicting the physiological limits of sport stress tests), but it is worth being noted that most of other contributions also propose interesting real data applications, including applications to:

- Economics (see Chapters **34**).
- Energy (see Chapter **15**).
- Environmetrics (see Chapter **11**).
- Medicine (see Chapters **6** and **22**).
- Mortality curves (see Chapters **17** and **31**).
- Phonetic curves (see Chapters **32**, **27** and **33**).
- Physical activity (see Chapters **13** and **35**).
- Traffic (see Chapter **8**).

1.6 Contributions on related fields

This includes contributions on:

- Big Data (see Chapter **2** for the combination of estimators obtained from different subsamples).
- High- and Infinite-Dimensional Statistics (see Chapter **5** for impact points selection in sparse models)
- High Dimensional Statistics (see Chapter **17** for the case of functional time series forecasting).

1.7 Concluding comments

Before closing this introductory presentation, it is worth stressing that the Journal of Multivariate Analysis has still been an important support for us. Once again, we have the chance to combine this fourth edition of IWFOS with the launching of a new special issue of this journal. Each participant whose contribution is exhibiting a high degree of methodological novelty as well as any other researcher (not necessarily taking part in IWFOS 2017) having new advances in FDA or related topics, are strongly encouraged to submit their work to this special issue. The novelty for this fourth edition is that it will also serve as starting point for an other special issue in the journal Computational Statistics for which we encourage any researcher (participant or not in IWFOS 2017) in submitting works with high computational and/or applied

novelty. These two special issues are a sample of the continuous evolving of IWFOS editions.

References

[1] Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis. Springer-Verlag, New York (2006)
[2] Goia, A., Vieu, P.: An introduction to recent advances in high/infinite dimensional statistics. J. Multivariate Anal. **46**, 1–6 (2016)
[3] Horváth, L., Kokoszka, P.: Inference for Functional Data with Applications. Springer, New York (2012)
[4] Hsing, T., Eubank, R.: Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. John Wiley & Sons, Chichester (2015)
[5] Li, R.: Editorial "Peter Hall". Ann. Statist. **44**, 5, 1817-1820 (2016)
[6] Müller, H.G.: Peter Hall, functional data analysis and random objects. Ann. Statist. **44**, 5, 1867-1887 (2016)
[7] Ramsay, J., Silverman, B.: Functional Data Analysis. Springer, New York (1997)
[8] Ramsay, J., Silverman, B.: Applied functional Data Analysis. Springer, New York (1997)
[9] Ramsay, J., Silverman, B.: Functional Data Analysis. (Second edition) Springer, New York (2005)
[10] Shi, J.Q., Choi, T.: Gaussian process regression analysis for functional data. CRC Press, Boca Raton, FL (2011)
[11] Zhang, J.T.: Analysis of variance for functional data. Monographs on Statistics and Applied Probability, 127. CRC Press, Boca Raton, FL (2014)

Chapter 2
Robust fusion methods for Big Data

Catherine Aaron, Alejandro Cholaquidis, Ricardo Fraiman and Badih Ghattas

Abstract We address one of the important problems in Big Data, namely how to combine estimators from different subsamples by robust fusion procedures, when we are unable to deal with the whole sample.

2.1 Introduction

Big Data covers a large list of different problems, see for instance [10, 11], and references therein. We address one of them, namely how to combine, using robust techniques, estimators obtained from different subsamples in the case where we are unable to deal with the whole sample. In what follows we will refer to this as Robust Fusion Methods (RFM).

To fix ideas, we start by describing one of the simplest problems in this area as a toy example. Suppose we are interested in the median of a huge set of iid random variables $\{X_1, \ldots, X_n\}$ with common density f, and we split the sample into m subsamples of length l, $n = ml$. We calculate the median of each subsample and obtain m random variables $Y_1, \ldots, Y_m$. Then we take the median of the set $Y_1, \ldots, Y_m$, i.e. we consider the well known median of medians. It is clear that it does not coincide

Catherine Aaron
Université Clermont Auvergne, Campus Universitaire des Cézeaux, France, e-mail: catherine.aaron@math.univ-bpclermont.fr

Alejandro Cholaquidis
Universidad de la República, Facultad de Ciencias, Uruguay e-mail: acholaquidis@cmat.edu.uy

Ricardo Fraiman (✉)
Universidad de la República, Facultad de Ciencias, Uruguay e-mail: rfraiman@cmat.edu.uy

Badih Gathas
Aix Marseille Université, CNRS, Centrale Marseille, I2M UMR 7373, 13453, Marseille, France e-mail: badihghattas@gmail.com

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_2

with the median of the whole original sample $\{X_1,\dots,X_n\}$, but it will be close. What else could we say about this estimator regarding efficiency and robustness?

The estimator is nothing but the median of m iid random variables but now with a different distribution given by the distribution of the median of l random variables with density f_X. Suppose for simplicity that $l = 2k+1$. Then, the density of the random variables Y_i is given by

$$g_Y(y) = \frac{(2k+1)!}{(k!)^2} F_X(t)^k (1-F_X(t))^k f_X(t). \tag{2.1}$$

In particular, if f_X is uniform on $(0,1)$, (2.1) is given by

$$h_Y(y) = \frac{(2k+1)!}{(k!)^2} t^k (1-t))^k \mathbf{1}_{[0,1]}(t), \tag{2.2}$$

a $Beta(k+1,k+1)$ distribution.

On the other hand, we have that asymptotically, for a sample of size n the empirical median $\theta = med(X_1,\dots,X_n)$ behaves as a normal distribution centred at the true median θ with variance $\frac{1}{4nf_X(\theta)^2}$, while the median of medians behaves asymptotically as a normal distribution centred at θ, the median of the median distribution, and with variance $\frac{1}{4mg_Y(\theta)^2}$. For the uniform case, both are centred at 1/2, $f_X(0.5) = 1$ and $g_Y(0.5) = (1/2)^{2k}(2k+1)!/(k!)^2$, so we can explicitly calculate the relative efficiency.

In Section 2.2 we generalize this idea and study the breakdown point, efficiency, and computational time of the robust fusion method. In Section 2.3 we tackle, as a particular case, the robust estimation of the covariance operator and show, in Section 2.3.3, the performance throughout a simulation study.

2.2 A general setup for RFM.

In this short note we present briefly a general framework for RFM methods for several multivariate and functional data problems. We illustrate our procedure considering only the problem of robust covariance operator estimation, based on a new simple robust estimator. Our approach is to consider RFM methods based on data depth functions. The idea is quite simple: given a statistical problem, (such as multivariate location, covariance operators, linear regression, principal components, among many others), we first split the sample into subsamples. For each subsample we calculate a robust estimator for the statistical problem considered. We will use them all to obtain an RFM estimator that is more accurate. More precisely, the RFM estimator is defined as the deepest point (with respect to the appropriate norm associated to the problem) among all the estimators obtained from the subsamples. Since we need to be able to calculate depths for large sample sizes and eventually high dimensional and infinite dimensional data, we will consider the spatial median corresponding to

maximizing the spatial depth function

$$D(x,P) = 1 - \left\| \mathbb{E}_P \left(\frac{X-x}{\|X-x\|} \right) \right\|, \tag{2.3}$$

where P is a probability in some Banach space $(E, \|\cdot\|)$ and $x \in E$, introduced by [2], formulated (in a different way) by [9], and extended to a very general setup by [1]. We want to address the consistency, efficiency, robustness and computational time properties of the RFM proposals.

To be more precise, the general algorithm is as follows. a) We observe $X_1, \ldots X_n$ iid random elements in a metric space E (for instance $E = \mathbb{R}^d$), b) we split the sample into subsamples $\{X_1, \ldots X_l\}, \{X_{l+1}, \ldots X_{2l}\}, \ldots, \{X_{(m-1)l+1}, \ldots X_{lm}\}$ with $n = ml$, c) we solve our statistical problem on each subsample with a robust procedure (for example, estimate a parameter θ on each subsample, obtaining $\hat{\theta}_1, \ldots, \hat{\theta}_m$), d) we take the fusion of the results at each subsample, (for instance $\hat{\hat{\theta}}$ can be the deepest point among $\hat{\theta}_1, \ldots, \hat{\theta}_m$.

2.2.1 Breakdown point

Breakdown point for the RFM. Following [4] we consider the finite-sample breakdown point.

Definition 2.1. Let $\hat{\theta}_n = \hat{\theta}_n(x)$ be an estimate of θ defined for samples $\mathbf{x} = \{x_1, \ldots, x_n\}$. Let us assume that θ takes values in $\Theta \subset \mathbb{R}^d$ (it can be $\Theta = \mathbb{R}^d$). Let $\mathscr{X}_p$ be the set of all data sets $\mathbf{y}$ of size n having $n-p$ elements in common with x:

$$\mathscr{X}_p = \{\mathbf{y} : card(\mathbf{y}) = n,\ card(\mathbf{x} \cap \mathbf{y}) = n - p\},$$

then $\varepsilon_n^*(\hat{\theta}_n, \mathbf{x}) = \frac{p^*}{n}$, where $p^* = \max\{p \geq 0 : \hat{\theta}_n(\mathbf{y})$ is bounded and also bounded away from $\partial\Theta\ \forall \mathbf{y} \in \mathscr{X}_p\}$.

Let us consider, for $n = ml$, the random walk S_n with $S_0 = 0$, and $S_j = B_1 + \ldots + B_j$ for $j = 1, \ldots, n$, B_i being iid Bernoulli(p) for $i = 1, \ldots, n$, where a one represents the presence of an outlier, while a zero represents no presence of an outlier. Then to compute the breakdown point for the median of medians, we need to count how many times the sequence $\{S_l, S_{2l} - S_l, \ldots S_n - S_{n-l}\}$ is larger than k (recall that $l = 2k+1$). Let us define, $U_{m,n} := card\{1 \leq j \leq m : S_{jl} - S_{(j-1)l} \leq k\}/m$, since the median has breakdown point 0.5 the fusion will break down if $U_{m,n}$ is greater than 0.5.

This will also be true if we take the median of any robust estimate with breakdown point equal to 0.5 calculated at each subsample.

To have a glance at the breakdown point, we performed 5000 replicates of the vector S_{30000} and calculated the percentage of times the estimator breaks down for $p = 0.45, 0.49, 0.495$ and 0.499. The results are in the following table.

Table 2.1: Percentage of estimator breakdowns for 5000 replications and different values of m for $n = 30000$; p is the proportion of outliers

m	$p = 0.45$	$p = 0.49$	$p = 0.495$	$p = 0.499$
5	0	0.0020	0.0820	0.3892
10	0	0.0088	0.1564	0.5352
30	0	0.0052	0.1426	0.5186
50	0	0.0080	0.1598	0.5412
100	0	0.0192	0.2162	0.6084
150	0	0.0278	0.2728	0.6780

Since the number Y of outliers in the subsamples of length l follows a Binomial distribution, Binom(l, p), as a direct application of Theorem 1 in [8] we can bound the probability, $q = \mathbb{P}(Y > l/2)$, of breakdown.

2.2.2 Efficiency of Fusion of M-estimators

In this section we will obtain the asymptotic variance of the RFM method, for the special case of M-estimators. Recall that M-estimators can be defined (see Section 3.2 in [6]) by the implicit functional equation $\int \psi(x, T(F))F(dx) = 0$, where $\psi(x;\theta) = (\partial/\partial\theta)\rho(x;\theta)$, for some function ρ. The estimator T_n is given by the empirical version of T, based on a sample $\mathscr{X}_n = \{X_1, \ldots, X_n\}$. It is well known that $\sqrt{n}(T_n - T(F))$ is asymptotically normal with mean 0, variance σ^2, and can be calculated in general as the integral of the square of the influence function. The asymptotic efficiency of T_n is defined as $\text{Eff}(T_n) = \frac{\sigma^2_{ML}}{\sigma^2}$, where σ^2_{ML} is the asymptotic variance of the maximum likelihood estimator. Then, the asymptotic variance of a M-estimator built from a sample $T_n^1, \ldots, T_n^p$ of p M-estimators of T can be calculated easily.

2.2.3 Computational time

We want to calculate the computational time of our robust fusion method for a sample $\mathscr{X}_n = \{X_1, \ldots, X_n\}$ iid of X, where we have split $\mathscr{X}_n$ into m subsamples of length l, then apply a robust estimator to every subsample of length l, and fuse them by taking the deepest point among the m subsamples. If we denote by compRE(l) the computational time required to calculate the robust estimator based on every subsample of length l, and compDeph(m) the computational time required to compute the deepest robust estimator based on the m estimators, then the computational time of our robust fusion method is $m\times$compRE(l)+compDeph(m).

2.3 Robust Fusion for covariance operator

The estimation of the covariance operator of a stochastic process is a very important topic that helps to understand the fluctuations of the random element, as well as to derive the principal functional components from its spectrum. Several robust and non-robust estimators have been proposed, see for instance [1], and the references therein. In order to perform RFM, we introduce a computationally simple robust estimator to apply to each of the m subsamples, that can be performed using parallel computing. It is based on the notion of impartial trimming applied on the Hilbert–Schmidt space, where covariance operators are defined. The RFM estimator is the deepest point among the m estimators corresponding to each subsample, where the norm in (2.3) is given by (2.4) below.

2.3.1 A resistant estimate of the covariance operator

Let $E = L^2(T)$, where T is a finite interval in $\mathbb{R}$, and $X, X_1, \ldots X_n, \ldots$ iid random elements taking values in $(E, \mathscr{B}(E))$, where $\mathscr{B}(E)$ stands for the Borel σ-algebra on E. Assume that $\mathbb{E}(X(t)^2) < \infty$ for all $t \in T$, and $\int_T \int_T \rho^2(s,t)dsdt < \infty$, so the covariance function is well defined and given by $\rho(s,t) = \mathbb{E}((X(t) - \mu(t))(X(s) - \mu(s)))$, where $\mathbb{E}(X(t)) = \mu(t)$.

For notational simplicity we assume that $\mu(t) = 0, \forall t \in T$. Under these conditions, the covariance operator, given by $S_0(f) = \mathbb{E}(\langle X, f\rangle X)$, is diagonalizable, with nonnegative eigenvalues λ_i such that $\sum_i \lambda_i^2 < \infty$. Moreover S_0 belongs to the Hilbert–Schmidt space $HS(E,E)$ of linear operators with square norm and inner product given by

$$\|S\|_{HS}^2 = \sum_{k=1}^{\infty} \|S(e_k)\|^2 < \infty, \quad \langle S_1, S_2\rangle_{\mathscr{F}} = \sum_{k=1}^{\infty} \langle S_1(e_k), S_2(e_k)\rangle, \tag{2.4}$$

respectively, where $\{e_k : k \geq 1\}$ is any orthonormal basis of E, and $S, S_1, S_2 \in HS(E,E)$. In particular, $\|S_0\|^2 = \sum_{i=1}^{\infty} \lambda_i^2$, where λ_i are the eigenvalues of S_0. Given an iid sample $X_1, \ldots, X_n$, define the Hilbert–Schmidt operators of rank one,

$$W_i : E \to E, \quad W_i(f) = \langle X_i, f\rangle X_i(.), \quad i = 1, \ldots n.$$

Let $\phi_i = X_i/\|X_i\|$. Then, $W_i(\phi_i) = \|X_i\|^2 \phi_i =: \eta_i \phi_i$.

The standard estimator of S_0 is just the average of these operators, i.e. $\hat{S}_n = \frac{1}{n}\sum_{i=1}^n W_i$, which is a consistent estimator of S_0 by the Law of Large Numbers in the space $HS(E,E)$. Our proposal is to consider an impartial trimmed estimate as a resistant estimator. The notion of impartial trimming was introduced in [5], while the functional data setting was considered in [3], from were it can be obtained the asymptotic theory for our setting. In order to perform the algorithm we will derive an exact formula for the matrix distances $\|W_i - W_j\|$, $1 \leq i \leq j \leq n$.

Lemma 1 *We have that*

$$d_{ij}^2 := \|W_i - W_j\|_{HS}^2 = \|X_i\|^4 + \|X_j\|^4 - 2\langle X_i, X_j\rangle^2 \quad \textit{for } 1 \le i \le j \le n. \tag{2.5}$$

2.3.2 The impartial trimmed mean estimator

Following [3], we define the impartial trimmed covariance operator estimator, which is calculated by the following algorithm.

Given the sample $X_1(t),\dots,X_n(t)$ (which we have assumed with mean zero for notational simplicity) and $0 < \alpha < 1$, we provide a simple algorithm to calculate an approximate impartial trimmed mean estimator of the covariance operator of the process $S_0 : E \to E$, $S_0(f)(t) = \mathbb{E}(\langle X, f\rangle X(t))$, that will be strongly consistent.

STEP 1: Calculate $d_{ij} = \|W_i - W_j\|_{HS}$, $1 \le i \le j \le n$, using Lemma 1.

STEP 2: Let $r = \lfloor (1-\alpha)n \rfloor + 1$. For each $i = 1,\dots n$, consider the set of indices $I_i \subset \{1,\dots,n\}$ corresponding to the r nearest neighbours of W_i among $\{W_1,\dots W_n\}$, and the order statistic of the vector $(d_{i1},\dots,d_{in})$, $d_i^{(1)} < \dots < d_i^{(n)}$.

STEP 3: Let $\gamma = \text{argmin}\{d_1^{(r)},\dots,d_n^{(r)}\}$.

STEP 4: The impartial trimmed mean estimator of S_0 is given by $\hat{S} =$ the average of the m nearest neighbours of W_γ among $\{W_1,\dots,W_n\}$, i.e the average of the observations in I_γ.

This estimator corresponds to estimating $\rho(s,t)$ by $\hat{\rho}(s,t) = \frac{1}{r}\sum_{j\in I_\gamma} X_j(s)X_j(t)$. Observe that Steps 1 and 2 of the algorithm can be performed using parallel computing.

2.3.3 Simulation results for the covariance operator

Simulations were done using a PC Intel core i7-3770 CPU, 8GO of RAM using 64 bit version of Win10, and R software ver. 3.3.0.

We vary the sample size n within the set $\{0.1e6, 1e6, 5e6, 10e6\}$ and the number of subsamples $m \in \{100, 500, 1000, 10000\}$. The proportion of outliers was fixed to $p = 13\%$ and $p = 15\%$. We replicate each simulation case $K = 5$ times and report a mean average of the results over these replicates.

We report the average time in seconds necessary for both a global estimate (*time0*, over the whole sample), and *time1* the estimate obtained by fusion (including computing the estimates over subsamples and aggregating them by fusion).
We compare the classical estimator (*Cov*), the mean of the classical estimators obtained from the subsamples (*AvCov*), the Fusion estimate of the classical estimator

(*Cov.RFM*), the global robust estimate (*CovRob*), the robust fusion estimate *RFM*, and the average of the robust estimates from the subsamples *AvRob*.

To generate the data, we have used a simplified version of the simulation model used in [7]:

$$X(t) = \mu(t) + \sqrt{2}\sum_{k=1}^{10} \lambda_k a_k \sin(2\pi kt) + \sqrt{2}\sum_{k=1}^{10} \nu_k b_k \cos(2\pi kt)$$

where $\nu_k = \left(\frac{1}{3}\right)^k, \lambda_k = k^{-3}$, and a_k and b_k are random standard Gaussian independent observations. The central observations were generated using $\mu(t) = 0$ whereas for the outliers we took $\mu(t) = 2 - 8\sin(\pi t)$. For t we used an equally spaced grid of $T = 20$ points in $[0,1]$.

The covariance operator of this process was computed for the comparisons:

$Cov(s,t) = \sum_{k=1}^{10} A_k(s)A_k(t) + B_k(s)B_k(t)$, where $A_k(t) = \sqrt{2}\lambda_k \sin(2\pi kt)$ and $A_k(t) = \sqrt{2}\nu_k \cos(2\pi kt)$.

The results are shown in the following two tables for two proportions of outliers, $p = 0.15$ and $p = 0.2$.

Table 2.2: Covariance operator estimator in presence of outliers. Using the classical and robust estimators over the entire sample, and aggregating by average or fusion of m subsamples estimates. $p = 0.15$, T=20

n	m	time0	time1	Cov	AvCov	Cov.RFM	CovRob	AvRob	RFM
0.05	20	553	18.20	24.3	24.3	24.7	5.16	5.21	5.52
0.05	50	543	7.81	24.3	24.3	24.9	5.20	5.24	5.60
0.05	100	528	4.79	24.3	24.3	25.2	5.20	5.17	5.58
0.05	1000	459	19.40	24.3	24.3	27.0	5.13	5.54	6.58
0.10	20	2300	69.00	24.2	24.2	24.4	5.14	5.22	5.43
0.10	50	2300	28.10	24.2	24.2	24.6	5.04	5.09	5.13
0.10	100	2290	15.20	24.2	24.2	25.0	5.06	5.15	5.43
0.10	1000	1850	21.60	24.3	24.2	26.1	5.21	5.35	6.13

Table 2.3: Covariance operator estimator in the presence of outliers. Using classical and robust estimators over the entire sample, and aggregating by average or fusion of m subsamples estimates. $p = 0.2$, T=20

n	m	time0	time1	Cov	AvCov	Cov.RFM	CovRob	AvRob	RFM
0.05	20	572	17.90	30.5	30.5	30.9	0.879	3.96	1.45
0.05	50	649	7.88	30.5	30.5	31.3	0.876	7.34	2.10
0.05	100	633	4.61	30.5	30.5	31.6	0.839	8.86	2.43
0.05	1000	478	19.50	30.5	30.5	32.3	0.864	13.10	7.08
0.10	20	1970	69.10	30.4	30.4	30.6	0.914	3.83	1.36
0.10	50	2030	28.10	30.4	30.4	31.1	0.921	4.32	1.55
0.10	100	2020	15.10	30.4	30.4	31.3	0.840	8.44	2.35
0.10	1000	1840	21.60	30.4	30.4	32.9	0.961	12.10	5.20

If the proportion of outliers is moderate $p = 15\%$, the average of the robust estimators still behaves well, better than RFM, but if we increase the proportion of outliers to $p = 0.2$, RFM clearly outperforms all the other estimators.

Acknowledgements We thank an anonymous referee for helpful suggestions on a first version.

References

[1] Chakraborty, A., Chaudhuri, P.: The spatial distribution in infinite dimensional spaces and related quantiles and depths. Ann. Statist. **42**(3) 1203–1231 (2014)
[2] Chaudhuri, P.: On a geometric notion of quantiles for multivariate data. J. Amer. Statist. Assoc. **91**(343) 862–872 (1996)
[3] Cuestas-Albertos, J. A., Fraiman, R.: Impartial means for functional data. Data Depth: Robust Multivariate Statistical Analysis, Computational Geometry & Applications. Eds. R. Liu, R. Serfling and D. Souvaine. Vol. 72 in the DIMACS Series of the American Mathematical Society, pp. 121–145 (2006)
[4] Donoho, D.L.: Breakdown properties of multivariate location estimators. Ph. D. qualifying papers, Dep. Statistics, Harvard University (1982)
[5] Gordaliza, A.: Best approximations to random variables based on trimming procedures. J. Approx. Theory **64**(2) 162180 (1991)
[6] Huber, P. J., Ronchetti, E. M.: Robust Statistics. Wiley, Hoboken, NJ (2009)
[7] Kraus, D., Panaretos, V.M.: Dispersion operators and resistant second-order functional data analysis. Biometrika **101**(1), 141–154 (2012)
[8] Short, M.: Improved inequalities for the Poisson and Binomial distribution and upper tail quantile functions. ISRN Probability and Statistics (2013)
[9] Vardi, Y., Zhang, C.: The multivariate L1-median and associated data depth. Proc. Nat. Acad. Sci. USA **97**(4) (2000)
[10] Wang, C., Chen, M.-H., Schifano, E., Wu, J., Yan, J.: Statistical methods and computing for big data. Stat. Interface **9**(4), 399414 (2016)
[11] Yu, B.: Let Us Own Data Science. IMS Bulletin Online **43**(7) (2014)

Chapter 3
Functional linear regression models for scalar responses on remote sensing data: an application to Oceanography

Nihan Acar-Denizli, Pedro Delicado, Gülay Başarır and Isabel Caballero

Abstract Remote Sensing (RS) data obtained from satellites are a type of spectral data which consist of reflectance values recorded at different wavelengths. This type of data can be considered as a functional data due to the continous structure of the spectrum. The aim of this study is to propose Functional Linear Regression Models (FLRMs) to analyze the turbidity in the coastal zone of Guadalquivir estuary from satellite data. With this aim different types of FLRMs for scalar response have been used to predict the amount of Total Suspended Solids (TSS) on RS data and their results have been compared.

3.1 Introduction

Functional Data Analysis (FDA) concerns with the data sets measured on a continuum such as a dense time interval, space or a spectrum. The data gathered from Remote Sensing (RS) sensors via transmission of electromagnetic energy is also a kind of spectral data. RS data are collected from the earth's surface in terms of reflectance values recorded at different number of wavelengths. They inform us in a fast and economical way about the environment. Therefore, they are used in many fields such as land-use mapping, agriculture, forestry and oceanography to make predictions [2, 3, 5, 11]. In oceanography, RS data are used to estimate ocean characteristic

Nihan Acar-Denizli (✉)
Mimar Sinan Güzel Sanatlar Üniversitesi, Istanbul, Turkey, e-mail: nihan.acar@msgsu.edu.tr

Pedro Delicado
Universitat Politècnica de Catalunya, Barcelona, Spain, e-mail: pedro.delicado@upc.edu

Gülay Başarır
Mimar Sinan Güzel Sanatlar Üniversitesi, Istanbul, Turkey, e-mail: gulay.basarir@msgsu.edu.tr

Isabel Caballero
ICMAN-CSIC, Cadiz, Spain e-mail: isabel.caballero@icman.csic.es

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_3

parameters such as Sea Surface Temperature (SST), Chlorophyll-a content (Chl-a) and Total Suspended Solids (TSS) [2, 3]. Recently, FDA gain importance in analyzing remote sensing sensor data sets [1, 4]. Although, there are many applications of multivariate analysis techniques on RS data in oceanography [2, 3, 10, 11], there are few studies that use FDA approach in this field [9].

The importance of this study is to propose FLRMs alternative to classical statistical methods to predict TSS parameter from RS curves at different time periods. In previous studies, mostly regression models with a combination of different band values or the band values which are most correlated to TSS measurements have been used to predict TSS parameter [2, 3]. FLRMs allows us to use the information recorded at all the bands rather than selecting single band or taking combination of the bands. In this study, the Remote sensing reflectance (Rrs) values, recorded in a spectrum that consists of eight different bands have been considered as functional predictors and TSS content, that are measured from collected in-situ samples have been taken, as scalar response vector. In order to determine the best prediction model several FLRMs for scalar responses have been constructed and their performances have been compared with the performance of classical statistical methods used in the literature. A 10 year data set has been constituted by matching the in-situ measurements with the satellite data recorded between the years 2002-2011. The work exhibits an approach of how to conduct analysis for processing and interpreting large-scale volume of heterogeneous data to improve the present knowledge as an essential piece of the future Big Earth Observation Data monitoring systems.

3.2 Methods

The general form of a functional scalar response model can be expressed by

$$Y = \int_T \chi(t)\beta(t)dt + \varepsilon, \tag{3.1}$$

where Y indicates the scalar response vector, ε is the error term, $\chi(t)$ and $\beta(t)$ define respectively the functional predictor and the parameter function that are defined on a continuous interval T.

To solve this problem, different techniques based on basis functions, eigenfunctions or nonparametric smoothing have been proposed to assess an interpretable estimate of the parameter function [6, 5, 22]. In this study we will focus on two different approaches. The first approach is to use B-Spline basis expansions to define the functional predictor and the model parameter function. The latter approach is based on dimension reduction method functional principal components analysis so that it is named as Functional Principal Components Regression (FPCR). The idea of FPCR is to predict scalar response vector $\mathbf{Y}$ on the functional predictors that are expanded in terms of the eigenfunctions of the empirical covariance operator which form an orthonormal basis in $L^2(T)$ [12, 6].

The main problem here is to determine the number of basis functions or components that will be used to expand data. In this study, Cross Validation (CV) criterion is preferred to choose the optimal number of basis functions. The estimates of the models are found by minimizing the Sum of Square Errors (SSE) as in the classical linear regression problem.

To compare the predictive performance of the mentioned models, an Adjusted version of Mean Error of Prediction (AMEP) based on Leave-One-Out Cross Validation (LOOCV) has been defined by the equation (3.2).

$$\text{AMEP} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_{-i})^2}. \tag{3.2}$$

The term $\bar{y}_{-i}$ in this equation indicates the mean after removing the ith observation from the data set.

3.3 Data

The data set consists of two parts. The in-situ data set is composed of TSS concentrations measured from the collected samples and the satellite data set is composed of the Remote sensing reflectance (Rrs) values recorded by MEdium Resolution Imaging Spectrometer (MERIS), one of the main instruments on board the European Space Agency (ESA)'s Envisat platform between the years 2002-2011. The data set has been constituted by matching the filtered satellite data with the in-situ considering the coordinate and the time that the sample has been collected. The coordinates have been matched considering the exact pixels that the sample is collected and the time difference between in-situ and satellite data has been constrained up to 1.5 hours.

3.3.1 In-Situ Data

The in-situ data consist of the records of TSS concentration which are obtained from the samples collected by the station of Junta de Andalucía and by the cruises of Reserva and Fluctuaciones in the Guadalquivir estuary. The surface samples taken into analysis were collected with a rosette sampler (5 m below water surface) with a distance from coast from 1km to 25 km offshore.

The samples were collected during different time periods. The sampling carried out by Junta de Andalucía covers the period between April 2008 and May 2011 where the samples of Reserva and Fluctuaciones were collected within the periods July 2002 - September 2004 and May 2005 - May 2007 respectively. Each sample is collected by one of the campaigns from a determined coordinate. The coordinate of the station Junta de Andalucía was fixed with the lattitude 36.78° N and longitude 6.37° W where the coordinates of the stations Reserva and Fluctuaciones were

chosen according to the campaign planning. The amount of TSS concentration in each sample has been measured according to the protocols mentioned in [3].

3.3.2 Satellite Data

The study area corresponds to the coastal region of the Gulf of Cadiz in the southwest coast of the Iberian Peninsula ($35.5^\circ - 37.5^\circ$ N latitude and $1^\circ - 10^\circ$ W longitude). The satellite data included within the Region Of Interest (ROI) was downloaded from the Ocean Colour Website (http://oceancolor.gsfc.nasa.gov) in hdf format. SeaDAS image analysis software (SeaWifs Data Analysis System, version 6, http://seadas.gsfc.nasa.gov/) and the interface VMware Workstation 12 Player (https://www.vmware.com/) were used to convert data from hdf format to ascii format. The RS data set consists of Level-2 Remote Sensing Reflectance (Rrs) (sr^{-1}) recorded at eight different wavelengths (413 nm, 443 nm, 490 nm, 510 nm, 560 nm, 620 nm, 665 nm, 681 nm) with 300 m full spatial resolution between the years 2002-2011. The data has been passed through a quality control process corresponding to the L2 flags given in [3] to remove the suspicious and low-quality data points. This filtering process is done by using MATLAB 7.12.0-R2011a software. Considering that the resolution of images is 300 m, the data set consists of 740×3330 pixel images which is equivalent to have 2464200 element vectors for each wavelength.

Statistical validation of satellite-derived products is an essential issue to verify the accuracy provided by the sensor. In this work, data match ups were made by matching the coordinates of Rrs values with the coordinates of the field measurements. Careful consideration of scales is critical when comparing remotely-sensed data with in situ observations, particularly because of the large spatio-temporal heterogeneity of estuarine and coastal water properties influencing those measurements [8]. In this sense, time difference between satellite overpasses and in situ sampling was reduced by a filter of < 1.5 hours from acquisition, thus preventing temporal biases to further evaluate the results of each data set; notwithstanding less number of match-up for validation purposes are available. If we use a wider time window of 4 or 5 hours, we get a major number of match-ups but more variability is encountered with the inconvenience of greater discrepancies between in-situ and RS observations.

3.4 Results

As a result of matching in a time window of 1.5 hour, totally 31 observations are obtained. 5 of them have been excluded from the analysis due to the measurement errors, 6 of them have been removed due to missing values at some wavelengths and 2 of them have not been included into the analysis due to their outlyingness. The analysis have been conducted on 18 observations: 8 observations from Junta

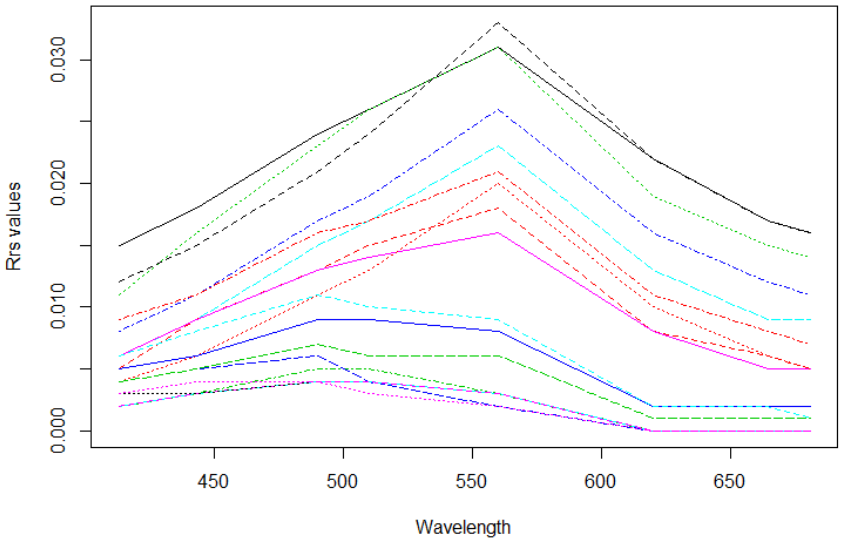


Fig. 3.1: The Rrs curves.

de Andalucía station, 7 observations from Reserva cruise and 3 observations from Fluctuaciones cruise.

The Rrs values at eight different wavelengths have been converted to a functional data object as given in Figure 3.1. Firstly, FLRM with B-Spline approach has been used to predict the TSS parameter. In this case, both the functional predictor and the functional parameter estimate has been smoothed by using B-Spline basis functions. The optimal number of basis functions has been chosen as 5 by the CV criterion [6].

The number of components that will be included in FPCR has been chosen based on their variability. The first 5 components for which the cumulative variance exceeds 95% have been taken into analysis.

The results of FLRMs have been compared with the results of classical approaches in the literature. As offered by [3], exponential regression models have been constructed between TSS and the Rrs values at the band most correlated with the response. The most correlated bands were found as 665 nm (r=0.65) and 681 nm (r=0.66). Therefore, two different single band exponential regression models have been used.

All the functional and exponential regression models were found significant ($p < 0.05$). The coefficient of determination (R^2), Standard Error (Std. Err.) and AMEP values based on LOOCV of the related models are given in Table 3.1.

Regarding R^2 values, FLRMs explain higher amount of variability of the response comparing to exponential regression models.

Table 3.1: R^2, Std. Err. and AMEP values of the models

Models	R^2	Std. Err.	AMEP
FLRM with B-Spline Basis	0.78	0.32	**0.42**
FPCR with the first 5 components	0.82	0.30	0.71
Exponential Regression with 665 nm	0.42	0.46	0.71
Exponential Regression with 681 nm	0.43	0.45	0.70

Among all the models, FLRM using 5 number of B-Spline basis functions has predicted TSS parameter better since the AMEP value of this model is the lowest. Although the AMEP value of FPCR with 5 components is not that low, we see that the predictive performance of the model is as good as the exponential regression models.

3.5 Conclusions

In this study, the performance of FLRMs to predict TSS on RS data has been compared to single band exponential regression models. The data set has been constituted under spatio-temporal filtering by matching exact coordinates in the time window of 1.5 hour difference. Although, the limited number of wavelengths and observations, it is seen that the FLRMs on RS data predict TSS content better than the classical exponential regression models offered in the literature. The best prediction model has been found as FLRM with B-Spline basis approach using 5 basis functions. To conclude, FLRMs estimate the TSS content in Guadalquivir estuary better than other classical approaches that have been used earlier in RS community.

There are several ways to explore in order to improve the prediction ability of the considered models. First, Figure 3.1 suggests the existence of two clusters of curves (well differentiated by Rrs values at wavelengths larger than or equal to 550 nm). These clusters may correspond to clear or low turbid to turbid water conditions in each scene. Then a dummy variable indicating if a day is considered *clear* or *turbid* could be included in the regression models. More observations with different concentrations of TSS will be required to have reliable estimations. Second, the studied observations present (as many environmental data) spatial and temporal dependence (observations from the same boat trip have been taken in close times and close sites). To take into account these dependence in the regression models could lead to more accurate predictions.

Acknowledgements This research was partially supported by the Spanish Ministry of Economy and Competitiveness, and European Regional Development Fund grant MTM2013-43992-R.

References

[1] Besse, P.C., Cardot, H., Faivre, R., Goulard, M.: Statistical modelling of functional data. Appl. Stoch. Model. Bus. **21**(2), 165–173 (2005)

[2] Caballero, I., Morris, Edward P., Ruiz, J., Navarro, G.: Assessment of suspended solids in the Guadalquivir estuary using new DEIMOS-1 medium spatial resolution imagery. Remote Sens. Environ. **146**, 148–158 (2014)

[3] Caballero, I., Morris, Edward P., Ruiz, J., Navarro, G.: The influence of the Guadalquivir River on the spatio-temporal variability of suspended solids and chlorophyll in the Eastern Gulf of Cadiz. Mediterr. Mar. Sci. **15**(4), 721–738 (2014)

[4] Cardot, H., Faivre, R., Goulard, M.: Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. J. Appl. Stat. **30**(10), 1185–1199 (2003)

[5] Faivre, R., Fischer, A.: Predicting crop reflectances using satellite data observing mixed pixels. J. Agric. Biol. Envir. S. **2**(1), 87–107 (1997)

[6] Febrero-Bande, M., Galeano, P., González-Manteiga, W.: Functional Principal Component Regression and Functional Partial Least-squares Regression: An Overview and a Comparative Study. Int. Stat. Rev. (2015) doi:10.1111/insr.12116

[7] Ferraty, F., Vieu, P.: Nonparametric functional data analysis: theory and practice. Springer, USA (2006)

[8] Fettweis, M.P., Nechad B.: Evaluation of in situ and remote sensing sampling methods for SPM concentrations, Belgian continental shelf (southern North Sea). Ocean Dynam. **61**(2), 157–171 (2011)

[9] Gong, M., Miller, C., Scott, E.: Functional PCA for remotely sensed lake surface water temperature data. Procedia Environ. Sci. **6**, 127–130 (2015)

[10] Nechad, B., Ruddick, KG., Park, Y.: Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. Remote Sens. Environ. **114**(4), 854–866 (2010)

[11] Nezlin, N., DiGiacomo, Paul M.: Satellite ocean color observations of stormwater runoff plumes along the San Pedro Shelf (southern California) during 1997–2003. Cont. Shelf Res. **25**(14), 1692–1711 (2005)

[12] Preda, C., Saporta, G.: PLS regression on a stochastic process. Comput. Stat. Data An. **48**(4), 149–158 (2005)

[13] Ramsay, J.O., Silverman, B.W.: Functional Data Analysis. Springer, USA (2005)

Chapter 4
A diagonal componentwise approach for ARB(1) prediction

Javier Álvarez-Liébana and M. Dolores Ruiz-Medina

Abstract This paper extends to the Banach-valued framework previous strong-consistency results derived, in the context of diagonal componentwise estimation of the autocorrelation operator of autoregressive Hilbertian processes, and the associated plug-in prediction. The Banach space B considered here is $B = \mathscr{C}([0,1])$, the space of continuous functions on $[0,1]$ with the supremum norm.

4.1 Introduction.

Functional time series theory emerges as a powerful tool in the statistical analysis of high dimensional data correlated in time. In the Hilbert-valued context, we refer to the reader to the book by Bosq [6], and the papers by Mas [14]; Guillas [9]; Kargin and Onatski [10], among others; and, more recently, to the papers [1, 2, 8, 18, 19]. In particular, in the autoregressive Hilbertian framework (ARH(1) framework), Álvarez-Liébana, Bosq and Ruiz-Medina [2] prove weak-consistency results, in the space of Hilbert-Schmidt operators, for a diagonal componentwise estimator of the autocorrelation operator, and its associated plug-in predictor (in the underlying Hilbert space H). For the same type of diagonal componentwise estimator of the autocorrelation operator, and its associated ARH(1) plug-in predictor, strong-consistency results in the space of bounded linear operators, and in the underlying Hilbert space H, are respectively obtained in Álvarez-Liébana and Ruiz-Medina [3].

Estimation and prediction in the context of Banach-valued autoregressive processes of order one (ARB(1) processes) have also been widely developed, when $B = \mathscr{C}([0,1])$. In the estimation of ARC(1) processes, and its associated ARC(1) plug-in prediction, strong-consistency results are derived in Pumo [16, 17] and Bosq [6], under suitable regularity conditions. In Labbas and Mourid [13], Kuelb's Lemma

Javier Álvarez-Liébana (✉)
University of Granada, Campus Fuentenueva, 18071 Granada. e-mail: javialvaliebana@ugr.es

M. Dolores Ruiz-Medina
University of Granada, Campus Fuentenueva, 18071 Granada. e-mail: mruiz@ugr.es

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_4

plays a key role to extend the results presented in [6, 16, 17]. The formulation of the previous results to the case of weakly dependent innovation process is considered in Dehling and Sharipov [7]. The estimation of ARB(p) processes, with p greater than one, is addressed, for example, in Mourid [15].

Here, we adopt the approach presented in [6, 13] in the framework of ARC(1) processes. Specifically, the extension, from a Banach space to a separable Hilbert space, of the autocorrelation operator of an ARB(1) process (in particular, ARC(1) process) is achieved. The diagonal formulation of the componentwise estimator of the extended autocorrelation operator is considered in this paper, as given in Álvarez-Liébana, Bosq and Ruiz-Medina [2]. A key feature of the diagonal design is the important dimension reduction achieved, with a better or equivalent performance in relation to the approaches presented in [4, 5, 6, 9], as shown in the simulation study undertaken in Álvarez-Liébana, Bosq and Ruiz-Medina [2]. The outline of this work is as follows. Preliminaries section describes the ARB(1) framework. A diagonal component-wise estimator of the extended autocorrelation operator is introduced in Section 4.2. Strong-consistency results, in the space of bounded linear operators on a Hilbert space, are detailed in Section 4.3. The formulation of such results in the Banach-valued context, considering the case of estimation of ARC(1) processes, in the norm of bounded linear operators over $\mathscr{C}([0,1])$, is analyzed in Section 4.4. Final comments and open research lines are given in Section 4.5.

4.2 Preliminaries: ARB(1) general framework and estimator of autocorrelation operator.

Let us consider $(B, \|\cdot\|_B)$ as a real separable Banach space, and let $X = \{X_n,\ n \in \mathbb{Z}\}$ be a zero-mean ARB(1) process, associated with $(\mu, \varepsilon = \{\varepsilon_n,\ n \in \mathbb{Z}\}, \rho)$, satisfying the following equation (see [6], p. 148):

$$X_n = \rho\left(X_{n-1}\right) + \varepsilon_n, \quad n \in \mathbb{Z}, \tag{4.1}$$

where ρ is the autocorrelation operator of X, belonging to the space $\mathscr{L}(B)$ of bounded linear operators on the Banach space B, such that $\|\rho\|_{\mathscr{L}(B)} < 1$. The innovation process ε is assumed to be a strong white noise, and to be also uncorrelated with the random initial condition, with $\sigma_\varepsilon^2 = \mathrm{E}\left[\|\varepsilon_n\|_B^2\right] < \infty$, for all $n \in \mathbb{Z}$. Note that $E\left[\|X_n\|_B^2\right] < \infty$, for all $n \in \mathbb{Z}$, under the above conditions assumed in the introduction of equation (4.1) (see [6], Theorem 6.1).

The autocovariance and cross-covariance operators of X, denoted as C_X and D_X, respectively, are defined as follows, for all $f, g \in B^*$:

$$C_X(f)(g) = \mathrm{E}\left[f(X_n)\, g(X_n)\right], \quad D_X(f)(g) = \mathrm{E}\left[f(X_n)\, g(X_{n+1})\right], \quad n \in \mathbb{Z}, \tag{4.2}$$

where B^* denotes the dual space of B.

Assumption A1. The operator C_X, defined in (4.2), is a nuclear operator given by

$$C_X(f) = \sum_{j=1}^{\infty} f(x_j)\, x_j, \quad \sum_{j=1}^{\infty} \|x_j\|_B^2 < \infty, \quad \forall f \in B^*, \tag{4.3}$$

where $\{x_j,\ j \geq 1\}$ denotes a sequence in B.

Following the approach presented in [13], Kuelb's Lemma is now introduced (see [11, 12] for more details). This lemma plays a key role in the ARB(1) framework, leading to a dense embedding of B into H, where, as before, H denotes a real separable Hilbert space.

Lemma 4.1. *(Kuelb's Lemma) Let $(B, \|\cdot\|_B)$ be a real separable Banach space. Thus, there exists an inner product $\langle\cdot,\cdot\rangle_0$ on B such that $\|\cdot\|_0$ is weaker than $\|\cdot\|_B$. Here, $\|\cdot\|_0$ represents the norm generated by $\langle\cdot,\cdot\rangle_0$, and H denotes the completion of B under $\|\cdot\|_0$, with $(H, \|\cdot\|_H := \|\cdot\|_0)$ being a real separable Hilbert space.*

In the following, H represents the real separable Hilbert space derived from Lemma 4.1. Let us associate to X a zero-mean ARH(1) process $X' = \{X'_n,\ n \in \mathbb{Z}\}$:

$$X'_n = \rho'\left(X'_{n-1}\right) + \varepsilon'_n, \quad X'_n,\ \varepsilon'_n \in H, \quad n \in \mathbb{Z}, \tag{4.4}$$

where H-valued processes $X' = \{X'_n,\ n \in \mathbb{Z}\}$ and $\varepsilon' = \{\varepsilon'_n,\ n \in \mathbb{Z}\}$ are given by

$$X'_n = \sum_{j=1}^{\infty} \langle X_n, e_j\rangle_H e_j, \quad \varepsilon'_n = \sum_{j=1}^{\infty} \langle \varepsilon_n, e_j\rangle_H e_j, \quad n \in \mathbb{Z}, \tag{4.5}$$

with $\{e_j,\ j \geq 1\}$ being an orthonormal basis of H. Hence, equation (4.5) is well defined, since $X_n, \varepsilon_n \in B \hookrightarrow H$, with $\hookrightarrow$ denoting the continuous embedding or inclusion of B into H. The following assumptions are imposed concerning the autocovariance and autocorrelation operator of ARH(1) process X', defined in equation (4.4).

Assumption A1B. The autocovariance operator $C_{X'} = \mathrm{E}\left[X'_n \otimes X'_n\right]$, for each $n \in \mathbb{Z}$, is a positive self-adjoint operator in the trace class, admitting the diagonal spectral decomposition:

$$C_{X'}(f)(g) = \sum_{j=1}^{\infty} C'_j \left\langle \phi_j, f\right\rangle_H \left\langle \phi_j, g\right\rangle_H, \quad \forall f, g \in H, \tag{4.6}$$

in terms of a complete orthonormal system of eigenvectors $\{\phi_j,\ j \geq 1\}$, such that

$$\sum_{j=1}^{\infty} C'_j < \infty, \quad C'_1 > \ldots > C'_j > C'_{j+1} > \ldots > 0. \tag{4.7}$$

Assumption A2. The extended autocorrelation operator ρ' is self-adjoint and Hilbert-Schmidt operator, admitting the following diagonal spectral decomposition:

$$\rho'(f)(g) = \sum_{j=1}^{\infty} \rho'_j \left\langle \phi_j, f\right\rangle_H \left\langle \phi_j, g\right\rangle_H, \quad \forall f, g \in H, \tag{4.8}$$

with $\left\{\rho'_j,\ j \geq 1\right\}$ being the eigenvalues of ρ' associated with the system of eigenvectors $\{\phi_j,\ j \geq 1\}$. In addition,

$$\rho' \in \mathscr{L}(H), \quad \left\|\rho'\right\|_{\mathscr{L}(H)} = \sup_{j\geq 1} \left|\rho'_j\right| < 1. \tag{4.9}$$

Under **Assumption A2**, the operator ρ' satisfies

$$\rho'(f)(g)=\sum_{j=1}^{\infty}\rho'_j\left\langle\phi_j,f\right\rangle_H\left\langle\phi_j,g\right\rangle_H,\quad \sum_{j=1}^{\infty}\left(\rho'_j\right)^2<\infty,\quad \forall f,g\in H. \tag{4.10}$$

As usual, $D_{X'}=\mathrm{E}\left[X'_n\otimes X'_{n+1}\right]$ is the cross-covariance operator of X', for any $n\in\mathbb{Z}$, and then $D_{X'}=\rho' C_{X'}$. Under **Assumptions A1B** and **A2**, we obtain

$$D_{X'}(f)(g)=\rho' C_{X'}(f)(g)=\sum_{j=1}^{\infty}D'_j\left\langle\phi_j,f\right\rangle_H\left\langle\phi_j,g\right\rangle_H,\quad \forall f,g\in H, \tag{4.11}$$

where $\left\{D'_j=\rho'_j C'_j,\ j\geq 1\right\}$ is the system of eigenvalues of $D_{X'}$, with respect to the eigenvectors $\left\{\phi_j,\ j\geq 1\right\}$. In an analogous way as done in [2, 3] approaches, under **Assumptions A1B** and **A2**, we may project (4.4) into $\left\{\phi_j,\ j\geq 1\right\}$, obtaining:

$$X'_{n,j}=\rho'_j X'_{n-1,j}+\varepsilon'_{n,j},\quad \rho'_j\in\mathbb{R},\quad \left|\rho'_j\right|<1, \tag{4.12}$$

where $X'_{n,j}=\left\langle X'_n,\phi_j\right\rangle_H$ and $\varepsilon'_{n,j}=\left\langle\varepsilon'_n,\phi_j\right\rangle_H$, for any $j\geq 1$ and $n\in\mathbb{Z}$.

Remark 4.1. As noted in [13], p. 769, the restriction to B of $C_{X'}$ and ρ' respectively coincides with C_X and ρ, as well as X coincides with X', considered as element of B, with the last one providing the continuous extension of X to H, since B is continuously embedded in H. Therefore, the diagonal spectral decomposition of $C_{X'}$ and ρ', under **Assumptions A1** and **A2**, as operators on H, also holds for their restrictions to B, in the weak-sense, but in the norm of H. The convergence results should be proved in the norm of B. That is the case of the strong-consistency results of diagonal componentwise ARB(1) predictors in this paper, when $B=\mathscr{C}([0,1])$, under the assumption that the diagonal spectral representation in H also holds in the norm of B. Note also that $\mathscr{A}_B=B\cap\mathscr{A}_H$, where $\mathscr{A}_B$ and $\mathscr{A}_H$ denote the Borel σ-algebras on B and H (see [11], Lemma 2.1).

For simplicity, in the current abstract, it is only addressed the case of $\left\{\phi_j,\ j\geq 1\right\}$ are unknown, which, on the other hand, it is the case of interest in practice. From Theorem 4.1, on pp. 98–99, and Corollary 4.1, on pp. 100–101, in [6], under **Assumption A1B**, for n sufficiently large,

$$C'_n=\frac{1}{n}\sum_{i=0}^{n-1}X'_i\otimes X'_i=\sum_{j=1}^{\infty}C'_{n,j}\phi_{n,j}\otimes\phi_{n,j},\ C'_{n,j}=\frac{1}{n}\sum_{i=0}^{n-1}\left(\widetilde{X}'_{i,j}\right)^2,\ \widetilde{X}'_{i,j}=\langle X'_i,\phi_{n,j}\rangle_H,\ j\geq 1, \tag{4.13}$$

where C'_n denotes the empirical estimator of the autocovariance operator $C_{X'}$, with $C'_{n,1}\geq\ldots\geq C'_{n,n}\geq 0=C'_{n,n+1}=C'_{n,n+2}=\ldots$ Recall that $C_n=\frac{1}{n}\sum_{i=0}^{n-1}f(X_i)X_i$, for any $f\in B^*$ and $n\geq 2$, represents the empirical covariance operator of B-valued process X. On the other hand, since $\left\{\phi_{n,j},\ j\geq 1\right\}$ is a complete orthonormal system of eigenvectors on H,

$$D'_n=\frac{1}{n-1}\sum_{i=0}^{n-2}X'_i\otimes X'_{i+1}=\sum_{l=1}^{\infty}\sum_{j=1}^{\infty}D^*_{n,j,l}\phi_{n,j}\otimes\phi_{n,l},\quad n\geq 2, \tag{4.14}$$

where $D^*_{n,j,l} = \langle D'_n(\phi_{n,j}), \phi_{n,l}\rangle_H = \frac{1}{n-1}\sum_{i=0}^{n-2} \widetilde{X}'_{i,j}\widetilde{X}'_{i+1,l}$, for each $j,l \geq 1$ and $n \geq 2$. In particular, we denote $D'_{n,j} = D^*_{n,j,j}$, for any $j \geq 1$ and $n \geq 2$.

Assumption A3. The random initial condition in (4.4) satisfies $\mathrm{E}\left[\left\|X'_0\right\|_H^4\right] < \infty$.

Assumption A4. $C'_{n,k_n} > 0$ a.s., for a truncation parameter $k_n < n$, with $\lim_{n\to\infty} k_n = \infty$.

Under **Assumptions A1B** and **A2-A4**, the following empirical component-wise estimator of operator ρ' on H is then established:

$$\widetilde{\rho}'_{k_n} = \sum_{j=1}^{k_n} \widetilde{\rho}'_{n,j}\phi_{n,j}\otimes\phi_{n,j} = \sum_{j=1}^{k_n} \frac{D'_{n,j}}{C'_{n,j}}\phi_{n,j}\otimes\phi_{n,j}, \quad n \geq 2, \tag{4.15}$$

where k_n is the truncation parameter introduced in **Assumption A4**.

In Section 4.3, the strong-consistency results derived in Álvarez-Liébana and Ruiz-Medina [3] are presented, which can directly be applied to the diagonal componentwise estimator of ρ', and its associated ARH(1) plug-in predictor for X' in the norm of H, as given in Álvarez-Liébana and Ruiz-Medina [3].

4.3 Strong-consistency results: ARH(1) framework.

The following auxiliary results are considered in [3] (see also [6], Lemma 4.2 on p. 103 and Theorem 4.8 on pp. 116–117) to derive strong-consistency results in the ARH(1) framework:

Lemma 4.2. *Let us set* $\Lambda'_k = \sup_{1\leq j\leq k}(C'_j - C'_{j+1})^{-1}$, *for any* $k \geq 1$, *where* $\|X'_0\|_H$ *is bounded. If* $\{k_n,\ n\in\mathbb{Z}\}$ *is a sequence of integers such that* $\Lambda'_{k_n} = o\left(n^{1/4}(\ln(n))^{\beta-1/2}\right)$, *when* $n\to\infty$, *then, under* ***Assumptions A1B*** *and* ***A3****,*

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{1\leq j\leq k_n} \|\phi'_{n,j} - \phi_{n,j}\|_H \to^{a.s.} 0, \quad n\to\infty, \quad \beta > 1/2, \tag{4.16}$$

where $\phi'_{n,j} = sgn\langle\phi_j,\phi_{n,j}\rangle_H\phi_j$, *for any* $j \geq 1$ *and* $n \geq 2$, *with* $sgn\langle\phi_{n,j},\phi_j\rangle_H = \mathbf{1}_{\langle\phi_{n,j},\phi_j\rangle_H\geq 0} - \mathbf{1}_{\langle\phi_{n,j},\phi_j\rangle_H<0}$ *and* $\mathbf{1}$ *being the indicator function.*

Lemma 4.3. *Under* ***Assumption A2*** *and the conditions of Lemma 4.2, for* $\beta > \frac{1}{2}$ *and n large enough,*

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{j\geq 1} \left|D'_{n,j} - D'_j\right| \to^{a.s.} 0, \quad n\to\infty. \tag{4.17}$$

Hence, strong-consistency of $\widetilde{\rho}'_{k_n}$, in the norm of $\mathscr{L}(H)$, is now provided in Proposition 4.1, under **Assumptions A1B** and **A2-A4**.

Proposition 4.1. *Let* k_n *a sequence of integers such that* $k_n C'_{k_n} < 1$, *for any* $n \geq \widetilde{n}_0$ *and* $\widetilde{n}_0$ *an integer large enough, such that, for* $\beta > 1/2$,

$$\Lambda'_{k_n} = o\left(n^{1/4}(\ln(n))^{\beta-1/2}\right), \quad \frac{1}{C'_{k_n}}\sum_{j=1}^{k_n} a'_j = \mathscr{O}\left(n^{1/4}(\ln(n))^{-\beta}\right), \quad n\to\infty, \tag{4.18}$$

where $a'_1 = 2\sqrt{2}\frac{1}{C'_1-C'_2}$ *and* $a'_j = 2\sqrt{2}\max\left(\frac{1}{C'_{j-1}-C'_j}, \frac{1}{C'_j-C'_{j+1}}\right)$, *for any* $j \geq 2$. *Then, under* ***Assumptions A1B*** *and* ***A2-A4****,*

$$\left\|\widetilde{\rho}'_{k_n} - \rho'\right\|_{\mathscr{L}(H)} \longrightarrow 0 \quad a.s. \tag{4.19}$$

4.4 Strong-consistency results: ARC(1) framework.

In this section, we study the particular framework of ARC(1) processes. That is, we consider the case of $B = \mathscr{C}([0,1])$, and the norm $\mathscr{L}(\mathscr{C}([0,1]))$ of the bounded linear operators on $\mathscr{C}([0,1])$, that will be denoted in the following as $\mathscr{L}(C)$. For this particular case, the following lemma builds H associated with that class of ARC(1) processes.

Lemma 4.4. *(See [13], Lemma 1, p. 769, for more details). Let B be the Banach space* $\mathscr{C}([0,1])$*, with its associated uniform norm, and the measure* $\varsigma = \sum_{n\in\mathbb{N}} z_n \delta_{t_n}$*, such that* $\{z_n,\ n\in\mathbb{N}\}$ *is a positive real sequence, with* $\sum_{n\in\mathbb{N}} z_n = 1$*, and* $\{t_n,\ n\in\mathbb{N}\}$ *is a real dense sequence in* $[0,1]$*. Thus,* $H = \mathscr{L}^2\left([0,1], \beta_{[0,1]}, \varsigma\right)$ *is the Hilbert space constructed according to Lemma 4.1, such that H is the continuous extension of* $\mathscr{C}([0,1])$.

As done in [6], from Lemma 4.4, the Lebesgue measure λ is adopted, with $H = \mathscr{L}^2\left([0,1], \beta_{[0,1]}, \lambda\right)$, denoted as L^2. The following particular assumption concerning ρ is established:

Assumption C1. Let X be the ARB(1) process given by (4.1), consider that ρ satisfies

$$\rho(x)(t) = \int_0^1 r(s,t)x(s)ds = \langle r(\cdot,t), x\rangle_{L^2}, \quad x\in\mathscr{C}([0,1]), \tag{4.20}$$

where $r(\cdot,\cdot)$ is a continuous kernel with $\|\cdot\|_{\mathscr{C}([0,1]^2)} < 1$. Thus, $\|\rho\|_{\mathscr{L}(C)} < 1$.

From Lemmas 4.1 and 4.4, and equations (4.4)-(4.5), the following ARH(1) could be associated to X, under **Assumptions A1B** and **A2**:

$$X'_n = \rho'\left(X'_{n-1}\right) + \varepsilon'_n, \quad X'_n,\ \varepsilon'_n \in H, \quad n\in\mathbb{Z}, \quad \left\|\rho'\right\|_{\mathscr{L}\left(L^2\right)} < 1, \tag{4.21}$$

$$X'_n = \sum_{j=1}^{\infty}\left(\int_0^1 X_n(s)e_j(s)ds\right)e_j, \quad \varepsilon'_n = \sum_{j=1}^{\infty}\left(\int_0^1 \varepsilon_n(s)e_j(s)ds\right)e_j, \tag{4.22}$$

where $\left\{e_j,\ j\geq 1\right\}$ is an orthonormal basis on H such that the empirical estimator C'_n, given in (4.13), is also defined as an integral operator (see [6], p. 223):

$$C'_n(x)(t) = \int_0^1 c_n(s,t)x(s)ds, \quad c_n(s,t) = \frac{1}{n}\sum_{i=0}^{n-1} X_i(s)X_i(t), \quad 0\leq s,t\leq 1, \tag{4.23}$$

where $\{\phi_{n,j},\ j \geq 1\}$ and $\left\{C'_{n,j},\ j \geq 1\right\}$ are defined in equation (4.13) above. From (4.23), C'_n can operate in L^2 as well as in $\mathscr{C}([0,1])$.

Assumption C2. The eigenvectors $\{\phi_{n,j},\ j \geq 1\}$ given in (4.13) are assumed to verify

$$\sup_{j\geq 1} \left\|\phi_j\right\|_{\mathscr{C}([0,1])} < \infty, \quad \sup_{j\geq 1} \left\|\phi_{n,j}\right\|_{\mathscr{C}([0,1])} \ < \ \infty, \tag{4.24}$$

$$\sup_{\|x\|_{\mathscr{C}([0,1])}<1} \left\| \rho(x) - \sum_{j=1}^{k_n} \rho'_j \langle x, \phi_j\rangle_{L^2} \phi_j \right\|_{\mathscr{C}([0,1])} \longrightarrow 0, \text{ as } k \to \infty. \tag{4.25}$$

The following diagonal component-wise estimator of ρ, as an operator on $\mathscr{C}([0,1])$, is then defined, under **Assumptions A1, A2, A3-A4** and **C1-C2**:

$$\widetilde{\rho}_{k_n} = \sum_{j=1}^{k_n} \widetilde{\rho}'_{n,j} \phi_{n,j} \otimes \phi_{n,j} = \sum_{j=1}^{k_n} \frac{D'_{n,j}}{C'_{n,j}} \phi_{n,j} \otimes \phi_{n,j}, \quad n \geq 2. \tag{4.26}$$

Remark 4.2. Note that condition (4.25) in **Assumption C2** could be supplied by

$$\sup_{\|x\|_{\mathscr{C}([0,1])}<1} \left\| \rho(x) - \sum_{j=1}^{k_n} \rho'_{n,j} \langle x, \phi_{n,j}\rangle_{L^2} \phi_{n,j} \right\|_{\mathscr{C}([0,1])} \xrightarrow{k\to\infty} 0, \ \sum_{j=1}^{k_n} \left\|\phi_{n,j} - \phi_j\right\|_{\mathscr{C}([0,1])} \xrightarrow{k\to\infty} 0. \tag{4.27}$$

where $\rho'_{n,j} = \langle \rho', \phi_{n,j}\rangle_{L^2}$, for any $j \geq 1$ and $n \geq 2$.

Under **Assumptions A1B, A2-A4** and **C1-C2**, the strong-consistency of $\widetilde{\rho}_{k_n}$, in the norm of $\mathscr{L}(C)$ is now derived.

Proposition 4.2. *Let $\{k_n,\ n \in \mathbb{N}\}$ a sequence of integers satisfying $k_n C'_{k_n} < 1$, for any $n \geq \widetilde{n}_0$ and $\widetilde{n}_0$ an integer large enough, such that, for $\beta > 1/2$, $\Lambda'_{k_n} = o\left(n^{1/4}(\ln(n))^{\beta-1/2}\right)$ and*

$$\frac{1}{C'_{k_n}} \sum_{j=1}^{k_n} a'_j = \mathscr{O}\left(n^{1/4}\left(\ln(n)\right)^{-\beta}\right), \quad n \to \infty, \tag{4.28}$$

*where $a'_1 = 2\sqrt{2}\frac{1}{C'_1 - C'_2}$ and $a'_j = 2\sqrt{2}\max\left(\frac{1}{C'_{j-1} - C'_j}, \frac{1}{C'_j - C'_{j+1}}\right)$, for any $j \geq 2$. Then, under **Assumptions A1B, A2-A4** and **C1-C2**,*

$$\|\widetilde{\rho}_{k_n} - \rho\|_{\mathscr{L}(C)} \longrightarrow^{a.s.} 0. \tag{4.29}$$

4.5 Final comments and open research lines.

As noted in the Introduction, the current abstract only focuses on the case where the eigenvectors of the autocovariance operator $C_{X'}$ are unknown, as it often occurs in practice. Strong-consistency of the plug-in predictor associated with $\widetilde{\rho}_{k_n}$ follows straightforwardly from Proposition 4.2, since, in general, for a Banach space B, the following inequality holds:

$$\left\|\left(\mathscr{A}-\mathscr{B}\right)\left(X_{n-1}\right)\right\|_B \leq \left\|\mathscr{A}-\mathscr{B}\right\|_{\mathscr{L}(B)} \left\|X_{n-1}\right\|_B, \quad \mathscr{A}, \mathscr{B} \in \mathscr{L}(B).$$

Both cases, when $\left\{\phi_j,\ j \geq 1\right\}$ are known and unknown, will be addressed in the presentation of this work, as well as weak-consistency results in different function spaces. A simulation study will also be undertaken. Currently, different particular spaces beyond $\mathscr{C}([0,1])$ are being considered.

Acknowledgements This work has been supported in part by project MTM2015–71839–P (co-funded by ERDF), of the DGI, MINECO, Spain.

References

[1] Álvarez-Liébana, J., Bosq, D., Ruiz-Medina, M. D.: Consistency of the plug-in functional predictor of the Ornstein-Uhlenbeck process in Hilbert and Banach spaces. Statistics & Probability Letters **117**, 12–22 (2016)

[2] Álvarez-Liébana, J., Bosq, D., Ruiz-Medina, M. D.: Asymptotic properties of a component-wise ARH(1) plug-in predictor. J. Multivar. Analysis **155**, 12–34 (2017)

[3] Álvarez-Liébana, J., Ruiz-Medina, M. D.: ARH(1) prediction from strongly-consistent diagonal componentwise functional parameter estimators (submitted).

[4] Antoniadis, A., Sapatinas, T.: Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. J. Multivariate Anal. **87**, 133–158 (2003)

[5] Besse, P. C., Cardot, H., Stephenson, D.B.: Autoregressive forecasting of some functional climatic variations. Scand. J. Statist. **27**, 673–687 (2000)

[6] Bosq, D.: Linear Processes in Function Spaces. Springer, New York (2000)

[7] Dehling, H., Sharipov, O. S.: Estimation of mean and covariance operator for Banach space valued autoregressive processes with dependent innovations. Stat. Inference Stoch. Process. **8**, 137–149 (2005)

[8] Febrero-Bande, M., Gonzalez-Manteiga, W.: Generalized additive models for functional data. Test **22**, 278–292 (2013)

[9] Guillas, S.: Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes. Statist. Probab. Lett. **55**, 281–291 (2001)

[10] Kargin, V., Onatski, A.: Curve forecasting by functional autoregression. J. Multivar. Analysis **99**, 2508–2526 (2008)

[11] Kuelbs, J.: Gaussian measures on a Banach spaces. J. Funct. Anal. **5**, 354–367 (1970)

[12] Kuelbs, J.: A strong convergence theorem for Banach valued random variables. Ann. Prob. **4**, 744–771 (1976)

[13] Labbas, A., Mourid, T.: Estimation et prévision d'un processus autorégressif Banach. C. R. Acad. Sci. Paris, Ser. I **335**, 767–772 (2002)

[14] Mas, A.: Normalité asymptotique de l'estimateur empirique de l'opérateur d'autocorrélation d'un processus ARH(1). C. R. Acad. Sci. Paris Sér. I Math. **329**, 899–902 (1999)

[15] Mourid, T.: Processus autorégressifs banachiques d'ordre supérieur. C. R. Acad. Sci. Paris Sér. I Math. **317**, 1167–1172 (1993)
[16] Pumo, B.: Estimation et prévision de processus autorégressifs fonctionnels. PhD thesis, Université de Paris 6, Paris (1992)
[17] Pumo, B.: Prediction of continuous time processes by $C_{[0,1]}$-valued autoregressive process. Stat. Inference Stoch. Processes **1**, 297–309 (1998)
[18] Ruiz-Medina, M. D.: Spatial autoregressive and moving average Hilbertian processes. J. Multivar. Analysis **102**, 292–305 (2011)
[19] Ruiz-Medina, M. D.: Spatial functional prediction from spatial autoregressive Hilbertian processes. Environmetrics **23**, 119–128 (2012)
[20] Saidi, F. B.: An extension of the notion of the orthogonality to Banach spaces. J. Math. Anal. and Applications **267**, 29–47 (2002)
[21] Shoja, A., Mazaheri, H.: General orthogonality in Banach spaces. Int. J. Math. Anal. **1**, 553–556 (2007)

Chapter 5
A general sparse modeling approach for regression problems involving functional data

Germán Aneiros and Philippe Vieu

Abstract This presentation aims to introduce an approach for dealing with sparse regression models when functional variables are involved in the statistical sample. The idea is not to restrict to any specific variable selection procedure, but rather to present a two-stage methodology allowing to adapt efficiently any multivariate procedure to the functional framework. These ideas can be applied to any kind of functional regression models, including linear, semi-parametric or non-parametric models.

5.1 Difference and similarities between High Dimensional and Functional problems in Statistics

In the recent years a wide scope of new statistical problems had emerged because of an increasing amount of situations in which larger and larger data are collected and stored. Usually, one can distinguish two kinds of problems according to the fact that the data come from a high dimensional variable $X = (X^1, \ldots, X^p)$ or from a continuously indexed one $\chi = \{\chi(t), t \in I\}$. The former situation is often called High Dimensional Statistics (HDS) or Big Data Analysis while the second one is known in the literature as Infinite Dimensional Statistics or Functional Data Analysis (FDA). A selective sample of general presentations on HDS involves [8], [29], [19] and [30], while a sample of books in FDA involves [33], [15], [21] and [22].

From a naïve point of view, one could have in mind that any functional variable $\chi = \{\chi(t), t \in I\}$ is in fact observed on a finite grid $t_1, \ldots, t_p$ in such a way that its study turns to be just the analysis of the high dimensional vector $\{\chi(t_1), \ldots, \chi(t_p)\}$.

Germán Aneiros
Departamento de Matemáticas, Universidade da Coruña, A Coruña, Spain, e-mail: ganeiros@udc.es

Philippe Vieu (✉)
Institut de Mathématiques, Université Paul Sabatier, Toulouse, France, e-mail: philippe.vieu@math.univ-toulouse.fr

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_5

But of course there are so many specifities for such a high dimensional vector (because of the underlying continuous structure) that FDA cannot be seen as a sub-field of HDS as this naïve point of view would suggest. This has been commonly accepted in the statistical community until a few years ago and both fields, HDS and FDA, have been widely developed independently one from each other. However, the naïve point of view cannot be rejected so directly, because without seeing one field as a sub-field of the other one, both have to front the same key question: is all the information in the data really useful? and if not, how to extract the most informative part of the data? In the last few years there has been some attention paid to the need for developing methods crossing both fields in order to capture advantages of each of them. Important steps in this direction were the 3rd issue of IWFOS meeting in Italy (see [7]) and the companion special issue of the Journal of Multivariate Analysis [17] (see also the discussion in the review paper [10]).

In the next Section 5.2 we will provide a discussion based on a necessarily short and selective set of works devoted to the specific question: how can we use/adapt the existing variable selection techniques widely studied in HDS for the important question arising in FDA of deciding which points among the discretization $t_1,\dots,t_p$ have to be selected for the analysis? Then, in Section 5.3 we will present a two-stage procedure whose main aims is not to construct a new discretized points selection method but rather to give a tool allowing direct and efficient adaptation to FDA of any existing selection method developed in HDS.

5.2 A short discussion on sparse functional regression models

Taking the point of view of looking at a random functional element χ as its discretized version $\{\chi(t_1),\dots,\chi(t_p)\}$, the question is to detect which ($\mathscr{S}$) among the real variables $X^j=\chi(t_j)$ are informative for some statistical purpose, as for instance for predicting some response Y. The natural idea is to construct models in which only a few (let say s) among the p variables are informative (that is, $\#\mathscr{S}=s$). In this framework people call the elements of $\mathscr{S}$ *impact points* (see [31] or [26]) or *most predictive points* (see [14]). Usually, one lets these parameters on n (that is, $s=s_n$ and $p=p_n$) and for main purpose one has $p_n >> n$ and $s_n << p_n$. This is called sparse regression modeling[1], and most often the properties of the associated statistical methods depend in a crucial way on the parameters p_n and s_n. Once again, a fast naïve approach would consist in seeing this search of informative points just a special variable selection problem from a multivariate vector. However, the nature of both problems exhibit at least one strong structural difference: in standard multivariate problems an increasing dimension p_n means a wider scope of information (some/most of it being probably unuseful) while in the functional setting an increasing dimension p_n just means a more precise information (just because it is linked with finer grid). In counterpart, the naïve point of view cannot be rejected totally

[1] Note that sparsity is a phenomenon that can affect functional data in many other ways than modeling. We will not enter here in this discussion (see for instance [5])

since despite of these strong structural differences the aims of both problems are similar.

There is a very extensive literature on sparse multivariate linear regression (see e.g. [11], [12] and [25] for SCAD-penalized estimators) with a few semi-linear extensions (see e.g. [18] for an adaptive penalized CQR approach, and [20], [23] and [1] for SCAD-penalized estimators) and a few nonparametric ones (see e.g. [32], [24], [6] and [35] for type-Lasso proposals; see also [36], [34], [28]) and [9] for other kinds of procedures). Since a few years, there has been various works aiming to adapt some of these multivariate selection procedures to the functional setting, most of them being concerned with functional linear regression (see [31] for estimation of an impact point in a continuous process; [27] for a proposal where principal components are included as additional explanatory variables in an augmented regression model, the corresponding parameters being estimated by means of selection procedures; [37] for penalized-Lasso estimators; and [26] for a recent contribution for estimating the number and locations of points of impact).

In the next Section 5.3, the aim is to present a general two-stage procedure which is able to take into account the functional specificity of the problem. These ideas are of double interest. Firstly, they are not linked with any specific selection approach and will be therefore applicable to any type of methods (LASSO, Dantzig, ...). Secondly, they are not linked with any specific regression model and will be therefore usable as well in linear, as in nonparametric or in semiparametric problems. The presentation will be restricted to methodological ideas; theoretical asymptotics as well as real data analysis will be only quickly discussed and referred to relevant additional literature.

5.3 The two-stage splitting ideas for impact points selection in functional regression

Consider a regression problem in which the covariate is a functional variable $\{\chi = \chi(t), t \in I\}$ and with some response Y:

$$Y = r(\chi) + \varepsilon.$$

Suppose that χ is observed on a finite grid $\{t_1, \ldots, t_{p_n}\}$ and denote the corresponding observed vector $X = (X^1, \ldots, X^{p_n}) = \{\chi(t_1), \ldots, \chi(t_{p_n})\}$. The question is to know which among the X^j are of interest for predicting Y. At this stage, we do not specify what is the model for the operator $r(\cdot)$ nor the nature of the response Y, since the ideas developed below are independent of these two points.

5.3.1 Splitting the data and first rough selection

The grid is split into several pieces in the following way

$$\{1,\dots,p_n\} = \cup_{k=1}^{w_n} E_k \text{ with } E_k = \{(2k-1)q_n/2 + m,\ m = -\frac{q_n}{2}+1,\dots,\frac{q_n}{2}\},$$

where w_n and q_n are integer sequences such that $w_n q_n = p_n$. Because of the continuous structure of the data, two variables X^j and X^k corresponding to close indexes j and k may have similar effects on the response, and so one may consider to select impact points only among the isolated covariates

$$\mathscr{S}_1 = \{X^{((2k-1)q_n+1)/2}\}_{k=1}^{w_n}.$$

5.3.2 Sharpening the procedure

Let say that, after this first rough procedure, one has selected some subset of variables $\{X^{((2k-1)q_n+1)/2}\}_{k\in K_0}$, where $K_0 \subset \{1,\dots,w_n\}$. Of course, there is the risk that the splitting procedure does not have some information between two of these selected points. To overcome this possible effect, it is possible to sharpen the method by performing again a second variable selection procedure among the new set of variables $\mathscr{S}_2$ defined to be those previously selected and their neighbours:

$$\mathscr{S}_2 = \{X^j, j \in \cup_{k\in K_0} E_k\}.$$

After this sharpening stage one has at hand a selected set of impact points $\tilde{\mathscr{S}}$.

5.3.3 Summary of the procedure

The interest of this general way of thinking is double: firstly, it can be developed for any kind of models for $r(\cdot)$ and, secondly, it may involve any kind of strategy for performing variable selection at each stage. Of course such a variable selection is linked with the model considered for the operator $r(\cdot)$. As a matter of conclusion the two-stage functional impact points selection procedure can be summarized in the following way:

i) Choose a specific model for the operator $r(\cdot)$ (see examples in Section 5.4);
ii) Choose a pilot variable selection procedure adapted to the model chosen in i);
iii) Apply the pilot procedure to the set $\mathscr{S}_1$ to get K_0. Construct $\mathscr{S}_2$;
iv) Apply again the pilot procedure to the set $\mathscr{S}_2$;
v) Get the final selected set of impact points $\tilde{\mathscr{S}}$.

5.4 About the properties of the functional two-stage selection procedure

Roughly speaking, in HDS literature the mathematical properties of a variable selection procedure are strongly depending on the dimension of the data: an increasing value of the parameter p_n has negative effects both on the rate of convergence of the method and on the restrictivity of the conditions needed to work. So, because the two-stage procedure acts on subsets of the original data, one may expect it to have much better rates of convergence and also to need less restrictive conditions for working. Of course these general ideas have to be deeply studied for each situation, that means for each specific model taken in i) and for each pilot procedure chosen in ii). To fix the ideas, we just discuss here three kinds of regression models including linear, semi-parametric and non-parametric situations. Precisely we consider the usual sparse linear model in which α^j are scalar parameters:

$$Y = \alpha_0 + \sum_{j=1}^{p_n} \alpha^j X^j + \varepsilon \text{ with } \mathscr{S} = \{j = 1, \ldots, p_n,\ \alpha^j \neq 0\}, \tag{5.1}$$

its nonparametric additive version in which f^j are smooth functional parameters:

$$Y = \alpha_0 + \sum_{j=1}^{p_n} f^j(X^j) + \varepsilon \text{ with } \mathscr{S} = \{j = 1, \ldots, p_n,\ f^j \neq 0\}, \tag{5.2}$$

and the partial linear one combining all together the four effects: linear, nonparametric, continuous and discretized, in which the α^j are scalar parameters, $m(\cdot)$ is a smooth operator and ζ is a covariate of functional nature:

$$Y = \alpha_0 + \sum_{j=1}^{p_n} \alpha^j X^j + m(\zeta) + \varepsilon \text{ with } \mathscr{S} = \{j = 1, \ldots, p_n,\ \alpha^j \neq 0\}. \tag{5.3}$$

Models (5.1), (5.2) and (5.3) are studied in [2], [3] and [4], respectively. Asymptotic results are stated showing how (pending to suitable conditions on the sequences q_n and w_n, and on the smoothness of the curves χ (and ζ, for the case of model (5.3))) the selected impact points $\hat{\mathscr{S}}$ tends to the true one $\mathscr{S}$. Convergence for the coefficients of the models (that are the α^j or the f^j) are also stated. Because the two-stage selection procedure does not act directly on the whole p_n-dimensional vector, it is shown that the rates of convergence improve, in general, the usual ones.

For instance, focusing first on the sparse linear model (5.1), [25] obtained that the SCAD-penalized least squares estimator of $\alpha = (\alpha^1, \ldots, \alpha^{p_n})^T$ converges at a rate $n^{-1/2} p_n^{1/2}$, assuming that $p_n = o(n^{1/2})$. Then, if one uses such SCAD method as pilot variable selection in our two-stage procedure, the corresponding rate of converge improves to $n^{-1/2} s_n^{1/2}$, while the condition on the dimensionality of the model changes to $p_n = o(q_n n^{1/2})$. It is worth noting that these are not minor improvements since in practice (i) s_n is always very much smaller than p_n and, (ii) the

fact that q_n could converge to ∞ allows to apply our procedure in scenarios of major dimensionality (this being of much interest because the technological progress for collecting and storing data provides curve datasets which are measured in more and more fine grids, leading to very high values of p_n). The case of variable selection in the sparse nonparametric additive model (5.2) was dealt in [24], among others. Because B-spline bases were used to approximate the additive components f^j, the group Lasso procedure was considered for variable selection, obtaining a rate of convergence $n^{-2/3}\log(m_n p_n)$, under $p_n = \exp(o(n^{2/3}))$ (m_n denotes the truncation algorithm in the B-spline representations). The rate of convergence of our two-stage proposal when uses such procedure as pilot variable selector is $n^{-2/3}\log(m_n)$, under the weaker assumption $p_n = q_n \exp(o(n^{2/3}))$. Again, the improvement from our proposal against the standard one is apparent. Finally, we focus on the more complex model (5.3). As far as we know, the only variable selection method for the sparse partial linear model (5.3) can be seen in [1]. Given the complexity of such model, and in the sake of simplicity, we will compare the behaviour of both the standard procedure and the proposed two-stage one restricted to fixed $s_n = s < \infty$. Then, when SCAD-based penalized estimators are considered, and assuming $p_n = o(n^{1/2})$, [1] proved that the rate of convergence for the standard method is $n^{-1/2}$. Although the proposed two-stage procedure maintains that same rate of convergence, the condition on the dimensionality is weaker ($p_n = o(q_n n^{1/2})$).

5.5 Future work

This presentation has introduced an approach for selecting covariates coming from a discretized functional variable. Although it was explicitly studied and applied only on three particular cases (sparse linear, nonparametric and partial linear regression models), the main idea of splitting is also expected to work in other regression models (as single index [16] and projection pursuit [13] models, ...). In addition, such idea could be also used in other problems dealing with functional data, as it is the case, for instance, of classification. These challengers (and other related ones) will be dealt in future researches.

Acknowledgements This research is partially supported by Grant MTM2014-52876-R from Spanish Ministerio de Economía y Competitividad.

References

[1] Aneiros, G., Ferraty, F., Vieu, P.: Variable selection in partial linear regression with functional covariate. Statistics **49**, 1322–1347 (2015)

[2] Aneiros, G., Vieu, P.: Variable selection in infinite-dimensional problems. Stat. Probab. Lett. **94**, 12–20 (2014)

[3] Aneiros, G., Vieu, P.: Partial linear modelling with multi-functional covariates. Comput. Stat. **30**, 647–671 (2015)
[4] Aneiros, G., Vieu, P.: Sparse nonparametric model for regression with functional covariate. J. Nonparametr. Stat. **28**, 839–859 (2016)
[5] Aneiros, G., Vieu, P.: Comments on: Probability enhanced effective dimension reduction for classifying sparse functional data. Test **25**, 27–32 (2016)
[6] Belloni, A., Chernozhukov, V., Wang, L.: Pivotal estimation via square-root Lasso in nonparametric regression. Ann. Stat. **42**, 757–788 (2014)
[7] Bongiorno, E., Goia, A., Salinelli, E., Vieu, P.: An overview of IWFOS'2014. In: Contributions in Infinite-Dimensional Statistics and Related Topics, 1–6. Esculapio, Bologna (2014)
[8] Bühlmann, P., van de Geer, S.: Statistics for High-Dimensional Data. Methods, Theory and Applications. Springer, Heidelberg (2011)
[9] Comminges, L., Dalalyan, A.: Tight conditions for consistency of variable selection in the context of high dimensionality. Ann. Stat. **40**, 2667–2696 (2012)
[10] Cuevas, A.: A partial overview of the theory of statistics with functional data. J. Stat. Plann. Inference **147**, 1–23 (2014)
[11] Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. **96**, 1348–1360 (2001)
[12] Fan, J., Peng, H.: Nonconcave penalized likelihood with a diverging number of parameters. Ann. Stat. **32**, 928–961 (2004)
[13] Ferraty, F., Goia,A., Salinelli, E., Vieu, P.: Functional projection pursuit regression. Test **22**, 293-320 (2013).
[14] Ferraty, F., Hall, P., Vieu, P.: Most-predictive design points for functional data predictors. Biometrika **97**, 807–824 (2010)
[15] Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis. Springer-Verlag, New York (2006)
[16] Goia, A., Vieu, P.: A partitioned single functional index model. Comput. Stat. **30**, 673-692 (2015)
[17] Goia, A., Vieu, P.: An introduction to recent advances in high/infinite dimensional statistics. J. Multivariate Anal. **46**, 1–6 (2016)
[18] Guo, J., Tang, M., Tian, M., Zhu, K.: Variable selection in high-dimensional partially linear additive models for composite quantile regression. Comput. Statist. Data Anal. **65**, 56–67 (2013)
[19] Härdle, W., Simar, L.: Applied Multivariate Statistical Analysis. (Fourth edition) Springer, Heidelberg (2015)
[20] Hong, Z., Hu, Y., Lian, H.: Variable selection for high-dimensional varying coefficient partially linear models via nonconcave penalty. Metrika **76**, 887–908 (2013)
[21] Horváth, L., Kokoszka, P.: Inference for Functional Data with Applications. Springer, New York (2012)
[22] Hsing, T., Eubank, R.: Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. John Wiley & Sons, Chichester (2015)

[23] Hu, Y., Lian, H.: Variable selection in a partially linear proportional hazards model with a diverging dimensionality. Stat. Probab. Lett. **83**, 61–69 (2013)
[24] Huang, J., Horowitz, J., Wei, F.: Variable selection in nonparametric additive models. Ann. Stat. **38**, 2282–2313 (2010)
[25] Huang, J., Xie, H.: Asymptotic oracle properties of SCAD-penalized least squared estimators. Asymptotics: Particles, Processes and Inverse Problems. In: IMS Lecture Notes-Monograph Series **55**, 149–166 (2007)
[26] Kneip, A., Poss, D., Sarda, P.: Functional linear regression with points of impact. Ann. Stat. **44**, 1–30 (2016)
[27] Kneip, A., Sarda, P.: Factor models and variable selection in high-dimensional regression analysis. Ann. Stat. **39**, 2410–2447 (2011)
[28] Lafferty, J., Wasserman, L.: RODEO: Sparse, greedy, nonparametric regression. Ann. Stat. **36**, 28–63 (2008)
[29] Liu, S., McGree, J., Ge, Z., Xie, Y.: Computational and Statistical Methods for Analysing Big Data with Applications. Elsevier/Academic Press, Amsterdam (2015)
[30] Matwin, S., Mielniczuk, J. (Eds): Challenges in Computational Statistics and Data Mining. Springer (2016)
[31] McKeague, I., Sen, B.: Fractals with point impact in functional linear regression. Ann. Stat. **38**, 2559–2586 (2010)
[32] Meier, L., van de Geer, S., Bühlmann, P.: High-dimensional additive modeling. Ann. Stat. **37**, 3779–3821 (2009)
[33] Ramsay, J., Silverman, B.: Functional Data Analysis. (Second edition) Springer, New York (2005)
[34] Vieu, P.: Choice of regressors in nonparametric estimation. Comput. Statist. Data Anal. **17**, 575–594 (1994)
[35] Wang, S., Su, L.: Simultaneous Lasso and Dantzig selector in high dimensional nonparametric regression. Int. J. Appl. Math. Stat. **42**, 103–118 (2013)
[36] Zhang, P.: Variable selection in nonparametric regression with continuous covariates. Ann. Stat. **19**, 1869–1882 (1991)
[37] Zhao, Y., Todd O., Reiss, P.: Wavelet-based LASSO in functional linear regression. J. Comput. Graph. Statist. **21**, 600–617 (2012)

Chapter 6
A time-dependent PDE regularization to model functional data defined over spatio-temporal domains

Eleonora Arnone, Laura Azzimonti, Fabio Nobile and Laura M. Sangalli

Abstract We propose a method for the analysis of functional data defined over spatio-temporal domains when prior knowledge on the phenomenon under study is available. The model is based on regression with Partial Differential Equations (PDE) penalization. The PDE formalizes the information on the phenomenon and models the regularity of the field in space and time.

6.1 Space-Time Regression with PDE Penalization

We propose a new method for the analysis of functional data defined over spatio-temporal domains. These data can be interpreted as time evolving surfaces or spatially dependent curves. The proposed method is based on regression with differential regularization and extends the models proposed in [16, 10, 3, 4]. We are in particular interested to the case when prior knowledge on the phenomenon under study is available. Analogously to [4], the prior knowledge is described in terms of a PDE. But, with respect to [4], we here also include the temporal dimension, considering

Eleonora Arnone (✉)
MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy, e-mail: eleonora.arnone@polimi.it

Laura Azzimonti
IDSIA - Department of Innovative Technologies, Università della Svizzera Italiana Galleria 1, via Cantonale, Switzerland e-mail: laura.azzimonti@supsi.ch

Fabio Nobile
MATHICSE-CSQI, École Polytechnique Fédérale de Lausanne, Route Cantonale, 1015 Lausanne, Switzerland e-mail: fabio.nobile@epfl.ch

Laura M. Sangalli
MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy e-mail: laura.sangalli@polimi.it

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_6

a time-dependent PDE that jointly models the spatial and temporal variation of the phenomenon.

Recently, various methods have been proposed to deal with spatially dependent functional data [12]. Starting from the pioneering work in [9], kriging prediction models for stationary spatial functional data have been developed in [7, 15, 8]. Universal kriging approaches have been presented in [6, 13, 14] and kriging with external drift in [10]. Moreover, methods based on differential regularizations has been recently developed in [2, 11, 5, 1]. In these works, the authors consider two roughness penalties that account separately for the regularity of the field in space and in time, using a tensor product approach. Differently from the latter works, we here use a unique regularizing term that jointly model the space-time variation of the phenomenon, on the base of problem-specific prior information. Specifically, the regularization involves the misfit of a time-dependent parabolic PDE modeling the phenomenon behavior, $\dot{f} + Lf = u$, where $\dot{f}$ is the time derivative of the spatio-temporal function f, and L is a differential operator in space. We consider various samplings designs, including geo-statistical and areal data. We show that the corresponding estimation problems are well posed and can be discretized in space by means of the Finite Element method, similarly to [16, 10, 3, 4], and in time by means of the Finite Difference method. The model can handle data distributed over domains having complex shapes such as domains with strong concavities and holes. Moreover, various types of boundary conditions can be considered, with a very flexible modeling of the behavior of the spatio-temporal field and the boundaries of the domain of interest.

6.2 Motivating application

As a motivating example, we want to study the blood flow velocity field in the common carotid artery, using data from Echo-Color Doppler (ECD). These data were analyzed for a single time instant, the systolic peak, in [4]. The echo doppler scan provides a time-dependent measure of the velocities of blood flow particles sampled within a bean in an artery section. Figure 6.1 shows the ECD signal registered in a centrally located bean at the cross-section of the common carotid artery located 2 cm before the artery bifurcation. The lower part of the ECD image displays the acquired velocity signal during the time lapse of about three heart beats. This signal represents the time-evolving histogram of the measured velocities in the beam: the gray-scaled intensity of pixels is proportional to the number of blood-cells in the beam moving at a certain velocity for any fixed time. Starting from the ECD signals over multiple beans, we would like to reconstruct the time-varying mean velocity of the blood-flow over the whole carotid section. A fundamental constraint in this application is given by the so-called no-slip boundary conditions; indeed the physics of the problem implies that, because of the friction between blood cells and arterial wall, blood-flow velocity is zero at the arterial wall. Including the prior knowledge about the blood fluid dynamics and appropriate boundaries conditions is in this context fundamental to achieve meaningful and physiological estimates.

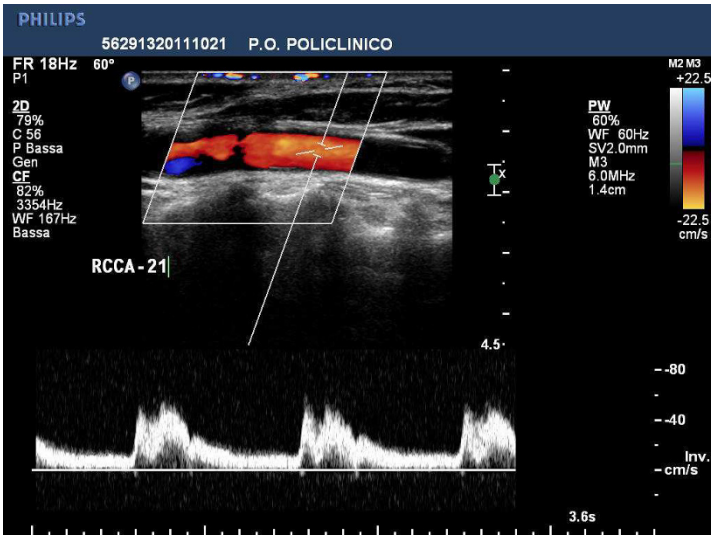


Fig. 6.1: ECD image corresponding to the central point of the carotid section located 2 cm before the carotid bifurcation.

References

[1] Aguilera-Morillo, M. C., Durbán, M., Aguilera, A. M.: Prediction of functional data with spatial dependence: a penalized approach. Stochastic Environ Res Risk Assess **31(1)**: 7–22 (2017)

[2] Augustin, N. H., Trenkel, V. M., Wood, S. N., Lorance, P.: Space-time modeling of blue ling for fisheries stock management. Environmetrics **24(2)**: 109–119 (2013)

[3] Azzimonti, L., Nobile, F., Sangalli, L. M., Secchi, P.: Mixed Finite Elements for Spatial Regression with PDE Penalization. SIAM/ASA Journal on Uncertainty Quantification **2(1)**: 305–335 (2014)

[4] Azzimonti, L., Sangalli, L. M., Secchi, P., Domanin, M., Nobile, F.: Blood flow velocity field estimation via spatial regression with PDE penalization. Journal of the American Statistical Association **110(511)**: 1057–1071 (2015)

[5] Bernardi, M. S., Sangalli, L. M., Mazza, G., Ramsay, J. O.: A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province. Stochastic Environmental Research and Risk Assessment, **31(1)**: 23–38, (2017)

[6] Caballero, W., Giraldo, R., Mateu, J.: A universal kriging approach for spatial functional data. Stochastic Environ Res Risk Assess **27(7)**: 1553–1563 (2013)

[7] Delicado, P., Giraldo, R., Comas, C., Mateu, J.: Statistics for spatial functional data: some recent contributions. Environmetrics **21(34)**: 224–239 (2010)

[8] Giraldo, R., Delicado, P., Mateu, J.: Ordinary kriging for functionvalued spatial data. Environ Ecol Stat **18(3)**: 411–426 (2011)
[9] Goulard, M., Voltz, M.: Geostatistical interpolation of curves: a case study in soil science. Geostatistics Troía'92, Springer: 805–816 (1993)
[10] Ignaccolo, R., Mateu, J., Giraldo, R.: Kriging with external drift for functional data for air quality monitoring. Stochastic Environ Res Risk Assess **28(5)**: 1171–1186 (2014)
[11] Marra, G., Miller, D. L., Zanin, L.: Modelling the spatiotemporal distribution of the incidence of resident foreign population. Statistica Neerlandica **66(2)**: 133–160 (2012)
[12] Mateu, J., Romano, E.: Advances in spatial functional statistics. Stochastic Environ Res Risk Assess **31(1)**: 1–6 (2017)
[13] Menafoglio, A., Secchi, P., Dalla Rosa, M.: A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space. ElectronJ Stat **7**: 2209–2240 (2013)
[14] Menafoglio, A., Guadagnini, A., Secchi, P.: A kriging approach based on Aitchison geometry for the characterization of particlesize curves in heterogeneous aquifers. Stochastic Environ Res Risk Assess **28(7)**: 1835-1851 (2014)
[15] Nerini, D., Monestiez, P., Manté, C.: Cokriging for spatial functional data. J Multivar Anal **101(2)**: 409–418 (2010)
[16] Ramsay, T.: Spline smoothing over difficult regions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **54(2)**: 307–319 (2002)
[17] Sangalli, L. M., Ramsay, J. O., Ramsay, T. O.: Spatial spline regression models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **75(4)**: 681–703 (2013)

Chapter 7
An asymptotic factorization of the Small–Ball Probability: theory and estimates

Jean-Baptiste Aubin, Enea G. Bongiorno and Aldo Goia

Abstract This work reviews recent results on an asymptotic factorization of the Small–Ball Probability of a $\mathscr{L}^2_{[0,1]}$–valued process, as the radius of the ball tends to zero. This factorization involves a volumetric term, a pseudo–density for the probability law of the process, and a correction factor. Estimators of the latter two factors are introduced and some of their theoretical properties considered.

7.1 Introduction

Since the seminal works of [9, 14], functional data analysis continues to massively attract the attention of researchers as proven by the recent monograph [5], special issues [8, 11] and activities [4] on the topic. In this framework, the Small–Ball Probality (SmBP) theory has been playing (and still now plays) an important role. It refers to the study of the asymptotic behaviour of $\mathbb{P}(X \in B(x,\varepsilon))$ as ε vanishes, where X is a random element taking its values in some topological space and $B(x,\varepsilon)$ denotes a suitable ball in such topology. From a theoretical point of view, researchers have mainly focused on different Gaussian processes and in providing the convergence rate (refer to the small tails probability theory; see [12, 13] and references therein). In functional regression, SmBP is a technical instrument used to express the convergence rate of estimators (see [9]). Recently, in the context of $\mathscr{L}^2_{[0,1]}$–valued random elements,

Jean-Baptiste Aubin
INSA-Lyon, ICJ, 20, Rue Albert Einstein, 69621 Villeurbanne Cedex, France, e-mail: jean-baptiste.aubin@insa-lyon.fr

Enea G. Bongiorno
Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, Via Perrone, 18, 28100, Novara, Italy e-mail: enea.bongiorno@uniupo.it

Aldo Goia (✉)
Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, Via Perrone, 18, 28100, Novara, Italy e-mail: aldo.goia@uniupo.it

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_7

the SmBP has been used to derive a concept of surrogate density and to introduce some non–parametric estimators for it (see [3, 6]). In particular, in [3] it has been shown that, for a fixed number d and as the radius ε of the ball tends to zero, the SmBP is asymptotically proportional to (a) the joint density of the first d principal components (PCs) evaluated at the center of the ball, (b) the volume of the d–dimensional ball with radius ε, and (c) a correction factor weighting the use of a truncated version of the process expansion. Under suitable assumptions on the decay rate of the eigenvalues of the covariance operator of the process, it has been shown that the correction factor in (c) tends to 1 as the number of considered dimension increases (see [3]). This fact provides a clear advantage in modelling the SmBP since justifies the use of the lonely term (a) as a surrogate density of the process.

In this work, after recalling in Section 7.2 the theoretical conditions that lead to the mentioned factorization, we illustrate in Section 7.3 how to estimate the terms (a) and (c) providing asymptotic properties. The model advantages and potential applications are discussed in the last Section 7.4.

7.2 Framework and Notations

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\mathscr{L}^2_{[0,1]}$ be the Hilbert space of square integrable real functions on $[0,1]$ endowed with the inner product $\langle g,h\rangle = \int_0^1 g(t)\,h(t)\,dt$ and the induced norm $\|g\|^2 = \langle g,g\rangle$. A Random Curve (RC) X is a measurable map defined on $(\Omega, \mathscr{F})$ taking values in $(\mathscr{L}^2_{[0,1]}, \mathscr{B})$, where $\mathscr{B}$ denotes the Borel sigma–algebra induced by $\|\cdot\|$. Suppose $\mathbb{E}\|X\|^2 < +\infty$ and denote by

$$\mu_X = \{\mathbb{E}[X(t)], t \in [0,1]\}, \quad \text{and} \quad \Sigma[\cdot] = \mathbb{E}[\langle X-\mu_X, \cdot\rangle (X-\mu_X)]$$

its mean function and covariance operator respectively. Consider the Karhunen–Loève expansion of X: denoting by $\{\lambda_j, \xi_j\}_{j=1}^{\infty}$ the decreasing to zero sequence of positive eigenvalues and their associated orthonormal eigenfunctions of Σ, it holds

$$X(t) = \mu_X(t) + \sum_{j=1}^{\infty} \theta_j \xi_j(t), \qquad 0 \le t \le 1, \tag{7.1}$$

where $\theta_j = \langle X-\mu_X, \xi_j\rangle$ are the so–called principal components (PCs) of X satisfying

$$\mathbb{E}[\theta_j] = 0, \qquad Var(\theta_j) = \lambda_j, \qquad \mathbb{E}[\theta_j \theta_{j'}] = 0, \qquad j \neq j'.$$

From now on and without loss of generality, suppose that $\mu_X = 0$. Moreover, assume that

(A-1) the first d PCs $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)'$ admit a strictly positive joint probability density f_d;

(A-2) there exists a strictly positive constant C such that $x_j^2 \leq C\lambda_j$ for any $j \geq 1$, with $x_j = \langle x, \xi_j \rangle$;

(A-3) f_d is sufficiently smooth (differentiable p times) and there exists a strictly positive constant C for which, for any $d \in \mathbb{N}$

$$\sup_{i,j \in \{1,\ldots,d\}} \sqrt{\lambda_i \lambda_j} \left| \frac{\partial^2 f_d(\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} \right| \leq C f_d(x_1, \ldots, x_d), \qquad \text{for any } \boldsymbol{\vartheta} \in D,$$

where $D = \left\{ \boldsymbol{\vartheta} \in \mathbb{R}^d : \sum_{j \leq d} (\vartheta_j - x_j)^2 \leq \rho^2 \right\}$ for some $\rho \geq \varepsilon$;

Now, consider the small ball probability of the process X defined by

$$\varphi(x, \varepsilon) = \mathbb{P}(\|X - x\| < \varepsilon), \qquad \text{for } \varepsilon > 0.$$

In [3], authors have proven that, for a given $d \in \mathbb{N}$ and under assumptions (A-1),…, (A-3),

$$\varphi(x, \varepsilon) \sim f_d(x_1, \ldots, x_d) \frac{\varepsilon^d \pi^{d/2}}{\Gamma(d/2+1)} C(x, \varepsilon, d), \qquad \text{as } \varepsilon \to 0, \tag{7.2}$$

where

$$C(x, \varepsilon, d) = \mathbb{E}\left[(1 - S_d)^{d/2} \, \mathbb{I}_{\{S_d \leq 1\}} \right],$$
$$S_d = S(x, \varepsilon, d) = \frac{1}{\varepsilon^2} \sum_{j \geq d+1} (\theta_j - x_j)^2,$$

and $\mathbb{I}_A$ is the indicator function of the event A. Roughly speaking, (7.2) means that, for a given positive integer d and as $\varepsilon \to 0$, the SmBP $\varphi(x, \varepsilon)$ behaves as the usual first order approximation of the SmBP in a d–dimensional space (i.e. the probability density function of the first d PCs evaluated at $(x_1, \ldots, x_d)$ times the volume of the d–dimensional ball of radius ε) up to the scale factor $C(x, \varepsilon, d)$ that balances the use of a truncated version of the process expansion (7.1).

To fully split the dependence on x and ε in factorization (7.2), the following assumption can be considered:

(A-4) The eigenvalues $\{\lambda_j\}_{j \in \mathbb{N}}$ decay hyper–exponentially, that is $d \sum_{j \geq d+1} \lambda_j = o(\lambda_d)$, as $d \to \infty$.

Under (A-1),…,(A-4), it can be proven that, as d tends to infinity and for a suitable choice of $\varepsilon = \varepsilon(d)$,

$$\begin{cases} C(x, \varepsilon, d) \to 1, \\ \varphi(x, \varepsilon) \sim f_d(x_1, \ldots, x_d) \frac{\varepsilon^d \pi^{d/2}}{\Gamma(d/2+1)}, \end{cases} \tag{7.3}$$

see [3]. A practical choice for ε is

$$\varepsilon^2(d) = \sqrt{d\lambda_d \sum_{j=d+1}^{\infty} \lambda_j}. \tag{7.4}$$

To take advantage of the above factorizations, in the next section some estimators for f_d and C are introduced and their basic properties stated. Further discussions and applications are discussed in the last section.

7.3 Estimates

Consider $(X_1,\dots,X_n)$ a sample of RCs distributed as X, $\overline{X}_n$ and $\widehat{\Sigma}_n$ the empirical versions of μ_X and Σ from which it is possible to estimate the empirical eigensystem $\{\widehat{\lambda}_j,\widehat{\xi}_j\}_{j\in\mathbb{N}}$ and $\{\widehat{\theta}_{i,j}=\langle X_i,\xi_j\rangle\}_{j\in\mathbb{N}}$ the estimated scores associated to X_i for any $i=1,\dots,n$. It is known that such estimators are consistent; see, for instance, [5].

For what concerns the surrogate density f_d, for a fixed d, let us introduce the kernel density estimate:

$$\widehat{f}_{d,n}\left(\widehat{\Pi}_d x\right) = \widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{H_n}\left(\left\|\widehat{\Pi}_d(X_i - x)\right\|\right) \tag{7.5}$$

where $K_{H_n}(\mathbf{u}) = \det(H_n)^{-1/2} K(H_n^{-1/2}\mathbf{u})$, K is a kernel function, H_n is a symmetric semi-definite positive $d\times d$ matrix and $\widehat{\Pi}_d$ denotes the projector onto the d–dimensional space spanned by $\{\widehat{\xi}_j\}_{j=1}^{d}$. Under regularity assumptions on f_d and on the kernel K, if one takes $H_n = h_n I$ with $h_n \to 0$ and $nh_n^d/\log n \to \infty$ as $n\to\infty$, the following result has been proven in [3].

Proposition 7.1. *Take the optimal bandwidth $c_1 n^{-1/(2p+d)} \le h_n \le c_2 n^{-1/(2p+d)}$ and $p > \max\{2, 3d/2\}$. Thus*

$$\mathbb{E}[(f_d(x) - \widehat{f}_n(x))^2] = O\left(n^{-2p/(2p+d)}\right),$$

as n goes to infinity and uniformly in $\mathbb{R}^d$.

Regarding the corrective factor $C(x,\varepsilon,d)$ an estimator is provided by the empirical one:

$$\widehat{C}_{n,d} = \widehat{C}_n(x,\widehat{\varepsilon},d) = \frac{1}{n}\sum_{i=1}^{n}\left(1-\widehat{S}_i(x,\widehat{\varepsilon},d)\right)^{d/2} \mathbb{1}_{\left\{\widehat{S}_i(x,\widehat{\varepsilon},d)\le 1\right\}},$$

with $\widehat{S}_i(x,\widehat{\varepsilon},d) = \widehat{\varepsilon}^{-2}\sum_{j\ge d+1}\left(\widehat{\theta}_{i,j} - \widehat{x}_j\right)^2$, $\widehat{\theta}_{i,j} = \langle X_i, \widehat{\xi}_j\rangle$, $\widehat{x}_j = \langle x, \widehat{\xi}_j\rangle$ and where $\widehat{\varepsilon}$ is the empirical version of (7.4). Asymptotics on such estimator have been provided in [1] and collected in the following proposition.

Proposition 7.2. *As n tends to infinity, $\widehat{\varepsilon}^2$ and $\widehat{C}_{n,d}$ are consistent estimator in the $L^1[\Omega,\mathscr{F},\mathbb{P};\mathbb{R}]$ metric. Moreover, $\sqrt{n}(\widehat{C}_{n,d} - C)$ is asymptotically normal distributed.*

7.4 Conclusions

This work collects some theoretical results concerning the factorization of the SmBP. These clarify those conditions under which it is possible to separate by means of distinct factors the spatial and volumetric components. The asymptotic (7.3) provides a modelling advantage: for d large enough, it justifies the use of factorized version of the SmBP since the corrective factor $C(x, \varepsilon, d)$ will be close to 1.
On the one hand, such approximation yields f_d a surrogate density of the process whose estimation can be tackled in a non–parametric manner (see [3, 6]) or parametrically (see [10] in the Gaussian mixture case). In [2], the estimate (7.5) is the starting point to build a pseudo–density oriented clustering algorithm where clusters are identified by the largest connected upper–surfaces containing only one mode. This technique was applied to different real datasets.
On the other hand, the convergence to 1 of $C(x, \varepsilon, d)$ holds theoretically only for $d \to \infty$. From the practical point of view, when d is fixed and in order to assess the goodness of f_d as a surrogate density, it is useful to evaluate how close to 1 is this correction factor C. This qualitatively suggests the dimension d to be used in practice (see [1]).

Acknowledgements The authors are grateful to R. Ignaccolo for the fruitful discussions they had at different stages of this work. E. Bongiorno and A. Goia are members of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM) that funded a visit by J.B. Aubin at Università del Piemonte Orientale.

References

[1] Aubin J.B., Bongiorno E.G.: Optimal local dimension for suitable Hilbert–valued processes. Preprint (2017)
[2] Bongiorno E.G., Goia A.: Classification methods for Hilbert data based on surrogate density. Comput. Statist. Data Anal. **99**, 204 – 222, (2016)
[3] Bongiorno E.G., Goia A.: Some insights about the small ball probability factorization for Hilbert random elements. Statist. Sinica (To Appear) (2017)
[4] Bongiorno E.G., Goia A., Salinelli E., Vieu P. (eds): Contributions in infinite-dimensional statistics and related topics, Società Editrice Esculapio, (2014)
[5] Bosq D.: Linear processes in function spaces, Lecture Notes in Statistics, vol 149. Springer-Verlag, New York, (2000)
[6] Delaigle A., Hall P.: Defining probability density for a distribution of random functions. Ann. Statist. **38**(2), 1171–1193, (2010)
[9] Ferraty F., Vieu P.: Nonparametric functional data analysis. Springer Series in Statistics, Springer, New York (2006)
[8] Goia A., Vieu P.: An introduction to recent advances in high/infinite dimensional statistics. J. Multivariate Anal. **146**, 1–6, (2016)

[5] Horváth L., Kokoszka P.: Inference for functional data with applications. Springer Series in Statistics, Springer, New York, (2012)
[10] Jacques J., Preda C.: Functional data clustering: a survey. Adv. Data Anal. Classif. **8**(3), 231–255, (2014)
[11] Kokoszka P., Oja H., Park B., Sangalli L.: Special issue on functional data analysis. Econometrics and Statistics, **1**, 99–10, (2017)
[12] Li W.V., Shao Q.M.: Gaussian processes: inequalities, small ball probabilities and applications. In: Stochastic processes: theory and methods, Handbook of Statist., vol 19, North-Holland, Amsterdam, 533–597, (2001)
[13] Lifshits M.A.: Lectures on Gaussian processes. Springer Briefs in Mathematics, Springer, Heidelberg, (2012)
[14] Ramsay J.O., Silverman B.W.: Functional data analysis, 2nd edn. Springer Series in Statistics, Springer, New York (2005)

Chapter 8
Estimating invertible functional time series

Alexander Aue and Johannes Klepsch

Abstract This contribution discusses the estimation of an invertible functional time series through fitting of functional moving average processes. The method uses a functional version of the innovations algorithm and dimension reduction onto a number of principal directions. Several methods are suggested to automate the procedures. Empirical evidence is presented in the form of simulations and an application to traffic data.

8.1 Introduction

Functional time series have come into the center of statistics research at the confluence of functional data analysis and time series analysis. Some of the more and most recent contributions in this area include Aston and Kirch [1, 2] and Aue et al. [7] who dealt with the detection and estimation of structural breaks in functional time series, Chakraborty and Panaretos [9] who covered functional registration and related it to optimal transport problems, Horváth et al. [13] and Aue and van Delft [8] who developed stationarity tests in the time and frequency domain, respectively, Hörmann et al. [12] who introduced methodology for the detection of periodicities, Hörmann et al. [11] and Aue et al. [5] who proposed models for heteroskedastic functional time series, van Delft and Eichler [20] who defined a framework for locally stationary time series, Kowal et al. [16] who developed Bayesian methodology for a functional dynamic linear model, Raña et al. [18, 19] who discussed outlier detection in functional time series and provided methodology for the construction of bootstrap

Alexander Aue (✉)
Department of Statistics, University of California, One Shields Avenue, Davis, CA 95616, USA, e-mail: aaue@ucdavis.edu

Johannes Klepsch
Center for Mathematical Sciences, Technische Universität München, 85748 Garching, Boltzmannstraße 3, Germany, e-mail: j.klepsch@tum.de

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_8

confidence intervals for nonparametric regression under dependence, respectively, and Paparoditis [17] who introduced a sieve bootstrap procedure.

A general framework for functional time series allowing for elegant derivations of large-sample results was put forward in Hörmann and Kokoszka [10]. This paper introduced a concept to measure closeness of functional time series to certain functional moving average processes. It was then exploited that the latter have non-trivial autocovariance operators only for finitely many lags in order to derive large-sample results concerning the validity of functional principal components analysis in a dependent setting and change-point analysis (see [4]), among others. The focus of Aue and Klepsch [6] was not on theoretical properties but on the more practical question of how to estimate an invertible functional time series. This was achieved by functional moving average model fitting. The fitting process involved an application of the functional innovations algorithm, whose population properties were derived in Klepsch and Klüppelberg [14]. This algorithm can be used to estimate the operators in the causal representation of a functional time series. The consistency of these estimates is the main result. For practical purposes, the proposed method requires the selection of the dimension reduction space through both model selection and testing approaches. Several methods are proposed and then evaluated in a simulation study and in an application to vehicle traffic data.

The remainder is organized as follows. Section 8.2 introduces the setting and the method for estimating functional moving average processes. Several algorithms for practical implementation are discussed in Section 8.3. Section 8.4 gives a glimpse on large-sample theory. Section 8.5 briefly covers empirical aspects.

8.2 Estimation methodology

Let $H = L^2[0,1]$ be the Hilbert space of square-integrable functions on $[0,1]$ equipped with the standard norm $\|\cdot\|$ defined by the inner product $\langle x,y\rangle = \int_0^1 x(s)y(s)ds$, for $x,y \in H$. Let $(\Omega,\mathscr{A},P)$ be a probability space and denote by $L_H^2 = L^2(\Omega,\mathscr{A},P)$ the space of square integrable random functions taking values in H, noting that L_H^2 is a Hilbert space with inner product $E[\langle X,Y\rangle]$, for $X,Y \in L_H^2$. A functional linear process with values in L_H^2 is given by the series expansion

$$X_j = \sum_{\ell=0}^{\infty} \psi_\ell \varepsilon_{j-\ell}, \qquad j \in \mathbb{Z}, \tag{8.1}$$

where $(\psi_\ell \colon \ell \in \mathbb{N}_0)$ is a sequence in the space of bounded linear operators acting on H and $(\varepsilon_j \colon j \in \mathbb{Z})$ a sequence of independent, identically distributed random functions in L_H^2. A functional time series $(X_j \colon j \in \mathbb{Z})$ is invertible if it admits the series expansion

$$X_j = \sum_{\ell=1}^{\infty} \pi_\ell X_{j-\ell} + \varepsilon_j, \qquad j \in \mathbb{Z}, \tag{8.2}$$

where $(\pi_\ell\colon \ell \in \mathbb{N})$ is a sequence of bounded linear operators such that $\sum_{\ell=1}^{\infty} \|\pi_\ell\|_{\mathscr{L}} < \infty$, with $\|A\|_{\mathscr{L}} = \sup_{\|x\|\le 1} \|Ax\|$ for A bounded and linear.

If $X \in L_H^2$ with $E[X] = 0$, then its covariance operator exists and admits a spectral representation; that is,

$$C_X(y) = E[\langle X, y\rangle X] = \sum_{i=1}^{\infty} \lambda_i \langle y, \nu_i\rangle \nu_i, \qquad y \in H, \tag{8.3}$$

where $(\lambda_i\colon i \in \mathbb{N})$ and $(\nu_i\colon i \in \mathbb{N})$ denote the eigenvalues and eigenfunctions, respectively. If $X, Y \in L_H^2$ with $E[X], E[Y] = 0$, the cross covariance operator exists and is given by

$$C_{XY}(y) = E[\langle X, y\rangle Y], \qquad y \in H. \tag{8.4}$$

Introducing $x \otimes y(\cdot) = \langle x, \cdot\rangle y$ for $x, y \in H$, the lag-h autocovariance operator of a stationary functional time series $(X_j\colon j \in \mathbb{Z})$ may be written as $C_{X;h} = E[X_0 \otimes X_h]$, for $h \in \mathbb{Z}$.

Let $\mathscr{V}_d = \overline{\text{sp}}\{\nu_1, \ldots, \nu_d\}$ be the subspace generated by the first d principal directions and let $P_{\mathscr{V}_d}$ be the projection operator from H to $\mathscr{V}_d$. For an increasing sequence $(d_i\colon i \in \mathbb{N}) \subset \mathbb{N}$ define $X_{d,j} = P_{\mathscr{V}_{d_i}} X_j$, $j \in \mathbb{Z}$, $i \in \mathbb{N}$ and denote by $\tilde{\mathscr{F}}_{n,k}$ the smallest subspace containing $X_{d_k,n}, \ldots, X_{d_1,n-k}$ that is closed with respect to bounded, linear operators. The best linear predictor of X_{n+1} given $\tilde{\mathscr{F}}_{n,k}$ is then

$$\tilde{X}_{n+1,k} = P_{\tilde{\mathscr{F}}_{n,k}}(X_{n+1}) = \sum_{i=1}^{k} \theta_{k,i}(X_{d_{k+1-i},n+1-i} - \tilde{X}_{n+1-i,k-i}), \tag{8.5}$$

where $\tilde{X}_{n-k,0} = 0$. This is the form required for the innovations algorithm developed in Klepsch and Klüppelberg [14] that provides a solution to recursively compute the coefficients $\theta_{k,1}, \ldots, \theta_{k,k}$ on the population level. However, for practical purposes, an estimated version of (8.5) is needed. Define

$$\hat{\mathscr{V}}_{d_i} = \overline{\text{sp}}\{\hat{\nu}_1, \ldots, \hat{\nu}_{d_i}\} \qquad \text{and} \qquad \hat{P}_{(k)} = \text{diag}(P_{\hat{\mathscr{V}}_{d_k}}, \ldots, P_{\hat{\mathscr{V}}_{d_1}}),$$

where $\hat{\nu}_i$ denotes the ith eigenfunction of the sample covariance operator $\hat{C}_X$. The best linear predictor can now be computed with the functional innovations algorithm given in Section 8.3. The large-sample behavior is presented in Section 8.4.

8.3 Algorithms

The following algorithm details how an estimate of the best linear predictor may be computed by recursion. Denote by A^* the adjoint of an operator A.

Functional innovations algorithm. The best linear predictor $\tilde{X}_{n+1,k}$ of X_{n+1} given $\tilde{\mathscr{F}}_{n,k}$ can be computed by the recursions

$$\tilde{X}_{n-k,0} = 0 \qquad \text{and} \qquad \hat{V}_1 = P_{\hat{\mathscr{V}}_{d_1}} \hat{C}_X P_{\hat{\mathscr{V}}_{d_1}},$$

$$\tilde{X}_{n+1,k} = \sum_{i=1}^{k} \hat{\theta}_{k,i}(X_{d_{k+1-i},n+1-i} - \tilde{X}_{n+1-i}^{k-i}),$$

$$\hat{\theta}_{k,k-i} = \left(P_{\hat{\mathscr{V}}_{d_{k+1}}} \hat{C}_{X;k-i} P_{\hat{\mathscr{V}}_{d_{i+1}}} - \sum_{j=0}^{i-1} \hat{\theta}_{k,k-j} \hat{V}_j \hat{\theta}_{i,i-j}^* \right) \hat{V}_i^{-1}, \qquad i = 1, \ldots, k-1,$$

$$\hat{V}_k = \hat{C}_{X_{d_{k+1}} - \tilde{X}_{n+1,k}} = \hat{C}_{X_{d_{k+1}}} - \sum_{i=0}^{k-1} \hat{\theta}_{k,k-i} \hat{V}_i \hat{\theta}_{k,k-i}^*. \tag{8.6}$$

Application of the algorithm requires the selection of the d_i and and also the FMA order q. The selection of the former can be achieved through the following portmanteau test for independence. Here all d_i are set to the same value.

Determining the principal subspace by testing for independence.

(1) Given functions $X_1, \ldots, X_n$, estimate $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$ and $\hat{\nu}_1, \ldots, \hat{\nu}_n$. Select d^* such that

$$\mathrm{TVE}(d^*) = \frac{\sum_{i=1}^{d^*} \hat{\lambda}_i}{\sum_{i=1}^{n} \hat{\lambda}_i} \geq P$$

for some prespecified $P \in (0,1)$.

(2) Let $f_h(\ell,\ell')$ and $b_h(\ell,\ell')$ denote the (ℓ,ℓ')th entries of $C_{\mathbf{X}^*;0}^{-1} C_{\mathbf{X}^*;h}$ and $C_{\mathbf{X}^*;h} C_{\mathbf{X}^*;0}^{-1}$, respectively, and $(\mathbf{X}_j^* \colon j \in \mathbb{Z})$ the process consisting of the $d+1$st to $d+l$th principal directions of C_X. If

$$Q_n^{d^*} = n \sum_{h=1}^{\bar{h}} \sum_{\ell,\ell'=d^*+1}^{d^*+l} f_h(\ell,\ell') b_h(\ell,\ell') > q_{\chi^2_{d^{*2}\bar{h}}},$$

set $d^* = d^* + 1$.

(3) If $Q_n^{d^*} \leq q_{\chi^2_{d^{*2}\bar{h}}}$, apply the functional innovations algorithm with $d_i = d^*$.

Once d is selected, the order of the resulting VMA process can be determined with a Ljung–Box test or an AICC criterion. Both are described next.

(I) Order selection with Ljung–Box test.

(1) Test the null hypothesis $H_0 \colon C_{\mathbf{X};h} = 0$ for all $h \in [\underline{h}, \bar{h}]$ with the test statistic

$$Q_{\underline{h},\bar{h}} = n^2 \sum_{h=\underline{h}}^{\bar{h}} \frac{1}{n-h} \mathrm{tr}\big(\hat{C}_{\mathbf{X};h}^{\top} \hat{C}_{\mathbf{X};0}^{-1} \hat{C}_{\mathbf{X};h} \hat{C}_{\mathbf{X};0}^{-1}\big),$$

which is asymptotically $\chi^2_{d^2(\bar{h}-\underline{h}-1)}$-distributed.

(2) Iteratively compute $Q_{1,\bar{h}}, Q_{2,\bar{h}}, \ldots$ and select q as the largest $\underline{h}$ such that $Q_{\underline{h},\bar{h}}$ is significant but $Q_{\underline{h}+h,\bar{h}}$ is insignificant for all h.

(II) Order selection with AICC criterion.

(1) Choose the order of the FMA process as the minimizer of

$$\mathrm{AICC}(q) = -2\ln L(\Theta_1,\dots,\Theta_q,\Sigma) + \frac{2nd(qd^2+1)}{nd-qd^2-2}$$

where $\Theta_1,\dots,\Theta_1$ are the VMA matrices and Σ the covariance matrix.

Another option is provided by the following FPE-type criterion that selects d and q jointly. Is is an adaptation of a similar criterion for FAR processes put forward in Aue et al. [3].

Determination of principal subspace and order selection with FPE criterion.

(1) Select (d,q) as minimizer of

$$\mathrm{fFPE}(d,q) = \frac{n+qd}{n}\mathrm{tr}(\hat{\mathbf{V}}_n) + \sum_{i>d}\hat{\lambda}_i,$$

where $\hat{\mathbf{V}}_n$ is the matrix version of $\hat{V}_n$ in (8.6).

8.4 Large-sample properties

The following theorem states that $\hat{\Gamma}_k$ is a consistent estimator for Γ_k.

Theorem 8.1. *If $(X_j\colon j\in\mathbb{Z})$ defined in (8.1) is such that $\sum_{m=1}^{\infty}\sum_{\ell=m}^{\infty}\|\psi_\ell\|_{\mathscr{L}}<\infty$ and $E[\|\varepsilon_0\|^4]<\infty$, then $(n-k)E[\|\hat{\Gamma}_k-\Gamma_k\|_{\mathscr{N}}^2]\le kU_X$, where $\|\cdot\|_{\mathscr{N}}$ denotes nuclear norm and U_X a constant that does not depend on n.*

To discuss the consistency of the estimators in the causal and invertible representations, further conditions are needed. As $n\to\infty$, let $k=k_n\to\infty$ and $d_k\to\infty$ such that

$$\begin{aligned} k^{1/2}(n-k)^{-1/2}\alpha_{d_k}^{-2} &\to 0,\\ k^{1/2}\alpha_{d_k}^{-1}\left(\sum_{\ell>k}\|\pi_\ell\|_{\mathscr{L}} + \sum_{\ell=1}^{k}\|\pi_\ell\|_{\mathscr{L}}\sum_{i>d_{k+1-\ell}}\lambda_i\right) &\to 0,\\ k^{3/2}\alpha_{d_k}^{-2}n^{-1}\left(\sum_{\ell=1}^{d_k}\delta_\ell^{-2}\right)^{1/2} &\to 0, \end{aligned} \tag{8.7}$$

where α_{d_k} is related to the spectral gaps of C_X and α_{d_n} is the infimum of the eigenvalues of the spectral density operator of $((\langle X_n,\nu_1\rangle,\dots,\langle X_n,\nu_{d_n}\rangle)^T : n\in\mathbb{N})$.

Theorem 8.2. *Under the assumptions of Theorem 8.1 and the above conditions, for all $x\in H$ and $i\in\mathbb{N}$ as $n\to\infty$, $\|(\hat{\beta}_{k,i}-\pi_i)(x)\|\xrightarrow{p}0$, and $\|(\hat{\theta}_{k,i}-\psi_i)(x)\|\xrightarrow{p}0$. If the*

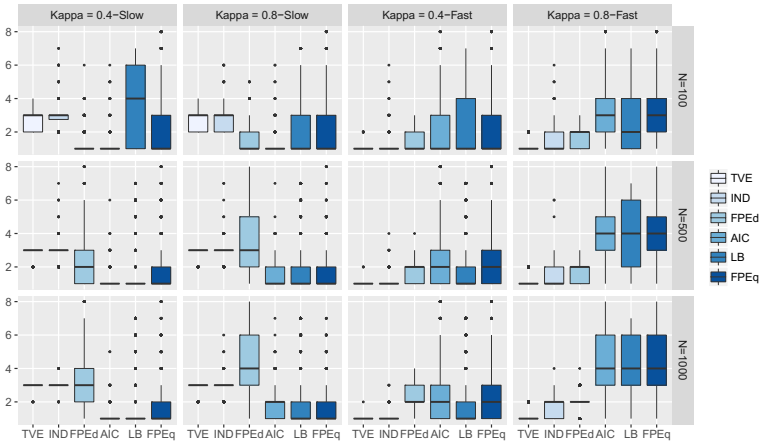


Fig. 8.1: Model selection for different FAR(1) processes. The left three plots in each small figure give the d chosen by total variation explained with $P = 0.8$ (TVE), the test for independence (IND) and the functional FPE criterion (FPEd). The right three plots in each small figure give the selected order q by AICC, Ljung–Box and fFPE.

operators $(\psi_\ell \colon \ell \in \mathbb{N})$ *and* $(\pi_\ell \colon \ell \in \mathbb{N})$ *are Hilbert–Schmidt, then the convergence is uniform.*

Detailed proofs of both theorems may be found in Aue and Klepsch [6].

8.5 Empirical results

As an illustration of the proposed fitting method, a simulation is provided in which an FAR(1) process is approximated through an FMA(q) process, where the process is generated as outlined in Aue et al. [3], choosing various norms κ of the FAR operator and fast and slow decays of eigenvalues of the covariance operator. Model selection results for the different methods are provided in Figure 8.1, noting that FMA models are fit to an FAR time series.

Figure 8.2 displays 1440 curves of average velocity per minute obtained at a fixed measurement station on A92 Autobahn in southern Germany. Klepsch et al. [15] indicate that this functional time series is stationary and that an FMA fit may be appropriate. Applying the functional innovations algorithm together with any of the proposed procedures to select d and q leads to a first-order dynamics. Additional information is provided in Aue and Klepsch [6].

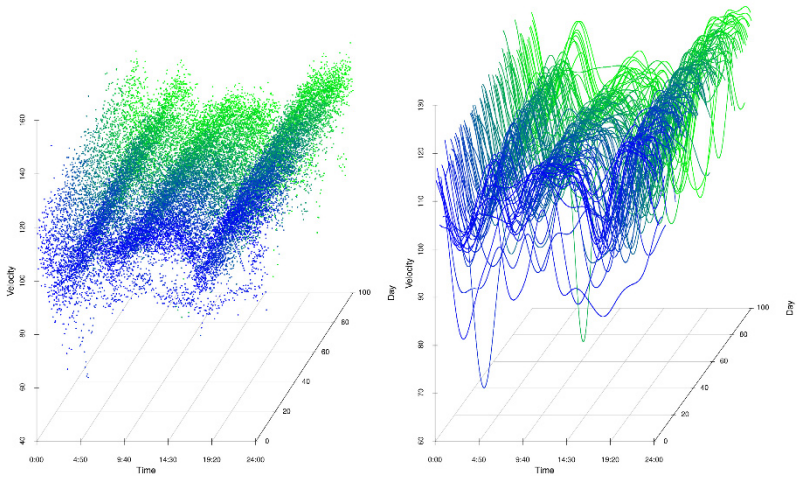


Fig. 8.2: Discrete velocity observations (left) and corresponding velocity functions (right).

Acknowledgements This research was supported by NSF grants DMS 1305858 and DMS 1407530. The velocity data was provided by Autobahndirektion Südbayern.

References

[1] Aston, J.A.D., Kirch, C.: Detecting and estimating changes in dependent functional data. J. Mult. Anal. **109**, 204–220 (2012)
[2] Aston, J.A.D., Kirch, C.: Evaluating stationarity via change-point alternatives with applications to FMRI data. Ann. Appl. Statist. **6**, 1906–1948 (2012)
[3] Aue, A., Dubart Norinho, D., Hörmann, S.: On the prediction of stationary functional time series. J. Am. Stat. Assoc. **110**, 378–392 (2015)
[4] Aue, A., Horváth, L.: Structural breaks in time series. J. Time Ser. Anal. **34**, 1–16 (2013)
[5] Aue, A., Horváth, L., Pellatt, D.: Functional generalized autoregressive conditional heteroskedasticity. J. Time Ser. Anal. **38**, 3–21 (2017)
[6] Aue, A., Klepsch, J.: Estimating functional time series by moving average model fitting. ArXiv preprint (2017). Available online at https://arxiv.org/abs/1701.00770
[7] Aue, A., Rice, G., Sönmez, O.: Detecting and dating structural breaks in functional data without dimension reduction. ArXiv preprint (2017). Available

online at https://arxiv.org/abs/1511.04020
[8] Aue, A., van Delft, A.: Testing for stationarity of functional time series in the frequency domain. ArXiv preprint (2017). Available online at https://arxiv.org/abs/1701.01741.
[9] Chakraborty, A., Panaretos, V.M.: Functional registration and local variations. ArXiv preprint (2017). Available online at https://arxiv.org/abs/1702.03556
[10] Hörmann, S., Kokoszka, P.: Weakly dependent functional data. Ann. Statist. **38**, 1845–1884 (2010)
[11] Hörmann, S., Horváth, L., Reeder, R.: A functional version of the ARCH model. Econometr. Th. **29**, 267–288 (2013)
[12] Hörmann, S., Kokoszka, P., Nisol, G.: Detection of periodicity in functional time series. ArXiv preprint (2016). Available online at https://arxiv.org/abs/1607.02017
[13] Horváth, L., Kokoszka, P., Rice, G.: Testing stationarity of functional time series. J. Econ. **179**, 66–82 (2014)
[14] Klepsch, J., Klüppelberg, C.: An innovations algorithm for the prediction of functional linear processes. J. Mult. Anal. **155**, 252–271 (2017)
[15] Klepsch, J., Klüppelberg, C., Wei, T.: Prediction of functional ARMA processes with an application to traffic data. Econ. Statist. **1**, 128–149 (2017)
[16] Kowal, D.R, Matteson, D.S., Ruppert, D.: A Bayesian multivariate functional dynamic linear model. J. Am. Statist. Assoc. (2016) doi: 10.1080/01621459.2016.1165104
[17] Paparoditis, E.: Sieve bootstrap for functional time series. ArXiv preprint (2016). Available online at https://arxiv.org/abs/1609.06029
[18] Raña, P., Aneiros, G., Vilar, J.M.: Detection of outliers in functional time series. Environm. **26**, 178–191 (2015)
[19] Raña, P., Aneiros, G., Vilar, J.M., Vieu, P.: Bootstrap confidence intervals for nonparametric regression under dependence. Electr. J. Statist. **10**, 1973–1999 (2016)
[20] Van Delft, A., Eichler, M.: Locally stationary functional time series. ArXiv preprint (2016). Available online at https://arxiv.org/abs/1602.05125

Chapter 9
On the Geometric Brownian Motion assumption for financial time series

Enea G. Bongiorno, Aldo Goia and Philippe Vieu

Abstract The Geometric Brownian Motion type process is commonly used to describe stock price movements and is basic for many option pricing models. In this paper a new methodology for recognizing Brownian functionals is applied to financial datasets in order to evaluate the compatibility between real financial data and the above modeling assumption. The method rests on using the volumetric term which appears in the factorization of the small–ball probability of a random curve.

9.1 Introduction

Modeling stock prices represents an important task in finance since this is the starting point for evaluating derivatives and other contracts, which have these prices as underlying. The most famous approach, dating back to Black and Scholes [1], states that the dynamic of prices behaves as a Geometric Brownian motion (GBM), with constant coefficients of drift and volatility. In the time, many variants have succeeded (see [5] for a review): in general, in such literature, it is common to assume that the prices follow a GBM, with drift and volatility which evolve during the time.

The problem to verify the compatibility of observed data with the GBM assumption is still an open problem: only indirect empirical evidences have been provided to

Enea G. Bongiorno
Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, e-mail: enea.bongiorno@uniupo.it

Aldo Goia (✉)
Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, e-mail: aldo.goia@uniupo.it

Philippe Vieu
Institut de Mathématiques de Toulouse, Université Paul Sabatier, e-mail: philippe.vieu@math.univ-toulouse.fr

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_9

support that modeling (for instance, by testing marginal Gaussianity, serial correlation of increments, and so on; see among many others, [7] and [8]).

In [2] a new approach to explore the nature of functional data has been introduced and discussed: starting from the possibility to factorize the small–ball probability of a random curve in a spatial factor and a volumetric one, the authors use this latter as the leading term to characterize the nature of the underlying process which has generated the observed curves.

The aim of this paper is to apply such new methodology to financial time series in order to verify the compatibility with the GBM assumption: after introducing in Section 9.2 the notation and summarize some important steps of the methodology introduced in [2], in Section 9.3 an application to real datasets is provided and the main results are illustrated.

9.2 Recognizing some Brownian functionals

Consider a random element X defined on a suitable probability space and mapping in $\mathscr{L}^2_{[0,1]}$, the space of square integrable functions on $[0,1]$, equipped with its natural inner product $\langle g,h\rangle = \int_0^1 g(t)\,h(t)\,dt$, and the induced norm $\|g\|^2 = \langle g,g\rangle$, $g,h \in \mathscr{L}^2_{[0,1]}$. In order to characterize the probability distribution of X, it is useful to known the behaviour of the small-ball probability of X, that is $P(\|X-x\| < \varepsilon)$, as ε tends to zero. The results on this topic available in the literature concern essentially some special classes of Gaussian processes and are presented in the form

$$\mathbb{P}(\|X-x\| < \varepsilon) \sim \psi(x)\,\phi(\varepsilon) \qquad \text{as } \varepsilon \to 0,$$

where $\psi(x)$ is a positive constant depending on x, which plays the role of the surrogate density of X, and $\phi(\varepsilon)$ representing the volumetric term independent on x. For some processes, the latter can be asymptotically approximated by $\varepsilon^{\alpha}\exp\left(-\gamma\varepsilon^{-\beta}\right)$, with α, β and γ non–negative constants; in particular, it is known that when X is a Brownian bridge process, $\alpha = 0$, $\gamma = 1/8$ and $\beta = 2$ (see [6]).

Suppose now to dispose of a sample $\{X_i, i = 1,\ldots,n\}$ of i.i.d. copies of X, from which one can obtain an estimate $\widehat{\phi}(\varepsilon)$ of $\phi(\varepsilon)$. The comparison of $\widehat{\phi}(\varepsilon)$ and $\phi(\varepsilon)$ by a suitable dissimilarity measure allows to evaluate the parameters involved, as the ones which minimize that dissimilarity. This idea has been developed in [2] where the goodness of the approach is shown by a simulations study.

In particular, if one assumes that the sample comes from a Brownian bridge, once $\widehat{\phi}(\varepsilon)$ is computed for $\varepsilon \in \mathscr{E}$, a suitable subset of $\mathbb{R}^+$, it is possible to estimate β as the minimizer $\widehat{\beta}$ of the *centered cosine dissimilarity (ccd)* between $\log\widehat{\phi}(\varepsilon)$ and $\varepsilon^{-\beta}$ over $\mathscr{E}$, where the ccd between g and h is defined by $1 - \langle g^\star, h^\star\rangle\left(\|g^\star\|\,\|h^\star\|\right)^{-1}$, with $g^\star = g - \int g(t)\,dt$ and all integrals are computed on $\mathscr{E}$. If $\widehat{\beta}$ is close to 2, then this represents an empirical evidence in favour of the correctness of assumption that the process is a Brownian bridge, whereas if $\widehat{\beta}$ is far from 2, there is not compatibility

with that specification. To have an idea about the variability of $\widehat{\beta}$, one can refer to confidence intervals estimated, by means of Monte Carlo approach, in [2] for various sample sizes.

The latter approach can be extended directly to other Gaussian processes which are functionals of the Browian bridge (as, for instance, the Wiener process and the Geometric Brownian Motion): in fact, it is enough to apply a suitable tranformation to the data to obtain a Brownian bridge.

9.3 Analysis of financial time series

In this Section we apply the methodology illustrated above to some time series of stock prices, in order to evaluate the compatibility of the data with the assumption that the ones come from a Geometric Brownian Motion.

To do this, we concentrate our attention on four well known financial indexes, used as underlying for a lot of derivatives, futures and other contracts: Dow Jones Industrial Average, NASDAQ composite, NIKKEI 225 and S&P 500.

Time series consist in daily closing prices from 12 March 1985 to 1 December 2016 for Dow Jones, from 5 June 1984 to 1 December 2016 for NIKKEI index, and from 13 April 1977 to 1 December 2016 for NASDAQ and S&P 500 indexes. Hence, overall we have 8 thousand daily prices for Dow Jones and NIKKEI indexes, and 10 thousand for NASDAQ and S&P 500. The plots of these time series are reported in Figure 9.1.

9.3.1 Modeling

Denote by $S(t)$ the price of a stock observed at time t. From the time series $\left\{S(t_j), j = 1,\ldots,N\right\}$ it is possible to build a sample of n discretized functional data X_i by dividing the interval $\mathscr{T} = [t_1, t_N]$ in n disjoint intervals $\mathscr{T}_i$ with constant width τ (positive and integer, so that $N = n\tau$) and cutting the whole trajectory as follows:

$$X_i(t_j) = S((i-1)\tau + t_j) \qquad t_j \in [0,\tau),\ i = 1,\ldots,n.$$

Accordingly with the financial literature (see e.g. [3], [5]) we assume that the underlying continuous process, from which data come, follows the GBM model:

$$X_i(t) = X_i(0)\exp\left\{\left(\mu_i - \frac{1}{2}\sigma_i^2\right)t + \sigma_i W(t)\right\} \qquad t \in [0,\tau)$$

where μ_i and σ_i are the specific drift term and the specific volatility rate of the period $\mathscr{T}_i$ and $W(t)$ is a Wiener process. In this way, we take into account the fact that volatility rate vary with time over $\mathscr{T}$, but can be considered constant over suitable subintervals.

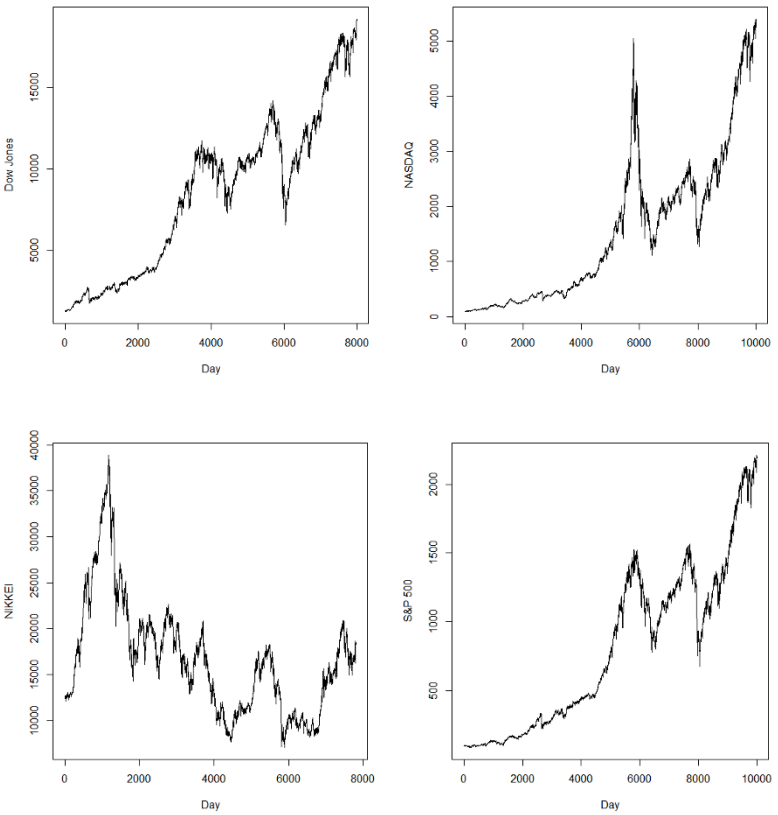


Fig. 9.1: Time series of daily prices of Dow Jones (left, top), NASDAQ (right, top), NIKKEI (left, bottom) and S&P 500 (rigth, bottom) indexes.

Since

$$W(t) = \left[\log\left(X_i(t)/X_i(0)\right) - \left(\mu_i - \sigma_i^2/2\right)t\right]/\sigma_i \tag{9.1}$$

and $W(t) - tW(\tau)$ is a Brownian Bridge, the methodology illustrated in Section 9.2 applies to verify the compatibility of data with the assumption.

9.3.2 Estimates and main results

The first step to operationalize the methodology is to cut the time series in order to obtain the samples of functional data: we decided to divide the whole intervals in subintervals of d days each one, with $d = 25, 50, 80, 100$ in order to evaluate the

Table 9.1: Estimated values of β for the the different stock prices indexes varying d. Into brackets the sample size.

d	Dow Jones	NASDAQ	NIKKEI	S&P 500
100	2.18 [159]	1.62 [199]	2.16 [159]	2.14 [199]
80	2.24 [199]	1.68 [249]	2.24 [199]	2.22 [249]
50	2.20 [160]	1.80 [200]	2.14 [160]	2.24 [200]
25	2.16 [320]	1.88 [400]	2.08 [320]	2.12 [400]

results with respect to the cutting criterion. In this way we should obtain samples of $n = 10000/d$ curves for NASDAQ and S&P 500 and of $n = 8000/d$ curves for Dow Jones and NIKKEI. Since, the larger d is, the smaller n is, for $d = 100, 80$ we add to the sample some curves built with the same cutting criterion but with starting point shifted by $d/2$ ahead. In this way we guarantee more accurate estimations.

For each sample the terms μ_i and σ_i are estimated from each curve by using the maximum likelihood estimates as follows:

$$\widehat{\mu}_i = \frac{1}{d}\sum_{j=1}^{d} X_i(t_j) \qquad \text{and} \qquad \widehat{\sigma}_i^2 = \frac{1}{d}\sum_{j=1}^{d} \left(X_i(t_j) - \widehat{\mu}_i\right)^2.$$

These values are used to transform the samples by means of (9.1). To the sake of illustration, the samples of curves when $d = 80$ are drawn in Figure 9.2.

For each case (varying d and the stock index), the volumetric part $\phi(\varepsilon)$ is estimated using the k–NN approach in [4]: here we used the box kernel (see Corollary 5.2 in [4]), and the number of neighbours equals the integer part of $n/2$. Finally, by means of the method described in Section 9.2, we get estimates of β which are reported in Table 9.1.

It is worth to noticing that $\widehat{\beta}$ is strictly positive: we can deduce that the log–volumetric part $\log(\phi)$ is proportional to $\varepsilon^{-\beta}$. Moreover, all the values are quite close to 2, that corresponds to the most of the Brownian functionals, and the smaller the length d of $\mathscr{T}_i$ is, the closer $\widehat{\beta}$ is to 2; this turns out to be coherent with the literature: the BGM can be used as a model for stock prices whenever the observation window is short enough to assume that drift and volatility terms are constant over that period. Indeed, if one has high–frequency data instead of daily ones (for instance observed at each 5 minutes), the effect of non–constant drift and volatility would be further relaxed and $\widehat{\beta}$ is very closed to 2. This is shown empirically in [2].

Acknowledgements E. Bongiorno and A. Goia are members of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM).

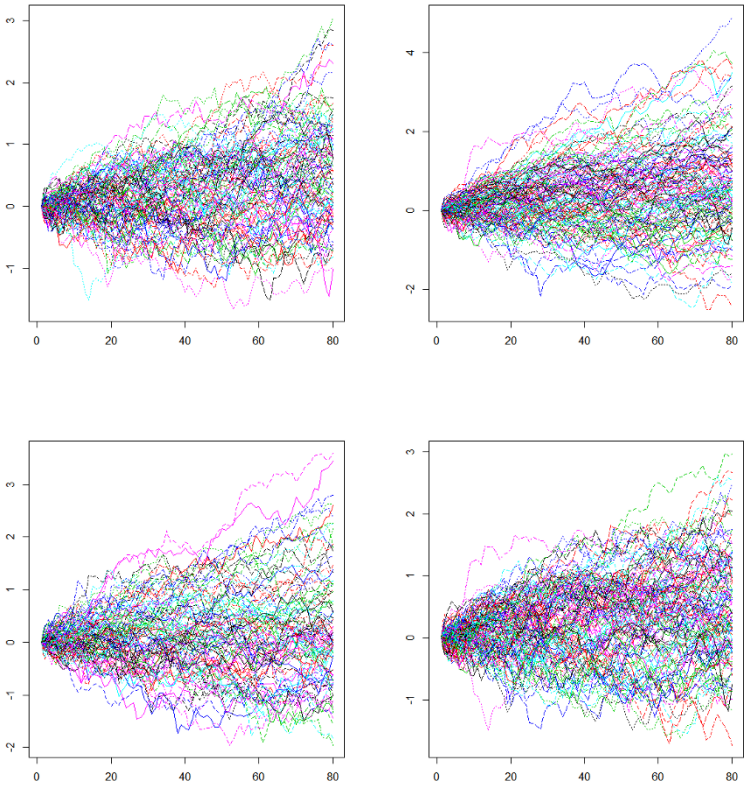


Fig. 9.2: The samples of functional data, derived from original time series, for stock prices Dow Jones (left, top), NASDAQ (rigth, top), NIKKEI (left, bottom) and S&P 500 (right, bottom) indexes, when $d = 80$.

References

[1] Black, F., Scholes, M.: The Pricing of Options and Corporate Liabilities. J. Polit. Econ. **81** (3), 637–654 (1973)
[2] Bongiorno, E. G., Goia, A., Vieu, P.: Evaluating the complexity of functional data. Preprint (2017)
[3] Campbell J. Y., Lo A. W., MacKinlay A.C.: The Econometrics of Financial Markets. Princeton University Press (2012)
[4] Ferraty, F., Kudraszow, N., Vieu, P.: Nonparametric estimation of a surrogate density function in infinite-dimensional spaces. J. Nonparametr. Stat. **24** (2), 447–464 (2012)

[5] Fusai, G., Roncoroni, A.: Implementing Models in Quantitative Finance: Methods and Cases. Springer Finance (2008)
[6] Li, W. V., Shao, Q. M.: Gaussian processes: inequalities, small ball probabilities and applications. In: Stochastic processes: theory and methods. Vol. **19** of Handbook of Statist. North-Holland, Amsterdam, 533–597 (2001)
[7] Marathe R. R., Ryan S. M.: On The Validity of The Geometric Brownian Motion Assumption. The Engineering Economist. **50** (2), 159–192 (2005)
[8] Yen G., Yen E. C.: On the Validity of the Wiener Process Assumption in Option Pricing Models: Contradictory Evidence from Taiwan, Rev. Quant. Finance Account. **12** (4), 327–340 (1999)

Chapter 10
Commutator of projectors and of unitary operators

Alain Boudou and Sylvie Viguier-Pla

Abstract We define and study the concept of commutator for two projectors, for a projector and a unitary operator, and for two unitary operators. Then we state several properties of these commutators. We recall that projectors and unitary operators are linked with the spectral elements of stationary processes. We establish relations between these commutators and some other tools related to the proximity between processes.

10.1 Introduction

This work relates to the field of the operatorial domain, dealing with projectors and unitary operators. These operators take a large place in the statistics of stationary processes. For example, the shift operator is a unitary operator, and a unitary operator is a linear combination of projectors. We define and study the concepts of commutator for two projectors, for a projector and a unitary operator, and for two unitary operators. These concepts are developped in the Hilbertian frame, and when the $\mathbb{C}-$Hilbert space H is of the type $L^2(\Omega, \mathscr{A}, P)$, our results apply to stationary processes. The commutativity of two stationary processes is a generalization of the notion of stationary correlation. When there is not a complete commutativity, we may extend the notion of commutativity, asking how to retrieve the part of each process which commutes. The commutator proposes an answer to this question. We recall that the product of convolution of spectral measures, such as defined in Boudou and

Alain Boudou
Equipe de Stat. et Proba., Institut de Mathématiques, UMR5219, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France e-mail: boudou@math.univ-toulouse.fr

Sylvie Viguier-Pla (✉)
Equipe de Stat. et Proba., Institut de Mathématiques, UMR5219, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France and Université de Perpignan via Domitia, LAMPS, 52 av. Paul Alduy, 66860 Perpignan Cedex 9, France e-mail: viguier@univ-perp.fr

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_10

Romain [5], needs an hypothesis of commutativity. Our work uses tools defined in Boudou and Viguier-Pla [8], such as the $r-$convergence and a distance-like measure of the gap between projectors.

Obviously, the commutator of two projectors is linked with the canonical analysis of the spaces generated by these projectors. When these spaces are complex, the practical interest of it is the domain of stationary processes, as seen above. Several authors work on spectral elements of processes, as, for example, in the large deviation field (Gamboa and Rouault [11]), the autoregressive processes (Bosq [2]), and for reduction of dimension (Brillinger [9], Boudou [3]). The joint study of two processes may lead to the comparison of these processes, by the way of the commutators. When these spaces are real, applications may be forseen by the search of common and specific subspaces of two spaces. Such problematics have been largely developped with other tools, as, for example, in the works of Flury and Gautschi [10], Benko and Kneip [1], or Viguier-Pla [12].

10.2 Prerequisites, recalls and notation

In this text, H is a $\mathbb{C}-$Hilbert space, (E,ξ) a measurable space, and $\mathscr{B}$ the Borel $\sigma-$field of $\Pi=[-\pi;\pi[$. We note S the measurable application $(\lambda,\lambda')\in\Pi\times\Pi\mapsto\lambda+\lambda'-2\pi\left[\frac{\lambda+\lambda'+\pi}{2\pi}\right]\in\Pi$, where $[x]$ denotes the integer part of x. When H' is a $\mathbb{C}-$Hilbert, $\mathscr{P}(H')$ denotes the set of the orthogonal projectors of H'.

Let us examine the notions of stationarity and of correlated stationarity.

Definitions 10.2.1. *A series $(X_n)_{n\in\mathbb{Z}}$ of elements of H is said to be stationary when, for any pair (n,m) of elements of $\mathbb{Z}$, we have $<X_n,X_m>=<X_{n-m},X_0>$.*
Two stationary series $(X_n)_{n\in\mathbb{Z}}$ and $(Y_n)_{n\in\mathbb{Z}}$ are said to be stationarily correlated when $<X_n,Y_m>=<X_{n-m},Y_0>$, for any pair (n,m) of elements of $\mathbb{Z}$.

Let us recall the notion of integral with respect to a random measure (r.m.).

Definition 10.2.2. *A r.m. Z is a vector measure, defined on $\mathscr{B}$ and taking values in H, such that $<ZA,ZB>=0$, for any pair (A,B) of disjoint elements of $\mathscr{B}$.*

Proposition 10.2.1. *If Z is a r.m., then the application $\mu_Z:A\in\mathscr{B}\mapsto\|ZA\|^2\in\mathbb{R}^+$ is a bounded measure. There exists one and only one isometry $\mathscr{I}_Z$ from $L^2(\Pi,\mathscr{B},\mu_Z)$ onto $H_Z=\overline{vect}\{ZA;A\in\mathscr{B}\}$ which, with A, associates ZA, for any A of $\mathscr{B}$.*

Definition 10.2.3. *If Z is a r.m., for any φ of $L^2(\Pi,\mathscr{B},\mu_Z)$, $\mathscr{I}_Z\varphi$ is named integral of φ with respect to the r.m. Z, and is denoted $\int\varphi dZ$.*

Let us now examine the association "stationary series$-$r.m.".

Proposition 10.2.2. *If Z is a r.m., then $(\int e^{i.n}dZ)_{n\in\mathbb{Z}}$ is a stationary series. Conversely, with any stationary series $(X_n)_{n\in\mathbb{Z}}$ of elements of H, we can associate a r.m. Z, and only one, defined on $\mathscr{B}$, taking values in H, such that $X_n=\int e^{i.n}dZ$, for any n of $\mathbb{Z}$.*

A spectral measure (s.m.) is a projector-valued application, defined on a $\sigma-$field, as any measure.

Definition 10.2.4. *A spectral measure (s.m.) $\mathscr{E}$, on ξ for H, is an application from ξ in $\mathscr{P}(H)$ such that*
i) $\mathscr{E}E = I_H$,
ii) $\mathscr{E}(A \cup B) = \mathscr{E}A + \mathscr{E}B$, for any pair (A,B) of disjoint elements of ξ,
iii) $lim_n \mathscr{E}A_n X = 0$, for any sequence $(A_n)_{n\in\mathbb{N}}$ of elements of ξ which decreasingly converges to $\emptyset$ and for any X of H.

With a s.m., we can associate an infinity of r.m.'s.

Proposition 10.2.3. *If $\mathscr{E}$ is a s.m. on $\mathscr{B}$ for H, then the application $Z^X_{\mathscr{E}} : A \in \mathscr{B} \mapsto \mathscr{E}AX \in H$ is a r.m..*

Just like the product of two probability measures is defined, we can define a product for two s.m.'s, with an hypothesis of commutativity.

Definition 10.2.5. *Two s.m.'s $\mathscr{E}_1$ and $\mathscr{E}_2$ on $\mathscr{B}$ for H commute when $\mathscr{E}_1 A_1 \mathscr{E}_2 A_2 = \mathscr{E}_2 A_2 \mathscr{E}_1 A_1$, for any A_1 and A_2 of $\mathscr{B}$.*

Proposition 10.2.4. *If the s.m.'s $\mathscr{E}_1$ and $\mathscr{E}_2$ commute, then there exists a s.m., and only one, denoted $\mathscr{E}_1 \otimes \mathscr{E}_2$, on $\mathscr{B} \otimes \mathscr{B}$ for H, such that $\mathscr{E}_1 \otimes \mathscr{E}_2 A_1 \times A_2 = \mathscr{E}_1 A_1 \mathscr{E}_2 A_2$, for any pair (A_1, A_2) of elements of $\mathscr{B}$.*

Proposition 10.2.5. *If the s.m.'s $\mathscr{E}_1$ and $\mathscr{E}_2$ commute, then the application $\mathscr{E}_1 * \mathscr{E}_2 : A \in \mathscr{B} \mapsto \mathscr{E}_1 \otimes \mathscr{E}_2 S^{-1} A \in \mathscr{P}(H)$ is a s.m. on $\mathscr{B}$ for H named product of convolution of the s.m.'s $\mathscr{E}_1$ and $\mathscr{E}_2$.*

With any s.m. on $\mathscr{B}$ for H, we can associate a unitary operator (u.o.).

Proposition 10.2.6. *If $\mathscr{E}$ is a s.m. on $\mathscr{B}$ for H, then the application $X \in H \mapsto \int e^{i.1} dZ^X_{\mathscr{E}} \in H$ is a u.o.. Conversely, if U is a u.o. of H, then there exists one, and only one, s.m. $\mathscr{E}$, on $\mathscr{B}$ for H, such that $UX = \int e^{i.1} dZ^X_{\mathscr{E}}$, for any X of H.*

From a u.o., we can generate a family of stationary series.

Proposition 10.2.7. *Assume that U is a u.o. of H of associated s.m. $\mathscr{E}$, then $(U^n X)_{n\in\mathbb{Z}}$ is a stationary series of associated r.m. $Z^X_{\mathscr{E}}$.*

When two o.u.'s commute, we can easily express the s.m. associated with their product.

Proposition 10.2.8. *Two u.o.'s U_1 and U_2, of respective associated s.m.'s $\mathscr{E}_1$ and $\mathscr{E}_2$, commute if and only if $\mathscr{E}_1$ and $\mathscr{E}_2$ commute. In that case, $\mathscr{E}_1 * \mathscr{E}_2$ is the s.m. associated with the u.o. $U_1 U_2$.*

For developments of these notions, the reader can refer to Boudou [4], Boudou and Romain [5], and Boudou and Romain [6].

We will end this section by recalls concerning a relation of partial order defined on $\mathscr{P}(H)$.

Definition 10.2.6. *We say that a projector P is smaller than a projector Q, and we note $P \ll Q$, when $P = PQ = QP$.*

Proposition 10.2.9. *If P and Q are two projectors such that $P \ll Q$, then $\|PX\| \leq \|QX\|$, for any X of H.*

The relation $\ll$ is of partial order, but it has the advantage that any family $\{P_\lambda ; \lambda \in \Lambda\}$ of projectors, finite or not, countable or not, has got a larger minorant, that is

a lower bound, denoted $\inf\{P_\lambda;\lambda\in\Lambda\}$, and a smaller majorant, that is an upper bound, denoted $\sup\{P_\lambda;\lambda\in\Lambda\}$. Then we have the following properties.

Proposition 10.2.10. *Let $\{P_\lambda;\lambda\in\Lambda\}$ be a family of projectors. Then*
i) $Im\,inf\{P_\lambda;\lambda\in\Lambda\}=\cap_{\lambda\in\Lambda}ImP_\lambda$;
ii) $(sup\{P_\lambda;\lambda\in\Lambda\})^\perp=inf\{P_\lambda^\perp;\lambda\in\Lambda\}$;
iii) $(inf\{P_\lambda;\lambda\in\Lambda\})^\perp=sup\{P_\lambda^\perp;\lambda\in\Lambda\}$;
iv) if P_1 and P_2 are two projectors which commute, then $inf\{P_1;P_2\}=P_1P_2$.

We can then obtain some properties which are similar to those existing in the classic analysis of sequences.

Definition 10.2.7. *Let $(P_n)_{n\in\mathbb{N}}$ be a sequence of projectors. If $sup\{inf\{P_m;m\geq n\};n\in\mathbb{N}\}=inf\{sup\{P_m;m\geq n\};n\in\mathbb{N}\}$, we say that $(P_n)_{n\in\mathbb{N}}$ $r-$converges to P, and we note it $lim_n^r P_n=P$.*

The $r-$convergence implies the point by point convergence, but the converse is not true.

From the relation $\ll$, we can measure the gap between two projectors.

Definition 10.2.8. *For any(P_1,P_2) of $\mathscr{P}(H)\times\mathscr{P}(H)$, we define*
$d(P_1,P_2)=sup(P_1,P_2)-inf(P_1,P_2)$.

Proposition 10.2.10. *For any(P_1,P_2) of $\mathscr{P}(H)\times\mathscr{P}(H)$, we have $Im(d(P_1,P_2))^\perp=Ker(P_1-P_2)$.*

This notion evoques the notion of distance, however, it is not a distance, as $d(P_1,P_2)$ is a projector. Its interest lies on the following property.

Proposition 10.2.10. *A sequence $(P_n)_{n\in\mathbb{N}}$ of projectors $r-$converges to P if and only if $lim_n^r d(P_n,P)=0$.*

Finally, let us define the notion of maximal equalizator of two u.o.'s.

Definition 10.2.11. *We name maximal equalizator of the u.o.'s U_1 and U_2 the projector, denoted R_{U_1,U_2}, on $\cap_{n\in\mathbb{Z}}Ker(U_1^n-U_2^n)$.*

These notions are developped in Boudou and Viguier-Pla [8].

10.3 Commutator of two projectors

Let us first define this notion of commutator.

Definition 10.3.1. *A projector K is a commutator of the projectors P_1 and P_2 when it commutes with P_1 and P_2 and when $P_1KP_2=P_2KP_1$.*

We can establish the following properties.

Proposition 10.3.1. *Let P_1 and P_2 be two projectors. Then*
i) the upper bound of a family of commutators of the projectors P_1 and P_2 is a commutator of the projectors P_1 and P_2;
ii) 0 is a commutator of the projectors P_1 and P_2;
iii) the upper bound of the family of the commutators of the projectors P_1 and P_2 is

the projector on $Ker(P_1P_2 - P_2P_1)$*, that is* $inf\{P_1, P_2\} + inf\{P_1, P_2^\perp\} + inf\{P_1^\perp, P_2\} + inf\{P_1^\perp, P_2^\perp\}$.

So we have the following definition.

Definition 10.3.2. *Let* P_1 *and* P_2 *be two projectors. We call maximal commutator of the projectors* P_1 *and* P_2 *the projector* C_{P_1,P_2} *on* $Ker(P_1P_2 - P_2P_1)$.

Of course, it is easy to establish that $C_{P_1,P_2} = I$ if and only if P_1 and P_2 commute.

The maximal commutator is a tool for measuring the degree of commutativity of the projectors P_1 and P_2. The larger it is, the larger $\mathrm{Ker}\,(P_1P_2 - P_2P_1)$ is. It is clear that when X belongs to $\mathrm{Ker}\,(P_1P_2 - P_2P_1) = \mathrm{Im}\,C_{P_1,P_2}$, $\|C^\perp_{P_1,P_2}X\| = 0$. So what can we speculate when X is close to $\mathrm{Ker}\,(P_1P_2 - P_2P_1)$, that is when $\|C^\perp_{P_1,P_2}X\|$ is small? We will bring an answer to this question, with the following property.

Proposition 10.3.2. *For any pair* (P_1, P_2) *of projectors, and for any* X *of* H*, we have* $\|P_1P_2X - P_2P_1X\| \leq 2\|C^\perp_{P_1,P_2}X\|$.

So when X is close to $\mathrm{Ker}(P_1P_2 - P_2P_1)$, then P_1P_2X is close to P_2P_1X.

10.4 Commutator of a projector and of a unitary operator

In the same way as we have defined a commutator of two projectors, we can define a commutator of a projector and of a u.o..

Definition 10.4.1. *A projector* K *is a commutator of the projector* P *and of the u.o.* U *when it commutes with* P *and* U*, and when* $PKU = UKP$.

We have got similar properties as those of the previous section.

Proposition 10.4.1. *Let* P *be a projector and* U *a u.o.. We can affirm that*
i) the upper bound of a family of commutators of the projector P *and of the u.o.* U *is a commutator of the projector* P *and of the u.o.* U*;*
ii) 0 *is a commutator of the projector* P *and of the u.o.* U*;*
iii) the upper bound of the family of commutators of the projector P *and of the u.o.* U *is the projector on* $\cap_{n\in\mathbb{Z}} Ker(PU^n - U^nP)$.

The following definition is a consequence of these properties.

Definition 10.4.2. *Let* P *be a projector and* U *a u.o.. We name maximal commutator of the projector* P *and of the u.o.* U*, and we note it* $C_{P,U}$*, the projector on* $\cap_{n\in\mathbb{Z}} Ker(PU^n - U^nP)$.

Of course, it is easy to verify that P and U commute if and only if $C_{P,U} = I$. The association "u.o.-s.m." being biunivocal, all these properties can express by means of the s.m. which is associated with a u.o.. We get then a relation between a commutator of a projector and of a u.o. and a family of commutators of two projectors.

Proposition 10.4.2. *If* P *is a projector and* U *a u.o. of associated s.m.* $\mathscr{E}$*, we can affirm that*
i) a projector K *is a commutator of the projector* P *and of the u.o.* U *if and only if, for any* A *of* $\mathscr{B}$*,* K *is a commutator of the projectors* P *and* $\mathscr{E}A$*;*

ii) $Im C_{P,U} = \cap_{A\in\mathscr{B}} Ker(P\mathscr{E}A - \mathscr{E}AP)$;
iii) $C_{P,U} = inf\{C_{P,\mathscr{E}A}; A \in \mathscr{B}\}$.

The last two points come from the fact that $\cap_{n\in\mathbb{Z}}\mathrm{Ker}(PU^n - U^nP) = \cap_{A\in\mathscr{B}}$ Ker $(P\mathscr{E}A - \mathscr{E}AP)$, and that $\mathrm{Im}\,C_{P,\mathscr{E}A} = \mathrm{Ker}(P\mathscr{E}A - \mathscr{E}AP)$.

If we remark that $\{U^{-n}PU^n; n \in \mathbb{Z}\}$ is a family of projectors and that $\mathrm{Ker}(PU^n - U^nP) = \mathrm{Ker}(U^{-n}PU^n - P) = (\mathrm{Im}\,d(U^{-n}PU^n, P))^{\perp}$, we can give to $C_{P,U}$ an ergodic definition.

Proposition 10.4.3. *For any projector P and for any u.o. U, we have*

$$C_{P,U} = inf\{(d(U^{-n}PU^n, P))^{\perp}; n \in \mathbb{Z}\}.$$

This last result can have the following interpretation. If all the elements of the family $\{U^{-n}PU^n; n \in \mathbb{Z}\}$ are close to P, that is, if for any n of $\mathbb{Z}$, $d(U^{-n}PU^n, P)$ is small, or even more, for any n of $\mathbb{Z}$, $(d(U^{-n}PU^n, P))^{\perp}$ is large, then it is the same for the lower bound $C_{P,U}$. This means that P and U are near to commute.

Proposition 10.4.3 lets us write $d(U^{-n}PU^n, P) \ll C_{P,U}^{\perp}$, so

$$\|(PU^n - U^nP)X\| = \|U^{-n}PU^nX - PX\| \leq 2\|d(U^{-n}PU^n, P)X\| \leq 2\|C_{P,U}^{\perp}X\|,$$

because for any pair of projectors (P, P'), we have $\|PX - P'X\| \leq 2\|d(P, P')X\|$ (cf. Boudou and Viguier-Pla [8]).

Thanks to a similar approach, Propositions 10.3.2 and 10.4.2 let us affirm that

$$\|PZ_{\mathscr{E}}^X A - Z_{\mathscr{E}}^{PX}A\| = \|P\mathscr{E}AX - \mathscr{E}APX\| \leq 2\|C_{P,\mathscr{E}A}^{\perp}X\| \leq 2\|C_{P,U}^{\perp}X\|.$$

Then the following stands.

Proposition 10.4.4. *For any projector P and for any u.o. U of associated s.m. $\mathscr{E}$, for any X of H, we have*
i) $\|PU^nX - U^nPX\| \leq 2\|C_{P,U}^{\perp}X\|$, *for any n of* $\mathbb{N}$;
ii) $\|PZ_{\mathscr{E}}^X A - Z_{\mathscr{E}}^{PX}A\| \leq 2\|C_{P,U}^{\perp}X\|$, *for any A of* $\mathscr{B}$.

So, if X is close to $\mathrm{Im}\,C_{P,U}$, that is if $\|C_{P,U}^{\perp}X\|$ is small, then the series $(PU^nX)_{n\in\mathbb{Z}}$ is "almost stationary", in such a way it is close to the stationary series $(U^nPX)_{n\in\mathbb{Z}}$. As for the application $P \circ Z_{\mathscr{E}}^{PX}$, it is almost a r.m., close to $Z_{\mathscr{E}}^{PX}$, r.m. associated with the stationary series $(U^nPX)_{n\in\mathbb{Z}}$.

Let us end this section by the resolution of the following problem.

Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary series, of associated r.m. Z, and P be a projector. We wish to define all the stationary series, stationarily correlated with $(X_n)_{n\in\mathbb{Z}}$, included in $\mathrm{Im}\,P$. Such series will be named "solution series". Then we have the following.

Proposition 10.4.5. *If U is a u.o. whose associated s.m. $\mathscr{E}$ is such that $Z_{\mathscr{E}}^{X_0} = Z$, then, for any X of $Im C_{P,U}$, we can affirm that $(U^nPX)_{n\in\mathbb{Z}}$ is a "solution series". Any "solution series" is of this type.*

Let us recall that when Z is a r.m. defined on $\mathscr{B}$, taking values in H, there exists at least one s.m. $\mathscr{E}$ on $\mathscr{B}$ for H such that $Z_{\mathscr{E}}^{X_0} = Z$, where $X_0 = \int e^{i.0} dZ$ (cf. Boudou [4]).

10.5 Commutator of two unitary operators

When two u.o.'s U and V commute, the s.m. which is associated with the u.o. UV is the product of convolution of the s.m.'s respectively associated with U and V. But what happens when $UV \neq VU$? The maximal commutator will bring a partial solution to this question.

Definition 10.5.1. *A projector K is a commutator of the u.o.'s U and V when it commutes with U and V, and when $UKV = VKU$.*

Then we can establish the following properties.

Proposition 10.5.1. *Let U and V be two u.o.'s. We can affirm that*
i) the upper bound of a family of commutators of the u.o.'s U and V is a commutator of the u.o.'s U and V;
ii) 0 is a commutator of the u.o.'s U and V;
iii) the upper bound of the family of commutators of the u.o.'s U and V is the projector on $\cap_{(n,m)\in\mathbb{Z}\times\mathbb{Z}} Ker(U^nV^m - V^mU^n)$.

So we can define the following.

Definition 10.5.2. *Let U and V be two u.o.'s. We name maximal commutator of the u.o.'s U and V, and we note it $C_{U,V}$, the projector on the space $\cap_{(n,m)\in\mathbb{Z}\times\mathbb{Z}} Ker(U^nV^m - V^mU^n)$.*

Of course, it is easy to verify that U and V commute if and only if $C_{U,V} = I$. The reader will notice the similarities between Definitions 10.3.1, 10.4.1 and 10.5.1, between Propositions 10.3.1, 10.4.1 and 10.5.1, and between Definitions 10.3.2, 10.4.2 and 10.5.2.

The commutator of two u.o.'s can be defined from the associated s.m.'s.

Proposition 10.5.2. *If U and V are two u.o.'s of respective associated s.m.'s $\mathscr{E}$ and α, we can affirm that*
i) a projector K is a commutator of U and V if and only if, for any pair (A,B) of elements of $\mathscr{B}$, K is a commutator of the projectors $\mathscr{E}A$ and αB;
ii) $Im C_{U,V} = \cap_{(A,B)\in\mathscr{B}\times\mathscr{B}} Ker(\mathscr{E}A\alpha B - \alpha B\mathscr{E}A)$;
iii) $C_{U,V} = inf\{C_{\mathscr{E}A,\alpha B}; (A,B) \in \mathscr{B}\times\mathscr{B}\} = inf\{C_{\alpha B,U}; B \in \mathscr{B}\}$.

To establish the last two points, we must notice that

$$\cap_{(n,m)\in\mathbb{Z}\times\mathbb{Z}} \mathrm{Ker}(U^nV^m - V^mU^n) = \cap_{(A,B)\in\mathscr{B}\times\mathscr{B}} \mathrm{Ker}(\mathscr{E}A\alpha B - \alpha B\mathscr{E}A).$$

Point *iii)* provides a relation between the three types of maximal commutators which we study. We can also establish a relation between the maximal commutator of two u.o.'s and the maximal equalizator of two u.o.'s.

Proposition 10.5.3. *Let U and V be two u.o.'s. We have $C_{U,V} = inf\{R_{V,U^{-n}VU^n}; n \in \mathbb{Z}\}$.*

For the proof, we have just to notice that

$$\mathrm{Im} C_{U,V} = \cap_n \cap_m \mathrm{Ker}(U^nV^m - V^mU^n) = \cap_n \cap_m \mathrm{Ker}(V^m - U^{-n}V^mU^n)$$
$$= \cap_n \cap_m \mathrm{Ker}(V^m - (U^{-n}VU^n)^m) = \cap_n \mathrm{Im} R_{V,U^{-n}VU^n} = \mathrm{Im}\inf\{R_{V,U^{-n}VU^n}; n \in \mathbb{Z}\}.$$

We can now approach the questions suggested at the beginning of the section. Let us recall some results that we can find in Boudou and Viguier [8].

Proposition 10.5.4. *Let U and V be two u.o.'s of associated s.m.'s $\mathscr{E}$ and α, and L be the application $X \in Im C_{U,V} \mapsto X \in H$. We have*
i) $L^*(X) = C_{U,V}X$, *for any* X *of* H; $L^*L = I_{Im C_{U,V}}$; $LL^* = C_{U,V}$; $L^*C_{U,V} = L^*$; $C_{U,V}L = L$;
ii) $U' = L^*UL$ *and* $V' = L^*VL$ *are u.o.'s of* $Im C_{U,V}$;
iii) for any A *of* $\mathscr{B}$, $\mathscr{E}'A = L^*\mathscr{E}AL$ *and* $\alpha'A = L^*\alpha AL$ *are projectors of* $Im C_{U,V}$;
iv) the applications $\mathscr{E}' : A \in \mathscr{B} \mapsto \mathscr{E}'A \in \mathscr{P}(Im C_{U,V})$ *and* $\alpha' : A \in \mathscr{B} \mapsto \alpha'A \in \mathscr{P}(Im C_{U,V})$ *are the s.m.'s respectively associated with the u.o.'s* U' *and* V'.

From the fact that

$$U'V' = L^*ULL^*VL = L^*UC_{U,V}VL = L^*VC_{U,V}UL = L^*VLL^*UL = V'U',$$

we can consider the s.m. $\mathscr{E}' \otimes \alpha'$ on $\mathscr{B} \otimes \mathscr{B}$ for $\mathrm{Im} C_{U,V}$ (as the s.m.'s $\mathscr{E}'$ and α' commute). For any pair (A,B) of elements of $\mathscr{B}$, we have

$$\mathscr{E}' \otimes \alpha'(A \times B) = \mathscr{E}'A\alpha'B = \inf\{L^*\mathscr{E}AL, L^*\alpha BL\} = L^*\inf\{\mathscr{E}A, \alpha B\}L.$$

If we notice that $U'V' = L^*UVL = L^*VUL$, we have the following.

Proposition 10.5.5. *There exists one and only one s.m.,* $\mathscr{E}' \otimes \alpha'$, *on* $\mathscr{B} \otimes \mathscr{B}$ *for* $Im C_{U,V}$, *such that* $\mathscr{E}' \otimes \alpha'(A \times B) = L^*inf\{\mathscr{E}A, \alpha B\}L$, *for any pair* (A,B) *of elements of* $\mathscr{B}$. *Its image by* S *is the s.m. associated with the u.o.* $L^*UVL = L^*VUL$.

Of course, when U and V commute, that is when $C_{U,V} = I$, we have $L = L^* = I$, $U' = U$, $V' = V$, $\mathscr{E}' = \mathscr{E}$, $\alpha' = \alpha$ and $\mathscr{E} \otimes \alpha(A \times B) = \inf\{\mathscr{E}A, \alpha B\} = \mathscr{E}A\alpha B$, for any (A,B) of $\mathscr{B} \times \mathscr{B}$.

Finally, we establish a link between the commutator and the equalizator of a same pair (U,V) of u.o.'s.

Proposition 10.5.6. *Let* U *and* V *be two u.o.'s, and* J *be the application* $X \in Im R^{\perp} \mapsto X \in H$. *Then*
i) J^*UJ *and* J^*VJ *are u.o.'s of* $Im R^{\perp}_{U,V}$;
ii) $C_{U,V} = JC_{J^*UJ,J^*VJ}J^* + R_{U,V}$.

References

[1] Benko, M., Kneip, A.: Common functional component modelling. Proceedings of 55th Session of the International Statistical Institute, Syndney, (2005)

[2] Bosq, D.: Linear processes in functions spaces: Theory and Applications. Lecture Notes Series, Vol. 149. Springer, Berlin (2000)

[3] Boudou, A.: Approximation of the principal components analysis of a stationary function. Statist. Probab. Letters **76** 571-578 (2006)

[4] Boudou, A.: Groupe d'opérateurs unitaires déduit d'une mesure spectrale - une application. C. R. Acad. Sci. Paris, Ser. I **344** 791-794 (2007)

[5] Boudou, A., Romain, Y.: On spectral and random measures associated to continuous and discrete time processes. Statist. Probab. Letters **59** 145-157 (2002)

[6] Boudou, A., Romain, Y.: On product measures associated with stationary processes. The Oxford handbook of functional data analysis, 423-451, Oxford Univ. Press, Oxford (2011)

[7] Boudou, A., Viguier-Pla, S.: Relation between unit operators proximity and their associated spectral measures. Statist. Probab. Letters **80** 1724-1732 (2010)

[8] Boudou, A., Viguier-Pla, S.: Gap between orthogonal projectors - Application to stationary processes. J. Mult. Anal. **146** 282-300 (2016)

[9] Brillinger, D.R.: Time Series: Data Analysis and Theory. Holt, Rinehart and Winston, New-York (1975)

[10] Flury, B.N., Gautschi, W.: An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. SIAM J. Stat. Comput. **7** 1 169-184 (1986)

[11] Gamboa, F., Rouault, A.: Operator-values spectral measures and large deviation. J. Statist. Plann. Inference **154** 72-86 (2014)

[12] Viguier-Pla, S.: Factor-based comparison of k populations. Statistics **38** 1-15 (2004)

Chapter 11
Permutation tests in the two-sample problem for functional data

Alejandra Cabaña, Ana M. Estrada, Jairo Peña and Adolfo J. Quiroz

Abstract We propose two kind of permutation tests for the two sample problem for functional data. One is based on nearest neighbours and the other based on functional depths.

11.1 Introduction

Let $X_1(t), \cdots, X_m(t)$ denote an i.i.d. sample of real valued curves defined on some interval J. Let $\mathscr{L}(X)$ be the common probability law of these curves. Likewise, let $Y_1(t), \cdots, Y_n(t)$, be another i.i.d. sample of curves, independent of the X sample and also defined on J, with probability law $\mathscr{L}(Y)$.

We want to test the null hypothesis, $H_0 : \mathscr{L}(X) = \mathscr{L}(Y)$ against the general alternative $\mathscr{L}(X) \neq \mathscr{L}(Y)$.

We discuss three different permutation tests: a functional Schilling test, Wilcoxon type test, and another test based on depths, that uses meta analysis ideas to assess significance. We compare their performance with the classical test by Kokoszka and Horváth, based on the principal components of the pooled covariance operator of the two samples in a simulated experiment and with real data.

Alejandra Cabaña (✉)
Universidad Autónoma de Barcelona, Barcelona, Spain, e-mail: acabana@mat.uab.cat

Ana M. Estrada, Jairo Peña and Adolfo J. Quiroz
Universidad de Los Andes, Bogotá, Colombia e-mail: am.estrada213@uniandes.edu.co, ji.pena45@uniandes.edu.co, aj.quiroz1079@uniandes.edu.co

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_11

11.2 Schilling's type test

This is an adaption to functional data of the k -nearest neighbours multivariate two-sample test of Schilling's tests based on nearest neighbors [10].

Let $N = m+n$ and $Z_1, \ldots, Z_N$ be the combined sample obtained by concatenating the X and Y samples.

Define the indicator function $I_i(r) = 1$ if Z_i and its r-nn belong to the same sample and else, $I_i(r) = 0$. Nearest neighbours are based on L^2 distance, and with probability 1 the are uniquely defined. In case of ties, we would decide at random.

In practice, if the functions have been registered in a common grid, say $0 = t_0 < t_1 < \ldots t_L = T$, a reasonable approximation to the distance between functions Z_i and Z_j is

$$d_{i,j} = \sum_{l=1}^{L} \Delta_l (Z_i(t_l) - Z_j(t_l))^2, \text{ where } \Delta_l = t_l - t_{l-1}$$

If the grid used is equally spaced, Δ_l can be omitted and the curves can be treated as points in R^L in order to compute faster the k-nearest-neighbours. When no common grid is available, represent the functions in the joint sample in terms of local polynomials, or some other basis functions, and the k-nearest-neighbours are identified by a quadratic algorithm (in the joint sample size N).

Define the statistic:

$$T_{N,k} = \frac{1}{Nk} \sum_{i=1}^{N} \sum_{r=1}^{k} I_i(r)$$

Under H_0 we expect $T_{N,k}$ to be small.

Observe that the expected value $\mathbf{E}\, T_{N,k} = \mathrm{e}\, I_i(r) = \frac{m(m-1)+n(n-1)}{N(N-1)}$ while the variance depends on the amount of pairs of points that are mutual neighbours and the amount of pairs that share a common neighbour.

Under the null hypothesis of equal distribution of $\{X_m(t)\}$ and $\{Y_n(t)\}$, permutations on the labelling of the Z_i do not alter the distribution of $T_{N,k}$ and hence can be computed with a standard permutation procedure.

Algorithm

1. For the concatenation of the samples Z, keeping the natural ordering, compute the $m \times n$ matrix of distances between its elements.
2. Assume the number of neighbours is fixed to k. Build a $N \times k$ matrix that in the i-th row contains the indices of the k nearest elements to the curve Z_i.
3. In order to compute the statistic it is enough to count how many of the indices in each of the first m rows are equal or less than m, (i.e. are originally X) and how many of the indices corresponding to $m+1, \ldots, N$ are greater than m.
4. Obtain the distribution of $T_{N,k}$ using the permutation procedure.

11.3 Depths-based tests

Another way of approaching the functional data analysis is introducing the notion of depth, that is related to a generalisation of the concept of ordering for functional data. The idea is to assign an order to each element of the sample related to its centrality within the whole set.

There are many measures of depth, and there is no agreement about their advantages. We will concentrate on the Fraiman-Muñiz depth [2].

Consider first a univariate sample $U_1,...,U_n$ and $U_{(1)},...,U_{(n)}$ be the corresponding order statistics. Assuming there are no ties, the natural depth of U_i is said to be $D_n(U_i) = \frac{1}{2} - \left|\frac{1}{2} - \left(\frac{j}{n} - \frac{1}{2n}\right)\right|$. This notion of depth assigns minimal and equal depth to the two extreme values of the sample, maximum to the innermost point(s) and changes linearly with the position the datum occupies in the sample.

For the case of functional data, consider the sample $X=\{X_i(t)\}$ defined in a common interval J. For each t we can compute the natural depth, $D_n(X_i(t))$, and then the depth for each $X_i(t)$ is:

$$I(X_i,X) = \int_J D_n(X_i(t))dt,$$

where, in practice, the integral is replaced by a sum over t for the time grid.

A Wilcoxon test based on this ordering is a natural option, and has been suggested by López-Pintado and Romo in [6], [7] based on their band-depth. Observe that for univariate samples, the Wilcoxon test statistic defined in terms of Fraiman-Muñiz depth corresponds to Ansay-Bradley's statistic, and hence is suitable for detecting differences in dispersion between the samples.

11.3.1 Meta-analysis based tests

Let $\mathscr{X} = \{X_1,\ldots,X_m\}$ denote the functional X sample and $\mathscr{Y} = \{Y_1,\ldots,Y_n\}$ the functional Y sample. For each $X_i \in \mathscr{X}$, we consider its depth with respect to the Y sample with the curve X_i added. We denote this depth $I(X_i, \mathscr{Y} \cup \{X_i\})$. This is a measure of how outlying the curve X_i is with respect to the Y sample. If "many" of the X_i turn out to be outlying with respect to $\mathscr{Y}$, that would be evidence against the null hypothesis of equality of distributions. Similarly we can measure how outlying is each curve Y_j with respect to the X-sample, $\mathscr{X}$, by computing $I(Y_j, \mathscr{X} \cup \{Y_j\})$. The first question is how to combine the values of $I(X_i, \mathscr{Y} \cup \{X_i\})$, for all $i \leq m$, in a single number that combines the information in all these depths. For this purpose, we rely in an idea coming from Meta-Analysis.

To the depth $I(X_i, \mathscr{Y} \cup \{X_i\})$ we associate an empirical p-value, Let p_i be an empirical p-value related to the *centrality* of the variable Y_i on the sample X,

$$\hat{p}_* = \frac{\sum_{j=1}^{m} I\left(D\left(X_j|X\right) \leq D\left(Y_*|X\right)\right)}{m+1} \qquad \begin{cases} p_i = \hat{p}_i \text{ if } \hat{p}_i \neq 0 \\ p_i = \frac{1}{m+1} \text{ if } \hat{p}_i = 0 \end{cases}$$

Under H_0 (ignoring ties), each p_i is uniformly distributed in $\{1/m, 2/m, \ldots, 1\}$, but they are not independent.

For symmetry, we consider as well the q_i p-values related to the *centrality* of the variable X_i on the sample Y.

$$\hat{q}_* = \frac{\sum_{j=1}^{n} I\left(D\left(Y_j|Y\right) \leq D\left(X_*|Y\right)\right)}{n+1} \qquad \begin{cases} q_i = \hat{q}_i \text{ if } \hat{q}_i \neq 0 \\ q_i = \frac{1}{n+1} \text{ if } \hat{q}_i = 0 \end{cases}$$

Observe that when H_0 does not hold, the depth of a curve in a family where it does not belong will be very low, and so would be its associated p-value. In that case the corresponding statistic will be very big.

Consider, as described in [4],

$$S_Y = \sum_{i=1}^{m} -\ln(p_i) \qquad S_X = \sum_{i=1}^{n} -\ln(q_i)$$

We want to associate a p-value to the pair (S_X, S_Y).

MA1: Meta-Analysis method 1

When the two samples display a difference in "scale", most, of the curves of the (say) X sample, turn out to be central with respect to the Y sample and S_X will not show a significant value. In such a situation, typically, several curves of the Y sample will turn out to be clearly outlying respect to the X sample, and the maximum will reach a significant value.

We propose the use of $S = \max\{S_X, S_Y\}$ as test statistic, and the use of a permutation procedure to obtain its distribution.

MA2: Meta-Analysis method 2

Better empirical performance is obtained using following result:

Lemma 1 *combining p-values Let p_X (p_Y) denote the p-value of S_X (S_Y), under the null permutation distribution, as obtained from procedure p-value if all subsets of size m were used (instead of just a sample of size B) and assuming the null hypothesis. Then*

1. $\Pr(p_X \leq t) \leq t$ *for any* $t \in (0,1)$*, and the same holds for* p_Y.
2. $\Pr(2\min(p_X, p_Y) \leq t) \leq t$ *for any* $t \in (0,1)$.

Proof. The null permutation distribution of S_X is a discrete distribution that can not be assumed uniform on its range (some values of S_X can appear more frequently than others when subsets are chosen at random). This is why part (i) of the Lemma is not obvious. Let F denote the null permutation c.d.f. of S_X and let $S_{X,\text{obs}}$ denote the observed value of S_X. Recall that large values of S_X are considered significant. Then, clearly, $p_X = 1 - F(S^-_{X,\text{obs}})$ and, for $t \in (0,1)$,

$$\Pr(p_X \leq t) = \Pr(F(S^-_{X,\text{obs}}) \geq 1-t) = \sum_{\{s:F(s)>1-t\}} \Pr(S_X = s) \leq t,$$

by definition of F.

Since p_X and p_Y are not independent, to prove (ii) we can use (i) together with the usual union bound:

$$\Pr(\min(p_X, p_Y) \leq t/2) \leq \Pr(p_X \leq t/2) + \Pr(p_Y \leq t/2) \leq t/2 + t/2 = t.$$

Part (2) of the Lemma tells us that an appropriate p-value for $2\min(p_X, p_Y)$ is the observed value of this statistic itself.

Thus, our second way of getting a p-value from S_X and S_Y is the following: Compute, approximately, p_X and p_Y for S_X and S_Y, respectively, using the permutation procedure and use $2\min(p_X, p_Y)$ as p-value.

11.4 Empirical comparison of powers and real data applications

In order to fix a standard, we have also computed the empirical power for Horvàth and Kokoszka's test for equality of mean functions. The null hypothesis is rejected for large values of the the statistic

$$U_{m,n} = \frac{mn}{m+n} \int \left(\hat{X}_m(t) - \hat{Y}_n(t)\right)^2 dt$$

Under some regularity conditions, the asymptotic distribution of $\hat{X}_m(t) - \hat{Y}_n(t)$ is a Gaussian process Γ whose covariance can be approximated by the pooled covariance operator of the two samples, hence, the distribution of $U_{m,n}$ can be approximated by the first d terms in the Karhunen-Loève expansion of $\int_0^1 \Gamma^2(s)ds \approx \sum_{i=1}^d \lambda_i N_i^2$, λ_i are the (ordered) eigenvalues of the pooled covariance estimator, and N_i are i.i.d. N(0,1).

11.4.1 A simulation experiment

We have simulated samples of functional data as realisations from a geometric Brownian motion process $f(t) = X_0 \exp\left\{rt - \frac{t\sigma^2}{2} + \sigma w_t\right\}$ where r and σ are, respectively,

the trend (drift) and volatility coefficients, w_t is a standard Wiener process and X_0 is the initial value.

We have tested $H_0 : \mathscr{L}(X) = \mathscr{L}(Y)$, against $\mathscr{L}(X) \neq \mathscr{L}(Y)$ where X was simulated with $r = 1, \sigma = 1, X_0 = 0$ and Y is any of the 'contaminated' samples with only one of the parameters variying at a time.

From Figure 11.1, we see that Wilcoxon's statistic performs very well against volatility variations but fails noticeably for the other alternatives considered in the experiment. On the other hand, Horvàth and Kokoszka's test (HK), being a test conceived for changes in the mean, shows the best performance against changes in the drift parameter, while its power numbers against changes in the origin (initial level) are good too. But HK results ineffective in picking the volatility changes.

The Meta Analysis methods have a power similar to HK against changes in the origin, while their power, although reasonable, is inferior to HK's when it comes to changes in drift. On the other hand, both Meta Analysis procedures display excellent power against the volatility alternatives, where HK fails. Schilling's statistic (with $k = 5$ and $K = 10$), shows very good power against all the alternatives considered in our experiment. Overall, Schilling's statistic displays the best performance in terms of power among the methods evaluated.

11.4.2 NO_2 Contamination in Barcelona

We have hourly measurements of nitrogen dioxide (a known pollutant formed in most combustion processes using air as the oxidant) in four neighbourhoods in Barcelona, namely Eixample, Palau Reial, Poblenou and Sants. The measurements were taken along the years 2014 and 2015 in automatic monitoring stations[1]

We have split the data sets into working days ($\approx$220 curves) and non-working days ($\approx$120 curves), each year.

There are many questions of interest, for instance, to assess whether the level of pollutants significantly different during working and non-working days, or if the levels of NO_2 changed from one year to the next in each of the neighbourhoods, or comparing the pollution levels among the different neighbourhoods. We include here the results of one of these many comparisons.

All tests show evidence of differences between working and non-working days in all four neigbourhoods, with Wilcoxon and Schilling-10 showing the strongest evidence of differences. Figure 11.2 shows the levels of contaminants in the neighbourhood of Sants in the years 2014 and 2015. The tests show that the level of pollutants did not change noticeably on non-working days, but significant changes are found from one year to the next on working days, with the Wilcoxon and Schilling procedures being the ones that find stronger evidence of change.

[1] available from http:// dtes.gencat.cat/icqa.

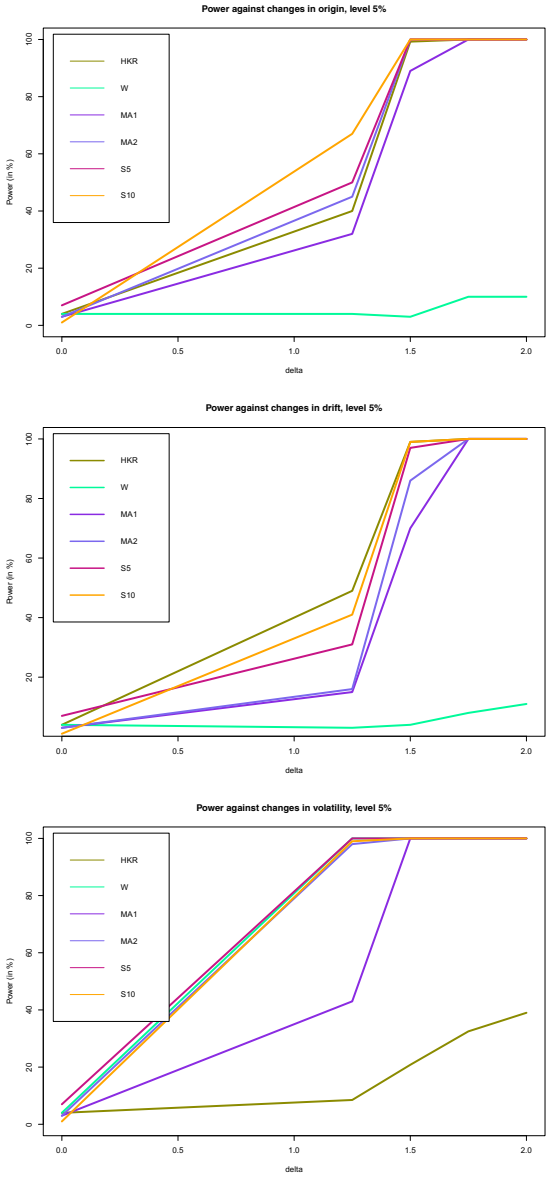


Fig. 11.1: Empirical power of the tests against changes of origen, drift and volatility in Geometric Brownian motion data; level $\alpha = 5\%$.

Acknowledgements The work of A. Cabaña was supported by Ministerio de Economía, Industria y Competitividad, Spain, project MTM2015-69493-R.

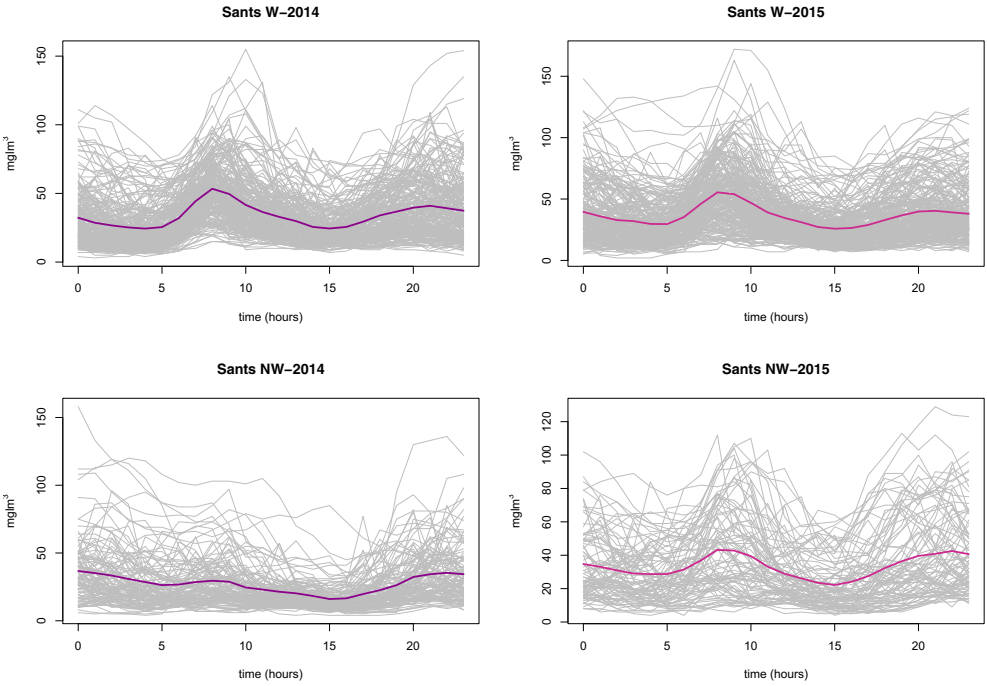


Fig. 11.2: NO_2 leves on working days (1st row) and non-working days (2nd row) in Sants, 2014-2015

References

[1] Estrada, A.M.: Nociones de Profundidad en Análisis de Datos Funcionales, MSc Thesis, Universidad de Los Andes, Colombia (2014)
[2] Fraiman, R., Muniz, G: Trimmed means for functional data. TEST **10**(2): 419–440 (2001)
[3] Good, P.: Permutation, Parametric and Bootstrap Tests of Hypothesis, Springer (2005)
[4] Hedges, L. V., Olkin, I.: Statistical Methods for Meta-Analysis, Academic Press (1985)
[5] Liu, R. Y., Parelius, J. M., Singh, K.: Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference. Ann. Statist. **27**(3), 783–840 (1999)
[6] López-Pintado, S., Romo, J.: Depth-based inference for functional data. Comput. Statist. Data Anal. **51**, 4957–4968 (2007)
[7] López-Pintado, S., Romo, J.: On the concept of depth for functional data. J. Amer. Statist. Assoc. **104**, 718–734 (2009)

[8] Peña, J.: Propuestas para el problema de dos muestras con datos funcionales., MSc Thesis, Universidad de Los Andes, Colombia (2012)
[9] Quiroz, A. J.: Graph-Theoretical Methods. In Encyclopedia of Statistical Sciences, 2nd Edition, S. Kotz, C. B. Read, N. Balakrishnan and B. Vidakovic, editors. Vol 5, pp 2910–2916, Wiley and Sons, New York (2006)
[10] Schilling, M. F. Multivariate two-sample tests based on nearest neighbours. J. Amer. Statist. Assoc. **81**:395, 799–806 (1986)

Chapter 12
Functional data analysis of "Omics" data: how does the genomic landscape influence integration and fixation of endogenous retroviruses?

Marzia A. Cremona, Rebeca Campos-Sánchez, Alessia Pini, Simone Vantini, Kateryna D. Makova and Francesca Chiaromonte

Abstract We consider thousands of endogenous retrovirus detected in the human and mouse genomes, and quantify a large number of genomic landscape features at high resolution around their integration sites and in control regions. We propose to analyze this data employing a recently developed functional inferential procedure and functional logistic regression, with the aim of gaining insights on the effects of genomic landscape features on the integration and fixation of endogenous retroviruses.

12.1 Introduction and motivations

Functional Data Analysis (FDA) has been instrumental to advances in many scientific domains. In recent years, FDA has also been applied to "Omics" research. One example are functional linear models used to screen variants (e.g., SNPs; single nucleotide polymorphisms) genome-wide, possibly accounting for a number of

Marzia A. Cremona (✉)
Department of Statistics, Pennsylvania State University, USA, e-mail: mac78@psu.edu

Rebeca Campos-Sánchez
Centro de Investigacin en Biologa Celular y Molecular, Universidad de Costa Rica, Costa Rica, e-mail: rcamposs@cariari.ucr.ac.cr

Alessia Pini
MOX - Department of Mathematics, Politecnico di Milano, Italy, e-mail: alessia.pini@polimi.it

Simone Vantini
MOX - Department of Mathematics, Politecnico di Milano, Italy, e-mail: simone.vantini@polimi.it

Kateryna D. Makova
Center for Medical Genomics and Department of Biology, Pennsylvania State University, USA, e-mail: kdm16@psu.edu

Francesca Chiaromonte
Center for Medical Genomics and Department of Statistics, Pennsylvania State University, USA, Sant'Anna School of Advanced Studies, Pisa, Italy, e-mail: chiaro@stat.psu.edu

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_12

covariates, to identify effects on a complex phenotype quantitated as a response curve (see [9]). Another example are analyses of the shapes of peaks produced by ChIP-seq experiments, which indicate the putative binding locations of proteins interacting with the genome under certain conditions or in certain tissues/cell types. Clustering of peak shapes can be used to identify meaningful groups or types of binding sites genome-wide (see [5]). While these examples demonstrate that FDA can lead to important novel insights in genetics and genomics, applications of FDA in this context are still very limited in number.

Here we present a functional inferential procedure for "Omics" data – the Interval Testing Procedure (ITP) – and its extension – the Interval-Wise Testing for functional data (IWT) – developed in [10] and [8], respectively, and generalized for our study. An R package implementing this generalized version of IWT is available on the Bioconductor website (see [3, 4]). We show an application of ITP/IWT and functional logistic regression to investigate features of the genomic landscape that may affect the integration and fixation of endogenous retroviruses (ERVs), based on their profiles around ERVs integration sites.

ERVs are the remnants of retroviral infections in the germ line. They occupy a large portion of many mammalian genomes ($\sim$8% and $\sim$10% of the human and mouse genomes, respectively) and distribute unevenly along them – contributing to shape genomic structure, evolution and function. Understanding whether a characteristic genomic landscape exists in the neighborhood of these elements is of utmost importance to unveil the biological mechanisms that govern them and the consequences they have on genome function and evolution, as well as on diseases (see e.g. [6, 2]). This entails quantitating a number of genomic features (e.g. non-B DNA structures, recombination rates or histone marks) in such neighborhoods and in control regions. Prior studies of this kind have produced important insights, but have been limited by the relatively low resolution at which data was available for some genomic features. As improved technologies and additional large-scale experiments make it possible to quantitate more features at higher resolution, it becomes paramount to use statistical methods that exploit high resolution information to learn how the genomic landscape affects the presence of ERVs on the genome. In this context, we propose to use FDA to extract signal from high resolution feature profiling, considering their local shapes. The biological results presented here are a selection of those published in our recent article [1].

12.2 Data

We consider a large collection of the most recently active ERVs in the human and mouse genomes, composed by fixed and *in vitro* HERV-Ks in human, and fixed and polymorphic ETns and IAPs in mouse. Importantly, we consider ERVs of different evolutionary ages (young *in vitro* and polymorphic ERVs, older fixed ERVs) with the aim of disentangling integration vs. fixation preferences. After pre-processing (we filter out elements with genome gaps and/or other elements in their 64-kb flanks), our

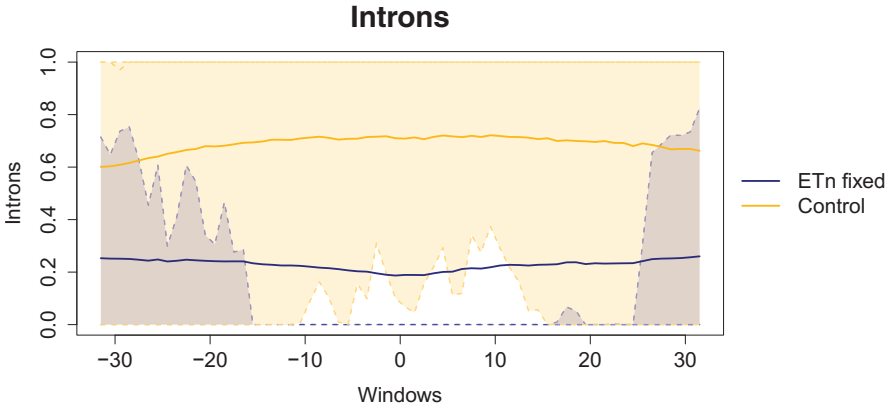


Fig. 12.1: Pointwise boxplot of the genomic feature "introns" in the 64-kb flanking regions of fixed ETns and in mouse control regions. The continuous lines represent the average signals in the two groups, while the dashed lines represent the quartiles in each window in the two groups of regions. Shaded areas between dashed lines correspond to the box between the first and third quartiles.

dataset comprises 826 fixed and 1,005 in vitro HERV-Ks in human, and 1,296 fixed and 217 polymorphic ETns, as well as 3,255 fixed and 1,788 polymorphic IAPs, in mouse.

In the $\pm$32kb flanking regions of these ERVs, we measure over 40 human and mouse genomic landscape features (e.g. DNA conformation, non-B DNA motifs, chromatin openness/modifications, gene expression) at 1-kb resolution. Similarly, we construct control sets of 1,142 mouse and 1,543 human regions, selecting contiguous 64-kb sequences that do not intersect the ERVs flanks nor other Long Terminal Repeats (LTR), and measuring the genomic landscape features over these regions. The genomic features we consider correspond to different types of "Omics" data and are quantitated in different ways (e.g. counts or coverages). Some are derived directly from the genomic sequences (e.g. microsatellites and CpG islands); others are generated by different high throughput techniques such as ChIP-seq or RNA-seq. Figure 12.1 shows, as an example, the pointwise boxplot of the genomic feature "introns" (measured in each 1 kb window as the share of the window occupied by introns) in the flanks of fixed ETns and in control regions. Additional details about the dataset can be found in [1].

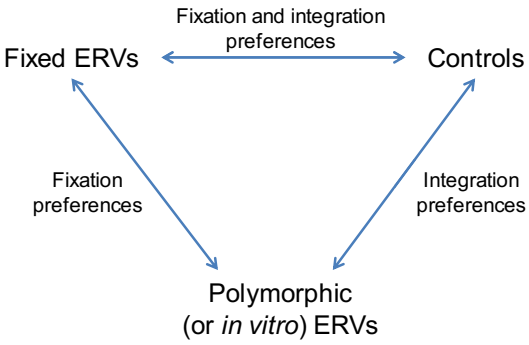


Fig. 12.2: Comparisons employed to disentangle integration and fixation effects.

12.3 Methods

For each genomic feature and each ERV flanking region (or control region), we consider the 64 measurements in consecutive 1-kb windows as a curve. We consider the set of all the curves corresponding to a particular type or ERVs, or to controls, as a random sample from a stochastic curve. This approach has the advantage that the consecutive ordering of the measurements along the genome is naturally embedded by the curves. We then compare the different stochastic curves using an extended version of the Interval Testing Procedure (ITP – developed in [10]) or the successive Interval-Wise Testing (IWT – [8, 3, 4]), as well as with functional logistic regression. In order to disentangle integration and fixation preferences of ERVs, we consider three different comparisons for each type of ERV and each genomic feature, as depicted in Figure 12.2.

Interval Testing Procedure (ITP) The ITP is a functional hypothesis test to assess differences between the distributions of two stochastic curves, that can detect the specific components of the curves that lead to reject the null hypothesis. As a first step, the ITP decomposes the curves on a truncated functional basis, thus representing each of them with K coefficients. Then, it performs a univariate non-parametric (permutation) test on each coefficient, as well as a multivariate test on each possible set of consecutive coefficients. Finally, the p-value for each component is adjusted in order to control the familywise error rate over sets of contiguous components, i.e. the probability of wrongly rejecting the null hypothesis over each set of successive coefficients where it is true. The resulting adjusted p-values allows one to attribute observed differences to specific coefficients, which translates into a selection of specific locations when the employed basis functions have compact support.

Interval-Wise Testing (IWT) The IWT is another functional inferential procedure representing an extension of the ITP. Indeed, similarly to the ITP, the IWT is able to select the subregions where the differences between two curve distributions are

significant. However, the IWT does not rely on any functional basis representation; it tests the curves directly. The IWT performs a functional permutation test on every subinterval of the domain of the curves. Then it computes an adjusted p-value curve that controls the interval-wise error rate, i.e. the probability of wrongly rejecting the null hypothesis over each subinterval where it is true, allowing the identification of the locations where the stochastic curves differ.

Generalization of ITP/IWT Exploiting the non-parametric nature of the tests, the generalized version of the ITP and IWT allows different test statistics to be employed in the permutation test, in order to highlight different characteristics of the curve distributions. In particular, we consider test statistics based on the mean difference, the median difference, the multi-quantile difference or the variance ratio. For example, the multi-quantile statistics employed by the IWT is defined for each subinterval S as

$$T_{multi-quantile}(S) = \sum_{q \in Q} \frac{1}{|S|} \int_S \left(y_1^q(t) - y_2^q(t)\right)^2 dt,$$

where $y_1^q(t)$ and $y_2^q(t)$ are the pointwise quantiles of order q of the curves in the two groups. In addition, we generalize the ITP and IWT to be able to detect not only the locations, but also the scales where the differences between the curves are significant. To this end, we compute adjusted p-value curves at all possible scales – from the smallest scale represented by a single component in the ITP and by a single position in the IWT (i.e. raw p-value curves, no correction), to the largest scale possible represented by the entire curve domain (i.e. correcting across the whole domain, like in the original version of ITP/IWT).

Functional logistic regression We employ single functional logistic regression to quantify the discriminatory power of the different genomic features identified in each of the comparisons in Figure 12.2. Then, multiple functional logistic regression is used to evaluate the predictive power of all the features simultaneously.

12.4 Results

Performing the functional test independently on each feature, we find that some of them show significant differences on the whole 64-kb region analyzed, while others present only localized differences, usually near the insertion site of the ERV (the center of the region) or everywhere but the insertion site. Moreover, some of the features appear to have an effect at all the considered scales (up to 64-kb), while others display significant differences especially at small scales. Interestingly, the three families of ERVs considered in the study show many common discriminatory features, but some features appear to be specific to a particular type of ERVs.

Functional logistic regression highlights few very strong, dominant predictors. For example, microsatellites are strongly overrepresented in the flanking regions of fixed and polymorphic ERVs compared to control regions (deviance explained up to

89%). In addition, even excluding these strong predictor from the multiple functional logistic regression (the ones with deviance explained >20%), we observe that the simultaneous predictive power of the remaining features is still quite high for the comparisons between ERVs and controls, with a deviance explained ranging from 24% to 79%.

These analyses provide important biological insights. In particular, we find that ERVs integrate preferentially in late-replicating, AT-rich regions with abundant microsatellites, mirror repeats, and repressive histone marks. Moreover, ERVs fixate preferentially in regions depleted of genes and evolutionarily conserved elements, and with low recombination rates – likely reflecting the fact that purifying selection and ectopic recombination act to remove ERVs from the genome. Interestingly, in addition to a negative effect on fixation of high recombination rates in both human and mouse genomes, there is a positive association between recombination hotspots and ERVs fixation in human, and one between hotspots and ERVs integration in mouse. A more detailed discussion of the biological implications of this study can be found in [1].

12.5 Conclusion

We proposed a novel approach for the analysis of diverse "Omics" data using FDA techniques, and we showed an application to the study of the genomic landscape of ERVs in the human and mouse genomes. Our analysis pipeline based on the ITP/IWT and the functional logistic regression allowed us to identify genomic scales and locations where various features display their influence, and to understand how they work in concert to provide signals essential for integration and fixation of ERVs. We expect the FDA techniques employed here to be very useful in several other "Omics" studies.

References

[1] Campos-Sánchez, R., Cremona, M. A., Pini, A., Chiaromonte, F., Makova, K. D.: Integration and fixation preferences of human and mouse endogenous retroviruses uncovered with functional data analysis. PLoS Comput. Biol., **12**(6):1–41, 06 (2016)

[2] Campos-Sánchez, R., Kapusta, A., Feschotte, C., Chiaromonte, F., Makova, K. D.: Genomic landscape of human, bat and ex vivo DNA transposon integrations. Mol. Biol. Evol., page msu138, (2014)

[3] Cremona, M. A., Pini, A., Chiaromonte, F., Vantini, S.: IWTomics: Interval-Wise Testing for omics data, (2016) R package version 0.99.9.

[4] Cremona, M. A., Pini, A., Makova, K. D., Chiaromonte, F., Vantini, S.: IWTomics: testing high resolution "Omics" data at multiple locations and scales. In

preparation, (2017)
[5] Cremona, M. A., Sangalli, L. M., Vantini, S., Dellino, G. I., Pelicci, P. G., Secchi, P., Riva, L.: Peak shape clustering reveals biological insights. BMC Bioinform., **16**(1):349, 10 (2015)
[6] Kvikstad, E. M., Makova, K. D.: The (r)evolution of SINE versus LINE distributions in primate genomes: sex chromosomes are important. Genome Res., **20**(5):600–613, (2010)
[7] Pini, A., Vantini, S.: The interval testing procedure: a general framework for inference in functional data analysis. Biometrics, (2016)
[8] Pini, A., Vantini, S.: Interval-wise testing for functional data. J. Nonparametr. Stat., In press, (2017)
[9] Reimherr, M., Nicolae, D.: A functional data analysis approach for genetic association studies. Ann. Appl. Stat., **8**(1):406–429, (2014)

Chapter 13
Functional data analysis in kinematics of children going to school

Manuel Escabias, Ana M. Aguilera, Jose M. Heredia-Jiménez and Eva Orantes-González

Abstract Traditionally gait analysis has examined discrete measures as descriptors of gait to compare different experimental situations. Functional data analysis (FDA) uses information from the entire curves and trajectories, thus revealing the nature of the movement. The aim of our study is to develop some repeated measures FDA methodologies to analyze kinematics of children's trunks while transporting a backpack and a trolley with different loads.

13.1 Introduction

As we can read in dictionary.com (http://www.dictionary.com/browse/kinematics), kinematics is the branch of mechanics that deals with pure motion, without reference to the masses or forces involved in it. With the name of gait analysis we identify the part of kinematics related with human motion. One of the main research areas of Sport Sciences is kinematics and gait analysis. Motion is characterized by the angular position of the different parts of the body along the gait cycle, providing one trajectory for each part of the body.

Recent advances in technology have automated much of the processes of capturing motion data electronically and then extracting the two or three-dimensional trajectories. Such technology and software allow to reconstruct the motions of body

Manuel Escabias
University of Granada, Department of Statistics and O.R. e-mail: escabias@ugr.es

Ana M. Aguilera (✉)
Univeristy of Granada, Department of Statistics and O.R. and IEMath-GR, e-mail: aaguiler@ugr.es

Jose M. Heredia-Jiménez
Univeristy of Granada, Department of Physical Education and Sport e-mail: herediaj@ugr.es

Eva Orantes-González
Univeristy of Granada, Department of Physical Education and Sport e-mail: maevor@ugr.es

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_13

segments and joints so that differences in motion patterns can be readily identified. As a result the most natural statistical methods to analyze this type of data are functional data analysis methods.

In spite of the continuous nature of trajectories, traditionally gait analysis has examined discrete measures such as angles at heel strike (HS), peak angles and range of motion (ROM) obtained during stance, or the timing of specific events (Donoghue *et al.*, 2008). Research commonly uses these parameters as descriptors of gait to compare different experimental situations. Given the huge amount of information that a fully comprehensive gait analysis can provide, studies reduce data to these descriptors to allow a more convenient and efficient analysis. However, these discrete values do not fully describe the relative movements of the limbs in attaining these positions. Functional data analysis uses information from the entire curves and trajectories, thus revealing the nature of the movement.

Ramsay and Silverman used gait data to explain some FDA methods as functional canonical correlation and functional principal component analysis in their classical book (Ramsay and Silverman, 2005). From then to now, FDA has been used in Biomechanics with different proposals. For example Ryan *et al.* (2006) used FDA to study vertical jump. Donoghue *et al.* (2008) used functional principal component analysis to compare different experimental situations in Running Kinematics in Chronic Achilles Tendon Injury. Several authors have used various PCA approaches as Sadeghi *et al* (1997) and Sadeghi (2003) to examine lower limb symmetry in several studies involving subjects with osteoarthritis and healthy subjects, or Deluzio *et al.* (1997). Daffertshofer *et al.* (2004) recognized the potential solution that a PCA approach could provide in analyzing coordination and human movement.

The aim of our study is to use existing FDA methods and propose some new approaches to analyze kinematics of children in their commuting to and from school. Children often use school trolleys or backpacks in their daily transportation to school. Recently a controversy exists about if it is better to use trolleys or backpacks, and also about the weight of them in order to avoid possible back pain. Some studies show that the use of the school trolleys allowed children to avoid supporting the load on their backs and also provided an easier mode of transportation but has an asymmetrical effort, and its use was related with a higher risk for children to suffer scoliosis (Ortega *et al.*, 2014) and forced postures of the shoulder and spine (Bort and Sim, 2002). Spatiotemporal gait parameters analysis is important to drive these kind of studies in order to determine the changes produced in the gait depending on the type of transportation (trolley or backpack) and the weight of them.

Some authors have studied kinematics of children's trunks while transporting a backpack and a trolley in different experimental situations (see for example Orantes-Gonzalez *et al.* (2015) or Schmidt and Docherty (2010)). Ir order to analyze the kinematics of gait while pulling a school trolley or by carrying a backpack with different loads in elementary school subjects, functional data analysis methods can be used instead of the classical ones that take into account only one measure instead of the whole observation.

13.2 Data

The experimental study to get the data was developed in the biomechanics laboratories of the Sport and Health Institute of the University of Granada (IMUDS). The data were provided by 53 participants (25 boys and 28 girls) between the ages of 8 and 11. Six experimental situations were considered: trolley and backpack used with different loads (10, 15, and 20% of body weight). 3D motion capture system (Qualisys AB, Göteborg, Sweden) was used to get the kinematics of gait based on the Calibrated Anatomical System Technique (Cappozzo *et al.*, 1995). Twenty six reflective markers were placed with adhesive tape on the children's skin on both sides of the lower body and the trunk following the recommendations proposed by van Sint Jan (2007). Nine infrared high-speed cameras at a capture rate of 250 Hz were used to collect the reflective markers. Markers were situated according to (Figure 13.1). Visual3D software (C-Motion Inc., Germantown, USA) was used to compute the gait kinematics. Each child walked one minute per condition along a 15 m walkway. The coordinate based algorithm (Zeni *et al.*, 2008) computed by Visual 3D, was used to calculate the spatiotemporal gait parameters. Velocity (m/s), cadence (steps/seg) and stride length (m) were normalized using the subject's height by following the equations proposed by Hoff (1996). In addition, stance phase, single support phase and double support phase were measured and expressed as a percentage of the gait cycle and step width was also computed as heel-to-heel distance.

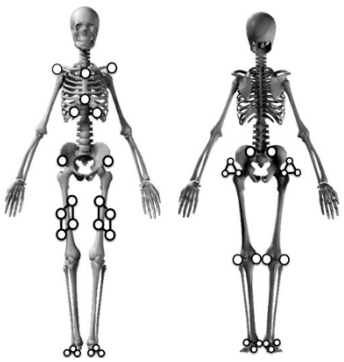


Fig. 13.1: Representation of used markers. Figure adapted from Marker Set Guidelines, C-Motion's Visual3D biomechanics research software.

The flexion/extension, adduction/abduction and internal/external rotation of thorax, pelvis, hip, knee and ankle were recorded. Pelvis angles were expressed as the absolute angles of the segments with regard to a global reference coordinate system. The hip angle was defined by the pelvis and femur; the knee angle was computed by the tight and shank; the ankle angle was computed by the foot and shank; and the hip and thorax segments were used to compute the thorax angle. The three dimensional

kinematic parameters were obtained at the data mean of both legs and normalized relative to the duration of the walking cycle of each subject.

Each subject completed 3 cycles walking over the platform and registered the angle for each join (ankle, foot progress, hip, knee, pelvis, thorax), each axis direction (X, Y and Z, except in foot progress that only has one direction) and each experimental condition (walk, backpack with 10% of the body weight of the subject, backpack with 15%, backpack with 20%, trolley with 10% of the body weight of the subject, trolley with 15% and trolley with 20%). For each subject the mean curve of the three cycles was calculated and selected as representative observation for the gait cycle of that subject in each join, axis direction and experimental condicion. The resulting data were 53 discretely observed curves for each join, axis direction and experimental condition. Functional data analysis was used in order to determine differences among gait curves in each one of the 3 axis directions (and possibly in the resultant) and each one of the 6 joins (18 comparisons) for the seven experimental conditions. We used basis expansion methods in terms of B-spline functions to turn the discrete observations of each curve into the proper curve. Mean curves for each join, direction and experimental condition can be seen in Figure 13.2.

The main purpose of this work is to find differences and possible relationships among the gait cycle under the different loads for each angle and join independently. Let us observe that we have repeated measures of the gait cycle under each experimental condition because the same children are considered to take the measures for all of them. The comparison of the means under each condition have been historically made by using a single measure as representative of the mean curve instead of the complete curve: sometimes the mean on the discrete values, the maximum, the minimum, etc. By using the complete curve, we can find differences in the whole curve or maybe parts of the curve that can detect periods of the cycle that may produce pain in that join. Figure 13.2 shows comparisons of mean curves among the different experimental conditions for each one of the joins and angles considered. We can visually observe differences in the mean curves for the case of X-axis for pelvis and thorax and Z-axis for hip, knee, pelvis and thorax.

In addition to the curves, we have scalar observations of related variables for each subject obtained from a questionnaire. Among the available variables we have: age, gender, degree, backpack (trolley) weight, subject height and weight, body mass index, fat percentage, back pain, and many others. The relationship between these variables and gait curves could be studied by using functional regression models.

13.3 Functional data analysis with repeated measures

Let us observe that for each join and angle direction (as for example X), we have $n \times J$ sample curves

$$\{X_{ij}(t) : t \in T, i = 1, \ldots, n, j = 1, \ldots, J\},$$

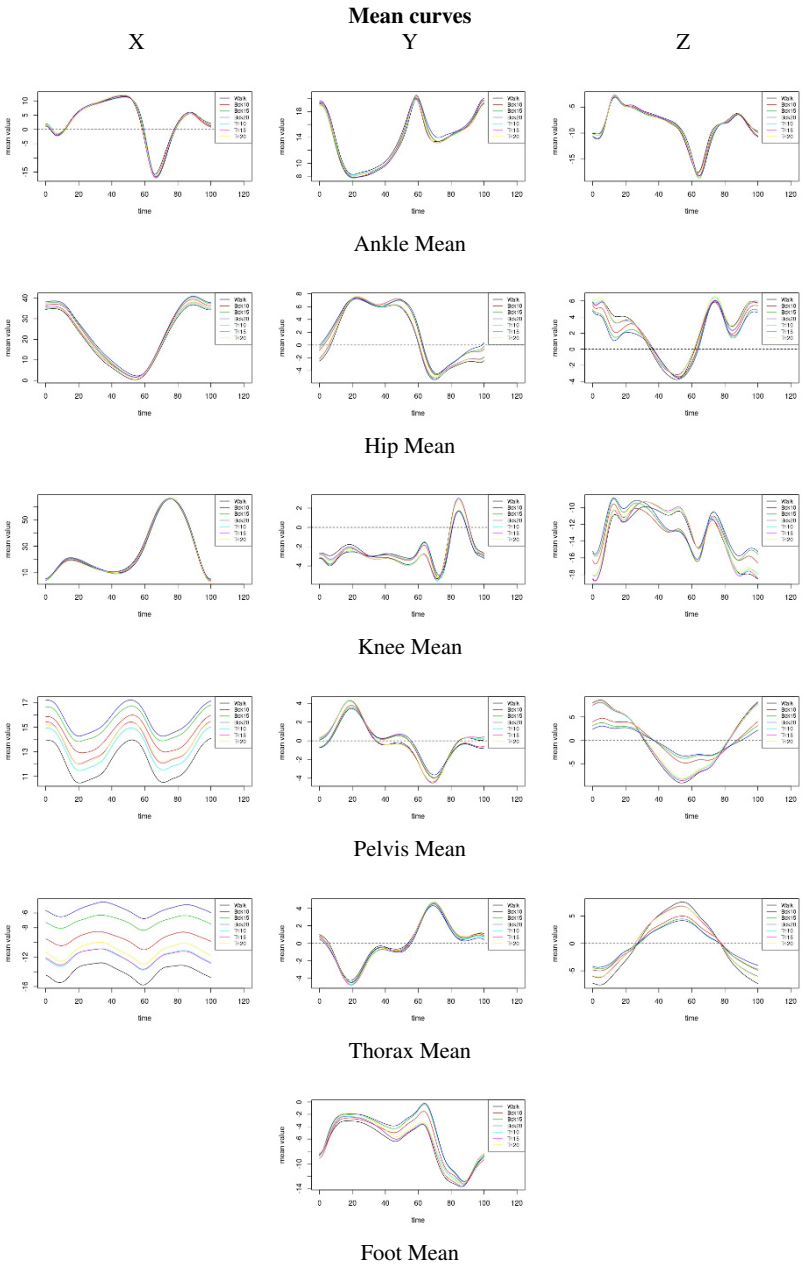


Fig. 13.2: Mean curves for each join, angle direction and experimental condition.

from n random subjects (in our experiment $n = 53$ children), drawn from a stochastic process $\{X_j(t) : t \in T\}$ representing the gait cycle under each one of the $J = 6$ considered conditions. Let us assume that these stochastic processes are second order, continuous in quadratic mean and with squared integrable trajectories on $L^2(T)$. The general aim of this research project is using FDA methodologies to explain the behaviour of gait cycle among the different conditions.

Next, different methodological possibilities to explore variability and testing the equality of the means curves for each join and angle direction under the different loading conditions are addressed. Let us take into account that it is an initial work and the results obtained so far are not definitive.

13.4 Functional principal component analysis

The main modes of variability in the functional data set could be studied by adapting different approaches for functional principal component analysis (FPCA) with longitudinal data to the case of functional data with repeated measures (Guo, 2004, Müller, 2005, Greven, 2011). The most basic way is developing a Functional PCA for the subjects of each condition so that for each fixed j we have

$$X_j(t) = \mu_j(t) + \sum_k \xi_k^j f_k^j(t) \;\; t \in T,$$

with the principal components ξ_k^j being zero mean uncorrelated random variables with maximum variance given by

$$\xi_k^j = \int_T (X_j(t) - \mu_j(t)) f_k^j(t) dt.$$

That is, $E[\xi_k^j] = 0$, $Var[\xi_k^j] = \lambda_k^j$, and for any j, $k_1 \neq k_2, E(\xi_{k_1}^j, \xi_{k_2}^j) = 0$, and $\{(f_k^j, \lambda_k^j)\}$ being the eigenfunctions and eigenvalues of the covariance function in the j-th group $C_j(s,t) = cov\{X_j(s), X_j(t)\}$.

Then, principal components under different conditions would be related for interpreting important differences in the gait cycle. A more sophisticated multilevel FPCA could also be obtained by a "total" subject decomposition of variability based on the "between" and "within" subject principal component decompositions as proposed in Crainiceanu et al. (2009).

13.5 ANOVA modelling

Let us consider the classical one-way ANOVA problem for functional data. As we have said before this research is motivated by a gait analysis where the problem is to compare the mean curves of the gait cycle measured in J samples (seven different

loading conditions for walking) for each of the three angle directions (X, Y, Z) and joins. The hypothesis to test is given by

$$H_0 : \mu_1 = \cdots = \mu_J,$$

where $\mu_j = E[X_j]$.

Literature has boarded this aim from different points of view by giving different proposals of ANOVA methods for functional data. Ramsay and Silverman (2005) illustrated the use of a pointwise ANOVA method to find differences in mean curves among different climate zones in their classical example of weather data. An asymptotic version of the well known ANOVA F-test to detect differences in the whole curve was proposed in Cuevas *et al.* (2004) and Cuesta-Albertos and Febrero-Bande (2010). Recently, Zhang (2014) has summarized different ANOVA methods for functional data in a very useful and illustrative book. The case of repeated measures which deals with the k-sample problem when the data are from the same subjects was investigated from both, the parametric and the nonparametric point of view, in Martínez-Camblor and Corral (2011).

The above mentioned methods will be adapted in this work to the case of testing the equality of means of dependent samples of functional data. The aim is to find those methods that better fit to our objective of finding global or local differences in mean curves that help analysts to detect possible problems in children, as, for example, back pain.

The starting point will be a direct analog of the classical F statistic for the one-way ANOVA model with repeated measures given by

$$F_n = \frac{\sum_{j=1}^{J} n||\bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot}||^2/(J-1)}{\left\{\left[\sum_{j=1}^{J}\sum_{i=1}^{n}||X_{ij} - \bar{X}_{\cdot j}||^2\right] - [J\sum_{i=1}^{n}||\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}||^2]\right\}/(n-1)(J-1)}$$

where

$$\bar{X}_{i\cdot}(t) = \frac{1}{J}\sum_{j=1}^{J} X_{ij}(t),\ \ \bar{X}_{\cdot j}(t) = \frac{1}{n}\sum_{i=1}^{n} X_{ij}(t),\ \ \bar{X}_{\cdot\cdot}(t) = \frac{1}{J\times n}\sum_{i=1}^{n}\sum_{j=1}^{J} X_{ij}(t),$$

and $||x||^2 = \int_T x^2(t)dt$.

Similar to the classical ANOVA F statistic for repeated measures, the numerator represents the "external" variability between the different samples or groups, and the denominator measures de "internal" variability within the samples minus the "subjects" variability due to the treatment of each subject as a level of an independent factor.

Acknowledgements This research has been funded by project P11-FQM-8068 from Consejería de Innovación, Ciencia y Empresa. Junta de Andalucía, Spain and project MTM2013-47929-P from Secretaría de Estado de Investigación, Desarrollo e Innovación, Ministerio de Economía y Competitividad, Spain. The work of Eva Orantes-González was supported by the Ministry of

Education, Culture and Sports of Spain (Ref. FPU13/00162, EST15/00019). The authors thank all of the participants: children and parents that took part in this study.

References

[1] Bort, S.N., Simó, P.A.: Carritos o mochilas en la edad escolar. Fisioterapia **24**(2), 63-72 (2002)
[2] Cappozzo, A., Catani, F., Croce, U.D., Leardini, A.: Position and orientation in space of bones during movement: anatomical frame definition and determination. Clin. biomech. **10**(4), 171-178 (1995)
[3] Cuesta-Albertos, J.A. and Febrero-Bande, M.: A simple multiway ANOVA for functional data. Test **19**(3), 537-557 (2010)
[4] Cuevas, A., Febrero, M. and Fraiman, R.: An anova test for functional data. Comput. Statist. Data Anal. **47**(1), 111-122 (2004)
[5] Daffertshofer, A., Lamoth, C.J.C., Meijer, O.G., Beek, P.J.: PCA in studying coordination and variability: a tutorial. Clin. Biomech. **19**(4), 415-428 (2004)
[6] Deluzio, K.J., Wyss, U.P., Zee, B., Costigan, P.A., Serbie, C.: Principal component models of knee kinematics and kinetics: normal vs. pathological gait patterns. Hum. Mov. Sci. **16**(2-3), 201-217(1997)
[7] Di, C.-Z., Crainiceanu, C.M., Caffo, B.S., Punjabi, N.M.: Multilevel functional principal component analysis. Ann. Appl. Stat. **3**(1), 458488 (2009)
[8] Donoghue, O.A., Harrison, A.J., Coffey, N., Hayes, K.: Functional Data Analysis of Running Kinematics in Chronic Achilles Tendon Injury. Medicine and science in sport and exercise **40**(7), 1323–1335 (2008)
[9] Greven, S., Crainiceanu, C., Caffo, B., Reich, D.: Longitudinal functional principal component analysis. Electron. J. Stat. **4**, 10221054 (2011)
[10] Guo, W.: Functional data analysis in longitudinal settings using smoothing splines. Stat. Methods. Med. Res. **13**(1), 49-62 (2004)
[11] Hof, A.L.: Scaling gait data to body size. Gait Posture **4**(3), 222-223 (1996)
[12] Martínez-Camblor, P., Norberto Corral, N.: Repeated measures analysis for functional data. Comput. Statist. Data Anal. **55**, 32443256 (2011)
[13] Müller, H-G.: Functional modelling and classification of longitudinal data. Scand. J. Stat. **32**, 223240 (2005)
[14] Orantes-González, E., Heredia-Jiménez, J., Soto-Hermoso, V.M.: The effect of school trolley load on spatiotemporal gait parameters of children. Gait and Posture **42**(3), 390-393 (2015)
[15] Ortega, F.Z., Sánchez, M.F., García, R.F., Schyke, C.E.J., Morales, L.Z.: Predictors of scoliosis in school-aged children. Gaceta Médica de México **150**, 524-530 (2014)
[16] Sadeghi, H., Local or global asymmetry in gait of people without impairments. Gait and Posture **17**(3), 197-204 (2003)
[17] Sadeghi, H., Allard, P., Duhaime, M.: Functional gait asymmetry in able-bodied subjects. Hum. Mov. Sci. **16**(2-3), 243-58(1997)

[18] Schmidt, J., Docherty, S.: Comparison of the posture of school children carrying backpacks versus pulling them on trolleys. Clin. Chiropr. **13**(4), 253-260 (2010)
[19] Ramsay, J. and Silverman, B.D.: Functional data analysis. Springer, Heidelberg (2005)
[20] Zang, J.T.: Analysis of Variance for Functional Data. CRC Press, (2014)
[21] Zeni J.A., Richards, J.G., Higginson, J.S.: Two simple methods for determining gait events during treadmill and overground walking using kinematic data. Gait Posture **27**(4), 710-714 (2008)

Chapter 14
Parameter estimation of the functional linear model with scalar response with responses missing at random

Manuel Febrero-Bande, Pedro Galeano and Wenceslao González-Manteiga

Abstract This contribution considers estimation of the parameters of the functional linear model with scalar response when some of the responses are missing at random. We consider two different estimation methods of the functional slope of the model and analyze their characteristics. Simulations and the analysis of a real data example provides some insight into the behavior of both estimation procedures.

14.1 Introduction

The functional linear model with scalar response is one of the most widely studied model in the literature on functional data analysis. The model establishes a linear relationship between a real response variable and a functional predictor variable. There exist several estimators of the functional slope of the model being the method based on functional principal components the most popular approach. The idea behind this method is that of expanding the functional predictor as well as the functional slope of the model in terms of the eigenfunctions linked to the largest eigenvalues of the functional predictor covariance operator, that allows the response to be written as a finite linear combination of the functional principal components scores. The associated coefficients are then estimated by least squares.

Manuel Febrero-Bande
Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, e-mail: manuel.febrero@usc.es

Pedro Galeano (✉)
Department of Statistics and UC3M-BS Institute of Financial Big Data, Universidad Carlos III de Madrid e-mail: pedro.galeano@uc3m.es

Wenceslao González-Manteiga
Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, e-mail: wenceslao.gonzalez@usc.es

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_14

Several papers have analyzed the properties of the functional principal components estimation method including [1, 2, 3, 4, 7, 9, 5], among others. See also [6], for a recent overview on the topic.

This contribution considers the case in which some of the responses are missing at random. This case has been little studied in the literature. [10] investigated the asymptotic properties of a kernel type estimator of the regression operator when there are responses missing at random, while [5] considered an imputation method of the missing responses. Here, we propose two estimators of the functional slope of the model. The first one is simplified estimator that only considers the complete pairs of observations. The second one is an imputed estimator that takes into account both the complete pairs and pairs completed with imputed responses.

The rest of this contribution is structured as follows. Section 14.2 presents the functional linear model with scalar response and the estimation method based on the functional principal components approach. Section 14.3 considers the problem of estimating the parameters of the model when there are responses that are missing at random and presents the two estimators that we propose of the functional slope of the model. Properties of the estimators, simulations and the analysis of real data are presented somewhere else.

14.2 The functional linear model with scalar response

Let $L^2(T)$, the separable Hilbertian space of squared integrable functions defined on the closed interval $T = [a,b] \subset \mathbb{R}$. Let χ be a functional random variable valued in $L^2(T)$ and let $\chi(t)$ be the value of χ at any point $t \in T$. We assume, for simplicity, that the functional random variable χ has zero mean function and a covariance operator Γ such that:

$$\Gamma(\eta) = E[(\chi \otimes \chi)(\eta)] = E[\langle \chi, \eta \rangle \chi]$$

for any $\eta \in L^2(T)$, where,

$$\langle \chi, \eta \rangle = \int_T \chi(t)\,\eta(t)\,dt$$

is the usual inner product in $L^2(T)$. We also assume that $E\left[\|\chi\|^2\right] < \infty$, where $\|\cdot\|$ denotes the usual norm in $L^2(T)$. Consequently, Γ has a sequence of non-negative eigenvalues, denoted by $a_1 > a_2 > \cdots > 0$, such that $\sum_{k=1}^{\infty} a_k < \infty$, associated with a sequence of orthonormal eigenfunctions, denoted by $\psi_1, \psi_2, \ldots$, such that $\Gamma(\psi_k) = a_k \psi_k$, for $k = 1, 2, \ldots$

The functional linear model with scalar response relates a real random variable y, defined on the same probability space that χ, with mean 0 and variance σ_y^2, with χ as follows:

$$y = \langle \chi, \beta \rangle + e = \int_T \chi(t)\,\beta(t)\,dt + e, \tag{14.1}$$

where $\beta \in L^2(T)$ is the functional slope of the model, and e is a real random variable with mean 0, finite variance σ_e^2, and uncorrelated with χ. In other words, we assume that the mean and variance of y conditional on χ are given by $E_\chi[y] = \langle \chi, \beta \rangle$ and $Var_\chi[y] = \sigma_e^2$, respectively.

As mentioned in the introduction, the functional principal components estimation method is the most popular approach to estimate the functional slope β of the model in (14.1). This is because the functional principal components allows the functional linear model to be more easily written. The functional principal components scores, given by $s_k = \langle \chi, \psi_k \rangle$, for $k = 1, 2 \ldots$, are uncorrelated univariate random variables with mean 0 and variance a_k that allows the Karhunen-Loève expansion of the functional random variable χ to be written as follows:

$$\chi = \sum_{k=1}^{\infty} s_k \psi_k. \tag{14.2}$$

Similarly, the functional slope β can be also written in terms of the eigenfunctions $\psi_1, \psi_2, \ldots$ as:

$$\beta = \sum_{k=1}^{\infty} b_k \psi_k, \tag{14.3}$$

where $b_k = \langle \beta, \psi_k \rangle$, for $k = 1, 2 \ldots$ are constant coefficients. Now, (14.2) and (14.3) allows the functional linear model with scalar response to be written as:

$$y = \sum_{k=1}^{\infty} b_k s_k + e,$$

which shows that the coefficients b_k can be written as:

$$b_k = \frac{Cov[y, s_k]}{a_k}, \tag{14.4}$$

for $k = 1, 2 \ldots$ where $Cov[y, s_k] = E[y s_k]$ is the covariance between the real response y and the k-th functional principal component score s_k.

Assume now that we are given a random sample of independent pairs, given by $\{(\chi_i, y_i),\ i = 1, \ldots, n\}$, drawn from the random pair (χ, y). Then, the functional slope β in the model (14.1) can be estimated with the functional principal component estimation method as follows. Let $\chi_C = \{\chi_1, \ldots, \chi_n\}$ and $y_C = \{y_1, \ldots, y_n\}$ be the complete sequences of predictors and responses, respectively. Then, the sample covariance operator of the complete sample χ_C, that converts any function $\eta \in L^2(T)$ into another function in $L^2(T)$ given by:

$$\widehat{\Gamma}_{\chi_C}(\eta) = \frac{1}{n} \sum_{i=1}^{n} \langle \chi_i, \eta \rangle \chi_i,$$

is an estimate of the covariance operator of χ, Γ. The sample covariance operator $\widehat{\Gamma}_{\chi_C}$ also has a sequence of non-negative eigenvalues, denoted by $\widehat{a}_{1,C} \geq \widehat{a}_{2,C} \geq \cdots$,

such that $\widehat{a}_{k,C} = 0$, for $k > n$, and a set of orthonormal eigenfunctions, denoted by $\widehat{\psi}_{1,C}, \widehat{\psi}_{2,C}, \ldots$, such that $\widehat{\Gamma}_{\chi_C}\left(\widehat{\psi}_{k,C}\right) = \widehat{a}_{k,C}\widehat{\psi}_{k,C}$, for $k = 1, 2, \ldots$ Additionally, the k-th sample functional principal component score of χ_i, $i = 1, \ldots, n$, based on the complete sample χ_C, is given by $\widehat{s}_{i,k,C} = \left\langle \chi_i, \widehat{\psi}_{k,C} \right\rangle$, for $k = 1, 2, \ldots$ The set of sample functional scores $\widehat{s}_{1,k,C}, \ldots, \widehat{s}_{n,k,C}$ has sample mean 0 and sample variance $\widehat{a}_{k,C}$. Now, the functional principal components estimate of the functional slope β is given by:

$$\widehat{\beta}_{k_C,C} = \sum_{k=1}^{k_C} \widehat{b}_{k,C}\widehat{\psi}_{k,C}, \tag{14.5}$$

where $\widehat{b}_{k,C}$ is an estimate of the coefficient b_k in (14.4) given by:

$$\widehat{b}_{k,C} = \begin{cases} \frac{1}{n\widehat{a}_{k,C}} \sum_{i=1}^{n} y_i \widehat{s}_{i,k,C} & \text{for } k = 1, \ldots, k_C \\ 0 & \text{for } k = k_C + 1, \ldots \end{cases}$$

and k_C is a certain threshold such that $\widehat{a}_{k_C,C} > 0$. Consequently, given a new value χ, say χ_{n+1}, the prediction of the corresponding response under the model (14.1), denoted by y_{n+1}, is given by:

$$\widehat{y}_{n+1,k_C,C} = \left\langle \chi_{n+1}, \widehat{\beta}_{k_C,C} \right\rangle.$$

See [9], [7] and [6], for finite sample properties of the slope estimate (14.5).

14.3 Estimation and prediction with responses missing at random

Assume now the situation in which we are given a random sample of independent triplets $\{(\chi_i, y_i, r_i), i = 1, \ldots, n\}$ drawn from the random triplet (χ, y, r), where r is a Bernoulli variable that acts as an indicator of the missing responses. Thus, for $i = 1, \ldots, n$, $r_i = 1$, if y_i is observed, and $r_i = 0$, if y_i is missing. Specifically, we assume a missing at random (MAR) mechanism, i.e.:

$$\Pr(r = 1|y, \chi) = \Pr(r = 1|\chi) = p(\chi),$$

where $p(\chi)$ is an unknown function operator of χ. As a consequence, the response y and the binary variable r are independent given the predictor χ.

Now, the goal is to estimate the functional slope β in (14.1) using the sample $\{(\chi_i, y_i, r_i), i = 1, \ldots, n\}$. For that, let $r_C = (r_1, \ldots, r_n)'$ be the complete sequence of missing indicators. Two different estimates are introduced next based on the missing indicators r_C.

The first estimate is the simplified functional principal component estimate, that uses only the complete pairs, i.e., those pairs with $r_i = 1$, for $i = 1, \ldots, n$. Then, let

$I_S = \{i : r_i = 1, i = 1, \ldots, n\}$, i.e., the indices of the complete pairs and let $n_S = \#I_S$, i.e., the number of observed complete pairs. Additionally, let $\chi_S = \{\chi_i : i \in I_S\}$ and $y_S = \{y_i : i \in I_S\}$, i.e., the sequences of predictors and responses, respectively, corresponding to the complete pairs. Then, the sample covariance operator of χ_S, that converts any function $\eta \in L^2(T)$ into another function in $L^2(T)$ given by:

$$\widehat{\Gamma}_{\chi_S}(\eta) = \frac{1}{n_S} \sum_{i=1}^{n} r_i \langle \chi_i, \eta \rangle \chi_i = \frac{1}{n_S} \sum_{i \in I_S} \langle \chi_i, \eta \rangle \chi_i,$$

is an estimate of Γ. As in the complete case developed in Section 14.2, $\widehat{\Gamma}_{\chi_S}$ has a sequence of non-negative eigenvalues, denoted by $\widehat{a}_{1,S} \geq \widehat{a}_{2,S} \geq \cdots$, such that $\widehat{a}_{k,S} = 0$, for $k > n_S$, and a set of orthonormal eigenfunctions, denoted by $\widehat{\psi}_{1,S}, \widehat{\psi}_{2,S}, \ldots$, such that $\widehat{\Gamma}_{\chi_S}\left(\widehat{\psi}_{k,S}\right) = \widehat{a}_{k,S}\widehat{\psi}_{k,S}$, for $k = 1, 2, \ldots$ Additionally, the k-th sample functional principal component score for χ_i, $i \in I_S$, based on the simplified sample χ_S, is given by $\widehat{s}_{i,k,S} = \left\langle \chi_i, \widehat{\psi}_{k,S} \right\rangle$, for $k = 1, 2, \ldots$ The set of sample functional components scores $\left\{\widehat{s}_{i,k,S} : i \in I_S\right\}$ have sample mean 0 and sample variance $\widehat{a}_{k,S}$. Now, the simplified functional component estimate of the functional slope β is given by:

$$\widehat{\beta}_{k_S,S} = \sum_{k=1}^{k_S} \widehat{b}_{k,S}\widehat{\psi}_{k,S}, \tag{14.6}$$

where $\widehat{b}_{k,S}$ is an estimate of the coefficient b_k in (14.4) given by:

$$\widehat{b}_{k,S} = \begin{cases} \frac{1}{n_S \widehat{a}_{k,S}} \sum\limits_{i \in I_S} y_i \widehat{s}_{i,k,S} & \text{for } k = 1, \ldots, k_S \\ 0 & \text{for } k = k_S + 1, \ldots \end{cases}$$

and k_S is a certain threshold such that $\widehat{a}_{k_S,S} > 0$. Prediction of the response y_{n+1} corresponding to a new predictor χ_{n+1} under the model (14.1), is given by:

$$\widehat{y}_{n+1,k_S,S} = \left\langle \chi_{n+1}, \widehat{\beta}_{k_S,S} \right\rangle.$$

The second estimate is the imputed functional principal component estimate, that uses both the complete pairs and the pairs obtained after imputing the missing responses with the estimate (14.6). Then, let $I_I = \{i : r_i = 0, i = 1, \ldots, n\}$, i.e., the indices of the pairs with missing responses and let $n_I = \#I_I$, i.e., the number of pairs with missing responses. Additionally, let $\chi_I = \{\chi_i : i \in I_I\}$ and $y_I = \{y_i : i \in I_I\}$, i.e., the sequences of predictors and responses, respectively, corresponding to the pairs with missing responses. Therefore, imputation of the missing responses using the simplified estimate $\widehat{\beta}_{k_S,S}$ in (14.6) can be done as follows:

$$\widehat{y}_{i,I} = \left\langle \chi_i, \widehat{\beta}_{k_S,S} \right\rangle,$$

for $i \in I_I$. Now, given the set of pairs $\{(\chi_i, y_{i,I}), i = 1, \ldots, n\}$ where:

$$y_{i,I} = r_i y_i + (1 - r_i)\widehat{y}_{i,I},$$

for $i = 1, \ldots, n$, the imputed functional principal component estimate of the functional slope β is given by:

$$\widehat{\beta}_{k_I,I} = \sum_{k=1}^{k_I} \widehat{b}_{k,I}\widehat{\psi}_{k,C}, \tag{14.7}$$

where $\widehat{b}_{k,I}$ is an estimate of the coefficient b_k in (14.4) given by:

$$\widehat{b}_{k,I} = \begin{cases} \frac{1}{n\widehat{a}_{k,C}} \sum_{i=1}^{n} y_{i,I}\widehat{s}_{i,k,C} & \text{for } k = 1, \ldots, k_I \\ 0 & \text{for } k = k_I + 1, \ldots \end{cases}$$

and k_I is a certain threshold such that $\widehat{a}_{k_I,C} > 0$. Two important comments are in order. First, $\widehat{\beta}_{k_I,I}$ depends on the eigenfunctions and eigenvalues of the sample covariance operator $\widehat{\Gamma}_{\chi_C}$ based on the complete set of predictors χ_C. Second, the threshold k_I in (14.7) does not necessarily coincides with the threshold k_S in (14.6). Prediction of the response y_{n+1} corresponding to a new predictor χ_{n+1} under the model (14.1), is given by:

$$\widehat{y}_{n+1,k_I,I} = \left\langle \chi_{n+1}, \widehat{\beta}_{k_I,I} \right\rangle.$$

Acknowledgements The first and third author acknowledges financial support from Ministerio de Economía y Competitividad grant MTM2013-41383-P. The second author acknowledges financial support from Ministerio de Economía y Competitividad grant ECO2015-66593-P.

References

[1] Cai, T.T., Hall, P.: Prediction in functional linear regression. Ann. Stat. **34**, 2159–2179 (2006)
[2] Cardot, H., Ferraty, F., Sarda, P.: Functional linear model. Stat. Probabil. Lett. **45**, 11–22 (1999)
[3] Cardot, H., Ferraty, F., Sarda, P.: Spline estimators for the functional linear model. Stat. Sinica. **13**, 571–591 (2003)
[4] Cardot, H., Mas, A., Sarda, P.: CLT in functional linear regression models. Probabil. Theory and Relat. Fields. **138**, 325–361 (2007)
[5] Crambes, C., Henchiri, Y.: Regression imputation in the functional linear model with missing values in the response. Manuscript.
[6] Febrero-Bande, M., Galeano, P., González-Manteiga, W.: Functional principal component regression and functional partial least-squares regression: an overview and a comparative study. Int. Stat. Rev. (2016) doi: 10.1111/insr.12116
[7] Ferraty, F., González-Manteiga, W., Martínez-Calvo, A., Vieu, P.: Presmoothing in functional linear regression. Stat. Sinica. **22**, 69–94 (2012)

[8] Hall, P., Horowitz, J. L.: Methodology and convergence rates for functional linear regression. Ann. Stat. **35**, 70–91 (2007)
[9] Hall, P., Hosseini-Nasab, M.: On properties of functional principal components analysis. J. Roy. Stat. Soc. B **68**, 109–126 (2006)
[10] Ling, N., Ling, L., Vieu, P.: Nonparametric regression estimation for functional stationary ergodic data with missing at random. J. Stat. Plan. Infer. **162**, 75–87 (2015)

Chapter 15
Variable selection in Functional Additive Regression Models

Manuel Febrero-Bande, Wenceslao González-Manteiga and Manuel Oviedo de la Fuente

Abstract This paper considers the problem of variable selection when some of the variables have a functional nature and can be mixed with other type of variables (scalar, multivariate, directional, etc). Our proposal begins with a simple null model and sequentially selects a new variable to be incorporated into the model. For the sake of simplicity, this paper only uses additive models. However, the proposed algorithm may assess the type of contribution (linear, non linear, ...) of each variable. The algorithm have showed quite promising results when applied to real data sets.

15.1 Introduction

The variable selection problem tries to find the best subset of covariates that better predicts or explains a response. In the classical approach, the covariates and the response are scalar (or multivariate) and the model established among them is linear.

The Stepwise regression, the most widely-used model selection technique throughout the 80's and the 90's, is rooted in the classical papers by [1], [12], [16] and [17]. The main idea is to use some diagnostic tools, directly derived from the linear model, to assess whether a new covariate must be included in the model or whether its contribution to the model is valuable. Usually, the construction of the final subset is done using mainly two strategies: the forward selection that begins with a simple

Manuel Febrero-Bande (✉)
Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, e-mail: manuel.febrero@usc.es

Wenceslao González-Manteiga
Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, e-mail: wenceslao.gonzalez@usc.es

Manuel Oviedo de la Fuente
Technological Institute for Industrial Mathematics and Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela e-mail: manuel.oviedo@usc.es

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_15

null model and tests at each step the inclusion of a new covariate in the model; and the backward one that starts with the full model including all candidate variables and at each step removes the most insignificant one. Also, it is possible to mix both strategies, testing at each step which variables can be included or excluded in the optimal regression subset of covariates. In any case, the stepwise regression is anchored in the diagnostics of the linear model; it is therefore blind to detect contributions other than the linear one among the covariates and the response.

The work by [19] proposing the LASSO estimator opens a new direction on variable selection procedures. The main innovation of the LASSO estimator is that it includes a l_1-type constraint for the coefficient vector β to force some parameters (components of β) to be equal to zero and thereby obtains the optimal subset of covariates such as those with non-zero coefficients. The effect of the constraint also helps in the optimization step and satisfactorily deals with the sparsity phenomenon. See [22] for a revision of the oracle properties of LASSO. In the literature we can find interesting examples following the same line but using different penalties or constraints: LARS ([4]), SCAD ([5]), COSSO ([9]), Additive models ([21]) and extensions to partial linear, additive or semiparametric models like PLM ([3]), APLM ([10]) or GAPLM ([20]). All these works have two common characteristics: each paper is based on a specific model and all the covariates must be included in the model at once. The latter leads to a highly demanding computing algorithms that are sometimes hard to implement. In particular, for high-dimensional or functional data problems, the previous steps commonly including variable standardization and/or variable representation may notably increase the complexity and the cost of the algorithms. See [8] for a nice review of some of the aforementioned methods.

Another type of strategy is a pure feature selection where the covariate is selected without a model. This is the approach employed in mRMR (minimum Redundancy Maximum Relevance, [14]). To enter into the model, a new candidate covariate must have a great relevancy with the response while maintaining a lower redundancy with the covariates already selected in the model. The main advantage of this approach is that it is an incremental rule; once a variate has been selected, it cannot be deselected in a later step. On the other hand, the measures for redundancy and relevancy must be chosen according to the regression model to be applied, to ensure good predictive results in the final model. The FLASH method proposed in [15] is a modification of the LASSO technique that sequentially includes a new variate changing the penalty at each step. This greedy increasing strategy is also employed by Boosting (see, for example, [2] or [7] in a functional data context). Boosting is not a purely feature selection method but rather a predictive procedure that selects at each step the best covariate/model respect to the unexplained part of the response. The final prediction is constructed as a combination of the different steps. All the previous solutions are not completely satisfactory in a functional data framework specially when the number of possible covariates can be arbitrarily large. In particular, we are interested in a automatic regression procedure capable of dealing with a large number of covariates of different nature, possibly with many closed relationships among them.

Our motivating example comes from the energy market. In Figure 15.1 the daily profile of Price and Energy (Electricity) Market Demand (both measured hourly) at

the Iberian Energy Market for 630 days of 2008 and 2009 are shown with a color code considering three groups (TWT (Tue, Wed & Thu), WEnd (Sat & Sun) and MF (Mon & Fri)). We are interested in predicting the price or the demand at certain hours of the following day. For doing that, we have a lot of different sources of information. These sources can include other variables related with energy market or generation, meteorological information, calendar effects or any transformation/filter of the preceding. Figures 15.2 and 15.3 show a small sample with some possibly covariates. Note that, in this case, the number of variables that can be included in the model is pretty high (over 150).

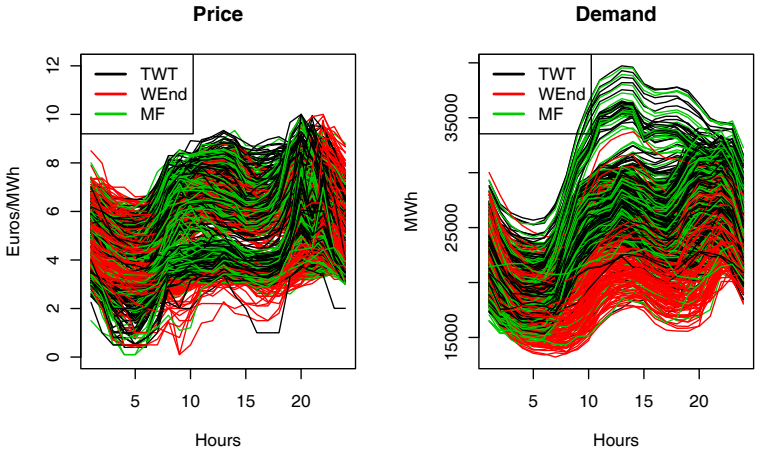


Fig. 15.1: Price and Energy Market Demand curves negotiated in the Iberian Energy Market. Period: 2008-2009 (Source omie.es).

The curves in Figure 15.2 are the components of the energy pool by generation type, which may be useful for predicting price or demand. Due to the Iberian market regulations, not all types of energy have the same chance to become part of the final energy pool consumed due to its price or availability. For instance, hydroelectric energy is only offered to the market when the price is high or when there are founded expectations on refilling the reservoir (using the weather forecasts).

The possible scalar covariates for our prediction problem can be related with calendar effects (like month or day-of-week), meteorological information (like Temperature or Wind Speed) and transformations from the functional variables (such as the market demand at midday $X(t)\mathbb{1}_{\{t=12h\}}$). This means we can create a lot of new variables from the original ones (for instance, using derivatives or considering certain subintervals), many of which may certainly be closely related.

Our aim is to select the covariates that are important for our additive regression model with scalar response:

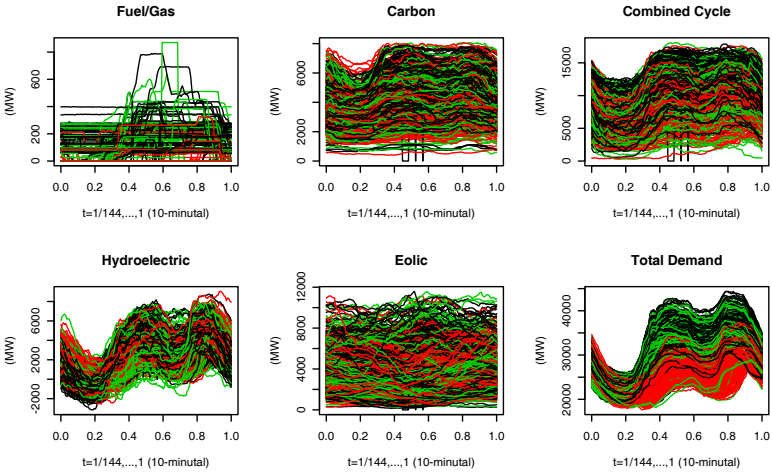


Fig. 15.2: Daily profile curves of amount of generated energy by type. Period: 2008-2009.

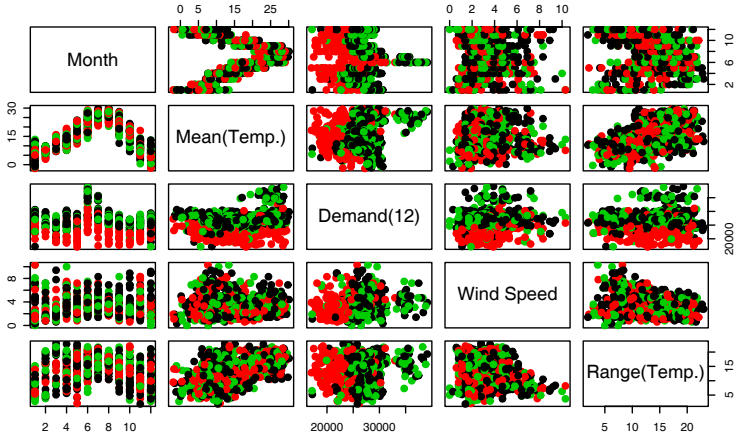


Fig. 15.3: Examples of available scalar covariates for the energy prediction problem.

$$Y_i = \alpha + \sum_{j=1}^{K} f_j\left(X_i^{(j)}\right) + \varepsilon_i, \quad i = 1,\ldots,n$$

where the covariates are chosen from the possibly huge set $S = \{X^1, X^2, \cdots, X^k, \cdots\}$. As the number of variates can be extraordinarily large, our idea is to construct the regression model in a sequential way, i.e. from the trivial model up to the one that includes all the useful information provided by the covariates in the set.

The rest of this paper is structured as follows. Section 15.2 presents our proposal and Section 15.3 includes the applications to our motivating example.

15.2 The procedure

Our proposal borrows some ideas from other variable selection methods. As in the case of pure feature methods, we separate the selection process from the estimation step picking the possible candidates using a measure not based in a concrete model. The method for selecting the covariate must be impartial in the sense that scalar, multivariate, directional or functional covariates can be comparable. Then, a model is chosen among those in a catalog that are applicable to the selected covariates and the response, ensuring that the model is correct i.e. the structural hypothesis of the model are fulfilled and its performance is superior to its competitors. Then, a test is performed to compare the new tentative model with the previous one to decide if the new covariate is definitely incorporated into the model. Finally, as in the case of Boosting, the incorporation of a subsequent covariates is done using the residuals of the actual model trying to capture the information not chosen by the previous ones. This sharply constrast with the rules of a pure feature method. This design is easy to implement and can be easily distributed along parallel machines.

As aforementioned, the key idea for the selection of a feature is to find a tool that can be homogeneously applied to variates of different nature. This paper is based on the exhaustive use of the distance correlation, $\mathscr{R}$, proposed by [18] that fulfills the following two conditions:

i) $\mathscr{R}(X,Y)$ is defined for X and Y in arbitrary finite dimensions.
ii) $\mathscr{R}(X,Y) = 0$ characterizes independence of X and Y.

These conditions mean that there is a way to measure the relationship among X and Y homogeneously for arbitrary dimensions of these vectors. Indeed, the work by [11] extends the properties for metric spaces of a certain type (*strong negative type*). In particular, Hilbert spaces or any space that can be embedded in a Hilbert space are of strong negative type. Also, as $\mathscr{R}$ characterizes independence, the distance correlation can detect relationships among variables other than the linear one. Finally, the computation is quite straightforward because only depends on the distances among data. Specifically, being $a_{kl} = d(X_k^j, X_l^j)$, the distance in X^j among cases k and l, and $A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot}$ and respectively, $b_{kl} = d(Y_k, Y_l)$, and $B_{kl} = b_{kl} - \bar{b}_{k\cdot} - \bar{b}_{\cdot l} + \bar{b}_{\cdot\cdot}$, the distance correlation is simply computed as:

$$\mathscr{R}(X,Y) = \frac{\sum_k \sum_l A_{kl} B_{kl}}{\sqrt{\sum_k \sum_l A_{kl}^2} \sqrt{\sum_k \sum_l B_{kl}^2}}$$

An example of the application can be found in the Table 15.1 applied to the classical Tecator dataset. The second derivative has the highest $\mathscr{R}$ among the possible functional covariates ($\mathscr{X}$, $\mathscr{X}'$, $\mathscr{X}''$ using in all cases the $\mathscr{L}_2$ distance) with respect to the content of Fat. In the scalar covariates, Water has the highest distance correlation although the two dimensional variable (Water, Protein) is clearly the best one because the triplet (Fat, Water, Protein) conforms a compositional vector.

Table 15.1: Distance correlation $\mathscr{R}$ computed for the classical tecator dataset including the first and second derivative of the absorbance curves and the scalar covariates.

$\mathscr{R}$	Fat	$\mathscr{X}(\mathscr{L}_2)$	$\mathscr{X}'(\mathscr{L}_2)$	$\mathscr{X}''(\mathscr{L}_2)$
Fat	1	0.196	0.785	0.914
$\mathscr{X}(\mathscr{L}_2)$	0.196	1	0.434	0.231
$\mathscr{X}'(\mathscr{L}_2)$	0.785	0.434	1	0.860
$\mathscr{X}''(\mathscr{L}_2)$	0.914	0.231	0.860	1
Water	0.967	0.232	0.792	0.881
Protein	0.717	0.052	0.532	0.717
(Water, Protein)	0.976	0.224	0.795	0.898

15.2.1 The algorithm

1. Let Y the response and $S = \left\{X^1, \ldots, X^p\right\}$ the set of all possible variables that can be included in the model.
2. Set $\hat{Y} = \bar{Y}$, and let $M^{(0)} = \emptyset$ the inital set of the variates included in the model. Set $i = 0$.
3. Compute $\hat{\varepsilon} = Y - \hat{Y}$.
4. Choose $X^j \in S \neq \emptyset$ such that $\mathscr{R}\left(\hat{\varepsilon}, X^j\right) \geq \mathscr{R}\left(\hat{\varepsilon}, X^i\right), \forall i \neq j$ and the p-value of the test of dependence among $\left\{X^j\right\}$ and $\hat{\varepsilon}$ is significant. IF NOT, END.
5. Update the sets M and S: $M^{(i+1)} = M^{(i)} \cup \left\{X^j\right\}$, and $S = S \setminus \left\{X^j\right\}$.
6. Compute the new model of Y using $M^{(i+1)}$ choosing the best one in the catalog of all possible correct models with the variates in $M^{(i+1)}$.
7. Analyze the contribution of X^j:
   IF this contribution is not relevant (typically comparing with the previous model)
   THEN $M^{(i+1)} = M^{(i+1)} \setminus \left\{X^j\right\}$
   ELSE $\hat{Y} = \bar{Y} + \sum_{k=1}^{\#M} \hat{f}_k\left(X^k\right)$
8. Set $i = i + 1$ and go to 3

15.3 Real data application

We have applied our proposal to the forecast of the Market Demand Energy in the Iberian Market and its Price using all the available information from different sources. All the information included as a functional variable is also included as scalar covariates (each discretization point is a new covariable). We have employed the FSAM/FAM model in all cases given its flexibility and availability not only for multivariate variables, but also for functional ones (see, for instance, [13] or [6]).

15.3.1 Energy Market Demand

We have information about $n = 635$ days during 2008 and 2009 of the Market Demand Energy at day t and hour H ($\mathrm{En}_t(H)$) in the IntraDay Electricity Iberian Market. Our aim is to forecast this variable with the following possible information up to day $t-1$:

- Energy Market Variables (source:www.omie.es): Daily profiles of Energy (En), Load (Lo) and Price (Pr) at $t-1$ and $t-7$.
- Generated energy type (source:demanda.ree.es): Daily profiles of Nuclear (Nu), Fuel/Gas (Fu), Carbon (Ca), Combined Cycle (Cc), Solar (So), Eolic (Eo), Hydroelectric (Hy), Cogeneration (Co), Rest (Re).
- Meteorological information at Madrid-Barajas airport (source:aemet.es): Temperatures (T_{Max}, T_{Min}, T_{Med}, $T_A = T_{Max} - T_{Min}$), Speed Wind, Precipitation,....
- Every discretization value of the functional variates ($t-1$ and $t-7$).
- Categorical: Year(YY), Month(MM), Day-of-Week: $\left\{\mathbb{1}_{\{Mon\}}, \ldots, \mathbb{1}_{\{Sun\}}\right\}$

The results of this application are summarized in Table 15.2 for four specific hours along the day. In all those models, the deviance is mostly explained by the effect of the two first variables. As an example, the evolution in deviance of the model for $\mathrm{En}_t(22)$ as the covariates enter into the model were (63.5%, 80.6%, 88.1% and 88.4%).

Table 15.2: Summary of models for energy with its selected variables (in order of entering)

Response	Covariates (in order)	Dev. expl.	σ_ε
$\mathrm{En}_t(3)$	En_{t-1}, MM, Pr_{t-1}	91.6%	723.96
$\mathrm{En}_t(12)$	En_{t-7}, En_{t-1}, MM, $\mathbb{1}_{\{Sun\}}$, $\mathbb{1}_{\{Sat\}}$, $\mathbb{1}_{\{Mon\}}$	91.1%	1159.67
$\mathrm{En}_t(18)$	En_{t-7}, $\mathrm{En}_{t-1}(23)$, MM, $\mathbb{1}_{\{Sun\}}$, $\mathbb{1}_{\{Sat\}}$, $\mathbb{1}_{\{Mon\}}$	91.1%	1238.152
$\mathrm{En}_t(22)$	En_{t-7}, $\mathrm{En}_{t-1}(22)$, $\mathbb{1}_{\{DoW\}}$, $\mathrm{En}_{t-1}(7)$	88.4%	1015.829

In these models, the first contributor tends to be the information of a week before rather than the one from the previous day except for the consumption in the early morning hours (03:00). This suggests a strongly weekly pattern in the market demand profile. The different behavior of the 03:00 model comes from the fact that the energy demand at that hour is quite regular and has no dependence on the activity of a given day of the week. The effect of the month enters in all models, except for 22:00 in which it is substituted by the information of the same hour of the preceding day.

15.3.2 Energy Price

The second application example corresponds to the negotiated price at day t and hour H ($\mathrm{Pr}_t(H)$) in the Iberian Energy Market. The set of possible covariates are the same as in the previous example.

Table 15.3: Summary of models for price with the selected variables (in order of entering)

Response	Covariates (in order)	Dev. expl.	σ_ε
$\mathrm{Pr}_t(3)$	Pr_{t-1}, $\mathbb{1}_{\{Sat\}}$, $\mathbb{1}_{\{Sun\}}$, T_{max}, T_{med}, Eo_{t-1}	88.6%	0.475
$\mathrm{Pr}_t(12)$	Pr_{t-1}, $\mathbb{1}_{\{Sun\}}$, $\mathbb{1}_{\{Sat\}}$, T_{max}, MM	96.0%	0.359
$\mathrm{Pr}_t(18)$	Pr_{t-1}, Lo_{t-7}, Pr_{t-7}, T_A	96.2%	0.387
$\mathrm{Pr}_t(22)$	$\mathrm{Pr}_{t-1}(22)$, Pr_{t-7}, Eo_{t-1}, Hy_{t-7}, Cc_{t-1}	96.5%	0.562

The main contributor in all price models (see the results in Table 15.3) is the price of the preceding day with values of distance correlation over 0.80 which indicates a strong persistence of this variable. Surprisingly now, the variables related with demand or generation have distance correlations with price below 0.16, contrary what is expected from the classical economic theory. The calendar effects (day-of-week, month), the meteorological information and the type of energy generated are all minor contributors of the model. In the case of meteorological variates this can be explained by the use of the information of a particular site to summarize the weather of the whole country. Surely, meteorological information is useful for predicting demand in small regions but its contribution for price (without demand) is quite unexpected.

Acknowledgements The authors acknowledge financial support from Ministerio de Economía y Competitividad grant MTM2013-41383-P.

References

[1] Akaike, H.: Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika **60**(2), 255–265 (1973)

[2] Bühlmann, P., Yu, B.: Boosting with the L_2 loss: regression and classification. J. Am. Stat. Assoc. **98**(462), 324–339 (2003)

[3] Du, P., Cheng, G., Liang, H.: Semiparametric regression models with additive nonparametric components and high dimensional parametric components. Comput. Stat. Data Anal. **56**(6), 2006–2017 (2012)

[4] Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Technical report, Department of Statistics (2002)

[5] Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. **96**(456), 1348–1360 (2001)

[6] Febrero-Bande, M., González-Manteiga, W.: Generalized additive models for functional data. Test **22**(2), 278–292 (2013)

[7] Ferraty, F., Vieu, P.: Additive prediction and boosting for functional data. Comput. Stat. Data Anal. **53**(4), 1400–1413 (2009)

[8] Hastie, T., Tibshirani, R., Wainwright, M.: Statistical learning with sparsity: the LASSO and generalizations. CRC Press (2015)

[9] Lin, Y., Zhang, H.H.: Component selection and smoothing in smoothing spline analysis of variance models. Ann. Stat. **34**(5), 2272–2297 (2006)

[10] Liu, X., Wang, L., Liang, H.: Estimation and variable selection for semiparametric additive partial linear models (ss-09-140). Stat. Sinica **21**(3), 1225 (2011)

[11] Lyons, R.: Distance covariance in metric spaces. Ann. Probab. **41**(5), 3284–3305 (2013)

[12] Mallows, C.L.: Some comments on C_p. Technometrics **15**(4), 661–675 (1973)

[13] Müller, H.G., Yao, F.: Functional additive models. J. Am. Stat. Assoc. (2012)

[14] Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on pattern analysis and machine intelligence **27**(8), 1226–1238 (2005)

[15] Radchenko, P., James, G.M.: Improved variable selection with Forward-Lasso adaptive shrinkage. Ann. Appl. Stat. **5**(1), 427–448 (2011).

[16] Schwarz, G., et al.: Estimating the dimension of a model. Ann. Stat. **6**(2), 461–464 (1978)

[17] Stone, M.: Comments on model selection criteria of Akaike and Schwarz. J. Roy. Stat. Soc. B Met. pp. 276–278 (1979)

[18] Székely, G.J., Rizzo, M.L., Bakirov, N.K., et al.: Measuring and testing dependence by correlation of distances. Ann. Stat. **35**(6), 2769–2794 (2007)

[19] Tibshirani, R.: Regression shrinkage and selection via the LASSO. J. Roy. Stat. Soc. B Met. pp. 267–288 (1996)

[20] Wang, L., Liu, X., Liang, H., Carroll, R.J.: Estimation and variable selection for generalized additive partial linear models. Ann. Stat. **39**(4), 1827 (2011)

[21] Xue, L.: Consistent variable selection in additive models. Stat. Sinica pp. 1281–1296 (2009)

[22] Zou, H.: The adaptive LASSO and its oracle properties. J. Am. Stat. Assoc. **101**(476), 1418–1429 (2006)

Chapter 16
Functional data analysis approach of Mandel's h and k statistics in Interlaboratory Studies

Miguel Flores, Salvador Naya, Javier Tarrío-Saavedra and Rubén Fernández-Casal

Abstract In this work, functional versions of Mandels's h and k statistics for outlier laboratory detection in interlaboratory studies (ILS) are presented. The critical values of h and k outlier test are approximated using bootstrap resampling, and their characteristic graphics are obtained. Thermogravimetric data are simulated to study the performance of the proposed d^H and d^K functional test statistics.

16.1 Introduction

The Interlaboratory Studies (ILS) are defined as the statistical quality control process used to evaluate the consistency (homogeneity) of laboratory experimental results, obtained using a well-defined experimental procedure. They are performed using the same specific controlled material and tested by different laboratories. The implementation of outlier detection test that identify the measurements or laboratories results that should be discarded is a necessary practice in analytical chemistry, biology, medicine an physics. To identify the inhomogeneous laboratories that provide results significantly different from the others, the use of h and k Mandel's statistics is

Miguel Flores (✉)
Departamento de Matemáticas, Facultad de Ciencias, Escuela Politécnica Nacional, Ladrón de Guevara E11-253, Quito-Ecuador, e-mail: miguel.flores@epn.edu.ec

Salvador Naya
Department of Mathematics, Higher Polytechnic University College, Universidade da Coruña, Mendizábal s/n, Ferrol-España, e-mail: salva@udc.es

Javier Tarrío-Saavedra
Department of Mathematics, Higher Polytechnic University College, Universidade da Coruña, Mendizábal s/n, Ferrol-España, e-mail: javier.tarrio@udc.es

Rubén Fernández-Casal
Department of Mathematics, Faculty of Computer Science, Universidade da Coruña, Campus de Elviña, A Coruña-España, e-mail: ruben.fcasal@udc.es

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_16

proposed by the ASTM E691 in the univariate or scalar case [9, 13, 7]. The h statistic measures the inter-laboratories consistency by comparing the replicate averages with the overall average. Moreover, the k statistic provides information about the intra-laboratory consistency by comparing the replicate standard deviations with respect to the repeatability standard deviation [9, 13, 7]. Higher h and k involves a less consistency.

Many types of experimental results in analytical chemistry are functional, thus, if h and k scalar statistics are applied, important information could be obviated [8]. This is the case of experimental results (curves) obtained by thermal analysis techniques such as thermogravimetry (TG), differential scanning calorimetry (DSC), dynamic mechanical analysis (DMA), thermomechanical analysis (TMA), and dielectric analysis (DEA).

The application of functional approaches prevents the typical information loss associated to the dimension reduction processes. A new FDA methodology for ILS has been proposed in [8]. The intra and inter-laboratory variability are estimated from a functional perspective, and they are compared to the results obtained in traditional reproducibility an repeatability studies. The FDA data depth concept was applied to detect the atypical TG and DSC curves (outlier detection). This procedure (combined with functional ANOVA) has identified the laboratories that provided non consistent (inhomogeneous) results. If data are functional, the drawback of scalar h and k application is that we need to extract a representative feature from curves. Depending on the extracted feature, the test result could be different [8]. Therefore, the use of FDA techniques is justified, the non-consistent laboratories can be identified using the whole curves.

In the present study, a FDA approach of h and k statistics is proposed to deal with reproducibility and repeatability studies. It allows to estimate the intra and inter-laboratory variability and location from a functional perspective. The results are obtained through the software R [10] and their packages, such as `fda.usc` library, used [4] for outlier detection based on functional data depth.

The FDA concepts and techniques used in this work can be consulted in the monographs of Ramsay and Silverman [12] and Ramsay and Silverman [11] can be consulted. In both cases all the techniques included are restricted to the space of L_2 functions (the Hilbert space of all square integrable functions over a certain interval). The book by Ferraty and Vieu [5] is another important reference incorporating non-parametric approaches as well as the use of other theoretical tools such as semi-norms and small ball probabilities that allow us to deal with normed or metric spaces.

This work is organized as follows. In Section 16.2, the main concepts of functional data are presented, the $H(t)$ and $K(t)$ statistics are defined for the functional case. In addition, the d^H and d^K test statistics are shown and described. Section 16.3 accounts for the description of bootstrap methodology applied to estimate the test critical values in the functional case. In Section 16.4, the performance of the proposed FDA procedure is analyzed by a simulation study.

16.2 $H(t)$ and $K(t)$ statistics for functional data

In the present section, the $H(t)$ and $K(t)$ statistics, as well the corresponding d^H and d^K test statistics, are introduced to detect laboratories that provide no consistent data in a ILS. Some notes are also included about the functional norm and the functional depth. These measures are used in FDA for the computation of location and dispersion functional estimates and for outlier detection [3, 1, 6].

16.2.1 Functional Data

Assume that the functional dataset $\{X_1(t), X_2(t), \ldots, X_n(t)\}$ was obtained as iid observations from a stochastic process $X(t)$, with continuous trajectories on the interval $[a,b]$, being $\mu(t)$ the functional mean and $\sigma^2(t) > 0$ the functional variance. We will consider the L_2-norm:

$$\|X\| = \left(\int_a^b X(t)^2 dt \right)^{\frac{1}{2}},$$

defining the distance between two functions as:

$$d(X(t), Y(t)) = \|X(t) - Y(t)\| = \left(\int_a^b (X(t) - Y(t))^2 dt \right)^{\frac{1}{2}}.$$

16.2.2 Data depth and outlier detection

The data depth concept explains how a datum is centered with respect to a set of observations from a given population. Therefore, the deepest datum will be that surrounded by the highest number of neighbors. In FDA context, deeper curves are identified as those closer to the center, which are usually estimated by the median [1]. Three of the most common approaches to calculate the functional depth are the depth of Fraiman and Muniz (or median depth) [6], the mode depth [1], and the depth based on random projections [2].

The functional data depth can be used for outlier detection. Febrero-Bande et al. [3] identify outliers in functional datasets, taking into account that depth and outlyingness are inverse notions (an outlier curve will have a significantly low depth). Therefore, a way to detect the presence of functional outliers is to look for curves with lower depths. In the present study, we use two procedures, proposed in [4] and included in the R package `fda.usc`, for detecting outliers: The first one is based on weighting, `outliers.depth.pond()`, and the second one is based on trimming, `outliers.depth.trim()`.

16.2.3 Functional statistics for ILS

In the ILS, a set of observations $\left\{X_1^l(t),\ldots,X_n^l(t)\right\}$ are obtained for each lab $l, l = 1,\ldots,p$. Each laboratory experimentally test n samples, obtaining n different curves. The functional $H_l(t)$ and $K_l(t)$ statistics are estimated for each laboratory and considering the null hypothesis that there is no statistical difference between laboratory measurements. The null hypotheses for R & R studies are described below:

The null hipothesis of reproducibility states that

$$H_0 : \mu_1(t) = \mu_2(t) = \cdots = \mu_p(t), \tag{16.1}$$

where $\mu_l(t), l = 1\ldots p$ are the populational functional mean for each laboratory l.

To test reproducibility of the laboratory results, the previous calculation of the $H(t)$ statistic is necessary. It is defined as

$$H_l(t) = \frac{X_i^l(t) - \bar{X}(t)}{S_l(t)}; l = 1,\ldots,p,$$

where $\bar{X}_l(t)$ y $S_l(t)$ are the mean and functional variance pointwise calculated for the l laboratory.

The null hipothesis of repeatability states that there are not differences in the laboratory variability:

$$H_0 : \sigma_1^2(t) = \sigma_2^2(t) = \cdots = \sigma_p^2(t), \tag{16.2}$$

where $\sigma_l(t), l = 1\ldots p$ are the theoretical functional variances corresponding to each laboratory l.

The repeatability test is based on the $K(t)$ statistic, defined as

$$K_l(t) = \frac{S_l(t)}{\sqrt{\bar{S^2}(t)}}; l = 1,\ldots,p,$$

where, $\bar{S^2}(t) = \frac{1}{p}\sum_{l=1}^{p} S_l^2(t)$.

On the one hand, in order to test the reproducibility hypothesis, we define the d^H test statistic as

$$d_l^H = \|H_l(t)\| = \left(\int_a^b H_l(t)^2 dt\right)^{\frac{1}{2}},$$

considering that d^H values corresponding to inhomogeneous laboratories will tend to be high. On the other hand, to test the repeatability hypothesis, we also define $d_l^K = \left\|K_(t)\right\|$, taking into account that higher values of d^K correspond to non consistent laboratories.

16.3 Bootstrap algorithm

A bootstrap algorithm to test if the d_l^H and d_l^K are significantly high is proposed. The proposed bootstrap procedure pretends to reproduce the distribution of these statistics under corresponding null hypothesis, (16.1) and (16.2) respectively. Assuming that a significance level α was fixed (typically $\alpha = 0.01$), the algorithm consist on the following steps:

1. Remove atypical observations, grouping all the curves in a single set (null hypothesis), and applying the procedure described in Manuel Febrero et al. [3].
2. Using the smoothed bootstrap proposed in [1], generate bootstrap samples of size $p \cdot n$ from the overall dataset without outliers. In each bootstrap sample, randomly assign the bootstrap observations to the laboratories.
3. For each bootstrap sample, compute the $H_l^*(t)$ and $K_l^*(t)$ functional statistics, and the corresponding d_l^{H*} and d_l^{K*} test statistics, for each laboratory $l = 1, \ldots, p$.
4. Approximate the critical values c_H and c_K as the as the empirical $100\,(1-\alpha)\,/p$ percentile of the distribution of the corresponding $p \cdot B$ bootstrap replicates of the test statistic.
5. Finally, compute the confidence bands for the $H(t)$ and $K(t)$statistics, determined by the envelope of bootstrap samples with a less norm than the corresponding critical value.

For each laboratory, the null hypotheses of reproducibility (16.1) (or repeatability (16.2)) will be rejected if $d_l^H = \left\|H_(t)\right\|$ (d_l^K) exceed the critical value c_H (c_K).

16.4 A simulation study

Each scenario is composed by p laboratories (each one has tested n samples). The results of each lab are simulated from a Gaussian process $Y(t) = \mu(t) + \sigma(t)\varepsilon(t)$, where $t \in [0,1]$ with $\mu(t) = \frac{1}{(1+exp(10(t-m)))}$ is the trend functions (generalized logistic model) and $\sigma(t)^2 = c_1 10^{-6}(1+(1-(\frac{t}{0.5}-1)^2)^3)$, the deterministic variance. Moreover, ε is a second order stationary process wit 0 mean and $\exp(-|s-t|/0.3)$ covariance.

Two scenarios are simulated in order to evaluate the performance of test statistics. The first one consists on varying the m parameter of $\mu(t)$, in order to evaluate the H statistic. For this purpose, 6 labs are simulated under the null hpothesis defined by $m = 0.5$ and the alternative defined by $m = 0.5 + \delta_H$. In the same way, the second scenario consists on varying the c_1 parameter of $\sigma(t)$, in order to evaluate the H statistic. For this purpose, 6 labs are simulated under the null hpothesis defined by $c_1 = 5$ and the alternative defined by $c_1 = 5 + \delta_K$.

The TG curves accounts for the mass of a material depending on time or temperate when the temperature. They provide information of the thermal stability of materials. TG curves where simulated taking into account real data retrieved from [8]

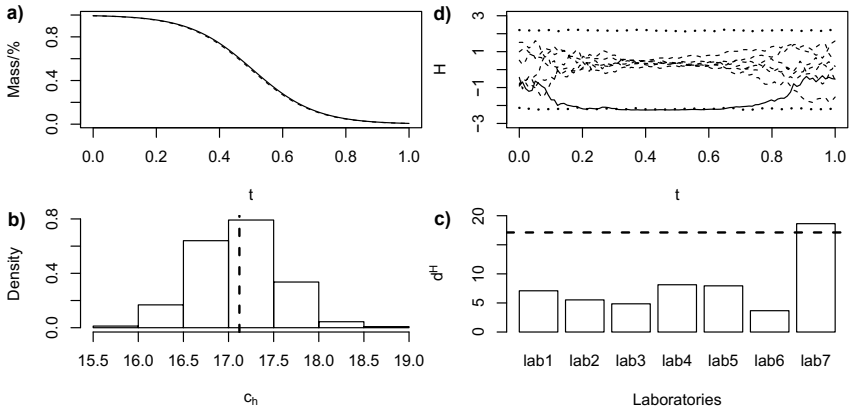


Fig. 16.1: The output for the $\delta = 1\%, \alpha = 1\%$ simulation scenario is summarized. Panel a: The simulated TG curves are shown. Panel b: Histogram of c_h distribution obtained in bootstrap resampling. The median is highlighted in dotted line. Panel c: Graphical output for d^H (for a specific sample of the simulation scenario). The d^H statistic for each laboratory is calculated and compare with the median of critical values (c_h) shown in panel b. Panel d: Functional H statistic corresponding to each laboratory (for a specific sample of the simulation scenario) is shown and compared to the limits that accounts for the 99% of resampled curves.

(Figure 16.1a), assuming similar variance structure (Figure 16.2a). In order to show an illustrative example of the ILS proposal, the simulated samples corresponding to $\delta = 1\%, \alpha = 1\%$ scenario are presented (Figures 16.1 and 16.2). Figure 1a show the case where the Lab 7 provides inhomogeneous results. In fact its m parameter is varied 1% with respect to the other labs. In the case of d^K study, the c_1 parameter corresponding to Lab 7 is varied 1%. In Figures 16.1 and 16.2, the graphical outputs for d^H and d^K are shown. Figure 16.1c and 16.1d shows that the proposed FDA approach detect the Lab 7 as an outlier when it is compared with critical values. Nevertheless, in Figures 16.2c and 16.2d, the changes in variance of Lab 7 are not detected. This is related to the lower power of d^K. It would need higher changes to detect a Lab 7 as an outlier (see Figure 16.3).

Acknowledgements The research of Salvador Naya, Javier Tarrío-Saavedra and Rubén Fernández-Casal has been partially supported by the Consellería de Cultura, Educación e Ordenación Universitaria of the Xunta de Galicia through the agreement for the Singular Research Center CITIC, and by Grant MTM2014-52876-R. The research of Miguel Flores has been partially supported by Grant PII-DM-002-2016 of Escuela Politécnica Nacional of Ecuador.

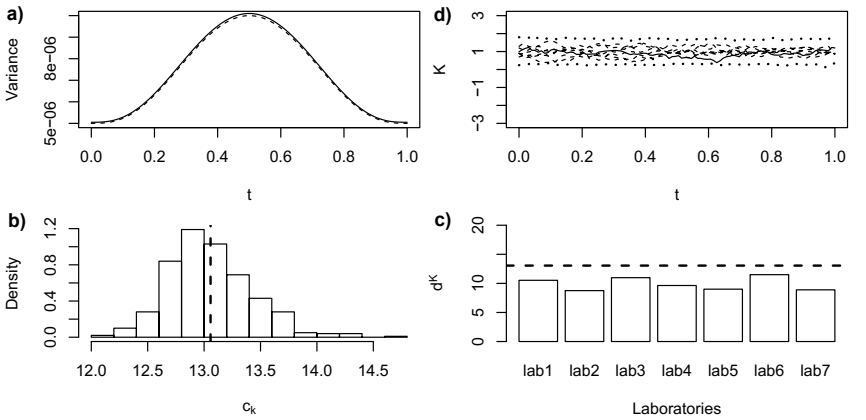


Fig. 16.2: The output for the $\delta = 1\%$, $\alpha = 1\%$ simulation scenario is summarized. Panel a: Variance assumed for simulated TG data under null and alternative hypothesis are shown. Panel b: Histogram of c_h distribution obtained in bootstrap resampling. The median is highlighted in dotted line. Panel c: Graphical output for d^K (for a specific sample of the simulation scenario). The d^K statistic for each laboratory is calculated and compare with the median of critical values (c_k) shown in panel b. Panel d: Functional K statistic corresponding to each laboratory (for a specific sample of the simulation scenario) is shown and compared to the limits that accounts for the 99% of resampled curves.

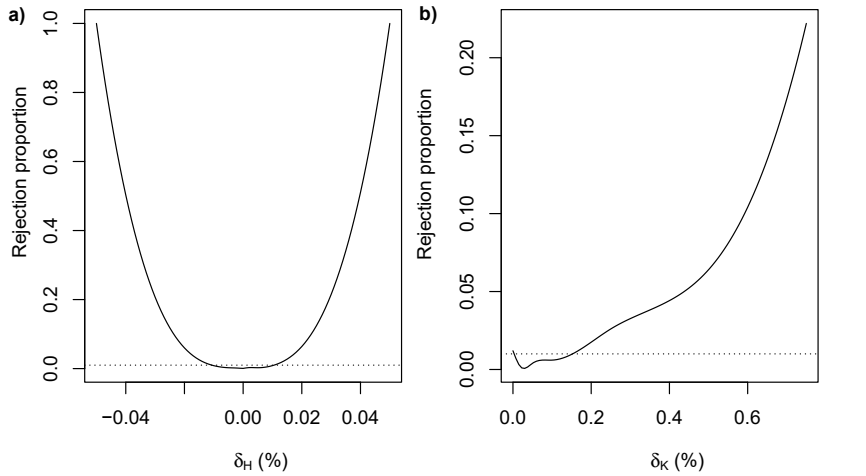


Fig. 16.3: Panel a: Power of d^H statistic. Power of d^K.

References

[1] Cuevas A., Febrero-Bande, M., Fraiman, R.: On the use of the bootstrap for estimating functions with functional data. Comput Stat Data Anal. **51**, 1063–1074 (2006)
[2] Cuevas A., Febrero-Bande, M., Fraiman, R.: Robust estimation and classification for functional data via projection-based depth notions. Comput Stat. **22**,81–96 (2007)
[3] Febrero-Bande M., Galeano P., González-Manteiga W.: Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. Environmetrics. **19**, 331–345 (2012)
[4] Febrero-Bande M., de la Fuente O.M.: Statistical computing in functional data analysis: the R package FDA.usc. J Stat Soft. **51**, 1–28 (2012)
[5] Ferraty F, Vieu P.: Nonparametric Functional Data Analysis. Springer-Verlag, New York (2006)
[6] Fraiman R., Muniz G.: Trimmed means for functional data. Test. **10**, 419–440 (2001)
[7] Mandel, J.: A new analysis of interlaboratory test results. In: ASQC Quality Congress Transaction–Baltimore. 360–366 (1985)
[8] Naya, S., Tarrío-Saavedra. J., López-Beceiro, J., Francisco Fernández, M., Flores, M., Artiaga, R.:Statistical functional approach for interlaboratory studies with thermal data. J. Therm. Anal. Calorim. **18**, 1229–1243 (2014)
[9] Practice for conducting and interlaboratory study to determine the precision of a test method. Annual Book of ASTM Standards. West Conshohocken, PA: ASTM International E691 (2004)
[10] R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing,Vienna, Austria, (2008). http://www.R-project.org.
[11] Ramsay J.O., Silverman B.W.: Applied Functional Data Analysis: Methods and Case studies. Springer-Verlag, New York (2002)
[12] Ramsay J.O., Silverman B.W.: Functional Data Analysis. 2nd edition. Springer-Verlag, New York (2005)
[13] Wilrich, P.-T.: Critical values of Mandels h and k, the Grubbs and the Cochran test statistic. AStA Advances in Statistical Analysis. **97**,1–10 (2013)

Chapter 17
High-dimensional functional time series forecasting

Yuan Gao, Hanlin L. Shang and Yanrong Yang

Abstract In this paper, we address the problem of forecasting high-dimensional functional time series through a two-fold dimension reduction procedure. Dynamic functional principal component analysis is applied to reduce each infinite-dimension functional time series to a vector. We use factor model as a further dimension reduction technique so that only a small number of latent factors are preserved. Simple time series models can be used to forecast the factors and forecast of the functions can be constructed. The proposed method is easy to implement especially when the dimension of functional time series is large. We show the superiority of our approach by both simulation studies and an application to Japan mortality rates data.

17.1 Introduction

Functional data are considered as realizations of smooth random curves. When curves are collected sequentially, they form functional time series $X_t(u), u \in \mathscr{I}$. To deal with infinite dimensional functions, there is a demand for efficient data reduction techniques. Functional principal component ananlysis (FPCA) is the most commonly used approach that serves this purpose. FPCA performs eigendecomposition on the underlying covariance functions. Most of the variance structures captured in a vector called principal component scores (see [13] for details). Existing FPCA method has

Yuan Gao (✉)
College of Business and Economics, Australian National University,
26C Kingsley St. ACTON 2601, Australia, e-mail: u5758483@anu.edu.au

Hanlin L. Shang
College of Business and Economics, Australian National University,
e-mail: hanlin.shang@anu.edu.au

Yanrong Yang
College of Business and Economics, Australian National University,
e-mail: yanrong.yang@anu.edu.au

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_17

been developed for independent observations, which is a serious weakness when we are dealing with time series data. In this paper, we adopt a dynamic FPCA approach ([10],[15]), where serial dependence between the curves are taken into account. With dynamic FPCA, functional time series are reduced to a vector time series, where the individual component processes are mutually uncorrelated principal component scores.

It is often the case that we collect a vector of N functions at a single time point t. If these N functions are assumed to be correlated, multivariate functional models should be considered. Classical multivariate FPCA concatenates the multiple functions into one to perform univariate FPCA ([13]). [9] suggested normalizing each random function as a preliminary step before concatenation. [6] studied functional version of principal component ananlysis, where multivariate functional data are reduced to one or two functions rather than vectors. However, existing models dealing with multivariate functional data either fail to handle data with a large N (as in the classical FPCA approach), or are hard to implement practically (as in [6]).

We propose a two-fold dimension reduction model to handle high-dimensional functional time series. By high-dimension, we allow that the dimension of the functional time series N to be as large as or even larger than the length of observed functional time series T. The dimension reduction process is straightforward and easy to implement:

1) Dynamic functional principal component analysis is performed on each set of functional time series, resulting in N sets of principal component scores of low dimension K (typically less than 5);

2) The N first principal component are fitted to a factor model, and is further reduced to a dimension of r ($r \ll N$). The same is done for the second principal component scores and thus K factor models are fitted. The vector of N functional time series is reduced to $r \times K$ vector time series.

Factor models are frequently used for dimension reduction. Some early application of factor analysis to multiple time series include [1], [16] and [7]. Time series in high-dimensional settings with $N \to \infty$ together with T are studied in [8], [4] and [11]. Among these, we adopt the method considered in [11], where the model is conceptually simple and asymptotic properties are established.

17.2 Research methods

In this section, we introduce dynamic FPCA and factor model for the two-fold dimension reduction process. Estimation and asymptotic properties are discussed. We also suggest methods for forecasting.

17.2.1 Dynamic functional principal component analysis

We consider stationary N-dimensional functional time series $\boldsymbol{X}_t : t \in \mathbb{Z}$, where $\boldsymbol{X}_t = (X_t^1(u), \ldots, X_t^N(u))^\top$, and each $X_t(u)$ takes values in the space $H := L^2(\mathscr{I})$ of real-valued square integrable functions on $\mathscr{I}$.

The space H is a Hilbert space, equipped with the inner product $\langle x, y \rangle := \int_{\mathscr{I}} x(u)y(u)du$. For each $i = 1, \ldots, N$, we assume X_t^i has a continuous mean function, $\mu^i(u)$ and an auto-covariance function at lag h, $\gamma_h^i(u,v)$, where

$$\begin{aligned} \mu^i(u) &= \mathrm{E}[X^i(u)], \\ \gamma_h^i(u,v) &= \mathrm{cov}[X_t^i(u), X_{t+h}^i(v)] \end{aligned} \tag{17.1}$$

The long-run covariance function is defined as

$$c^i(u,v) = \sum_{h=-\infty}^{\infty} \gamma_h^i(u,v) \tag{17.2}$$

Using $c^i(u,v)$ as a kernel, we define the operator C by:

$$C^i(x)(u) = \int_{\mathscr{I}} c^i(u,v)x(v)dv, \quad u, v \in \mathscr{I} \tag{17.3}$$

The kernel is symmetric, non-negative definite. Thus by Mercer's Theorem, the operator C admits an eigendecomposition

$$C^i(x) = \sum_{k=1}^{\infty} \lambda_k \langle x, \upsilon_k \rangle \upsilon_k, \tag{17.4}$$

where $(\lambda_l : l \geq 1)$ are C's eigenvalues in descending order and $(\upsilon_l : l \geq 1)$ the corresponding normalized eigenfunctions. By Karhunen-Loève Theorem, X can be represented with

$$X_t^i(u) = \sum_{k=1}^{\infty} \beta_{t,k}^i \upsilon_k^i(u) \tag{17.5}$$

where $\beta_{t,k}^i = \int_{\mathscr{I}} X_t^i(u)\upsilon_k^i(u)du$ is the kth principal component score at time t. The infinite dimensional functions can be approximated by the first K principal component scores:

$$X_t^i(u) = \sum_{k=1}^{K} \beta_{t,k}^i \upsilon_k^i(u) + \varepsilon_t^i(u) \tag{17.6}$$

17.2.2 Factor model

With the first step dimension reduction, we now have principal component scores $\beta_{t,k}^i$, where $i = 1, \ldots, N$. Following the early work by [10] and [5], we consider the

following factor model for $\boldsymbol{\beta}_{i,t} = (\beta_{t,k}^1, \ldots, \beta_{t,k}^N)^\top$. For each $k = 1, \ldots, K$, let

$$\boldsymbol{\beta}_{t,k} = \boldsymbol{A}_k \boldsymbol{\omega}_{t,k} + \boldsymbol{e}_{i,t}, \quad t = 1, \ldots T, \tag{17.7}$$

where $\boldsymbol{\beta}_{t,k}$ is the vector that contains the kth principal component score of all N functional time series. $\boldsymbol{\omega}_{t,k}$ is an $r \times 1$ unobserved factor time series. $\boldsymbol{A}_k$ is an $N \times r$ unknown constant factor loading matrix, and $\boldsymbol{e}_{t,k}$ is idiosyncratic error with mean 0 and variance $\sigma_{i,t}^2$. K factor models are fitted to the principal component scores.

With two-fold dimension reduction, the original functional time series can be approximated by

$$X_t^i(u) = \sum_{k=1}^{K} [\boldsymbol{A}_k \boldsymbol{\omega}_{t,k}]_i \upsilon_k^i(u) + \theta_t^i(u), \tag{17.8}$$

where $[\boldsymbol{A}_k \boldsymbol{\omega}_{t,k}]_i$ is the ith element in the vector $\boldsymbol{A}_k \boldsymbol{\omega}_{t,k}$, and $\theta_t^i(u)$ is the error term from two steps of approximation.

17.2.3 Estimation

We want to estimate $\boldsymbol{A_k}$, $\boldsymbol{\omega}_{t,k}$ and $\upsilon_k^i(u)$ in (17.8). In the dynamic FPCA step, the long-run covariance function $c^i(u,v)$ can be estimated by

$$\widehat{c}^i(u,v) = \sum_{h=-\infty}^{\infty} W(\frac{h}{q}) \widehat{\gamma}_h^i(u,v) \tag{17.9}$$

where

$$\widehat{\gamma}_h^i(u,v) = \begin{cases} \frac{1}{T}\sum_{j=1}^{T-h}(X_j^i(u) - \bar{X}_j(u))(X_{j+h}(v) - \bar{X}(v)), & h \geq 0 \\ \frac{1}{T}\sum_{j=1-h}^{T}(X_j^i(u) - \bar{X}_j(u))(X_{j+h}(v) - \bar{X}(v)), & h < 0 \end{cases}$$

where W is a weight function with $W(0) = 1, W(u) = W(-u), W(u) = 0$ if $|u| > m$ for some $m > 0$, and W is continuous on $[-m, m]$. Some possible choices include Bartlett, Parzen, Tukey-Hanning, Quadratic spectral and Flat-top functions ([2, 3]). q is a bandwidth parameter. [14] proposed a plug-in procedure to select q. $\widehat{c}_i(u,v)$ is used as the kernel of the operator $\widehat{C}$, with which we can estimate $\widehat{\upsilon}_k^i(u)$ by performing eigendecomposition on $\widehat{C}$. $\widehat{\upsilon}_k^i(u)$ is the normalized eigenfunction that corresponds to the kth largest eigenvalue. The empirical principal component scores $\widehat{\beta}_{t,k}^i = \int_{\mathscr{I}} X_t^i(u) \widehat{\upsilon}_k^i(u) du$, can be calculated by numerical integration.

The estimates $\widehat{\beta}_{t,k}^i$ are fitted to a factor model. The estimation of latent factors for high-dimensional time series can be found in [11]. A natural estimator for $\boldsymbol{A}_k$ is defined as $\widehat{\boldsymbol{A}}_k = (\widehat{\boldsymbol{a}}_1, \ldots, \widehat{\boldsymbol{a}}_r)$, where $\widehat{\boldsymbol{a}}_j$ is the jth eigenvector of $\widehat{\boldsymbol{Q}}_k$, and

$$\widehat{\boldsymbol{Q}}_k = \sum_{h=1}^{h_0} \widehat{\boldsymbol{\Sigma}}_{h,k} \widehat{\boldsymbol{\Sigma}}_{h,k}^\top, \quad \widehat{\boldsymbol{\Sigma}}_h = \frac{1}{T-h} \sum_{h=1}^{T-h} (\widehat{\boldsymbol{\beta}}_{t+h,k} - \tilde{\boldsymbol{\beta}}_k)(\widehat{\boldsymbol{\beta}}_{t,k} - \tilde{\boldsymbol{\beta}}_k)^\top, \tag{17.10}$$

where $\tilde{\boldsymbol{\beta}}_k = \frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{\beta}}_{t,k}$. Thus we estimate the kth factor by

$$\widehat{\boldsymbol{\omega}}_{t,k} = \widehat{\boldsymbol{A}}_k^{\top}\widehat{\boldsymbol{\beta}}_{t,k} \tag{17.11}$$

The estimator for the original function $X_t^i(u)$ is

$$\widehat{X}_t^i(u) = \sum_{k=1}^{K}[\widehat{\boldsymbol{A}}_k\widehat{\boldsymbol{\omega}}_{t,k}]_i\widehat{\upsilon}_k^i(u), \quad i=1,\ldots,N, \quad t=1,\ldots,T \tag{17.12}$$

where $[\widehat{\boldsymbol{A}}_k\widehat{\boldsymbol{\omega}}_{t,k}]_i$ is the ith element of the vector $\widehat{\boldsymbol{A}}_k\widehat{\boldsymbol{\omega}}_{t,k}$.

17.2.4 Forecasting

With two-fold dimension reduction, information of serial correlation is contained in the factors $\boldsymbol{\omega}_{t,k}$. To forecast N-dimensional functional time series, we could instead make forecast on the estimated factors. Scalar or vector time series models could be applied. We suggest univariate time series models, autoregressive moving average (ARMA) models, for instance, since the factors are mutually uncorrelated. Consequently, we need to fit $r \times K$ ARMA models on the factors. The prediction of the functions could be calculated:

$$\mathrm{E}[\widehat{X}_{t+h|t}^i(u)] = \sum_{k=1}^{K}[\widehat{\boldsymbol{A}}_k\widehat{\boldsymbol{\omega}}_{t+h|t,k}]_i\widehat{\upsilon}_k^i(u), \quad i=1,\ldots,N, \quad t=1,\ldots,T, \tag{17.13}$$

where $\widehat{X}_{t+h|t}^i(u)$ is h-step ahead forecast at time t.

17.3 Empirical studies

Japanese sub-national mortality rates in 47 prefectures is used to demonstrate the effectiveness of our proposed method. Available at [12], the data set contains yearly age-specific mortality rates in a span of 40 years from 1975 to 2014. The dimension of the functional time series is 47, which is greater than the sample size 40. With the two-fold dimension reduction model, we use the first three principal component scores for each population, and the first three factors. The problem of choosing the appropriate number of scores and factors will be discussed in detail.

We compare the forecast accuracy of our proposed method with the independent functional time series model, where each sub-national population is forecast individually. Both point forecast and interval forecast errors are calculated and it is found that the proposed method outperforms the independent model in most of the prefectures.

References

[1] Anderson, T.: The use of factor analysis in the statistical analysis of multiple time series. Psychometrika. **28**, 1-25 (1963)

[2] Andrews, D.: Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica. **59**, 817-858 (1991)

[3] Andrews, D., Monahan, J.C.: An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica. **60**, 953-966 (1992)

[4] Bai, J.: Inferential theory for factor models of large dimensions. Econometrica. **71**, 135-171 (2003)

[5] Bai, J.: Panel data models with interactive fixed effects. Econometrica. **77**, 4, 1229-1279 (2009)

[6] Berrendero, J., Justel, A., Svarc, M.: Principal components for multivariate functional data. Comput. Statist. Data Anal. **55**, 9, 2619-2634 (2011)

[7] Brillinger, D.: Time Series Data Analysis and Theory, extended ed. Holder-Day, San Francisco (1981)

[8] Chamberlain, G.: Funds, factors, and diversification in arbitrage pricing models. Econometrica. **70**, 191-221 (1983)

[9] Chiou, J.M., Chen, Y.T., Yang, Y.F.: Multivariate functional principal component analysis: A normalization approach. Statist. Sinica. **24**, 4, 1571-1596 (2014)

[10] Hörmann, S., Kidziński, L., Hallin, M.: Dynamic functional principal components. J. R. Stat. Soc. Ser. B Stat. Methodol. **77**, 2, 319-348 (2015)

[11] Lam, C., Yao, Q., Bathia, N.: Estimation of latent factors for high-dimensional time series. Biometrika. **98**, 4, 901-918 (2011)

[12] National Institute of Population and Social Security Research: Japan Mortality Database. Available at http://www.ipss.go.jp/p-toukei/JMD/index-en.html. (2016)

[13] Ramsay, J.O., Silverman, B.W.: Functional Data Analysis. Springer, New York (2005)

[14] Rice, G., Shang, H.L.: A plug-in bandwidth selection procedure for long run covariance estimation with stationary functional time series. J. Time Series Anal. in proceeding (2016)

[15] Panaretos, V.M., Tavakoli, S.: Cramér-Karhunen-Loève representation and harmonic principal component analysis of functional time series. Stochastic Process. Appl. **123**, 7, 2779-2807 (2013)

[16] Priestley, M., Rao, T., Tong, J.: Applications of principal component analysis and factor analysis in the identification of multivariable systems. IEEE Trans. Auto. Contr. **19**, 6 730-734 (1974)

Chapter 18
Essentials of backward nested descriptors inference

Stephan F. Huckemann and Benjamin Eltzner

Abstract Principal component analysis (PCA) is a popular device for dimension reduction and their asymptotics are well known. In particular, principal components through the mean span the data with decreasing residual variance, as the dimension increases, or, equivalently maximize projected variance, as the dimensions decrease, and these spans are nested in a backward and forward fashion – all due to Pythagoras Theorem. For non-Euclidean data with no Pythagorian variance decomposition available, it is not obvious what should take the place of PCA and how asymptotic results generalize. For spaces with high symmetry, for instance for spheres, backward nested sphere analysis has been successfully introduced. For spaces with less symmetry, recently, nested barycentric subspaces have been proposed. In this short contribution we sketch how to arrive at asymptotic results for sequences of random nested subspaces.

18.1 Introduction

From the early days of statistics of non-Euclidean data, *Procrustes analysis* proposed by [3] for shape data as a generalization of PCA, has been a successful device of choice. In essence, data are mapped to a tangent space of a Fréchet mean and PCA is performed in that tangent space. Notably, in that setting, not only the PCs but also the base point of the tangent space is random. Because all tangent spaces are the same in a Euclidean space, this complication is non-existent for asymptotics of classical PCA, derived by [1, 10, 9] and others.

Beyond lacking a general asymptotic theory, one may view this and similar methods (e.g. [2]) also as non-satisfactory, because these tangent space PCs neither

Stephan F. Huckemann (✉) and Benjamin Eltzner
Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, University of Göttingen, Goldschmidstr. 7, 37077 Göttingen, Germany, e-mail: huckeman@math.uni-goettingen.de and e-mail: beltzne@uni-goettingen.de

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_18

minimize residual variance nor maximize projected variance with respect to an invariant distance. To this end *geodesic PCs* that are geodesics minimizing intrinsic residual variance have been considered by [6, 5]. These are non-nested in the sense that the *intrinsic mean* μ is in general not located on the first PC, cf. [6]. From a dimension reduction viewpoint, however, nestedness appears as a desirable feature, where one seeks a sequence of subspaces $\{p_j\}_{j=0}^m$ of the data space Q, where each subspace approximates the data best, in a certain sense, over a family of admissible subspaces, that is nested.

$$\{\mu\} = p^0 \subset p^1 \subset \ldots \subset p^m = Q\,. \tag{18.1}$$

On a sphere, if all the subspaces are small subspheres, this is realized by *principal nested sphere* (PNS) analysis by [7]. For general spaces, if each subspace is a barycentric center of an nested sequence of points, this is realized by *barycentric subspaces* (BS) by [8].

In the following we formulate a general setup for (18.1) and state asymptotic results from which inferential bootstrap procedures can be derived, as detailed in [4]. In particular we have shown that the geometric assumptions below are satisfied for PNS and for the *intrinsic mean on a first principal component geodesic*, cf. [4, 6, 5]. It is still an open problem, to explore under which conditions these assumptions hold also for barycentric subspaces.

18.2 Setup

In the following, *smooth* refers to existing continuous 2nd order derivatives.

For a topological space Q we say that a continuous function $d : Q \times Q \to [0,\infty)$ is a *loss function* if $d(q,q') = 0$ if and only if $q = q'$. We say that a set $A \subset Q$ is d-bounded if $\sup_{a,a' \in A} d(a,a') < \infty$. Moreover, we say that $B \subset Q$ is *d-Heine Borel* if all closed d-bounded subsets of B are compact.

Definition 18.1. A separable topological space Q, called the *data space*, admits *backward nested families of descriptors* (BNFDs) if

1. there is a collection P_j $(j = 0,\ldots,m)$ of topological separable spaces with loss functions $d_j : P_j \times P_j \to [0,\infty)$;
2. $P_m = \{Q\}$;
3. every $p \in P_j$ $(j = 1,\ldots,m)$ is itself a topological space and gives rise to a topological space $\emptyset \neq S_p \subset P_{j-1}$ which comes with a continuous map

$$\rho_p : p \times S_p \to [0,\infty)\,;$$

4. for every pair $p \in P_j$ $(j = 1,\ldots,m)$ and $s \in S_p$ there is a measurable map called *projection*

$$\pi_{p,s} : p \to s\,.$$

For $j \in \{1,\dots,m\}$ and $k \in \{1,\dots,j\}$ call a family

$$f = \{p^j,\dots,p^{j-k}\},\ \text{with } p^{l-1} \in S_{p^l}, l = j-k+1,\dots,j$$

a *backward nested family of descriptors (BNFD) from P_j to P_{j-k}*. The space of all BNFDs from P_j to P_{j-k} is given by

$$T_{j,k} = \left\{ f = \{p^{j-l}\}_{l=0}^{k} : p^{l-1} \in S_{p^l}, l = j-k+1,\dots,j \right\} \subseteq \prod_{l=0}^{k} P_{j-l}\,.$$

For $k \in \{1,\dots,m\}$, given a BNFD $f = \{p^{m-l}\}_{l=0}^{k}$ set

$$\pi_f = \pi_{p^{m-k+1},p^{m-k}} \circ \ldots \circ \pi_{p^m,p^{m-1}} : p^m \to p^{m-k}$$

which projects along each descriptor. For another BNFD $f' = \{p'^{j-l}\}_{l=0}^{k} \in T_{j,k}$ set

$$d^j(f,f') = \sqrt{\sum_{l=0}^{k} d_j(p^{j-l},p'^{j-l})^2}\,.$$

In case of PNS, the nested projection π_f is illustrated in Figure 18.1 (a).

Definition 18.2. Random elements $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} X$ on a data space Q admitting BNFDs give rise to *backward nested population* and *sample means* (abbreviated as BN means)

$$\{E^{f^j} : j = m,\dots,0\}, \quad \{E_n^{f_n^j} : j = m,\dots,0\}$$

recursively defined via $E^m = \{Q\} = E_n^m$, i.e. $p^m = Q = p_n^m$ and

$$E^{f^{j-1}} = \operatorname{argmin}_{s \in S_{p^j}} \mathbb{E}[\rho_{p^j}(\pi_{f^j} \circ X, s)^2], \qquad f^j = \{p^k\}_{k=j}^m$$

$$E_n^{f_n^{j-1}} = \operatorname{argmin}_{s \in S_{p_n^j}} \sum_{i=1}^{n} \rho_{p_n^j}(\pi_{f_n^j} \circ X_i, s)^2, \qquad f_n^j = \{p_n^k\}_{k=j}^m\,.$$

where $p^j \in E^{f^j}$ and $p_n^j \in E^{f_n^j}$ is a measurable choice for $j = 1,\dots,m$.

We say that a BNFD $f = \{p^k\}_{k=0}^m$ gives *unique* BN population means if $E^{f^j} = \{p^j\}$ with $f^j = \{p^k\}_{k=j}^m$ for all $j = 0,\dots,m$.

Each of the $E^{f^{j-1}}$ and $E_n^{f_n^{j-1}}$ is also called a *generalized Fréchet mean*.

Note that by definition there is only one $p^m = Q \in P_m$. For this reason, for notational simplicity, we ignore it from now on and begin all BNFDs with p^{m-1} and consider thus the corresponding $T_{m-1,k}$.

Definition 18.3 (Factoring Charts). Let $j \in \{0,\dots,m-1\}, k \in \{1,\dots,j\}$. If $T_{j,k}$ and P^{j-k} carry smooth manifold structures near $f' = (p'^j,\dots,p'^{j-k}) \in T_{j,k}$ and

$p'^{j-k} \in P^{j-k}$, respectively, with open $W \subset T_{j,k}, U \subset P^{j-k}$ such that $f' \in W$, $p'^{j-k} \in U$, and with local charts

$$\psi : W \to \mathbb{R}^{\dim(W)},\ f = (p^j, \ldots, p^{j-k}) \mapsto \eta = (\theta, \xi), \quad \phi : U \to \mathbb{R}^{\dim(U)},\ p^{j-k} \mapsto \theta$$

we say that the *chart* ψ *factors*, if with the projections

$$\pi^{P^{j-k}} : T_{j,k} \to P^{j-k},\ f \mapsto p^{j-k}, \quad \pi^{\mathbb{R}^{\dim(U)}} : \mathbb{R}^{\dim(W)} \to \mathbb{R}^{\dim(U)},\ (\theta, \xi) \mapsto \theta$$

we have

$$\phi \circ \pi^{P^{j-k}}|_W = \pi^{\mathbb{R}^{\dim(U)}}|_{\psi(W)} \circ \psi\,.$$

18.3 Assumptions

For the following assumptions suppose that $j \in \{1, \ldots, m-1\}$.

Assumption 18.1 *For a random element X in Q, assume that $\mathbb{E}[\rho_{p^j}(\pi_f \circ X, s)^2] < \infty$ for all BNFDs f ending at p^j, $s \in S_{p^j}$.*

In order to measure a difference between $s \in S_p$ and $s' \in S_{p'}$ for $p, p' \in P_j$ define the orthogonal projection of $s \in S_p$ onto $S_{p'}$ as

$$S^s_{p'} = \operatorname*{argmin}_{s' \in S_{p'}} d_{j-1}(s, s')\,.$$

In case of PNS this is illustrated in Figure 18.1 (a).

Assumption 18.2 *For every $s \in S_p$ there is $\delta > 0$ such that*

$$|S^s_{p'}| = 1$$

whenever $p, p' \in P_j$ with $d_j(p, p') < \delta$.

For $s \in S_p$ and $p, p' \in P_j$ sufficiently close let $s^{p'} \in S^s_{p'}$ be the unique element. Note that in general

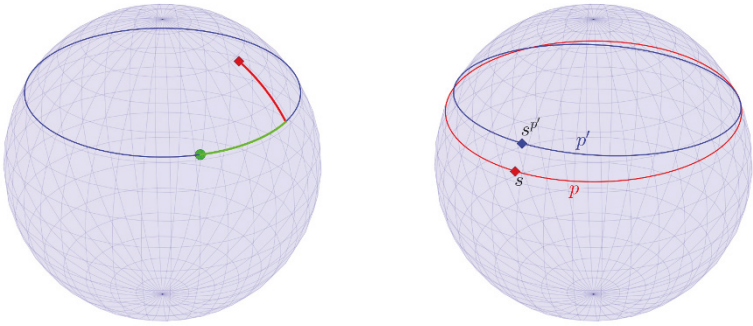
$$(s^{p'})^p \neq s\,.$$

In the following assumption, however, we will require that they will uniformly not differ too much if p is close to p'. Also, we require that $s^{p'}$ and s be close.

Assumption 18.3 *For $\varepsilon > 0$ there is $\delta > 0$ such that*

$$d_{j-1}(s^{p'}, s) < \varepsilon \text{ and } d_{j-1}\big((s^{p'})^p, s\big) < \varepsilon \quad \forall s \in S_p$$

whenever $p, p' \in P_j$ with $d_j(p, p') < \delta$.



(a) Nested projection (b) Projection of descriptors

Fig. 18.1: *PNS illustration. Left: Projection of X (filled diamond) in $Q = \mathbb{S}^2$ onto small circle p and further onto s (filled circle). Right: Projection $s^{p'}$ (on the top circle) onto $S_{p'}$ (which is p' in this case) of s (on the lower circle) on S_p (which is p in this case).*

We will also require the following assumption, which, in conjunction with Assumption 18.3, is a consequence of the triangle inequality, if d_{j-1} is a metric.

Assumption 18.4 *Suppose that $d_j(p_n, p) \to 0$ and $d_{j-1}(s_n, s) \to 0$ with $p, p_n \in P_j$ and $s \in S_p, s_n \in S_{p_n}$. Then also*

$$d_{j-1}(s_n, s^{p_n}) \to 0$$

Moreover, we require uniformity and coercivity in the following senses.

Assumption 18.5 *For all $\varepsilon > 0$ there are $\delta_1, \delta_2 > 0$ such that*

$$\left| \rho_p(\pi_f(q), s) - \rho_{p'}(\pi_{f'}(q), s') \right| < \varepsilon \quad \forall q \in Q$$

for all BNFDs $f, f' \in T_{m-1,m-j-1}$ ending in $p, p' \in P_j$, respectively, with $d(f, f') < \delta_1$ and $s \in S_p, s' \in S_{p'}$ with $d_{j-1}(s, s') < \delta_2$.

Assumption 18.6 *If $p_n, p \in P_j$ and $s_n \in S_{p_n}, s \in S_p$ with $d_{j-1}(s_n, s) \to \infty$, then for every $C > 0$ we have that*

$$\rho_{p_n}(\pi_{f_n} q, s_n) \to \infty$$

for every $q \in Q$ with $\rho_p(\pi_f q, s) < C$ and BNFDs $f, f_n \in T_{m-1,m-j-1}$ ending at p, p_n respectively.

Remark 18.4 *Due to continuity, Assumptions 18.1 and 18.5 hold if Q is compact and Assumption 18.6 if each P_j is compact.*

Again, let $j \in \{1,\dots,m-1\}$.

Assumption 18.7 *Assume that $T_{m-1,m-j}$ carries a smooth manifold structure near the unique BN population mean $f'^{j-1} = (p'^{m-1},\dots,p'^{j-1})$ such that there is an open set $W \subset T_{m-1,m-j}$, $f'^{j-1} \in W$ and a local chart*

$$\psi : W \to \mathbb{R}^{\dim(U)},\ f^{j-1} = (p^{m-1},\dots,p^{j-1}) \mapsto \eta\,.$$

Further, assume that for every $l = j,\dots,m$ the mapping

$$\eta \mapsto f^{l-1} \mapsto \rho_{p^l}(\pi_{f^l} \circ X, p^{l-1})^2 := \tau^l(\eta, X)$$

has first and second derivatives, such that for all $l = j,\dots,m$,

$$\mathrm{Cov}\big[\mathrm{grad}_\eta\, \tau^l(\eta', X)\big], \text{ and } \mathbb{E}\big[\mathrm{Hess}_\eta\, \tau^l(\eta', X)\big]$$

exist and are in expectation continuous near η', i.e. for $\delta \to 0$ we have

$$\mathbb{E}\left[\sup_{\|\eta-\eta'\|<\delta} \Big\|\mathrm{grad}_\eta\, \tau^l(\eta, X) - \mathrm{grad}_\eta\, \tau^l(\eta', X)\Big\|\right] \to 0\,,$$

$$\mathbb{E}\left[\sup_{\|\eta-\eta'\|<\delta} \Big\|\mathrm{Hess}_\eta\, \tau^l(\eta, X) - \mathrm{Hess}_\eta\, \tau^l(\eta', X)\Big\|\right] \to 0\,.$$

Finally, assume that the vectors $\mathbb{E}\big[\mathrm{grad}_\eta\, \tau^{j+1}(\eta', X)\big], \dots, \mathbb{E}\big[\mathrm{grad}_\eta\, \tau^m(\eta', X)\big]$ are linearly independent.

18.4 Asymptotic Theorems

The proofs of the two asymptotic theorems can be found in [4].

Theorem 18.8. *Let $k \in \{0,\dots,m-1\}$ and consider random data $X_1,\dots,X_n \overset{iid}{\sim} X$ on a data space Q admitting BN descriptor families from P_m to P_k, unique BN population means $\{p^m,\dots,p^k\}$ and BN sample means $\{E_n^{f_n^m},\dots,E_n^{f_n^k}\}$ due to a measurable selection $p_n^j \in E_n^{f_n^j}$ giving rise to BNFDs $f_n^j = \{p_n^l\}_{l=j}^m$, $j = k,\dots,m$. If Assumptions 18.1 – 18.6 are valid for all $j = k,\dots,m-1$, and every $\overline{\cup_{n=1}^\infty E_n^{f_n^j}}$ is a.s. d_j-Heine Borel ($j = k,\dots,m$) then $\{E_n^{f_n^m},\dots,E_n^{f_n^k}\}$ converges a.s. to $\{p^m,\dots,p^k\}$ in the sense that $\exists \Omega' \subset \Omega$ measurable with $\mathbb{P}(\Omega') = 1$ such that for all $j = k,\dots,m$, $\varepsilon > 0$ and $\omega \in \Omega'$, $\exists N = N(\varepsilon,\omega)$ with*

$$\bigcup_{r=n}^{\infty} E_r^{f_r^j} \subset \{p \in P_j : d_j(p^j, p) \le \varepsilon\} \quad \forall n \ge N,\ \omega \in \Omega'\,. \tag{18.2}$$

Remark 18.5 *In fact, for the proof we require that the "distances" d_j vanish on the diagonal $d_j(p,p) = 0$ for all $p \in P_j$; they need not be definite, i.e. it is not necessary that $d_j(p,p') = 0 \Rightarrow p = p'$.*

Moreover, note that the d_j-Heine Borel property holds trivially in case of unique sample descriptors.

Theorem 18.9. *Let $j \in \{1,\dots,m-1\}$ and consider random data $X_1,\dots,X_n \overset{iid}{\sim} X$ on a data space Q admitting BNFDs from P_{m-1} to P_{j-1}, a unique BN population mean $f'^{j-1} = \{p'^{m-1},\dots,p'^{j-1}\}$ and BN sample means $\{E_n^{f_n^{m-1}},\dots,E_n^{f_n^{j-1}}\}$ due to a measurable selection $p_n^l \in E_n^{f_n^l}$, $f_n^{j-1} = \{p_n^{m-1},\dots,p_n^{j-1}\}$, $l = j-1,\dots,m-1$.*

(i) Assuming that Assumption 18.7 hold as well as (18.2) for all $j \in \{j-1,\dots,m-1\}$, we have that

$$\sqrt{n} H_\psi\big(\psi^{-1}(f_n^{j-1}) - \psi^{-1}(f'^{j-1})\big) \to \mathcal{N}(0, B_\psi)$$

with a chart ψ as specified in Assumption 18.7 as well as

$$H_\psi = \mathbb{E}\left[\operatorname{Hess}_\eta \tau^j(\eta', X) + \sum_{l=j+1}^{m} \lambda^l \operatorname{Hess}_\eta \tau^l(\eta', X)\right] \text{ and}$$

$$B_\psi = \operatorname{Cov}\left[\operatorname{grad}_\eta \tau^j(\eta', X) + \sum_{l=j+1}^{m} \lambda^l \operatorname{grad}_\eta \tau^l(\eta', X)\right],$$

with the notation from Assumption 18.7 where $\lambda^{j+1},\dots\lambda^m \in \mathbb{R}$ are suitable such that

$$\operatorname{grad}_\eta \mathbb{E}\big[\tau^j(\eta, X)\big] + \sum_{l=j+1}^{m} \lambda^l \operatorname{grad}_\eta \mathbb{E}\big[\tau^l(\eta, X)\big]$$

vanishes at $\eta = \eta'$.

(ii) If additionally $H_\psi > 0$, then f_n^{j-1} satisfies a Gaussian $\sqrt{n}$-CLT

$$\sqrt{n}\big(\psi^{-1}(f_n^{j-1}) - \psi^{-1}(f'^{j-1})\big) \to \mathcal{N}(0, \Sigma_\psi), \quad \Sigma_\psi = H_\psi^{-1} B_\psi H_\psi^{-1}.$$

(iii) If additionally the chart ψ factors as in Definition 18.3, then also p_n^{j-1} satisfies a Gaussian $\sqrt{n}$-CLT

$$\sqrt{n}\big(\phi^{-1}(p_n^{j-1}) - \phi^{-1}(p'^{j-1})\big) \to \mathcal{N}(0, \Sigma_\phi), \quad \Sigma_\phi = \big(\Sigma_{\psi\, ik}\big)_{i,k=1}^{\dim(P_{j-1})}$$

with the notation of Definition 18.3.

Acknowledgements The authors gratefully acknowledge DFG HU 1575/4, DFG SFB 755 and the Niedersachsen Vorab of the Volkswagen Foundation.

References

[1] Anderson, T.: Asymptotic theory for principal component analysis. Ann. Math. Statist. **34** (1), 122–148 (1963)
[2] Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S. C.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging **23** (8), 995–1005 (2004)
[3] Gower, J. C.: Generalized Procrustes analysis. Psychometrika **40**, 33–51 (1975)
[4] Huckemann, S.F, Eltzner, B.: Backward Nested Descriptors Asymptotics with Inference on Stem Cell Differentiation, arXiv preprint: 1609.00814 (2016)
[5] Huckemann, S.F, Hotz, T., Munk, A.: Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). Statist. Sinica **20** (1), 1–100 (2010)
[6] Huckemann, S.F, Ziezold, H.: Principal component analysis for Riemannian manifolds with an application to triangular shape spaces. Advances of Applied Probability (SGSA) **38** (2), 299–319 (2006)
[7] Jung, S., Dryden, I.L., Marron,J.S.: Analysis of principal nested spheres. Biometrika **99** (3), 551–568 (2012)
[8] Pennec, X. Barycentric subspace analysis on manifolds. arXiv preprint:1607.02833 (2016)
[9] Ruymgaart, F. H., Yang, S.: Some applications of Watsons perturbation approach to random matrices. J. Multivariate Anal. **60** (1), 48–60 (1997)
[10] Watson, G: Statistics on Spheres. University of Arkansas Lecture Notes in the Mathematical Sciences, Vol. 6. New York, Wiley (1983)

Chapter 19
Two-sample tests for multivariate functional data

Qing Jiang, Simos G. Meintanis and Lixing Zhu

Abstract We consider two–sample tests for functional data with observations which may be uni– or multi–dimensional. The new methods are formulated as L2–type criteria based on empirical characteristic functions and are convenient from the computational point of view.
Keywords: Functional data, Empirical characteristic function, Two–sample problem

19.1 Introduction

Suppose that we observe data X_{1ij} and X_{2ij} arising from two different groups. For each fixed i, X_{1ij}, is viewed as realization of a curve $x_{1i}(t)$ observed at distinct time points t_{1ij}, $j = 1,...,m_{1i}$, and we index the curves by $i = 1,...,n_1$, for the first group. Likewise suppose that X_{2ij} is realization of a curve $x_{2i}(t)$, observed at times t_{2ij}, $j = 1,...,m_{2i}$, and indexed by $i = 1,...,n_2$, for the second group. The observation times t_{1ij}, t_{2ij} are assumed to belong to some closed bounded interval $\mathbb{T}$, and we often take $\mathbb{T} = [0,1]$. Although we work under the assumption of independence between groups, we allow for noise in the observations. Specifically we consider the model

$$X_{1ij} = x_{1i}(t_{1ij}) + \varepsilon_{1ij}, \;\; X_{2ij} = x_{2i}(t_{2ij}) + \varepsilon_{2ij}, \tag{19.1}$$

Qing Jiang
School of Statistics, Beijing Normal University e-mail: 201631011006@mail.bnu.edu.cn

Simos G. Meintanis (✉)
Department of Economics, National and Kapodistrian University of Athens, Athens, Greece and Unit for Business Mathematics and Informatics, North–West University, Potchefstroom, South Africa e-mail: simosmei@econ.uoa.gr

Lixing Zhu
School of Statistics, Beijing Normal University and Department of Mathematics, Hong Kong Baptist University, Hong Kong e-mail: lzhu@hkbu.edu.hk

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_19

where $x_{11}(t),...,x_{1n_1}(t)$, are independent and identically distributed as $x_1(t)$, and independent of the the errors $\varepsilon_1(t)$, for $t \in \mathbb{T}$, and likewise $x_{21}(t),...,x_{2n_2}(t)$ are iid as $x_2(t)$, and independent of $\varepsilon_2(t)$. The errors are also assumed to be mutually independent with zero means. We wish to test the null hypothesis

$$\mathscr{H}_0 : x_1(t) \stackrel{d}{=} x_2(t), \text{ for each } t \in \mathbb{T}, \tag{19.2}$$

where $\stackrel{d}{=}$ stands for equality in law.

Earlier works for the two–sample problem with functional data include testing for common location ([6, 5, 16]), and for common covariance matrix ([11, 8]), while [1] considers the more general problem of testing for common principal components. The framework of the current paper though is much in the spirit of [4] where the two–sample problem was first studied in its full generality of the null hypothesis (19.2); see also [12]. However we deviate from this paper by proposing procedures which instead of the empirical distribution function, utilize the empirical characteristic function (ECF). Apart from other favorable features which will become apparent along the paper, note that ECF–based procedures for scalar data are readily extended to multidimensional observations which is not always true if one employs classical procedures based on the empirical distribution function.

19.2 Test Statistics

19.2.1 Univariate case

Our approach for testing the null hypothesis $\mathscr{H}_0$ in (19.2) will be based on the fact that $\mathscr{H}_0$ is tantamount to the identity

$$\varphi_{x_1(t)}(u) = \varphi_{x_2(t)}(u), \ \forall\, u \in \mathbb{R}, \text{ and each } t \in \mathbb{T}, \tag{19.3}$$

and vice versa. Here, as well as elsewhere below, $\varphi_{z(t)}(u) := \mathbb{E}(e^{\imath u z(t)})$, $(\imath = \sqrt{-1})$, will denote the characteristic function (CF) of the stochastic quantity $z(t)$. Based on this fact, [10] and [9] develop two–sample testing procedures for multivariate data. Here we follow this approach and in line with [4], we assume that the curves $x_{1i}(t)$ and $x_{2i}(t)$ may be recovered following non–parametric techniques and write $\widehat{x}_{1i}(t)$ and $\widehat{x}_{2i}(t)$ for the resulting curve estimators. Consider the corresponding ECFs

$$\widehat{\varphi}_{1t}(u) = \frac{1}{n_1}\sum_{i=1}^{n_1} e^{\imath u \widehat{x}_{1i}(t)}, \quad \widehat{\varphi}_{2t}(u) = \frac{1}{n_2}\sum_{i=1}^{n_2} e^{\imath u \widehat{x}_{2i}(t)}, \tag{19.4}$$

computed from $\widehat{x}_{11}(t),...,\widehat{x}_{1n_1}(t)$ and $\widehat{x}_{21}(t),...,\widehat{x}_{2n_2}(t)$, respectively. Then in view of (19.3) we suggest the test statistic

$$D_w = \int_{\mathbb{T}} \int_{\mathbb{R}} \delta_t(u) w(u) du dt, \tag{19.5}$$

where

$$\delta_t(u) = \left| \widehat{\varphi}_{1t}(u) - \widehat{\varphi}_{2t}(u) \right|^2, \tag{19.6}$$

and $w > 0$ denotes a weight function satisfying $\int_{\mathbb{R}} w(u) du < \infty$.

19.2.2 Multivariate case

The latent curves $x_k(t) = (\chi_{k1}(t), ..., \chi_{kp}(t))'$, $k = 1, 2$, may also be multidimensional. This is a new area where functional data are observed over time t, but realizations are complex geometrical structures in dimension $p > 1$; see [3], [7], and [2], for recent contributions on statistical techniques for multivariate functional data. Following the lines of the previous section for testing the null hypothesis (19.2) we will consider a criterion analogous to that in (19.5). However, in order to avoid nonparametric estimation which is problematic in high dimension we will have to modify our assumptions regarding model (19.1). Specifically we adopt the model

$$X_{1i}(t) = x_{1i}(t) + \varepsilon_{1i}(t), \;\; X_{2i}(t) = x_{2i}(t) + \varepsilon_{2i}(t), \tag{19.7}$$

and we assume that observations are collected over time $t := t_j$, $j = 1, ..., m$, for both groups, with m being large, i.e., we have a common sampling design between the two groups which is dense. Moreover we assume that sampling noise is equidistributed between the two groups, $\varepsilon_1(t) \stackrel{d}{=} \varepsilon_2(t)$, with a common CF that never vanishes. Under these assumptions and using the Fourier identities $\varphi_{X_{k(t)}}(u) = \varphi_{x_{k(t)}}(u) \varphi_{\varepsilon_{k(t)}}(u)$, $k = 1, 2$, resulting from model (19.7), we conclude that the null hypothesis $\mathscr{H}_0$ in (19.2) holds if and only if

$$\varphi_{X_1(t)}(u) = \varphi_{X_2(t)}(u), \; \forall \, u \in \mathbb{R}^p, \text{ and each } t \in \mathbb{T}. \tag{19.8}$$

In view of this fact we propose the test statistic

$$\Delta_W = \int_{\mathbb{R}^p} \delta(u) W(u) du, \tag{19.9}$$

with $W : \mathbb{R}^p \mapsto (0, \infty)$ and satisfying $\int_{\mathbb{R}^p} W(u) du < \infty$, where

$$\delta(u) = \frac{1}{m} \sum_{j=1}^{m} \left| \phi_{1j}(u) - \phi_{2j}(u) \right|^2, \tag{19.10}$$

with

$$\phi_{1j}(u) = \frac{1}{n_1} \sum_{i=1}^{n_1} e^{\imath u' X_{1ij}}, \quad \phi_{2j}(u) = \frac{1}{n_2} \sum_{i=1}^{n_2} e^{\imath u' X_{2ij}}, \tag{19.11}$$

being the ECFs computed directly from the observed data $X_{k1j},...,X_{kn_kj}$, and which correspond to the CFs $\varphi_{X_1}(u)$ and $\varphi_{X_2}(u)$, respectively, considered at fixed time points t_j, for each $j = 1,...,m$.

19.3 Computations and Interpretations

19.3.1 Univariate case

Our procedures enjoy the advantage of computational simplicity. To see this we first proceed from (19.6) and by using simple algebra and trigonometric identities we get

$$\delta_t(u) = \frac{1}{n_1^2}\sum_{i,\ell=1}^{n_1}\cos(u(\widehat{x}_{1i}(t)-\widehat{x}_{1\ell}(t))) + \frac{1}{n_2^2}\sum_{i,\ell=1}^{n_2}\cos(u(\widehat{x}_{2i}(t)-\widehat{x}_{2\ell}(t))) - \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{\ell=1}^{n_2}\cos(u(\widehat{x}_{1i}(t)-\widehat{x}_{2\ell}(t))) \tag{19.12}$$

Then by making use of the previous equation in (19.5) we conclude that the test statistic can be written as

$$D_w = \frac{1}{n_1^2}\sum_{i,\ell=1}^{n_1} I_{w,\mathbb{T}}(\widehat{x}_{1i},\widehat{x}_{1\ell}) + \frac{1}{n_1^2}\sum_{i,\ell=1}^{n_2} I_{w,\mathbb{T}}(\widehat{x}_{2i},\widehat{x}_{2\ell}) - \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{\ell=1}^{n_2} I_{w,\mathbb{T}}(\widehat{x}_{1i},\widehat{x}_{2\ell}) \tag{19.13}$$

where

$$I_{w,\mathbb{T}}(z_1,z_2) = \int_{\mathbb{T}}\int_{\mathbb{R}}\cos(u(z_1(t)-z_2(t)))w(u)dudt. \tag{19.14}$$

The weight function $w(\cdot)$ in (19.14) may be chosen in a way that avoids numerical integration in the inner integral $\int\cos(u(z))w(u)du$ but for further details on this we refer to the next subsection. Then again having computed $\int\cos(u(z))w(u)du := g(z(t))$, say, one also has to compute the outer integral $\int g(z(t))dt$, over $\mathbb{T}$. However even in the simplest case of local linear smoothers[4], the closed form obtained for $\widehat{x}_{ki}(t)$ is quite complicated and therefore one needs to resort to numerical integration. Despite this, integration in closed bounded domains is a well studied numerical problem and there exist several routines available for this purpose. Hence we do not expect any complications to be associated with this part of our procedure. For simplicity we take $\mathbb{T} = [0,1]$

The choice for the weight function $w(\cdot)$ is usually based upon computational considerations. In fact if $w(\cdot)$ integrates to one (perhaps after some scaling) and satisfies $w(-u) = w(u)$ then the inner integral in (19.14) can be interpreted as the CF of a symmetric around zero random variable having density $w(\cdot)$. In this connection $w(\cdot)$ can be chosen as the density of any such distribution. Typically we consider a fixed family of weight functions, say $w := w_\gamma$ indexed by a parameter $\gamma > 0$. For instance a weight function $w_\gamma(u)$ which is proportional to $e^{-\gamma u^2}$, corresponds to the

Gaussian density, but for computational purposes any density with a simple CF will do. In fact, one might wonder whether there is a weight function which is optimal in some sense. The issue is still open but based on earlier results it appears that the issue of the choice of the weight function is similar to the corresponding problem of choosing a kernel and a bandwidth in nonparametric estimation: the asymptotics of the test are qualitatively invariant with respect to w_γ. Moreover most weight functions (kernels) render similar finite–sample behavior of the resulting test statistic, which is very competitive compared to classical procedures based on the empirical distribution function. Nevertheless there is some sensitivity of the ECF–tests with respect to the "bandwidth" parameter γ; see [10] and [9]. This is a highly technical problem that has been tackled only under the restrictive scenario of testing goodness–of–fit for a given parametric distribution, and even then a good choice of γ depends on the direction away from the null hypothesis; see [15]. Thus in our context the approach to the weight function is in some sense pragmatic: we use the Gaussian weight function which has become a standard, and investigate the behavior of the criterion over a grid of values of the weight parameter γ. However in our Monte Carlo study, alternative weight functions will also be tried.

19.3.2 Multivariate case

We proceed from (19.10) and by using (19.12) we obtain

$$\delta(u) = \frac{1}{m}\sum_{j=1}^{m}\delta_j(u), \tag{19.15}$$

where

$$\begin{aligned}\delta_j(u) = {} & \frac{1}{n_1^2}\sum_{i,\ell=1}^{n_1}\cos(u'(X_{1ij}-X_{1\ell j})) + \frac{1}{n_2^2}\sum_{i,\ell=1}^{n_2}\cos(u'(X_{2ij}-X_{2\ell j})) \\ & -\frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{\ell=1}^{n_2}\cos(u'(X_{1ij}-X_{2\ell j}))\end{aligned} \tag{19.16}$$

Consequently the test statistic can be written as

$$\begin{aligned}\Delta_W = {} & \frac{1}{m}\sum_{j=1}^{m}\Big(\frac{1}{n_1^2}\sum_{i,\ell=1}^{n_1} I_W(X_{1ij}-X_{1\ell j}) + \frac{1}{n_1^2}\sum_{i,\ell=1}^{n_2} I_W(X_{2ij}-X_{2\ell j}) \\ & -\frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{\ell=1}^{n_2} I_W(X_{1ij}-X_{2\ell j})\Big)\end{aligned} \tag{19.17}$$

where

$$I_W(x) = \int_{\mathbb{R}^p}\cos(u'x)W(u)du. \tag{19.18}$$

As already mentioned, the weight function $W(\cdot)$ in (19.18) may be chosen in a way that avoids numerical integration, which is problematic in higher dimension. To see this recall that the CF of any spherical random variable Z is given by $\varphi_Z(u) = \Psi(\|u\|)$, for some, family specific, scalar function $\Psi(\cdot)$, where $\|u\|$ denotes the usual Euclidean norm. Hence if $f_Z(z)$ denotes the density corresponding to $\varphi_Z(u)$ we have

$$\int_{\mathbb{R}^p} \cos(u'z) f_Z(z) dz = \Psi(\|u\|).$$

The last equation implies that if $f_Z(\cdot)$ is used as weight function $W(\cdot)$ in (19.18), then the resulting test statistic, say Δ_Ψ, reduces to

$$\Delta_\Psi = \frac{1}{m}\sum_{j=1}^{m}\Big(\frac{1}{n_1^2}\sum_{i,\ell=1}^{n_1}\Psi(\|X_{1ij}-X_{1\ell j}\|)+\frac{1}{n_1^2}\sum_{i,\ell=1}^{n_2}\Psi(\|X_{2ij}-X_{2\ell j}\|) - \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{\ell=1}^{n_2}\Psi(\|X_{1ij}-X_{2\ell j}\|)\Big). \tag{19.19}$$

The test criterion in (19.19) is further advanced by considering specific families of spherically symmetric distributions with a simple CF. Such a family of distributions is the family of spherical stable distributions with $\Psi(u) = e^{-u^\alpha}$, where $0 < \alpha \le 2$, stands for the characteristic exponent. Interesting special cases of spherical stable distributions are the multivariate Cauchy and normal distributions corresponding to $\alpha = 1$ and $\alpha = 2$, respectively. Other convenient choices are the multivariate Laplace distribution with $\Psi(u) = (1+u^2)^{-1}$ and some special cases of the family of multivariate Kotz–type distributions.

We will elaborate here on the case of the spherical stable distribution as weight function in (19.18). Note that if this function is used in (19.19), it yields the test criterion

$$\Delta_\alpha = \frac{1}{m}\sum_{j=1}^{m}\Big(\frac{1}{n_1^2}\sum_{i,\ell=1}^{n_1}e^{-\|X_{1ij}-X_{1\ell j}\|^\alpha}+\frac{1}{n_2^2}\sum_{i,\ell=1}^{n_2}e^{-\|X_{2ij}-X_{2\ell j}\|^\alpha} - \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{\ell=1}^{n_2}e^{-\|X_{1ij}-X_{2\ell j}\|^\alpha}\Big). \tag{19.20}$$

Interestingly there is a connection between (19.20) and another two–sample test statistic in the literature. To see this let Z follow a spherical stable distribution with characteristic exponent α and choose the density of the random variable $Z/\gamma^{1/\alpha}$ as weight function in (19.17)–(19.18), for some $\gamma > 0$. To get a formula for the resulting criterion recall that the CF of the last random variable is given by $e^{-\frac{\|u\|^\alpha}{\gamma}}$ which yields a test statistic, say $\widetilde{\Delta}_{\alpha,\gamma}$, analogous to the criterion in (19.20) but with $\|\cdot\|^\alpha$ being replace by $\|\cdot\|^\alpha/\gamma$ throughout eqn. (19.20). Now if we take a two–term expansion $e^{-\|x\|^\alpha/\gamma} = 1 - \|x\|^\alpha/\gamma + o(\gamma^{-1})$, $\gamma \to \infty$, in the new test statistic $\widetilde{\Delta}_{\alpha,\gamma}$, this will lead after some algebra to

$$\lim_{\gamma\to\infty}\gamma\widetilde{\Delta}_{\alpha,\gamma}=\frac{1}{m}\sum_{j=1}^{m}\Big(\frac{2}{n_1n_2}\sum_{i=1}^{n_1}\sum_{\ell=1}^{n_2}\|X_{1ij}-X_{2\ell j}\|^{\alpha}-\frac{1}{n_1^2}\sum_{i,\ell=1}^{n_1}\|X_{1ij}-X_{1\ell j}\|^{\alpha}$$
$$-\frac{1}{n_2^2}\sum_{i,\ell=1}^{n_2}\|X_{2ij}-X_{2\ell j}\|^{\alpha}\Big). \tag{19.21}$$

The criterion in the r.h.s. of (19.21) is the so–called *energy statistic* of [13] adapted to the functional context. We mention in this connection that energy statistics have gained considerable popularity lately as they have been employed not just for two–sample testing but also for testing for independence as well as in nonparametric analysis of variance. The reader is referred to the review of [14] for more information on energy statistics.

19.4 Resampling procedures

The null distribution of the test statistics considered depends, among other things, on the underlying stochastic properties of the random variables $x_1(t)$ and $x_2(t)$ involved. In order to deal with this issue we apply appropriate resampling procedures for computing critical points and actually carrying out the tests. To this end, let $\mathscr{D}=\mathscr{D}(\xi_1,...,\xi_n)$ be a generic notation for a test statistic which depends on a sample of size n of observations $\xi_j,\ 1\le j\le n$. Clearly in our case $n=n_1+n_2$. We will apply the permutation procedure whereby we randomly generate a permutation $b=\{b_1,...,b_n\}$ of $\{1,...,n\}$, and compute the test statistic $\mathscr{D}_b=\mathscr{D}(\xi_{b_1},...,\xi_{b_n})$. The procedure is repeated a number of times $b=1,...,\mathrm{B}$, and the critical point of the test of size α is determined as the corresponding $(1-\alpha)$ quantile $D_{((1-\alpha)\mathrm{B})}$ of the values $\mathscr{D}_b$, $b=1,...,\mathrm{B}$. The null hypothesis is then rejected if $\mathscr{D}>\mathscr{D}_{((1-\alpha)\mathrm{B})}$.

Suppose that data $X_{k1j},\ldots,X_{kn_kj}$ are observed at fixed time points t_j, for each $j=1,\ldots,m$. For univariate data, the critical point of the test statistic in (19.5) is computed as in [4], i.e., by permuting $\{\widehat{x}_{11},...,\widehat{x}_{1n_1},\widehat{x}_{21},...,\widehat{x}_{2n_2}\}$. In turn with multivariate data, permutations for the criterion in (19.9) are performed on $\{X_{11j},\ ...,\ X_{1n_1j},\ X_{21j},\ ...,\ X_{2n_2j}\}$, for each $j=1,...,m$.

In the univariate case, we generate data $\{(t_{kij},x_{ki}(t_{kij})):j\}_{i=1}^{n_k},k=1,2$, mainly as in [4]. For completeness we describe the data as follows: the sampling design for the curves is assumed balanced $(m_{1i}=m_{2i}=m),\forall i$, and regular. Specifically, suppose that $t_{ki\cdot},k=1,2,i=1,\ldots,n_k$ are discrete uniform fixed time points on $[0,1]$. It is assumed that $x_{1i}(t)=\sum_{k=1}^{15}e^{-k/2}N_{k1i}\psi_k(t)$ and

$$x_{2i}(t)=\sum_{k=1}^{15}e^{-k/2}N_{k21i}\psi_k(t)+\delta\sum_{k=1}^{15}k^{-2}N_{k22i}\psi_k^*(t)$$

where $N_{k1i},N_{k21i},N_{k22i}$ are i.i.d. standard normal variables, $\delta\ge 0$ controls the deviation from the null hypothesis ($\delta=0$ under H_0). Here $\psi_1(t)\equiv 1,\forall t$ and $\psi_k(t)=\sqrt{2}\sin\{(k-1)\pi t\}$ are orthonormal basis functions. Also

$$\psi_k^*(t) = \begin{cases} 1 & \text{if } k = 1 \\ \sqrt{2}\sin\{(k-1)\pi(2t-1)\} & \text{if } k \text{ is odd and } k > 1 \\ \sqrt{2}\cos\{(k-1)\pi(2t-1)\} & \text{if } k \text{ is even} \end{cases}$$

are orhonormal basis functions. Two scenarios are considered: *i*) $m = 20$ points per curve, and *ii*) $m = 100$ points per curve. Figure 19.1 (without sampling noise) illustrates the ECF test results for significance level $a = 0.05$. The simulation results are based on 500 samples, and the critical values of the test are obtained from 1000 permutation samples.

Figure 19.1 illustrates that the level is well respected under the null hypothesis and the power increases for larger values of m, n and δ in two conditions. The number of observations per curve, m, has limited impact on the power and the conclusion is consistent with [4]. However, compared with their results, the empirical power of the ECF test increases at a faster rate than the CVM test of [4]. This should not be surprising, as we do not need to estimate all basis functions by smoothing the data when observations are without noise or sampling noise is equidistributed. When observations are without noise, we can directly estimate the test statistic Δ_α in (19.20). Also the Fourier identities makes it consistent to transform the null hypothesis (19.2) to equation (19.8) when sampling noise is equidistributed.

19.5 Conclusion

We suggest a new procedure for testing the two–sample null hypothesis with functional data. The procedure is an adaptation to the functional– data set up of earlier methods for the same problem with perfectly observed i.i.d. data. Here we present only the main ideas of the new methods and a small Monte Carlo study. A detailed study of the asymptotic as well as the finite–sample behavior of the methods is currently under investigation and will be reported elsewhere.

References

[1] Benko, M., Härdle, W., Kneip, A.: Common functional principal components. Ann. Stat. **37**, 1–34 (2009)
[2] Berrendero, J.R., Justel, A., Svarc, M.: Principal components for multivariate functional data. Comput. Stat. Data An. **55**, 2619–2634 (2011)
[3] Chiou, J.M., Müller, H.G.: Linear manifold modeling of multivariate functional data. J. Roy. Stat. Soc. B. **76**, 605–626 (2014)
[4] Hall, P., Van Keilegom, I.: Two–sample tests in functional data analysis starting from discrete data. Stat. Sinica. **17**, 1511–1531 (2007)
[5] Horváth, L., Kokoszka, P.: Inference for Functional Data with Applications. Springer Series in Statistics, Springer, New York (2012)

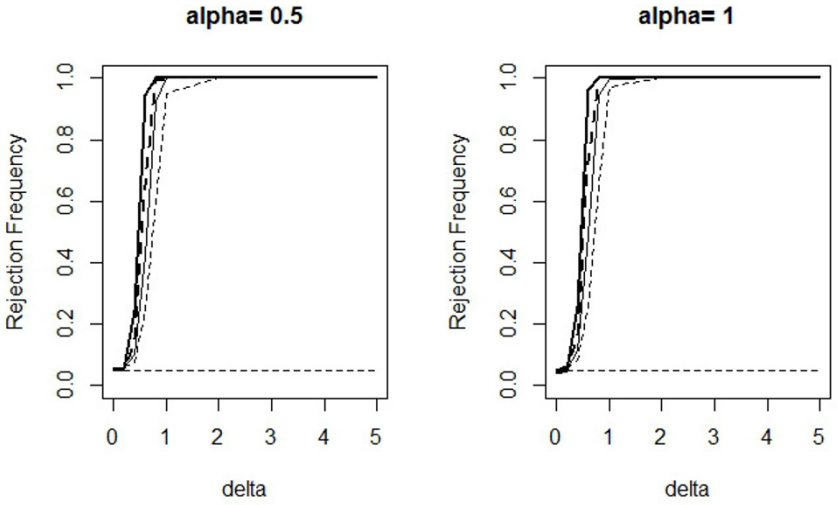


Fig. 19.1: Rejection frequencies of the ECF test for data without noise when $n = 15$ (dashed line) and $n = 25$ (solid line) corresponding to level of significance is $a = 5\%$ for the test statistic in (20) with characteristic exponent α. The thin lines correspond to $m = 20$, the thick lines to $m = 100$.

[6] Horváth, L., Kokoszka, P., Reeder, R.: Estimation of the mean of functional time series and a two sample problem. J. Roy. Stat. Soc. B. **75**, 103–122 (2013)
[7] Jacques, J., Preda, C.: Model–based clustering for multivariate functional data. Comput. Stat. Data An. **71**, 92–106 (2014)
[8] Kraus, D., Panaretos, V.M.: Dispersion operators and resistant second–order functional data analysis. Biometrika. **99**, 813–832 (2012)
[9] Hušková, M., Meintanis, S.G.: Tests for the multivariate k-sample problem based on the empirical characteristic function. J. NonParametr. Statist. **20**, 263–277 (2008)
[10] Meintanis, S.G.: Permutation tests for homogeneity based on the empirical characteristic function. J. Nonparametr. Statist. **17**, 583–592 (2005)
[11] Panaretos, V.M., Kraus, D., Maddocks, J.H.: Second–order comparison of Gaussian random functions and the geometry of DNA minicircles. J. Am. Stat. Assoc. **105**, 670–682 (2010)
[12] Pomann, G.M., Staicu, A.M., Ghosh, S.: Two–sample hypothesis testing for functional data with application to a diffusion tensor imaging study of multiple sclerosis. J. Roy. Stat. Soc. C-App. **65**, 395–414 (2016)
[13] Székely G., Rizzo, M.: Hierarchical clustering via joint between–within distances: Extending Ward's minimum variance method. J. Classif. **22**, 151–183 (2005)

[14] Székely G., Rizzo, M.: Energy statistics: A class of statistics based on distances. J. Stat. Plan. Infer. **143**, 249–272 (2013)
[15] Tenreiro, C.: On the choice of the smoothing parameter for the BHEP goodness-of-fit test. Computat. Statist. Dat. Anal. **53**, 1038–1053 (2009)
[16] Zhang, C., Peng, H., Zhang, J.T.: Two–sample tests for functional data. Commun. Stat.–Theor. M. **39**, 559–578 (2010)

Chapter 20
Functional quantile regression: local linear modelisation

Zoulikha Kaid and Ali Laksaci

Abstract A nonparametric local linear estimator of the conditional quantiles of a scalar response variable Y given a random variable X taking values in a semi-metric space. We establish the almost complete consistency and the asymptotic normality of this estimate. We prove that the asymptotic proprieties of this estimate are closely related to some topological characteristics of the data. Finally, a Monte Carlo study is carried out to evaluate the performance of this estimate.

20.1 Introduction

In the last two decades the statistical analysis of functional random variable has attracted considerable interest (see [17, 21, 11, 13], for an overview on this topic). In this context the conditional quantiles estimation has received extensive attention in the literature. See, for instance [4], [14] for some parametric approach or [10], [5] for the nonparametric modeling. Noting that most of nonparametric studies for functional quantiles regression are based on the classical kernel estimation method. In this paper we consider an alternative approach based on the local linear method. Recall that the local linear modeling has various advantage over the classical kernel method. In particular, the local linear method has superior bias properties than the kernel method (cf. [9] for more discussions on this subject). Moreover, the kernel method can be treated as particular case of the local linear method. Precisely, the classical kernel method is so-called local constant method. In functional data analysis, the local linear estimation has been introduced by [3]. They studied the L_2

Zoulikha Kaid
Laboratoire de Statistique et Processus Stochastiques, Université Djillali Liabès, Sidi Bel-Abbès, BP. 89, Sidi Bel-Abbès 22000, Algeria , e-mail: kaedzoulekha@yahoo.com

Laksaci Ali (✉)
Laboratoire de Statistique et Processus Stochastiques, Université Djillali Liabès, Sidi Bel-Abbès, BP. 89, Sidi Bel-Abbès 22000, Algeria e-mail: alilak@yahoo.fr

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_20

consistency of a local linear estimator of the regression operator when the covariates are Hilbertian. [1] proposes an alternative regression local linear estimator which can be used for more general functional regressor. The authors show the almost complete convergence (with rate) of the proposed estimate. Another local linear estimate of the regression operator was considered by [2]. The latter estimate is obtained by inverting the local covariance operator of the functional explanatory variable. [7] provides pointwise and uniform almost consistency of the functional local linear estimate of the conditional density function. They applied their results to develop a local linear estimate of the conditional mode function. We return to [8] for the local linear estimating of the functional cumulative distribution function. Recently, Messaci et al. [16] have constructed a local linear estimator of the conditional quantile function obtained by inverting the estimator proposed by [8]. However, unlike to the local constant estimate, the inverse of the local linear estimate of the functional cumulative distribution function is is very difficult to obtain in practice. Thus, to avoid this problem we construct an alternative estimate based on the L_1 approach. Specifically, we use the functional local linear procedure proposed by [1] to construct the local linear estimate of the quantile regression function. Our estimator keeps the robustness of the quantile regression function and the advantages of the local linear approach. As asymptotic results we show the almost complete consistency and the asymptotic normality of the constructed estimator. We specify for both results the convergence rate of the estimate. The convergence rates of this estimate are expressed by means of some functional probability function such as the small probability function and the variation function. It should be noted that the quntile regression function is one of the most studied models in nonparametric regression analysis (see for instance [19], and [18] for previous results and [12], [6] and [20] for recent advances and references).

This contribution is organized as follows: In the following section, we present the steps that allow us to construct our local linear estimator. The main results is given in Section 3.

20.2 The model and the estimator

Let (X,Y) be a couple of random variables in $\mathscr{F} \times \mathbb{R}$, where $\mathscr{F}$ is a semi-metric space, of eventually infinite dimension. In this context, X can be a functional random variable. We denote by d the semi-metric on $\mathscr{F}$. We assume that the regular version of the conditional probability of Y given X exists and has a continuous, bounded density with respect to Lebesgue measure over $\mathbb{R}$. For all $x \in \mathscr{F}$, we denote by F^x the conditional distribution function of Y given $X = x$ (resp. by f^x the conditional density of Y given $X = x$).
In the following, we fix a point x in $\mathscr{F}$ and we denote by N_x a neighbor of this point. For $\alpha \in]0,1[$, the conditional quantile of order α, denoted $t_\alpha(x)$ is solution with respect to (w.r.t.) t of the following equation

$$\mathbb{E}\left[L_\alpha(Y-t)\,|X=x\right]=0 \tag{20.1}$$

where $L_\alpha(x)=\mathbb{1}_{x>0}-(1-\alpha)$ and $\mathbb{1}$ is indicator function.
As previously indicated, our main purpose of this work is to study the functional local linear estimate of the quantile regression. Recall that the local polynomial smoothing method is based on the assumption that is the functional operator is smooth enough to be locally well approximated by a polynomial. In functional statistics, there are several ways for extending this approach (see, [3] or [1] for some examples). In this work we adopt the fast version proposed by [1] for which the regression function $t_\alpha(x)$ is approximated by

$$t_\alpha(z)=a+b\beta(x,z)$$

where a and b are estimated by $\widehat{a}$ and $\widehat{b}$ solution of

$$\min_{(a,b)\in\mathbb{R}^2}\sum_{i=1}^{n}L_\alpha(Y_i-a-b\beta(X_i,x))K(h^{-1}\delta(x,X_i)) \tag{20.2}$$

where $\beta(.,.)$ is a known function from $\mathscr{F}^2$ into $\mathbb{R}$ such that, $\forall\xi\in\mathscr{F}$, $\beta(\xi,\xi)=0$, with K is kernel and $h=h_{K,n}$ is a sequence of positive real numbers and $\delta(.,.)$ is a function of $\mathscr{F}\times\mathscr{F}$ such that $d(.,.)=|\delta(.,.)|$. It follows that

$$t_\alpha(x)=a \quad \text{and} \qquad \widehat{t_\alpha}(x)=\widehat{a}.$$

We point out that if $b=0$ then we obtain from (20.2) the functional Nadaraya-Watson estimator studied by [15]. It is worth to pointing out that unlike the local linear regression model this estimator cannot explicitly determined. So, the establishment of the asymptotic proprieties of our estimate is very difficult, it requires the Bahadur representation of $t_\alpha(x)$.

20.3 Main results

In order to derive the almost complete convergence (*a.co.*) and the asymptotic normality of the local linear estimate $\widehat{t_\alpha}(x)$ of $t_\alpha(x)$. For this, we denote by $\phi_x(r_1,r_2)=\mathbb{P}(r_2\leq\delta(x,X)\leq r_1)$ and we assume the following hypotheses:

(H1) For any $r>0$, $\phi_x(r):=\phi_x(-r,r)>0$ and there exists a function $\chi_x(\cdot)$ such that:

$$\forall t\in(-1,1),\ \lim_{h\to0}\frac{\phi_x(th,h)}{\phi_x(h)}=\chi_x(t)$$

(H2) There exists $\delta>0$, $\forall(t_1,t_2)\in[t_p(x)-\delta,t_p(x)+\delta]^2$, $\forall(x_1,x_2)\in\mathscr{N}_x^2$,

$$|F^{x_1}(t_1)-F^{x_2}(t_2)|\leq C_0\left(d^b(x_1,x_2)+|t_1-t_2|^k\right),\quad \text{for } C_0,b,k>0.$$

(H3) The function $\beta(.,.)$ is such that:

$$\begin{cases} \forall z \in \mathscr{F},\ C_1\,|\delta(x,z)| \le |\beta(x,z)| \le C_2\,|\delta(x,z)|,\ \text{where } C_1>0, C_2>0, \\ \\ \qquad\qquad \sup_{u\in B(x,r)} |\beta(u,x)-\delta(x,u)| = o(r) \\ \\ \text{and} \qquad h\int_{B(x,h)} \beta(u,x)dP(u) = o\left(\int_{B(x,h)} \beta^2(u,x)\,dP(u)\right) \end{cases}$$

where $B(x,r)=\{z\in\mathscr{F} : |\delta(x,z)|\le r\}$ and $dP(x)$ is the probability distribution of X.

(H4) The kernel K is a positive, differentiable function which is supported within $(-1,1)$ such that

$$\begin{pmatrix} K(1)-\int_{-1}^1 K'(t)\chi_x(t)dt & K(1)-\int_{-1}^1 tK'(t)\chi_x(t)dt \\ \\ K(1)-\int_{-1}^1 tK'(t)\chi_x(t)dt & K(1)-\int_{-1}^1 t^2K'(t)\chi_x(t)dt \end{pmatrix}$$

is a positive definite matrix.

(H5) The bandwidth h satisfies: there exists an integer n_0, such that

$$\forall n>n_0,\ -\frac{1}{\phi_x(h)}\int_{-1}^1 \phi_x(zh,h)\frac{d}{dz}\left(z^2K(z)\right)dz > C_3 > 0,$$

$$\lim_{n\to\infty}\frac{\log n}{n\phi_x(h)} = 0.$$

Conditions (H1)-(H5) are are rather classical in this context of functional local linear analysis. More precisely, (H1) and (H3)-(H5) are the same as those used by [1] for the nonparametric regression operator. The mild condition (H2) is also similar to condition (A1) in [15] for the multivariate local linear quantile regression.

Our main results is given in the following Theorem

Theorem 20.1. *Under hypotheses (H1)-(H5) and if $f^x(t_\alpha(x))>0$ then*

$$|\widehat{t_\alpha}(x)-t_\alpha(x)| = O\left(h^b\right)+O\left(\left(\frac{\log n}{n\,\phi_x(h)}\right)^{1/2}\right) \qquad a.co.$$

and

$$\left(\frac{n\phi_x(h)}{\sigma^2(x)}\right)^{1/2}\left(\widehat{t_\alpha}(x)-t_\alpha(x)\right)\xrightarrow{\mathscr{D}}\mathscr{N}(0,1) \qquad as \quad n\to\infty$$

where $\xrightarrow{\mathscr{D}}$ denotes the convergence in distribution, $\sigma^2(x)=\dfrac{\alpha(1-\alpha)a_2(x)}{(f^x(t_\alpha(x)))^2a_1^2(x)}$ with

$$a_j(x)=K^j(1)-\int_{-1}^1 (K^j)'(s)\chi_x(s)ds \qquad for \quad j=1,2. \tag{20.3}$$

References

[1] Barrientos-Marin, J., Ferraty, F., Vieu, P.: Locally modelled regression and functional data. J. Nonparametr. Stat., **22**, No. 5, 617–632 (2010)

[2] Berlinet, A., Elamine, A., Mas, A.: Local linear regression for functional data. Ann. Inst. Statist. Math., **63**, 1047–1075 (2011)

[3] Baíllo, A., Grané, A.: Local linear regression for functional predictor and scalar response. J. Multivariate Anal., **100**, 102–111 (2009)

[4] Cardot, H.; Crambes, Ch., Sarda, P.: Quantile regression when the covariates are functions. J. Nonparametr. Stat., **17**, 841–856 (2005)

[5] Dabo-Niang, S., Kaid, Z., Laksaci, A.: Asymptotic properties of the kernel estimate of spatial conditional mode when the regressor is functional. *AStA Adv. Stat. Anal.* **99**, 131–160 (2015)

[6] Dabo-Niang, S., Laksaci, A.: Nonparametric quantile regression estimation for functional dependent data. Comm. Statist. Theory Methods, **41** , 1254–1268 (2015)

[7] Demongeot, J., Laksaci, A., Madani, F., Rachdi, M.: Functional data: local linear estimation of the conditional density and its application. Statistics, **47**, 26–44 (2013)

[8] Demongeot, J., Laksaci, A., Rachdi, M., Rahmaani S.: On the local linear modelization of the conditional distribution for functional data, Sankhya A **76** 328–355 (2014)

[9] Fan, J., Gijbels, I. Local polynomial modelling and its applications. London, Chapman & Hall (1996)

[10] Ferraty, F., Vieu, P. Nonparametric functional data analysis. Theory and Practice. Springer Series in Statistics. New York (2006)

[11] Goia, A., Vieu, P.: An introduction to recent advances in high/infinite dimensional Statistics. J. Multivariate Anal., **146**, 1-6 (2016)

[12] Hallin, M., Lu, Z., Yu, K.: Local linear spatial quantile regression. Bernoulli, **15** , 659–686 (2009)

[13] Hsing, T., Eubank, R.: Theoretical foundations of functional data analysis, with an introduction to linear operators. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester (2015)

[14] Kato, K.: Estimation in functional linear quantile regression. Ann. Statist. **40**, 3108–3136 (2012)

[15] Laksaci, A., Lemdani, M., Ould Saïd, E.: A generalized L^1-approach for a kernel estimator of conditional quantile with functional regressors: Consistency and asymptotic normality. Statist. Probab. Lett., **79**, 1065-1073 (2009)

[16] Messaci, F., Nemouchi, N., Ouassou, I., Rachdi, M.: Local polynomial modelling of the conditional quantile for functional data. Stat. Methods Appl. **24** , 597–622 (2015)

[17] Ramsay, J.O., Silverman, B.W.: Applied functional data analysis. Methods and case studies. Springer Series in Statistics. New York (2002)

[18] Samanta, M.: Non-parametric estimation of conditional quantiles. Statist. Probab. Lett., **7**, 407-412 (1989)

[19] Stone, C.J.: Consistent nonparametric regression. Discussion. Ann. Statist., **5**, 595-645 (1977)
[20] Wang, K., Lin, L.: Variable selection in semiparametric quantile modeling for longitudinal data. Comm. Statist. Theory Methods **44** , 2243–2266, (2015)
[21] Zhang, J.: Analysis of variance for functional data. Monographs on Statistics and Applied Probability, 127. CRC Press, Boca Raton, FL (2014)

Chapter 21
Uniform in the smoothing parameter consistency results in functional regression

Lydia Kara-Zaïtri, Ali Laksaci, Mustapha Rachdi and Philippe Vieu

Abstract This paper focuses on uniform in bandwidth and uniform in nearest neighbors consistencies of both kernel and kNN type estimators involving functional data. We established in previous works results in this topic for a selection of nonparametric conditional operators. Our interest here is to adapt that approach for studying the generalized nonparametric regression function.

21.1 Introduction

The main tool for studying a response variable given an explanatory one is certainly the regression operator. The investigations on the asymptotic behavior of estimators of such an operator over functional data have been popularized for some decades now, mainly by:[21] for linear models, and [10] for non-parametric ones. There have been considerably progresses since then. For a selection of recent general books, we can mention those by [11, 23, 12]. In addition, there is a large list of contributions on kernel type estimators through various statistical papers, like [9] for general aspects, [18] for dependent data, [20, 1] for bandwidth choice, [7, 8] for uniform convergence and [4] for adaptive estimation. Another type of kernel regression estimator that is less studied, namely the k nearest neighbors estimator, has been adapted in the

Lydia Kara-Zaïtri (✉)
Université Djillali Liabès, LSPS, Sidi Bel Abbès, Algeria,
e-mail: kara_zaitri@hotmail.fr

Ali Laksaci
Université Djillali Liabès, LSPS, Sidi Bel Abbès, Algeria, e-mail: alilak@yahoo.fr

Mustapha Rachdi
Université Grenoble Alpes, AGIM Team, AGEIS EA 7407, Grenoble, France,
e-mail: mustapha.rachdi@univ-grenoble-alpes.fr

Philippe Vieu
Université Paul-Sabatier, IMT, Toulouse, France, e-mail: vieu@math.univ-toulouse.fr

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_21

functional framework with [3]. Whereas [2] developed its rate of convergence, [16] established asymptotic results when the response variable is also functional, and [15] studied its uniform convergence.

Recently, a new asymptotic aspect caught our attention, namely estimator uniformity on the smoothing parameter. This aspect has been rather treated in the multivariate case this last decade. We refer to [6, 17, 5] for general results in classical kernel type estimators, and [19] for kNN ones. In our functional framework, we brought in [13, 14] new almost complete [1] uniform in bandwidth/nearest-neighbors convergence results with rates to kernel type estimators. Four nonparametric operator models were processed: classical regression, conditional distribution, conditional density and conditional hazard functions. In this paper, we extend our work to the study of the generalized regression operator. The theoretical results that this sort of convergence generates are very significant, since they allow to choose data-driven bandwidths.

The paper is organized as follows. We first set up in Section 21.2 the trail frame and present our model. Then, we study in Section 21.3 the traditional kernel estimator of the generalized regression function. The asymptotic properties of its kNN type estimator are presented in Section 21.4. Finally, we devote Section 21.5 to general comments and open issues. We do mention that our assumptions unify both cases of finite and infinite dimension models. In fact, our results are established under general topological properties and smoothness condition for the regression operator. The technical difficulties due to uniform aspect considerations are overcome by means of Kolmogrov's entropy properties defined bellow.

Definition 21.1. Let (C,d) be a subset of a normed space of real functions.
The covering number $\mathscr{N}(\varepsilon,C,d)$ is the minimal number of open balls of radius ε (with respect to the measure d) needed to cover the set C. The centers of the balls need not belong to C.
The quantity $\log(\mathscr{N}(\varepsilon,C,d))$ is called the Kolmogorov's ε-entropy of the set C. It can be considered as a tool that allows measuring the complexity of sets in the sens that high entropy means that a large amount of information is needed to describe the set.

21.2 Study framework

Consider a semi-metric space (F,d) where F is not necessarily of finite dimension. Let $\{(X_i,Y_i)\}_{i=1,\ldots,n}$ be n pairs of independent and identically distributed as (X,Y)

[1] Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of real random variables. We say that (Z_n) converges almost-completely (a.co.) towards zero if for all $\varepsilon>0$, $\sum_{n=1}^{\infty}\mathbf{P}(|Z_n|>\varepsilon)<\infty$. The rate of convergence is of order u_n (with $u_n\to 0$) and we write $z_n=O_{a.co.}(u_n)$ if there exists $\varepsilon>0$ such that $\sum_{n=1}^{\infty}\mathbf{P}(|Z_n|>\varepsilon u_n)<\infty$.

which is a random vector valued in $F \times \mathbb{R}$, and take a fixed point x in F.

We are concerned on the generalized regression function defined for a known real-valued Borel function L by :

$$m(x) = E\left(L(Y)/X = x\right) , \tag{21.1}$$

Note that this model includes many important variants of regressors. Indeed, one obtains the classical regression operator when take $L(Y) = Y$. If one replaces $L(Y)$ by $\mathbf{1}_{]-\infty,y]}(Y)$ one has the conditional distribution function, whereas taking $L(Y) = \mathbf{1}_{]y,+\infty[}(Y)$ gives the conditional survival function.

In all our approaches, instead of assuming X to admit a density function, we will use small ball probability considerations to control the concentration of the data. So we would assume that :

$$\mathbb{P}\left(X \in B(x,r)\right) =: \phi_x(r) > 0 , \tag{21.2}$$

where $B(x,r) = \{x' \in F, d(x,x') \leq r\}$.
This condition on the concentration function, widely commented in [10], is a fundamental tool in the infinite dimensional setting. It ensures that data are not scattered in F, which guarantees a certain regularity for the estimator.

In order to control the bias, we need first to ensure a regular variation to this concentration function around zero. We then suppose that:

$$\text{For all } s \in (0,1), \lim_{r\to 0} \frac{\phi_x(sr)}{\phi_x(r)} =: \tau_x(s) < \infty. \tag{21.3}$$

The nature of the regression operator has also an impact on the bias, since the more it is regular and the less its estimator is biased. So we suppose that for N_x, fixed neighborhood of x, there exists $\beta > 0$ and $C_1 > 0$ such that :

$$\text{for all } x_1, x_2 \in N_x , \ |m_L(x_1) - m_L(x_2)| \leq C_1 d^{\beta}(x_1, x_2), \tag{21.4}$$

Furthermore, we assume that there exist $m \geq 2$ and $C_2 > 0$, such that:

$$\mathbb{E}[|L(Y)|^m | X] < C_2 < \infty, \text{ almost-surely.} \tag{21.5}$$

Those four assumptions are common to all studies on the asymptotic behavior of nonparametric regression estimators, and aren't very restrictive conditions.

21.3 Generalized regression kernel estimator

In this section, we consider the kernel estimator proposed by [7] defined as:

$$\widehat{m}_h(x) = \frac{\sum_{i=1}^n K\{h^{-1}d(x,X_i)\}L(Y_i)}{\sum_{i=1}^n K\{h^{-1}d(x,X_i)\}} \tag{21.6}$$

where K is an asymmetric kernel, and $h = h_n$ is a sequence of positive real numbers belonging to an interval $[a_n, b_n]$ which may decrease in length to zero when n tends to infinity.

In order to establish the uniform in bandwidth convergence, we need also some conditions on the kernel K. So assume that :

- The kernel K is a decreasing function supported within $[0,1/2]$ which has a continuous derivative, such that:

$$\text{There exist } 0 < C_3 \leq C_4 < \infty : \ 0 < C_3 \mathbf{1}_{[0,1/2]}(\cdot) \leq K(\cdot) \leq C_4 \mathbf{1}_{[0,1/2]}(\cdot) \tag{21.7}$$

and

$$K(1/2) - \int_0^{1/2} K'(s)\tau_x(s)ds > 0. \tag{21.8}$$

- The class of functions:

$$\mathscr{K} = \{\cdot \mapsto K(\gamma^{-1}d(x,\cdot)), \gamma > 0\} \text{ is a pointwise measurable class}^2 \tag{21.9}$$

such that

$$\sup_Q \int_0^1 \sqrt{1 + \log \mathscr{N}(\varepsilon \|\mathbb{F}\|_{Q,2}, \mathscr{K}, d_Q)} d\varepsilon < \infty, \tag{21.10}$$

where the supremum is taken over all probability measures Q and $Q(\mathbb{F}^2) < \infty$, where $\mathbb{F}$ is the envelope[3] function of the set $\mathscr{K}$. Here, d_Q is the $L_2(Q)$-metric, $\|\cdot\|_{Q,2}$ the $L_2(Q)$-norm, and $\mathscr{N}(\varepsilon, \mathscr{K}, d_Q)$ the covering number defined in Section21.1.

Assumptions (21.7) and (21.8) are a necessary condition for the construction of an efficient estimator. To facilitate the calculations we chose support interval $[0,1/2]$, but any other choice of the support would lead to the same results.
Assumptions (21.9) and (21.10) are typically used for uniform in bandwidth aspects. For more details refer to [22, 6, 5].

Theorem 21.1. *Suppose that assumptions (21.2)-(21.5), (21.7)-(21.10) hold, and that the sequence of numbers (a_n) satisfies :*

$$\frac{\log n}{n \min(a_n, \phi_x(a_n))} \to 0.$$

[2] A class of functions C is said to be a pointwise measurable class if there exists a countable subclass C_0 such that, for any function $g \in C$, there exists a sequence of functions $(g_m)_{m\in\mathbb{N}}$ in C_0 such that: $|g_m(z) - g(z)| = o(1)$.

[3] An envelope function G for a class of functions C is any measurable function such that: $\sup_{g\in C} |g(z)| \leq G(z)$, for all z.

Then we have :

$$\sup_{a_n\leq h\leq b_n}|\widehat{m}_h(x)-m(x)|=O\left\{b_n^{\beta}\right\}+O_{a.co.}\left\{\sqrt{\frac{\log n}{n\phi_x(a_n)}}\right\}.$$

21.4 Generalized regression kNN type estimator

The uniform in bandwidth feature provided in Theorem 21.1 allows to control both the bias and the variance terms of estimator regardless of its bandwidth shape. Indeed, such results assert that the estimator efficiency no longer depends on the smoothing parameter, which may just as well be random.
Considering this result, we can now study the asymptotic behavior of kNN-type estimator without encountering the technical difficulties that its random bandwidth usually generates.

Consider in this section the kNN-type estimator of the generalized regression function studied in [15] and defined bye :

$$\widehat{m}_k(x)=\frac{\sum_{i=1}^n K\{H_{k,x}^{-1}d(x,X_i)\}L(Y_i)}{\sum_{i=1}^n K\{H_{k,x}^{-1}d(x,X_i)\}} \tag{21.11}$$

where

- K is an asymmetric kernel,
- $k=k_n$ is a sequence of natural numbers belonging to an interval $[k_{1,n},k_{2,n}]$ which may decrease in length to zero when n tends to infinity,
- and $H_{k,x}$ is a positive real random variable depending on k and the sample $X_1,X_2,\ldots,X_n$ defined by :

$$H_{k,x}=\min\left\{h\in\mathbb{R}^+\,;\,\sum_{i=1}^n\mathbf{1}_{B(x,h)}(X_i)=k\right\}.$$

Since the variable $H_{k,x}$ takes the k^{th} highest value of the quantities $d(x,X_i)$ for $i=1,\cdots,n$, the kernel takes into account only Y_i whose correspondent X_i belong to the k nearest neighbors. So, the random feature of H allows to choose a more suitable neighborhood for x than the standard parameter.

Theorem 21.2. *Suppose that assumptions (21.2)-(21.5) hold, that the kernel K satisfies (21.7)-(21.10), and that the sequence of numbers* $(k_{1,n})$ *is such that :*

$$\frac{\log n}{\min\left\{n\phi_x^{-1}\left(\frac{k_{1,n}}{n}\right),k_{1,n}\right\}}\to 0.$$

Then we have :

$$\sup_{k_{1,n}\leq k\leq k_{2,n}} |\widehat{m}_k(x) - m(x)| = O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta}\right\} + O_{a.co.}\left\{\sqrt{\frac{\log n}{k_{1,n}}}\right\}.$$

21.5 Conclusion and prospects

Our work shows that the consistency of estimators given functional explanatory variable is uniform in bandwidth. Furthermore, it shows that the consistency of kNN estimators is uniform in the number of neighbors. Our contribution is pioneering since that, as long as we know, the asymptotic uniformity in the bandwidth has been treated so far only in the multidimensional case. All results are brought under classical assumptions that are not very restrictive, and that are substantially similar to those assumed for pointwise convergence results.

An optimal bandwidth is a one that depends on the local structure of the data and their concentration. So its selection refers automatically to a random bandwidth, hence the interest to use the kNN estimators. The impact of our uniform in nearest neighbors results in practice is very significant since it allows an easy choice of data-driven number k of neighbors which is a crucial parameter for the behavior of the estimator.

We conclude that, we may also foresee possible extensions to dimension reduction. As far as we know, the semi-parametric models are not sufficiently studied, in particular with kNN type estimators, and uniform in bandwidth ideas should be also interesting in this setting.

References

[1] Benhenni, K., Ferraty, F., Rachdi, M., Vieu, P.: Local smoothing regression with functional data. Comput. Statist. **22**, 353–369 (2007)
[2] Biau, G., Cérou, F., Guyader, A.: Rates of convergence of the functional k-nearest neighbor estimate. IEEE Trans. Inform. Theory, **56**, 2034–2040 (2010)
[3] Burba, F., Ferraty, F., Vieu, P.: k-nearest neighbor method in functional nonparametric regression. J. Nonparametr. Statist. **21**, 453–469 (2009)
[4] Chagny, G., Roche, A.: Adaptive estimation in the functional nonparametric regression model. J. Multiv. Anal. **146**, 105–118 (2016)
[5] Deheuvels, P., Ouadah, S.: Uniform-in-bandwidth functional limit laws. J. Theoret. Probab.**26**, 697–721 (2013)
[6] Einmahl, U., Mason, D.: Uniform in bandwidth consistency of kernel-type function estimators. Ann. Statist. **33**, 1380–1403 (2005)

[7] Ferraty, F., Laksaci, A., Tadj, A., Vieu, P.: Rate of uniform consistency for nonparametric estimates with functional variables. J. Statist. Plann. Inference, **140**, 335–352 (2010)
[8] Ferraty, F., Laksaci, A., Tadj, A., Vieu, P.: Kernel regression with functional response. Electronic J. of Statistics, 5, 159-171 (2011)
[9] Ferraty, F., Mas, A., Vieu, P.: Nonparametric regression on functional data: inference and practical aspects. Aust. N. Z. J. Statist. **49**, 267–286 (2007)
[10] Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis, Theory and Practice. Springer-Verlag, New York (2006)
[11] Horváth, L., Kokoszka, P.: Inference for Functional Data With Applications. Springer, New York (2012)
[12] Hsing, T., Eubank, R.: Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators. Wiley, Chichester (2015)
[13] Kara-Zaitri, L.Z., Laksaci, A., Rachdi, M., Vieu, P.: Uniform in bandwidth consistency for various kernel estimators involving functional data. J. of Nonparametric Statistics, 1–23 (2016)
[14] Kara-Zaitri, L.Z., Laksaci, A., Rachdi, M., Vieu, P.: Data-driven kNN estimation in nonparametric functional data analysis. J. of multivar. Analysis. **153**, 176–188 (2017)
[15] Kudraszow, N., Vieu, P.: Uniform consistency of kNN regressors for functional variables. Statist. Probab. Lett. **83**, 1863–1870 (2013)
[16] Lian, H.: Convergence of functional k-nearest neighbor regression estimate with functional responses. Electron. J. Statist. **5**, 31–40 (2011)
[17] Mason, D., Swanepoel, J.: A general result on the uniform in bandwidth consistency of kernel-type function estimators. Test, **20**, 72–94 (2011)
[18] Masry, E.: Nonparametric regression estimation for dependent functional data: Asymptotic normality. Stoch. Proc. and their Appl. **115**, 155–177 (2005)
[19] Ouadah, S.: Uniform-in-bandwidth nearest-neighbor density estimation. Statist. Probab. Lett. **83**(8), 1835-1843 (2013)
[20] Rachdi, M., Vieu, P.: Nonparametric regression for functional data: automatic smoothing parameter selection. J. Statist. Plann. Inference. **137**, 2784–2801 (2007)
[21] Ramsay, J.O., Silverman, B.: Functional Data Analysis. Springer, New York (2005)
[22] Van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes: With Applications to Statistics. Springer, New York (1996)
[23] Zhang, J.: Analysis of Variance for Functional Data. Chapman & Hall/CRC, London (2013)

Chapter 22
Functional data analysis of neuroimaging signals associated with cerebral activity in the brain cortex

Eardi Lila, John A. D. Aston, Laura M. Sangalli

Abstract We consider the problem of performing principal component analysis of functional data observed over two-dimensional manifolds. The method is illustrated via the analysis of neuroimaging signals associated with cerebral activity in the brain cortex.

22.1 Motivating application

We are interested in the analysis of neuroimaging signals referred to the cerebral cortex, the highly convoluted thin sheet of neural tissue that constitutes the outermost part of the brain and where most neural activity is focused. In particular we study hemodynamic signals recorded by functional magnetic resonance imaging (fMRI), that offer an indirect measure of the activity in the cortex based on the changes in deoxy-hemoglobin concentration related to energy use by brain cells. When analyzing these signals, neglecting the morphology of the cortex may lead to totally inaccurate estimates, since functionally distinct areas, that are far apart along the cortex, may in turn be close in 3D space, due to its highly convoluted nature.

Eardi Lila
Cambridge Centre For Analysis, University of Cambridge, Wilberforce Road, Cambridge Cb3 0Wb, UK
e-mail: e.lila@maths.cam.ac.uk

John A. D. Aston
Statistical Laboratory, University of Cambridge, Wilberforce Road, Cambridge Cb3 0Wb, UK
e-mail: j.aston@statslab.cam.ac.uk

Laura M. Sangalli (✉)
MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
e-mail: laura.sangalli@polimi.it

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_22

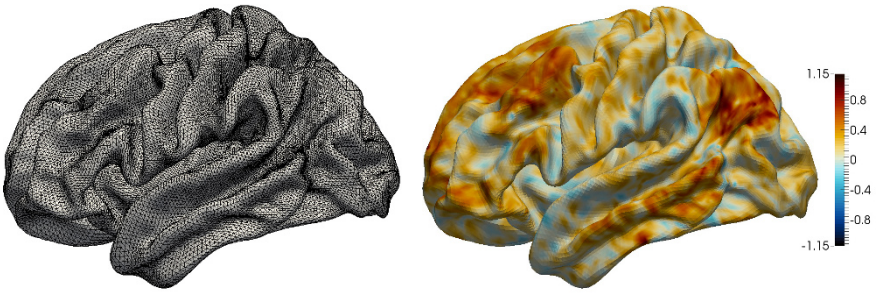


Fig. 22.1: Left: Triangulated surface approximating the left hemisphere of a cerebral cortex. Right: functional connectivity map obtained from fMRI signal. Figure adapted from [8].

Classical tools such as non-parametric smoothing models have already been adapted to deal with this kind of data, see e.g. [4, 2, 12, 1, 6, 13].

Here we consider the problem of exploring the variability across multiple neuroimaging signals, corresponding to different subjects, that is key to unveil common activation patterns and connectivity patterns. The considered dataset comes from the Human Connectome Project Consortium [5] and consists of resting state fMRI scans from about 500 healthy volunteers. The data of the various subjects are all mapped to a common template cortical surface, to allow multi-subject statistical analysis. The left panel of Figure 22.1 shows a triangulated surface that approximates the smooth two-dimensional Riemannian manifold representing the left hemisphere of the template cerebral cortex, where all subjects data are projected. We focus our analysis on functional connectivity maps that can be computed from the fMRI scans. For each subject, the functional connectivity map highlights the areas of the cortex that are functionally connected to a selected region of interest; see right panel of Figure 22.1.

22.2 Principal component analysis of functional data observed over two-dimensional manifolds

In [8] we propose a method for principal component analysis (PCA) of functional signals defined over two-dimensional manifold domains. The method is inspired by approaches based on regularized functional PCA, developed for the case of functional data observed over one-dimensional domains by, e.g. [9, 11, 7]. In particular, we consider a smoothing penalty coherent with the geodesic distance over the manifold, and involving the Laplace-Beltrami operator associated with the manifold. The method leverages on the regression techniques over two-dimensional manifolds developed in [6, 3, 13], and can be applied to data observed over any two-dimensional

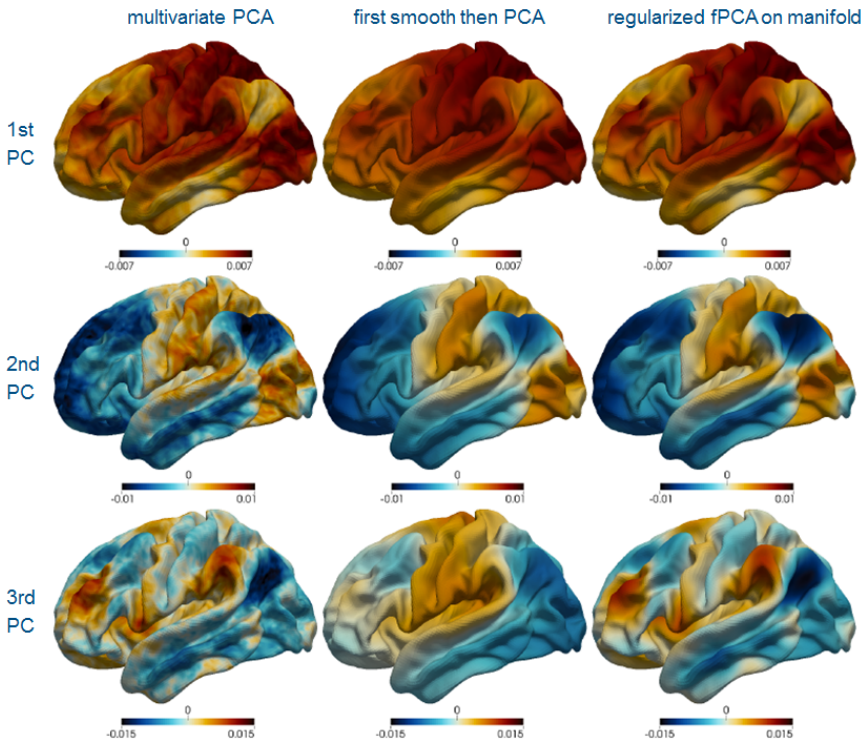


Fig. 22.2: From top to bottom, the first, second and third principal component functions estimated respectively by multivariate PCA (left), smoothing on manifolds followed by multivariate PCA (center) and the proposed regularized fPCA (right). Figure adapted from [8]

Riemannian manifold topology. The problem is made numerically tractable by the use of surface finite elements.

Extensive simulation studies [8] show that the proposed method outperform standard multivariate PCA and also provides superior estimates to those that can be obtained following a classical pre-smoothing approach, where each functional datum is smoothed previous to performing the multivariate PCA.

Figure 22.2 shows the first three estimated principal components of the functional connectivity maps computed on the dataset from the Human Connectome Project. The figure compares the estimates provided by standard multivariate PCA (in the first column), the pre-smoothing approach (center column) and the proposed regularized fPCA on manifold (last column). As highlighted by the figure, especially by the estimate of the third principal component, the proposed method combines the desired smoothness with the ability to capture the strongly localized features of the main

modes of variability, that are apparent with the multivariate PCA, but are instead lost in the pre-smoothing approach.

References

[1] Chung, M., Hanson, J., Pollak, S.: Statistical analysis on brain surfaces. In Handbook of Neuroimaging Data Analysis, Edited by Hernando Ombao, Martin Lindquist, Wesley Thompson, John Aston, Series: Handbooks of Modern Statistical Methods, Chapman & Hall/CRC, 234–255 (2016)

[2] Chung, M. K., Robbins, S. M., Dalton, K. M., Davidson, R. J., Alexander, A. L., and Evans, A. C.: Cortical thickness analysis in autism with heat kernel smoothing. NeuroImage, **25**, 1256–1265 (2005)

[3] Dassi, F., Ettinger, B., Perotto, S., Sangalli, L.M.: A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain. Appl. Numer. Math., **90**, 111–131 (2015)

[4] Duchamp, T. and Stuetzle, W.: Spline smoothing on surfaces. J. Comput. Graph. Statist., **12**(2), 354–381 (2003)

[5] Essen, D. V., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Penna, S. D., Feinberg, D., Glasser, M., Harel, N., Heath, A., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S., Prior, F., Schlaggar, B., Smith, S., Snyder, A., Xu, J., and Yacoub, E.: The human connectome project: A data acquisition perspective. NeuroImage, **62**(4), 2222–2231 (2012)

[6] Ettinger, B., Perotto, S., Sangalli, L.M.: Spatial regression models over two-dimensional manifolds. Biometrika, **103** (1), 71–88 (2016)

[7] Huang, J. Z., Shen, H., and Buja, A.: Functional principal components analysis via penalized rank one approximation. Electron. J. Stat., **2**, 678–695 (2008)

[8] Lila, E., Aston, J. A. D. and Sangalli, L. M.: Smooth Principal Component Analysis over two-dimensional manifolds with an application to Neuroimaging. Ann. Appl. Stat., **10**(4), 1854–1879 (2016)

[9] Rice, J. A. and Silverman, B.W.: Estimating the mean and covariance structure nonparametrically when the data are curves. J. R. Stat. Soc. Ser. B Stat. Methodol., **53**(1), 233–243 (1991)

[10] Sangalli, L. M. and Ramsay, J. O. and Ramsay, T. O.: Spatial spline regression models. J. R. Stat. Soc. Ser. B Stat. Methodol. **75**(4), 681–703 (2013).

[11] Silverman, B. W.: Smoothed functional principal components analysis by choice of norm. Ann. Statist., **24**(1), 1–24 (1996)

[12] Seo, S. Chung, M. K., Vorperian, H. K.: Heat Kernel Smoothing Using Laplace-Beltrami Eigenfunctions. Medical Image Computing and Computer-Assisted Intervention, **13**(3), 505–512 (2010)

[13] Wilhelm, M., Dede', L., Sangalli, L.M., Wilhelm, P.: IGS: an IsoGeometric approach for Smoothing on surfaces. Computer Methods in Applied Mechanics and Engineering, **302**, 70–89 (2016)

Chapter 23
On asymptotic properties of functional conditional mode estimation with both stationary ergodic and responses MAR

Nengxiang Ling, Yang Liu and Philippe Vieu

Abstract This contribution deals with functional conditional mode estimation given a functional explanatory variable with both stationary ergodic and responses missing at random (MAR). More precisely, we propose the estimators for functional conditional density and conditional mode respectively in this case. The main results of the work are the establishment of the asymptotic properties of such estimators.

23.1 Introduction

Over the past few years, the progress of the computing techniques, both in terms of memory and computational capacities, allows us to deal with increasing bulky data which is regarded as observations of curve, surface or image, etc. or called functional data. Functional data analysis (FDA) has received considerable interest in the statistical literature.

In this work, we pay attention to the estimation of nonparametric conditional mode via the functional conditional density. This problem has been interesting in the last few years. For example, [7] focused on kernel methods and almost sure convergence with rate was stated. This precursor work has been extended in many directions, including asymptotic properties (see [5, 6]), local linear estimation (see [16]) or semi-parametric extensions to single index setting (see [14]). For more discussions on nonparametric functional estimation context via some conditional features includ-

Nengxiang Ling (✉)
School of Mathematics, Hefei University of Technology, China, e-mail: hfut.lnx@163.com

Yang Liu
School of Mathematics, Hefei University of Technology, China, e-mail: liuyang0806@foxmail.com

Philippe Vieu
Institut de Mathématiques, Université Paul Sabatier, Toulouse, France, e-mail: philippe.vieu@math.univ-toulouse.fr

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_23

ing the estimation of conditional mean, conditional median and conditional mode, one can refer to the monograph by [9] and the references therein. The aims of our contribution in this paper is double.

Firstly, it is worth being pointed that all the results involved in the whole nonparametric literature (and not only on conditional mode) are concerned with completely observed samples. However, in many practical works including for instance sampling survey, pharmaceutical tracing or reliability, data are often uncompletely observed and part of the responses are missing at random (MAR). The literature in multivariate setting for MAR samples is rather developed (see, among other, [2]; [15]; [4] and references therein). When the explanatory variable is infinite dimensional, as far as we know the only contribution dealing with MAR sample is by [8] and concerns the simple (parametric) problem of estimating the reponse mean. The first aim of our contribution is to develop a functional methodology for dealing with MAR samples in nonparametric problems (namely in conditional mode estimation).

The second aim of our contribution is to allow for high dependence between the sample data. In the usual nonparametric literature, dependence is often modelled by some kind of mixing assumption (see [10], for large discussion on multivariate situations and [9], for first advances on functional setting). Mixing is some kind of asymptotic independence assumption which is commonly used for seak of simplicity but which can be unrealistic in situations where there is strong dependence betwen the data. Extending nonparametric functional ideas to general dependence structure is a rather underdeveloped field. Some interesting contributions involve those by [1] for long memory structure and those of [11, 12] for ergodic sequences, but these papers were only based on fully observed samples. The second novelty of our contribution is to develop our nonparametric conditional mode methodology by using the erodicity ideas developed in [11, 12] but with suitable adaptation to the MAR situation.

23.2 Model and methodology

23.2.1 Modelling with responses MAR and estimators

Let $\{(X_i,Y_i),1\leq i\leq n\}$ be a sequence of stationary ergodic functional samples with identically distribution as (X,Y), where X takes values in a some semi-metric abstract space H with a semi-metric $d(.,.)$ and Y takes values in R. In the case of response MAR, one has incomplete sample of size n from (X,Y,δ) which is classically denoted by $\{(X_i,Y_i,\delta_i),1\leq i\leq n\}$, where $\delta_i=1$ if Y_i is observed, and $\delta_i=0$ otherwise. The Bernoulli random variable δ is supposed to be such that

$$P(\delta=1|X=\chi,Y=y)=P(\delta=1|X=\chi)=p(\chi), \tag{23.1}$$

here $p(\chi)$ is a functional operator. This last condition models the fact that the censoring process δ is, conditionally on X, independent of the response Y. As

motivated for instance in [17] and [3] in the finite dimensionality case, the new functional estimator of $f(y|\chi)$ adapted to response MAR can be defined as

$$\widehat{f}_n(y|\chi) = \frac{\sum_{i=1}^{n} \delta_i K(\frac{d(\chi,X_i)}{h}) \frac{1}{g} H(\frac{1}{g}(y-Y_i))}{\sum_{i=1}^{n} \delta_i K(\frac{d(\chi,X_i)}{h})} := \frac{\widehat{r}_n(\chi,y)}{\ell_n(\chi)}, \tag{23.2}$$

where

$$\widehat{r}_n(\chi,y) = \frac{1}{nE\Delta_1(\chi)} \sum_{i=1}^{n} \delta_i \lambda_i(y) \Delta_i(\chi) \tag{23.3}$$

and

$$\ell_n(\chi) = \frac{1}{nE\Delta_1(\chi)} \sum_{i=1}^{n} \delta_i \Delta_i(\chi), \tag{23.4}$$

with $\Delta_i(\chi) = K\left(\frac{d(\chi,X_i)}{h}\right)$ and $\lambda_i(y) = \frac{1}{g} H\left(\frac{1}{g}(y-Y_i)\right)$.
Of course, as a matter of consequence, a new estimate of the conditional mode $\theta(\chi)$ can be constructed by:

$$\widehat{\theta}_n(\chi) = \arg\sup_{y \in S} \widehat{f}_n(y|\chi). \tag{23.5}$$

where S is a fixed compact subset of R.

23.2.2 Some notations and assumptions

Suppose that $\mathscr{F}_i$ and $\mathscr{G}_i$ are the σ-fields generated by $((X_1,Y_1),...,(X_i,Y_i))$ and $((X_1,Y_1),...,(X_i,Y_i),X_{i+1})$ respectively, and write $F_\chi(h) = P(d(\chi,X) \leq h) = P(X \in B(\chi,h))$ and $F_\chi^{\mathscr{F}_{i-1}}(h) = P(d(\chi,X) \leq h|\mathscr{F}_{i-1}) = P(X \in B(\chi,h)|\mathscr{F}_{i-1})$ for any fixed $\chi \in H$ and $h > 0$, where $B(\chi,h) = \{y|d(\chi,y) \leq h, y \in H\}$. Our results are shown under some assumptions that we gathered below for making the reading easier.
(A1) K is a nonnegative bounded kernel of class $\mathscr{C}^1$ over its support $[0,1]$. The derivative $K'(t) < 0$, $\left|\int_0^1 (K^j)'(u)du\right| < \infty$ for $j = 1,2$.
(A2) There exists a sequence of nonnegative random functional $(f_{i,1}(\chi))_{i\geq 1}$ a.s. bounded by a sequence of deterministic quantities $(b_i(\chi))_{i\geq 1}$ accordingly, a sequence of random functions $(g_{i,\chi})_{i\geq 1}$, a deterministic nonnegative bounded functional f_1 and a nonnegative real function ϕ tending to zero as its argument tends goes to zero, such that
(i) $F_\chi(h) = \phi(h) f_1(\chi) + o(\phi(h))$, as $h \to 0$.
(ii) For any $i \in N$, $F_\chi^{\mathscr{F}_{i-1}}(h) = \phi(h) f_{i,1}(\chi) + g_{i,\chi}(h)$ with $g_{i,\chi}(h) = o_{a.s}(\phi(h))$ as $h \to 0$, $\frac{g_{i,\chi}(h)}{\phi(h)}$ a.s. bounded and $n^{-1} \sum_{i=1}^{n} g_{i,\chi}^j(h) = o_{a.s}(\phi^j(h))$ as $n \to \infty$, $j = 1,2$.
(iii) $n^{-1} \sum_{i=1}^{n} {f^j}_{i,1}(\chi) \to {f_1}^j(\chi)$ a.s. as $n \to \infty$, j=1,2.

(iv) There exists a nondecreasing bounded function τ_0 such that, uniformly in $t \in [0,1]$,
$\frac{\phi(ht)}{\phi(h)} = \tau_0(t) + o(1)$ as $h \downarrow 0$ and $\int_0^1 (K^j)'(t)\tau_0(t)dt < \infty$ for $j \geq 1$.
(v) $n^{-1}\sum_{i=1}^n b_i(x) \to D(x) < \infty$ as $n \to \infty$.
(A3) For any $m \geq 1$, $E(|\lambda_i^{(l)}(y)|^m|\mathscr{G}_{i-1}) = E(|\lambda_i^{(l)}(y)|^m|X_i)$, a.s. for $l = 0,2$.
(A4) Let N_χ be a fixed neighborhood of $\chi \in H$, $\exists C > 0, \exists \beta_i > 0, i = 1,2$ such that

$$\left| f^{(j)}(y_1|\chi_1) - f^{(j)}(y_2|\chi_2) \right| \leq C(d(\chi_1,\chi_2)^{\beta_1} + |y_1 - y_2|^{\beta_2}),$$

for $\forall(\chi_1,\chi_2) \in N_\chi \times N_\chi$, $\forall(y_1,y_2) \in S \times S$ and $j = 0,2$.
(A5) (i) $\exists \xi > 0$ and only $y_0 \in S$ such that $f(.|\chi)$ is strictly increasing on $(y_0 - \xi, y_0)$ and strictly decreasing on $(y_0, y_0 + \xi)$.
(ii) $f(.|\chi)$ is twice continuously differentiable around $\theta(\chi)$ with $|f^{(2)}(\theta(\chi)|\chi)| \neq 0$, where $f^{(q)}(\theta(\chi)|\chi)(q = 1,2)$ is the qth derivative of $f(y|\chi)$ with respect to y.
(iii) $p(\chi)$ is continuous in a neighborhood of χ.
(A6) The kernel H is a positive bounded with twice differentiable function such that
(i) $\int H^2(t)dt < \infty$ and $\int |t|^{\beta_2} H(t)dt < \infty$.
(ii) $H^{(1)}(t)$ is bounded with $\int (H^{(1)}(t))^2 dt < \infty$.
(iii) $H^{(2)}(t)$ is Hölder continuous of order one.

23.3 Asymptotic properties

23.3.1 Preliminary propositions for conditional density estimation

Proposition 3.1 Assume that the assumptions (A1)-(A3) and(A4) hold true. If

$$\frac{\log n}{ng^2\phi(h)} \to 0 \ \ as \ \ n \to \infty, \tag{23.6}$$

and

$$\exists \zeta > 0, \quad n^\zeta g^2 \to \infty, \ as \ \ n \to \infty, \tag{23.7}$$

then, we have

$$\sup_{y \in S} \left| \widehat{f}_n(y|\chi) - f(y|\chi) \right| = O_{\text{a.s.}}(h^{\beta_1} + g^{\beta_2}) + O_{\text{a.s.}}\left(\frac{\log n}{ng^2\phi(h)} \right)^{\frac{1}{2}}. \tag{23.8}$$

Proposition 3.2 Suppose that the assumptions (A1)-(A4), (A6)(i), (3.1) and (3.2) hold true, if in addition

$$\sqrt{ng\phi(h)}(h^{\beta_1} + g^{\beta_2}) \to 0, \ \ as \ \ n \to \infty, \tag{23.9}$$

then we have

$$\sqrt{ng\phi(h)}\left(\widehat{f}_n(y|\chi)-f(y|\chi)\right)\xrightarrow{D}N(0,\sigma^2(\chi,y)),\quad as\quad n\to\infty, \tag{23.10}$$

where "$\xrightarrow{D}$" means convergence in distribution and $\sigma^2(\chi,y)=\frac{M_2}{M_1^2}\cdot\frac{f(y|\chi)}{p(\chi)f_1(\chi)}\int H^2(t)dt$ with $M_j=K^j(1)-\int_0^1\left(K^j\right)'(t)\tau_0(t)dt$ for $j=1,2$.

23.3.2 Asymptotic properties of conditional mode estimate

Theorem 3.3 Under the conditions of Proposition 3.1, if (A5) holds true, then, we have

$$\left|\widehat{\theta}_n(\chi)-\theta(\chi)\right|=O_{\text{a.s.}}\left(h^{\beta_1}+g^{\beta_2}\right)^{\frac{1}{2}}+O_{\text{a.s.}}\{\frac{\log n}{ng^2\phi(h)}\}^{\frac{1}{4}}. \tag{23.11}$$

Theorem 3.4 Under the conditions of Proposition 3.2, (A5) and (A6), if, in addition, $ng^3\phi(h)\to\infty$, then we have

$$\sqrt{ng^3\phi(h)}\left(\widehat{\theta}_n(\chi)-\theta(\chi)\right)\xrightarrow{D}N\left(0,\gamma^2(\chi,\theta(\chi))\right),\quad as\quad n\to\infty, \tag{23.12}$$

where

$$\gamma^2(\chi,\theta(\chi))=\frac{M_2}{M_1^2}\cdot\frac{f\left(\theta(\chi)|\chi\right)}{p(\chi)f_1(\chi)\left(f^{(2)}(\theta(\chi)|\chi)\right)^2}\int(H^{(1)}(t))^2dt.$$

23.4 Further research

This contribution has investigated conditional mode prediction in functional setting by adapating the nonparametric methodology in two directions: strong dependent (i.e. ergodic) and missing at random (i.e. MAR) data. Based on the experience of this work on conditional mode, our guess is that most of techniques using nonparametric functional kernel smothers could be efficiently extended to ergodic and/or MAR settings. For more details in this direction, one can see [13] and its references.

Acknowledgements Ling's work is supported by the National Social Science Fund of China (14ATJ005,16BJT023).

References

[1] Benhenni, K., Hedli-Griche, S., Rachdi, M., Vieu, P. (2008). Consistency of the regression estimator with functional data under long memory conditions. Statist. Probab. Lett., **78**: 1043-1049, (2008)

[2] Cheng, P.: Nonparametric estimation of mean functionals with data missing at random, J. Amer. Statist. Assoc., **89**: 81-87, (1994)

[3] Cheng, P. E., Chu, C. K.: Kernel estimation of distribution functions and quantiles with missing data. Statist. Sinica, **6**: 63-78, (1996)

[4] Efromovich, S.: Nonparametric Regression with Predictors Missing at Random, J. Amer. Statist. Assoc., **106**: 306-319, (2011)

[5] Ezzahrioui,M., Ould-Saïd, E.: Asymptotic normality of nonparametric estimators of the conditional mode function for functional data. J. Nonparametr. Stat., **20**: 3-18, (2008)

[6] Ezzahrioui, M., Ould-Saïd, E.: Some asymptotic results of a non-parametric conditional mode estimator for functional time-series data. Stat. Neerl., **64**: 171-201, (2010)

[7] Ferraty, F., Laksaci, A., Vieu, P.: Functional times series prediction via conditional mode. Comptes Rendus Mathematique, **340**: 389-392, (2005)

[8] Ferraty, F., Sued, M, Vieu, P.: Mean estimation with data missing at random for functional covariables, Statistics, **47**: 688-706, (2013)

[9] Ferraty, F, Vieu, P.: Nonparametric functional data analysis, Theory and Practice, Springer Series in Statistics. Springer, New York, (2006)

[10] Györfi, L., Härdle, W., Sarda, P, Vieu, P.: Nonparametric curve estimation from time series. Lecture Notes in Statistics, 60. Springer-Verlag, Berlin, (1989)

[11] Laib, N, Louani, D.: Nonparametric kernel regression estimation for functional stationary ergodic data: asymptotic properties. J. Multivariate Anal., **101**: 2266-2281, (2010)

[12] Laib, N, Louani, D.: Rates of strong consistencies of the regression function estimator for functional stationary ergodic data, J. Statist. Plann. Inference, **141**: 359-372, (2011)

[13] Ling, N.X, Liu, Y., Vieu, P.: Conditional mode estimation for functional stationary ergodic data with missing at random, Statistics, **50** : 991-1013, (2016)

[14] Ling, N. X, Xu, Q.: Asymptotic normality of conditional density estimation in the single index model for functional time series data, Statist. Probab. Lett., **82**: 2235-2243, (2012)

[15] Little, R, Rubin, D.: Statistical Analysis with Missing Data, Second Edition, Wiley, New York, (2002)

[16] Rachdi M, Laksaci A, Demongeot J, Abdali A, Madani F.: Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data. Comput. Statist. Data Anal., **73**: 53-68, (2014)

[17] Wang, Q. H.: Probability density estimation with data missing at random when covariables are present, J. Statist. Plann. Inference, **138**: 568-587, (2008)

Chapter 24
Predicting the physiological limits of sport stress tests with functional data

Marcos Matabuena, Mario Francisco-Fernández and Ricardo Cao

Abstract This work aims at illustrating the enormous potential of continuous monitoring of the athlete, jointly with the application of statistical techniques for functional data in the analysis and control of sports performance. It is shown that using low intensity exercise, one can predict the performance of a group of athletes without forcing them to fatigue. This is the first indirect methodology proposed in the scientific literature that allows to estimate in a precise way physical fitness without producing fatigue. The areas of application of this procedure are not only limited to sport science. They are diverse and include, among others, medicine and education.

24.1 Introduction

The increase in the accuracy of the current electronic devices allows the collection of data in continuous time. The functional nature of these observations suggests avoiding the use of classical multivariate techniques and requires assuming that the data are realizations of stochastic processes in continuous time. Functional data analysis (FDA) [26, 29] studies this type of observations.

In sport science, new electronic devices capable of recording significant amounts of athletes data appear every day. Some examples are accelerometers, heart rate meters, and respiratory and blood variable meters.

Marcos Matabuena (✉)
Universidade de Vigo, Spain, e-mail: marcos.matabuena.tic@gmail.com

Mario Francisco-Fernández
Research group MODES, INIBIC, CITIC, Departamento de Matemáticas, Facultade de Informática, Universidade da Coruña, 15071 A Coruña, Spain, e-mail: mariofr@udc.es

Ricardo Cao
Research group MODES, INIBIC, CITIC, Departamento de Matemáticas, Facultade de Informática, Universidade da Coruña, 15071 A Coruña, Spain, e-mail: rcao@udc.es

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_24

Nowadays, one of the main goals of the sport training science is to use all this available information to be able to control the subject's sport training correctly or even, being more ambitious, to design algorithms to recommend appropriate training routines. For this, the goals of the subject, his current condition and individual features would be taken into account. A thorough review of the current state of the art of the application of machine learning, statistics and artificial intelligence techniques in sports can be found in [13] and [17]. Additionally, some proposals from a statistical and mathematical point of view to predict the effects of training sessions on physical fitness changes are presented in [2], or more recently in [20].

However, having a large amount of data is not a guarantee of success regardless of the statistical techniques used. It is especially important to develop methodologies to organize and use the information correctly. In sports, for example, it is necessary to use new metrics to measure the athlete's activity, to study with caution the influential variables in the physical activity and to establish mechanisms that predict the physical fitness of the athlete. Some studies and proposals on the topics above can be found in [1, 3, 4, 8, 14, 19, 23], but they are still insufficient.

In sport science, the main methodological difficulty is to predict the physical fitness using indirect mechanisms that do not cause fatigue to the athlete. This is perhaps the most important open issue in sport training science. Solving this problem, coaches and athletes could know the individual effect of training sessions on fitness, something only possible to achieve through experience. They can also have a valuable tool to control the accumulated fatigue, thus avoiding possible conditions of overtraining and estimating the appropriate intensities of the training or competition. This leads to optimize the performance achieved by the athlete in each moment. If this problem is not solved, we will never know the actual effects of the training loads on the subject, independently of the available amount of data. Therefore, without solving the previous issue, any analysis and applied decision-making procedure will not select the best combination of possible trainings.

The problem of predicting physical fitness by indirect mechanisms also has special interests outside the framework of sports. For example, it is very important for patients undergoing rehabilitation treatments after illness [6, 28]. It is also useful in schools to evaluate students' physical fitness consistently, in the general population, in the control of healthy habits [5, 24], or in public examinations where the tests used are usually exhausting and often unassuming due to the large amount of events that the participants have to overcome.

In the field of sports, only two applications of functional data are highlighted: a descriptive analysis of the lactic acid curves and an analysis of knee coordination in vertical jumps. For more information, see [16], [22].

In this work, a new practical application of FDA using data associated with sport stress tests is presented. A procedure to predict the maximal oxygen consumption is designed. This is one of the best indicators of sport performance in endurance sports [30]. To do this, a low intensity and progressive test on a treadmill, where some functional variables are observed, is carried out.

24.2 Material and methods

The population of the study is made up of 341 subjects with heterogeneous characteristics, coming from different sports: athletics, badminton, handball, basketball, cycling, football, judo, wrestling, canoeing, rowing, squash, taekwondo, tennis, triathlon and sailing.

The test procedure is as follows: It starts at a speed of 6 km/h and the speed is increased 0.25 km/h every 15 seconds during 6 minutes. In those 6 minutes of effort, some multivariate functional variables are measured every 5 seconds:

- Speed
- $EqCO_2$
- EqO_2
- O_2/HR
- HR
- VE in L/min
- VO_2 in ml/min
- VCO_2 in ml/min
- RER
- VO_2 per kg of body weight

A functional regression model was used to predict the maximal oxygen consumption. For this purpose, the available functional observations of the first 6 minutes of the sample of the individuals previously described were used. These same individuals continued the effort to exhaustion with the same speed increases as those established up to the minute 6, and their VO_2 max was measured.

The average duration of the stress test was around 13 minutes. The maximal intensity obtained in the first 6 minutes compared to the maximal speed reached by athletes in the test is on average the 61%, which represents an insignificant effort. This percentage approximately corresponds to the intensity that the athletes can withstand if they were prepared to compete for a walking\running race of 100 km.

24.3 Statistical procedure

The statistical software *R*, namely the fda.usc library [10], was used to perform the statistical data analysis.

First, the functional data have been smoothed using nonparametric methods [12]. Then, dimension reduction techniques, such as functional principal components [27], have been applied. Finally, several functional regression models have been employed: multiple linear regression model [7, 27], PLS regression [25], functional generalized kernel additive models [9], functional regression with scalar response using nonparametric kernel estimation [12, 11], and functional generalized spectral additive models [21]. The functional generalized spectral additive model was the one finally used since the best results were obtained.

This model is briefly presented now. Let us assume that $\mathbf{X} = (X_1, \ldots, X_p)$ is a p-dimensional functional random variable in $L^2[a,b]$ and consider a response variable Y. The relationship between the variables $\mathbf{X}$ and Y in a generalized spectral additive model (*GSAM*) is

$$E(Y|X) = \beta_0 + \sum_{k=1}^{p} f_k(X_k) \tag{24.1}$$

where every f_k is a smooth function.

On the other hand, using Karhunen-Loeve decomposition, every functional variable X_k (with $k = 1, \ldots, p$) can be expressed as follows:

$$X_k(t) = \mu(t) + \sum_{j=1}^{\infty} x_{kj} v_{kj}(t),$$

where $v_{kj}(t)$ is the j-th eigenfunction for the k-th functional variable and x_{kj} is a score term. Consequently, an approximation of equation (24.1) can be easily found (see [21] for details):

$$E(Y|X) \approx \beta_0 + \sum_{k=1}^{p} \sum_{m=1}^{r_k} f_k(x_{km}),$$

where x_{mk} is the m-th principal component score of the k-th functional covariate and r_k is the number of principal components considered for the k-th functional covariate.

The estimation of the smooth functions is carried out using the technique known as principal component analysis with conditional expectation (PACE). This method selects, in an automatic way, the number of eigenfunctions to be chosen for every functional covariate by means of the criterion AIC [21].

For our dataset, the response variable in the regression model is the oxygen consumption (a scalar variable), while the explanatory variables have been the scalar variables, maximal heart rate, weight, height, and two functional variables, oxygen consumption (Figure 24.1) and heart rate (Figure 24.2). The variables have been selected after performing several fits and analyzing the error obtained in each of them.

24.4 Results

The results of the best regression model yield $R^2 = 0.846$ and the average absolute error is, in this case, 2.398. In Table 24.1 the adjusted R^2 of all models are shown. Additionally, Figure 24.3 presents a scatter plot of observed vs predicted values using the functional generalized spectral additive model. The point cloud is very close to the diagonal which shows that fit is very good. These results are satisfactory: in comparison with other tests proposed in the literature to predict VO_2, the error made is very low. The other tests have errors larger than the one shown here. In addition, they force the athlete to make a maximal effort, while here, only 6 minutes of the

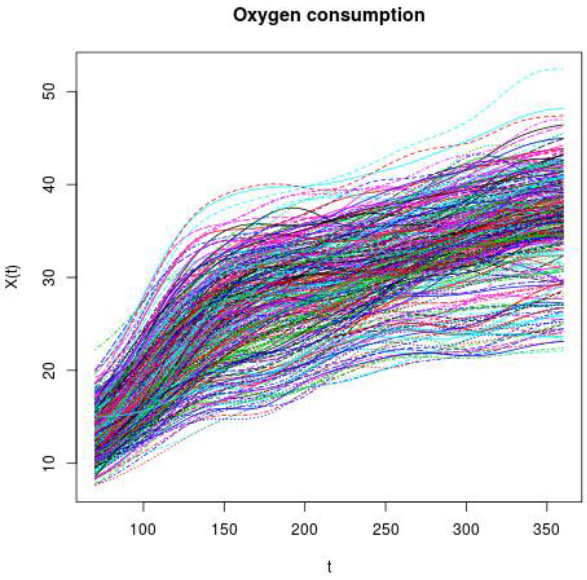


Fig. 24.1: Oxygen consumption in the first 6 minutes of the effort.

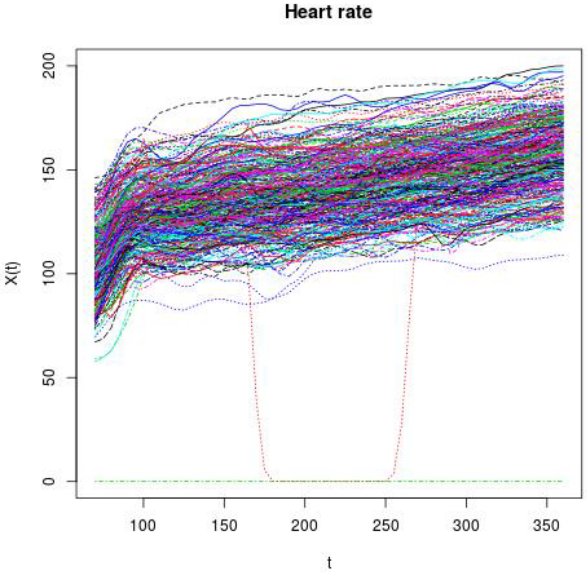


Fig. 24.2: Heart rate in the first 6 minutes of exercise.

stress test are being used. This activity does not cause suffering and fatigue to the athlete.

The test has been validated with a significant amount of individuals (341) of different level, weight, size and age. This is very important because other tests are validated with a set of students with similar characteristics and, therefore, can only be applied to a restrictive set of individuals. Because of this, good fits are often obtained: many authors select a sample with hardly any variability and, therefore, any approximation will be good. This is observed, for instance, in the comparative study shown in [15].

Our results show that with statistical techniques based on functional data and using the athlete's monitoring with continuous data, we can characterize the athlete's performance with indirect methods in a very reliable way. This is a real innovation in sport training.

Table 24.1: Adjusted R^2 for the regression models compared

Model	R^2 (adjusted)
Functional generalized kernel additive model	0.708
Functional generalized spectral additive model	0.846
Functional nonparametric kernel regression model with scalar response	0.603
Multiple linear regression model	0.726
PLS regression model	0.463

24.5 Conclusions

In this work, we have proven that with the joint use of techniques based on functional data and continuous monitoring, we get a good fit in the proposed test.

The intensities of the test have been too low. However, the performance of our proposal has been better than those of all tests proposed in the literature to predict the maximal consumption of oxygen, in particular better than the Leger test [18, 19], the most widely used and highly cited in the literature.

The use of techniques based on functional data represents a very hopeful way to obtain indirect tests that measure the fitness of the subject without generating fatigue and, therefore, can be used every day. In this case, the correlation between the oxygen consumption reached in the first 6 minutes of effort and oxygen consumption reached is 0.54. With a functional approach, the predictive capacity of the model increases significantly, being, for example, the heart rate an unimportant variable if we fit a regression model with scalar covariables, but not in the functional case. This indicates the need of using functional models with this type of data if we want to achieve precision in the final result.

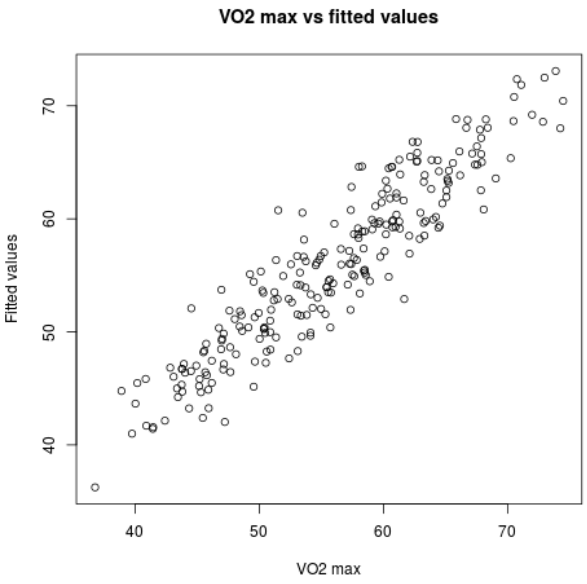


Fig. 24.3: Real values vs predicted values using the functional generalized spectral additive model

In future studies, the optimal measurement intensities will be studied to obtain the best results in the prediction of oxygen consumption, searching for the benefit between the improved prediction (the higher intensity and duration of the test, the better fit) and the athlete's fatigue. In this work, we have been too conservative in this regard, but we are still able to beat other existing tests to predict maximal oxygen consumption, one of the best predictors of sports performance [30]. This makes us to be optimistic about the potential of this research line.

References

[1] Achten, J., Jeukendrup, A. E.: Heart rate monitoring. Sports Medicine, **33**(7), 517–538 (2003)

[2] Banister, E. W., Calvert, T. W., Savage, M. V., Bach, T.: A systems model of training for athletic performance. Australian Journal of Sports Medicine, **7**(3), 57–61 (1975)

[3] Bassett, D. R., Howley, E. T.: Limiting factors for maximum oxygen uptake and determinants of endurance performance. Medicine and Science in Sports and Exercise, **32**(1), 70–84 (2000)

[4] Billat, V. L., Flechet, B., Petit, B., Muriaux, G., Koralsztein, J. P.: Interval training at VO2max: effects on aerobic performance and overtraining markers. Medicine and Science in Sports and Exercise, **31**(1), 156–163 (1999)
[5] Blair, S. N.: Physical inactivity: the biggest public health problem of the 21st century. British Journal of Sports Medicine, **43**(1), 1–2 (2009)
[6] Cahalin, L. P., Mathier, M. A., Semigran, M. J., Dec, G. W., DiSalvo, T. G.: The six-minute walk test predicts peak oxygen uptake and survival in patients with advanced heart failure. CHEST Journal, **110**(2), 325–332 (1996)
[7] Cardot, H., Ferraty, F., Sarda, P.: Functional linear model. Statist. Probab. Lett., **45**(1), 11–22 (1999)
[8] Conconi, F., Grazzi, G., Casoni, I., Guglielmini, C., Borsetto, C., Ballarin, E., Mazzoni, G., Patracchini, M., anfredini, F.: The Conconi test: Methology after 12 years of application. International Journal of Sports Medicine, **17**, 509-519 (1996)
[9] Febrero-Bande, M., González-Manteiga, W.: Generalized additive models for functional data. TEST, **22**(2), 278–292 (2013)
[10] Febrero-Bande, M., Oviedo de la Fuente, M.: Statistical computing in functional data analysis: the R package fda. usc. Journal of Statistical Software, **51**(4), 1–28 (2012)
[11] Ferraty, F., Vieu, P.: Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. J. Nonparametr. Stat., **16**(1-2), 111–125 (2004)
[12] Ferraty, F., Vieu, P. (2006). Nonparametric functional data analysis: theory and practice. Springer Science & Business Media.
[13] Fister, I., Ljubi, K., Suganthan, P. N., Perc, M.: Computational intelligence in sports: challenges and opportunities within a new research domain. Applied Mathematics and Computation, **262**, 178–186 (2015)
[14] Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Dodge, C.: A new approach to monitoring exercise training. The Journal of Strength Conditioning Research, **15**(1), 109-115 (2001)
[15] García, G. C., Secchi, J. D.: Test course navette de 20 metros con etapas de un minuto. Una idea original que perdura hace 30 años. Apunts. Medicina de l'Esport, **49**(183), 93–103 (2014)
[16] Harrison, A. J., Ryan, W., Hayes, K.: Functional data analysis of joint coordination in the development of vertical jump performance. Sports Biomechanics, **6**(2), 199–214 (2007)
[17] Kampakis, S.: Predictive modelling of football injuries. arXiv preprint arXiv:1609.07480 (2016)
[18] Leger, L. A., Lambert, J.: A maximal multistage 20-m shuttle run test to predict dot VO2 max. European Journal of Applied Physiology and Occupational Physiology, **49**(1), 1–12 (1982)
[19] Leger, L. A., Mercier, D., Gadoury, C., Lambert, J.: The multistage 20 metre shuttle run test for aerobic fitness. Journal of Sports Sciences, **6**(2), 93–101 (1988)

[20] Matabuena, M., Rodríguez, R.: A new approach to predict changes in physical condition: A new extension of the classical Banister model. ArXiv e-prints arXiv:1612.08591 (2016)
[21] Müller, H. G., Yao, F.: Functional additive models. J. Amer. Statist. Assoc., **103**(484), 1534–1544 (2012)
[22] Newell, J., McMillan, K., Grant, S., McCabe, G.: Using functional data analysis to summarise and interpret lactate curves. Computers in Biology and Medicine, **36**(3), 262–275 (2006)
[23] Noakes, T. D., Myburgh, K. H., Schall, R.: Peak treadmill running velocity during the V O2 max test predicts running performance. Journal of Sports Sciences, **8**(1), 35–45 (1990).
[24] Penedo, F. J., Dahn, J. R.: Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Current Opinion in Psychiatry, **18**(2), 189–193 (2005)
[25] Preda, C., Saporta, G.: Clusterwise PLS regression on a stochastic process. Comput. Statist. Data Anal., **49**(1), 99–108 (2005)
[26] Ramsay, J. O.: When the data are functions. Psychometrika, **47**(4), 379–396 (1982)
[27] Ramsay, J. O. Silverman, B.W.: Functional data analysis. John Wiley & Sons, Inc. (2006)
[28] Solway, S., Brooks, D., Lacasse, Y., Thomas, S.: A qualitative systematic overview of the measurement properties of functional walk tests used in the cardiorespiratory domain. CHEST Journal, **119**(1), 256–270 (2001)
[29] Wang, J. L., Chiou, J. M., Müller, H. G.: Review of functional data analysis. arXiv preprint arXiv:1507.05135 (2015)
[30] Weltman, A., Snead, D., Seip, R., Schurrer, R., Weltman, J., Rutt, R., Rogol, A.: Percentages of maximal heart rate, heart rate reserve and VO2max for determining endurance training intensity in male runners. International Journal of Sports Medicine, **11**(03), 218–222 (1990)

Chapter 25
An overview of consistency results for depth functionals

Stanislav Nagy

Abstract Data depth is a nonparametric tool which may serve as an extension of quantiles to general data. Any viable depth must posses the uniform strong consistency property of its sample version. In this overview, a concise summary of the available uniform consistency results for most of the depths for functional data is given. Extensions of this theory towards random surfaces, imperfectly observed, and discontinuous functional data are studied.

25.1 Functional Data Depth

In nonparametric analysis of functional data, the concept of data depth gained considerable attention in the past years. Assume that $\mathscr{M}$ is a functional (or multivariate) measurable space, and denote by $\mathscr{P}(\mathscr{M})$ the space of all probability measures on $\mathscr{M}$. Data depth is a mapping D which to any $x \in \mathscr{M}$ and $P \in \mathscr{P}(\mathscr{M})$, assigns a number $D(x;P) \in [0,1]$ indicating how much "centrally located" x is with respect to (w.r.t.) P. Functions with high depth values form loci of points deeply inside the mass of P, and may serve as "centre-outwards" analogues of quantile regions for functional data. For instance, the function at which the depth $D(\cdot;P)$ is maximized generalizes the median to $\mathscr{M}$-valued data. On the other hand, functions whose depth is close to zero may be considered as lying on the outskirts of the distribution, or being potentially outlying w.r.t. P.

In recent literature, many depths for random curves have been proposed. Most of them fall into three large groups of functionals:

- *integrated depths* ([5, 3]), taking the form of an integral

Stanislav Nagy (✉)
Department of Mathematics, KU Leuven, Belgium
Department of Probability and Mathematical Statistics, Charles University, Czech Rep., e-mail: stanislav.nagy@kuleuven.be

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_25

$$FD(x;P) = \int_{\Phi} D(\phi(x);\phi(P))\,\mathrm{d}\,\mu(\phi) \quad \text{for } x \in \mathscr{M}, P \in \mathscr{P}(\mathscr{M}), \tag{25.1}$$

where Φ is a subset of the d-th Cartesian product of the dual of the sample space $\mathscr{M}$, D is a depth suitable for d-dimensional data, and μ is a measure on Φ,

- *infimal depths* ([10]), where the integral in (25.1) is replaced by an infimum

$$ID(x;P) = \inf_{\phi \in \Phi} D(\phi(x);\phi(P)) \quad \text{for } x \in \mathscr{M}, P \in \mathscr{P}(\mathscr{M}), \tag{25.2}$$

- *band depths* ([8]), extending the so-called simplicial depth to functional data in $\mathscr{M} = \mathscr{C}$, the space of continuous functions $x\colon [0,1] \to \mathbf{R}$

$$BD(x;P) = \mathsf{P}(x \in B(X_1, X_2)) \quad \text{for } x \in \mathscr{C}, P \in \mathscr{P}(\mathscr{C}). \tag{25.3}$$

Here, $(\Omega, \mathscr{F}, \mathsf{P})$ is the probability space on which all random variables are defined, $X_1, X_2 \sim P$ are independent, and $B(x_1, x_2)$ is the band of $x_1, x_2 \in \mathscr{C}$ defined as

$$B(x_1,x_2) = \{y \in \mathscr{C}\colon \min\{x_1(t), x_2(t)\} \le y(t) \le \max\{x_1(t), x_2(t)\} \text{ for all } t\}.$$

For FD and ID, various choices of the depth D in (25.1) and (25.2) have been proposed in the literature. Here, we focus only on the important *halfspace depth* (see [4])

$$D(u;Q) = \inf_{H \in \mathscr{H}(u)} Q(H), \quad \text{for } u \in \mathbf{R}^d, Q \in \mathscr{P}\left(\mathbf{R}^d\right), \tag{25.4}$$

where $\mathscr{H}(u)$ is the collection of all halfspaces in $\mathbf{R}^d$ that contain u. Analogous study for other depths D in $\mathbf{R}^d$ is straightforward.

The first integrated depth was proposed in [5] for $\mathscr{M} = \mathscr{C}$. There, $d = 1$ and Φ is the set of Dirac functionals $\phi_t(x) = x(t)$ for $t \in [0,1]$. The measure μ can then be taken to be the Lebesgue measure on $[0,1]$, and FD takes the basic form

$$FD(x;P) = \int_0^1 D(x(t);P_t)\,\mathrm{d}\,t, \quad \text{for } x \in \mathscr{C}, P \in \mathscr{P}(\mathscr{C}), \tag{25.5}$$

where by $P_t \in \mathscr{P}(\mathbf{R})$ we mean the marginal distribution of P at $t \in [0,1]$. For a different choice of D in (25.5) one gets, for instance, the modified band depth in [8].

Likewise, the basic representative of infimal depths is the depth proposed in [10]

$$ID(x;P) = \inf_{t \in [0,1]} D(x(t);P_t), \quad \text{for } x \in \mathscr{C}, P \in \mathscr{P}(\mathscr{C}). \tag{25.6}$$

In a natural extension of the functional depths from $\mathscr{C}$, consider now the case of multivariate functional data and continuous random surfaces. These are defined as random variables taking values in $\mathscr{C}^K$, the space of continuous mappings $x\colon U \to \mathbf{R}^K$ for $U \subset \mathbf{R}^L$ compact, where $K, L \ge 1$. The integrated depth (25.5) then translates into

$$FD(x;P) = \int_U D(x(t);P_t)\,\mathrm{d}\,t, \quad \text{for } x \in \mathscr{C}^K, P \in \mathscr{P}\left(\mathscr{C}^K\right), \tag{25.7}$$

for D the depth (25.4) in $\mathbf{R}^K$, $x(t) \in \mathbf{R}^K$ and $P_t \in \mathscr{P}\left(\mathbf{R}^K\right)$, see also [2]. The infimal depth (25.6) is also easily generalized to $\mathscr{C}^K$-valued random variables

$$ID(x;P) = \inf_{t \in U} D(x(t);P_t), \quad \text{for } x \in \mathscr{C}^K, P \in \mathscr{P}\left(\mathscr{C}^K\right),$$

for D as above. Finally, the band depth (25.3) may be extended to the multivariate band depth ([9]) by considering the multivariate version of bands

$$B(x_1,\ldots,x_{K+1}) = \left\{y \in \mathscr{C}^K \colon y(t) \in co\,(x_1(t),\ldots,x_{K+1}(t)) \text{ for all } t \in U\right\}.$$

Here, by *co* we mean the closed convex hull mapping. This way, the band depth naturally becomes

$$BD(x;P) = \mathsf{P}\left(x \in B(X_1,\ldots,X_{K+1})\right), \quad \text{for } x \in \mathscr{C}^K, P \in \mathscr{P}\left(\mathscr{C}^K\right),$$

where $X_1,\ldots,X_{K+1} \sim P$ are independent.

Let $P_n \in \mathscr{P}(\mathscr{M})$ denote the empirical measure of a random sample $X_1,\ldots,X_n$ from $P \in \mathscr{P}(\mathscr{M})$, and let $M \subset \mathscr{M}$. We say that the functional depth D is *uniformly consistent* for P over M, if

$$\sup_{x \in M} |D(x;P_n) - D(x;P)| \xrightarrow[n\to\infty]{\text{a.s.}} 0. \tag{25.8}$$

If (25.8) is true for any $P \in \mathscr{P}(\mathscr{M})$, we say that D is *universally consistent* over M. Any reasonable depth D must be at least uniformly consistent over compact sets in $\mathscr{M}$. Otherwise, the depth level sets, being the quantile-like regions of interest, may not be consistently estimated using random samples from P.

Herein, we study the consistency properties of the elementary depths (25.3) in Section 25.2, (25.6) in Section 25.3, and (25.5) in Section 25.4. In Section 25.5 we focus on data depth in connection with imperfectly observed functional data, and in the final Section 25.6 we provide a result on the consistency of integrated depths for discontinuous random functions. It will be seen that most of these results can be extended also to general depths for functional data of integrated, and infimal types.

25.2 Band Depth

Quite surprisingly, it can be shown that the renown band depth (25.3) with sample version

$$BD(x;P_n) = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \mathsf{P}\left(x \in B(X_{i_1},X_{i_2})\right) \quad \text{for } x \in \mathscr{C},$$

is not uniformly consistent over compact subsets of $\mathscr{C}$ w.r.t. rather simple infinite-dimensional distributions (see [7, 6], and [1]).

In [6], one possible remedy for these issues has been proposed in the form of the so-called *adjusted band depth*. This depth acts as a "smoothed version" of *BD*, and for $w\colon [0,\infty) \to [0,1]$ non-increasing such that $w(0) = 1$, $\lim_{t\to\infty} w(t) = 0$, it is defined as

$$aBD(x;P) = \mathrm{E}\, w\left(\inf_{y\in B(X_1,X_2)} \|x-y\|\right) \quad \text{for } x \in \mathscr{C}, P \in \mathscr{P}(\mathscr{C}),$$

where $B(x_1,x_2)$ and X_1, X_2 are as for *BD*, and $\|\cdot\|$ is the (uniform) norm on $\mathscr{C}$. Instead of computing the probability that x is inside the band $B(X_1,X_2)$, the adjusted band depth evaluates the expected distance of x from the band. For w the indicator of zero, $aBD \equiv BD$. For w quickly decreasing to zero, *aBD* behaves very similarly to *BD*, and in practice gives virtually the same results. An advantage of using *aBD* over *BD* is its universal consistency.

Theorem 25.1 ([6, 12]). *For w continuous aBD is universally consistent over* $\mathscr{C}$.

Continuity of w is crucial in the proof of Theorem 25.1. On the other hand, for w continuous it is possible to generalize this result considerably also to $\mathscr{C}^K$-valued random variables. The depth in this case takes the form

$$aBD(x;P) = \mathrm{E}\, w\left(\inf_{y\in B(X_1,\dots,X_{K+1})} \|x-y\|\right) \quad \text{for } x \in \mathscr{C}^K, P \in \mathscr{P}\left(\mathscr{C}^K\right), \qquad (25.9)$$

for $\|\cdot\|$ the (uniform) norm on $\mathscr{C}^K$, and the assertion of Theorem 25.1 holds true also in $\mathscr{C}^K$. Furthermore, it is not necessary to restrict to the case of continuous functions, or the uniform norm in (25.9). As demonstrated in [12], the adjusted band depth (for w continuous) can be shown to be universally consistent also when the random functions take values in $\mathscr{L}_2^K$, the space of square-integrable functions $x\colon U \to \mathbf{R}^K$.

25.3 Infimal Depth

For the infimal depths, theoretical results are rather scarce in the literature. As far as we know, the only uniform consistency result is available for the depth (25.6).

Theorem 25.2 ([6]). *Let* $P \in \mathscr{P}(\mathscr{C})$ *be a mixture of* $P^{(1)}$ *and* $P^{(2)}$ *such that*

- *all the marginal distributions* $P_t^{(1)}$ *of* $P^{(1)}$ *have continuous distribution functions, and*
- $P^{(2)}$ *is concentrated in a finite-dimensional subspace of* $\mathscr{C}$.

Then, the depth ID in (25.6) *is uniformly consistent for P over* $\mathscr{C}$.

Unfortunately, *ID* is not universally consistent. This can be seen on the example of P being the standard Wiener measure in $\mathscr{C}$. Note that P violates the conditions imposed on $P^{(1)}$ in Theorem 25.2 only for $t = 0$. Though, it is easy to see that already

for $x \equiv 0$, $ID(x;P) = 1/2$, but $ID(x;P_n) = 0$ almost surely for any n, and the depth cannot be consistent for P.

Theorem 25.2 is based on a functional extension of the bracketing Glivenko-Cantelli theorem for the empirical distribution function processes of the collection $\left\{P_t^{(1)} : t \in [0,1]\right\}$. Its generalization to $\mathscr{C}^K$-valued functional data appears to be difficult, and out of reach using the current methodology. Reasonable extensions of ID to spaces of integrable functions are obviously not possible, due to the fact that the functional values $x(t)$ are not well defined for $x \in \mathscr{L}_2^K$ (recall that functions in $\mathscr{L}_2^K$ are only equivalence classes of almost everywhere equal functions).

25.4 Integrated Depth

For integrated depths such as (25.5), most of the earlier consistency results rely on the use of uniform central limit theorems in functional setups. This approach, however, works only for compact subsets of $\mathscr{C}$, and convoluted additional assumptions need to be imposed on P in order to prove consistency.

In [14], a different method is used to resolve this problem. Following [3], simple measure-theoretic tools are used in [14] to obtain the weak universal consistency

$$\begin{aligned}
\mathrm{E} \sup_{x \in \mathscr{C}} |FD(x;P_n) - FD(x;P)| &= \mathrm{E} \sup_{x \in \mathscr{C}} \left| \int_0^1 D(x(t);P_{n,t}) - D(x(t);P_t)\,\mathrm{d}t \right| \\
&\le \mathrm{E} \int_0^1 \sup_{x \in \mathscr{C}} |D(x(t);P_{n,t}) - D(x(t);P_t)|\,\mathrm{d}t \\
&\le \int_0^1 \mathrm{E} \sup_{u \in \mathbf{R}} |D(u;P_{n,t}) - D(u;P_t)|\,\mathrm{d}t \\
&\xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

The first inequality is of Jensen's type[1]; the second involves Fubini's theorem. The universal consistency of D from (25.4) in $\mathbf{R}$ ([4]) along with the dominated convergence theorem provide the final assertion. For the derivation above it is, however, vital to verify the measurability of the integrand function

$$I\colon\ \Omega \times [0,1] \to [0,1]\colon\ (\omega,t) \mapsto \sup_{x \in \mathscr{C}} |D(x(t);P_{n,t}(\omega)) - D(x(t);P_t)|\,.$$

Note that the measurability of I, being a supremum of an *uncountable* set of discontinuous functions, is by no means easy to show. Furthermore, the technique used above provides only weak universal consistency of FD. Seemingly, the omission of E from all the inequalities above would be enough to conclude that FD is (strongly) universally consistent, using the same argument. This is, unfortunately, not true, because of the last step when using the dominated convergence theorem. Indeed,

[1] Of course, $P_{n,t} \in \mathscr{P}(\mathbf{R})$ stands for the marginal distribution of P_n at $t \in [0,1]$.

here it is *not true* that the almost sure convergence of all the marginals $I(\cdot,t)$ of the random process I implies the almost sure convergence of I *at all* points $t \in [0,1]$. For the special case of the integrated depths, it is possible to overcome all these difficulties, and the following result can be stated.

Theorem 25.3 ([14]). *The depth FD in* (25.5) *is universally consistent over* $\mathscr{C}$.

This result does not only resolve the question of consistency of integrated depths, it is also extremely easy to generalize. As the most straightforward extension, it is easy to show that also the integrated depth (25.7) for functions from $\mathscr{C}^K$ is universally consistent over $\mathscr{C}^K$. Moreover, integrals with respect to measures other than the Lebesgue measure on U can be considered, and also for these depths the results apply ([14]). Finally, integrated depths for $\mathscr{L}_2^K$-valued functional data, or even for general Borel measurable random functions, can be considered, see Section 25.6.

25.5 Depth for Imperfectly Observed Random Functions

Consider a sequence $X_1, X_2, \ldots$ of independent random functions from $P \in \mathscr{P}(\mathscr{C})$. In practice, the functional values of any single datum X_i are never observed at each $t \in [0,1]$, simply because this set is infinite. Instead, for each X_i the researcher is usually provided only with

$$(X_i(T_{i,1}), \ldots, X_i(T_{i,m_i})) \in \mathbf{R}^{m_i} \quad \text{for } T_{i,j} \in [0,1], j = 1, \ldots, m_i, \tag{25.10}$$

where $T_{i,j}$ may be random, and the lengths m_i vary across $i = 1, 2, \ldots$. For the statistical analysis, some kind of reconstruction of the unobserved functional datum $X_i \in \mathscr{C}$ must then be performed.

If it can be believed that the functional values (25.10) are observed without noise (or if the noise present in the measurements is negligible), one simple reconstruction method that can be used is the linear interpolation of the known functional values (setting the curves to be constant near the endpoints of the domain). This way, we obtain a sequence $\widetilde{X}_1, \widetilde{X}_2, \ldots$ of independent reconstructions of the random curves that are observed only partially, and we may define an empirical measure $\widetilde{P}_n \in \mathscr{P}(\mathscr{C})$ as the uniform measure supported in the first n functions from this sequence. The same technique can be used also for functions from $\mathscr{C}^K$, for $U = [0,1]$.

Assume that as the sampling process continues (i.e. $n \to \infty$), the largest span between two adjacent points in the set $\{T_{n,1}, \ldots, T_{n,m_n}\} \subset [0,1]$ vanishes in probability. This is true if, for instance, the observation points for the n-th curve are a random sample of size m_n from a distribution on $[0,1]$ with a density bounded away from 0, and $m_n \xrightarrow[n\to\infty]{} \infty$. Under these conditions, it can be shown that $\widetilde{P}_n$ converges weakly to P in space $\mathscr{P}(\mathscr{C})$ (or $\mathscr{P}(\mathscr{C}^K)$) with probability one ([13]). As a corollary, it is possible to obtain that also if the true empirical measure P_n is replaced by $\widetilde{P}_n$ in depths (25.7), (25.6), or (25.9), the resulting functionals are still uniformly consistent.

Theorem 25.4 ([13, 12]). *Let $P \in \mathscr{P}\left(\mathscr{C}^K\right)$ for $U = [0,1]$. Assume that the random curves are observed as described above. Then*

$$\sup_{x \in \mathscr{C}^K} \left| aBD(x; \widetilde{P}_n) - aBD(x; P) \right| \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

Assume further that $P_t \in \mathscr{P}\left(\mathbf{R}^K\right)$ is absolutely continuous for each $t \in [0,1]$. Then

$$\sup_{x \in \mathscr{C}^K} \left| FD(x; \widetilde{P}_n) - FD(x; P) \right| \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

For any $P \in \mathscr{P}(\mathscr{C})$ such that the conditions of Theorem 25.2 are satisfied, also

$$\sup_{x \in \mathscr{C}} \left| ID(x; \widetilde{P}_n) - ID(x; P) \right| \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

Further complications arise when the functional values (25.10) are observed in presence of non-negligible additive noise. Then, we deal with the so-called noisy functional data. Clearly, interpolation is not acceptable for such data. Instead, some kind of approximation of the unknown curves X_i must be performed. For kernel approximation, these problems are addressed in [12]. Therein, results in the spirit of Theorem 25.4 are obtained also for noisy random functions, under some additional mild assumptions. Furthermore, in that paper, a comprehensive treatment of the rates of convergence of the depth FD under various scenarios is provided.

25.6 Depth for Discontinuous Random Functions

Finally, let us assume that the functional observations may contain discontinuities. In this case, one may consider the data to live in

$$\mathscr{B}^K = \left\{ x \colon U \to \mathbf{R}^K \text{ Borel measurable} \right\}.$$

This functional space is extremely large, and not many results can be found for data from $\mathscr{B}^K$. To give a proper definition of a random variable in $\mathscr{B}^K$, one needs to resort to the concept of *Borel measurable function* ([15]) defined as a mapping

$$X \colon \Omega \times U \to \mathbf{R}^K, \quad \text{jointly Borel measurable.}$$

The amount of measurability required from X is greater than what is usually assumed for stochastic processes (and random functions). Interestingly, this is still *not enough* for X to be a measurable mapping from $(\Omega, \mathscr{F}, \mathsf{P})$ to $\mathscr{B}^K$ ([15]). However, if one restricts to subspaces of $\mathscr{B}^K$ such as that of semi-continuous functions, or càdlàg functions for $U = [0,1]$, a random variable $X \sim P \in \mathscr{P}\left(\mathscr{B}^K\right)$ can be defined (for details see [11]). Surprisingly, also at this great level of generality, universal consistency of the integrated depth for $\mathscr{B}^K$-valued data given by

$$FD(x;P) = \int_U D(x(t);P_t)\,\mathrm{d}t, \quad \text{for } x \in \mathscr{B}^K, P \in \mathscr{P}\left(\mathscr{B}^K\right), \tag{25.11}$$

can be shown, for D from (25.4).

Theorem 25.5 ([11]). *The depth FD in* (25.11) *is universally consistent over* $\mathscr{B}^K$.

Acknowledgements This research is supported by the IAP research network no. P7/06 of the Federal Science Policy (Belgium). The author is a Research Assistant of the Research Foundation - Flanders, and acknowledges support from this foundation.

References

[1] Chakraborty, A., Chaudhuri, P.: On data depth in infinite dimensional spaces. Ann. Inst. Statist. Math., **66**(2):303–324, (2014)
[2] Claeskens, G., Hubert, M., Slaets, L., Vakili, K.: Multivariate functional halfspace depth. J. Amer. Statist. Assoc., **109**(505):411–423, (2014)
[3] Cuevas, A., Fraiman, R.: On depth measures and dual statistics. A methodology for dealing with general data. J. Multivariate Anal., **100**(4):753–766, (2009)
[4] Donoho, D. L., Gasko, M.: Breakdown properties of location estimates based on halfspace depth and projected outlyingness. Ann. Statist., **20**(4):1803–1827, (1992)
[5] Fraiman, R., Muniz, G.: Trimmed means for functional data. Test, **10**(2):419–440, (2001)
[6] Gijbels, I., Nagy, S.: Consistency of non-integrated depths for functional data. J. Multivariate Anal., **140**:259 – 282, (2015)
[7] Kuelbs, J., Zinn, J.: Half-region depth for stochastic processes. J. Multivariate Anal., **142**:86–105, (2015)
[8] López-Pintado, S., Romo, J.: On the concept of depth for functional data. J. Amer. Statist. Assoc., **104**(486):718–734, (2009)
[9] López-Pintado, S., Sun, Y., Lin, J. K., Genton, M. G.: Simplicial band depth for multivariate functional data. Adv. Data Anal. Classif., **8**(3):321–338, (2014)
[10] Mosler, K.: Depth statistics. In Robustness and complex data structures, pages 17–34. Springer, Heidelberg, (2013)
[11] Nagy, S.: Integrated depth for measurable functions and sets. Statist. Probab. Lett., **123**:165–170, (2017)
[12] Nagy S., Ferraty, F.: Data depth for measurable noisy random functions. Submitted.
[13] Nagy, S, Gijbels, I., Hlubinka, D.: Weak convergence of discretely observed functional data with applications. J. Multivariate Anal., **146**:46 – 62, (2016)
[14] Nagy, S., Gijbels, I., Omelka, M., Hlubinka, D.: Integrated depth for functional data: statistical properties and consistency. ESAIM Probab. Stat., **20**:95–130, (2016)
[15] Tsirelson, B.: Measurability and continuity. Lecture Notes. Tel Aviv University, (2012) http://www.tau.ac.il/ tsirel/Courses/MeasCont/main.html

Chapter 26
Statistical functional depth

Alicia Nieto-Reyes and Heather Battey

Abstract This presentation is a summary of the paper [14], which formalizes the definition of statistical functional depth, with some extensions on the matter.

26.1 Introduction

Depth functions assign order to elements in a space $\mathscr{X}$ based on a probability measure P on $\mathscr{X}$. The assignment of order is non-trivial beyond the case of $\mathscr{X} = \mathbb{R}$, yet order statistics lay the foundation for a number of classical frequentist procedures, such as rank-based inference and outlier detection. Particularly, in functional datasets outlier detection provides a means to detect a fraudulent signature amongst a sample of genuine signatures. Similar applications of functional outlier detection are abundant in medical statistics, where for instance, the ability to detect abnormal time course gene expression curves, electrocardiograms, etc. serves as a useful diagnostic tool.

In Figure 26.1, we elicit the essential idea of statistical depth of a distribution P on univariate space, $\mathscr{X} = \mathbb{R}$, multidimensional space, $\mathscr{X} = \mathbb{R}^p$, and function space $\mathscr{X} = \mathfrak{F}$. Whilst the details are unimportant at this stage, the depth is discernible in the colour, with the depth median in dark red and the observations with lowest depth in dark blue. As is visually apparent from the plots in Figure 26.1, the depth induces a centre outward ordering of the observations from the median, going from dark red to dark blue through red, orange, yellow, green, cyan and blue.

For multivariate spaces, the notion of depth is formalized in [16], using some properties defined in [10]. The most prominent example is the Tukey depth [15], which has a computationally effective approximation: the random Tukey depth [3].

Alicia Nieto-Reyes (✉)
Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria. Avd/ Los Castros s/n. 39005 Santander, Spain. e-mail: alicia.nieto@unican.es

Heather Battey
Department of Mathematics, Imperial College London e-mail: h.battey@imperial.ac.uk

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_26

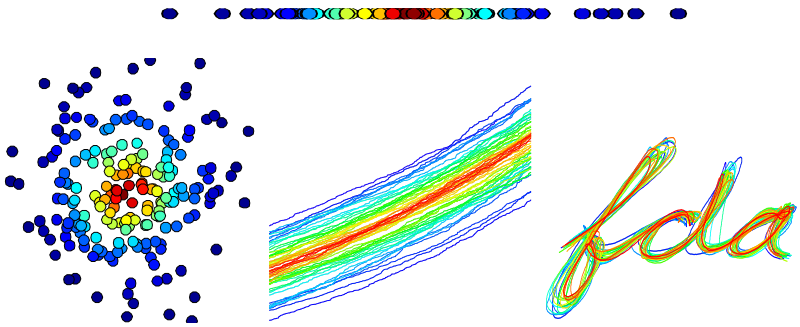


Fig. 26.1: Top row: observations from a univariate normal distribution. Bottom row: observations from a bivariate normal distribution (left), an example of univariate functional data (middle) and bivariate functional data (right). In each case, the observations are coloured according to the rank of their depth value.

Relatively recently there have been several extensions of examples of multivariate depth to cover high dimensional spaces and functional spaces. In [14] the notion is formalized for functional spaces, where six axiomatic properties are established which are needed for a functional depth to be useful. The formalization of this definition varies from multivariate to functionals spaces to recognise topological features such as continuity, smoothness and contiguity of the functional space.

The properties that form the definition of functional depth provide not only a sophisticated extension of those defining the multivariate depth, but also implicitly address several common or inherent difficulties associated with functional data. Particularly, we focus on the difference between the multivariate and functional notions when handling a property that regards the depth value at the center of symmetry of symmetric distributions (see Section 26.3).

The robustness of the empirical depth to the presence of outliers is commonly considered one of its important characteristics, despite the absence of a formal guarantee in the multivariate definition of depth. As an incidental product, functional depths satisfying this definition automatically produce a qualitatively robust estimator of the population depth. Functional depth constructions adhering to these axioms intrinsically address the delicate challenge of inherent partial observability of functional data, providing a minimal guarantee on the performance of the empirical depth counterpart beyond the idealised and practically infeasible case of full observability. Another challenge, automatically dealt (when present) by the definition, regards functional data with little variability and a big overlap over part of the domain. Intuitively, observations over such part of the domain should have less importance in the assignment of depth than others.

Some functional depth proposals, the h-depth [5], the random Tukey depth [3, 4], the band and modified band depth [11] and the half-region and modified half-region depth [12], are discussed with reference to the aforementioned axioms in [14]. The

literature on examples of functional depth is, however, wider, including among other the integrated depth [8, 2], the integrated dual depth [6], the spatial depth [1], a depth based on distances [13] and one based on tilting [9].

In Section 26.2 we include the notion of functional depth and in Section 26.3 we regard the importance of the proposed properties and its implications in the application of the functional depth.

26.2 Definition

We provide here the axiomatic definition of statistical functional depth [14], where minor clarifications are added to rule out pathological cases.

Definition 26.1. [*Statistical functional depth*] Let $(\mathfrak{F}, \mathscr{A}, P)$ be a probability space. Let $\mathscr{P}$ be the space of all probability measures on $\mathfrak{F}$. The mapping

$$D(\cdot,\cdot) : \mathfrak{F} \times \mathscr{P} \longrightarrow \mathbb{R}$$

is a *statistical functional depth* if it satisfies properties [P-1.] to [P-6.], below.

P-1. *Distance invariance*. $D(f(x), P_{f(X)}) = D(x, P_X)$ for any $x \in \mathfrak{F}$ and $f : \mathfrak{F} \to \mathfrak{F}$ such that for any $y \in \mathfrak{F}$,

$$d(f(x), f(y)) = a_f \cdot d(x,y),$$

with $a_f \in \mathbb{R}\backslash\{0\}$.

P-2. *Maximality at centre*. For any $P \in \mathscr{P}$ possessing a unique centre of symmetry $\theta \in \mathfrak{F}$ with respect to some notion of functional symmetry,

$$D(\theta, P) = \sup_{x \in \mathfrak{F}} D(x, P).$$

P-3. *Strictly decreasing with respect to the deepest point*. For any $P \in \mathscr{P}$ such that $D(z,P) = \max_{x \in \mathfrak{F}} D(x,P)$ exists with $D(z,P) = D(z',P)$ implying $d(z,z') = 0$, $D(x,P) < D(y,P) < D(z,P)$ holds for any $x, y \in \mathfrak{F}$ such that

$$\min\{d(y,z), d(y,x)\} > 0 \text{ and } \max\{d(y,z), d(y,x)\} < d(x,z). \tag{26.1}$$

P-4. *Upper semi-continuity in x*. $D(x,P)$ is upper semi-continuous as a function of x, i.e., for all $x \in \mathfrak{F}$ and for all $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\sup_{y \in \mathfrak{F}_x \,:\, d(x,y) < \delta} D(y,P) \leq D(x,P) + \varepsilon, \tag{26.2}$$

where

$$\mathfrak{F}_x := \Big\{ y \in \mathfrak{F} : d(y,x) < d(y,\theta) \text{ or } \max\{d(y,\theta), d(y,x)\} < d(x,\theta) \Big\}$$

for $\theta = \arg\sup_{x\in\mathfrak{F}} D(x,P)$.

P-5. *Receptivity to convex hull width across the domain.* $D(x,P_X) < D(f(x),P_{f(X)})$ for any $x \in \mathfrak{C}(\mathfrak{F},P)\backslash 0$ with $D(x,P) < \sup_{y\in\mathfrak{F}} D(y,P)$ and $f:\mathfrak{F}\to\mathfrak{F}$ such that $f(y(v)) = \alpha(v)y(v)$ with $\alpha(v)\in(0,1)$ for all $v \in L_\delta$ and $\alpha(v) = 1$ for all $v \in L_\delta^c$.

$$L_\delta := \arg\sup_{H\subseteq\mathscr{V}} \left\{ |H| : \sup_{x,y\in\mathfrak{C}(\mathfrak{F},P)} d(x(H),y(H)) \le \delta \right\}$$

for any $\delta \in [\inf_{v\in\mathscr{V}} d(L(v),U(v)), d(L,U))$ such that $\lambda(L_\delta) > 0$ and $\lambda(L_\delta^c) > 0$, with $|H|$ denoting the length of H.

P-6. *Continuity in P.* For all $x\in\mathfrak{F}$, for all $P\in\mathscr{P}$ and for every $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$ such that

$$|D(x,Q) - D(x,P)| < \varepsilon$$

P-almost surely for all $Q\in\mathscr{P}$ with $d_{\mathscr{P}}(Q,P) < \delta$ P-almost surely, where $d_{\mathscr{P}}$ metricises the topology of weak convergence.

26.3 Implications of the functional depth properties

Here we give discussion the definition of statistical functional depth with implications. A thorough discussion is given in [14].

P-1. *Distance invariance*.

This property concerns the fact that the depth of a dataset or space, with respect to a distribution on that space, should remain the same under certain transformations such as recentering around the deepest element.

It has implications when applying the functional depth, for instance, in that ,because of this property, the depth will not be altered if we change the units of measurement.

P-2. *Maximality at centre*.

The existing notions of functional symmetry are pointwise and so direct extensions of multivariate notions. Thus, to compare different existing examples of funtional depth it is proposed in [14] to check the following property, named *Maximality at Gaussian process mean*.

For P a zero-mean, stationary, almost surely continuous Gaussian process on $\mathscr{V}$, $D(\theta,P) = \sup_{x\in\mathfrak{F}} D(x,P) \neq \inf_{x\in\mathfrak{F}} D(x,P)$, *where* θ *is the zero mean function.*

P-3. *Strictly decreasing with respect to the deepest point*.

This property is a generalization of two multivariate properties, monotonicity relative to the deepest point and vanishing at infinity. Though it is more strict than the multivariate monotonicity property. The reason is directly link with the fact that

depth is commonly applied to classification problems and therefore ties in the depth values are not recommended. See [7] for an analysis of the problem in multivariate spaces.

P-4. *Upper semi-continuity in x.*

This property relates the depth to the cumulative distribution function and as a consequence to the fact that is commonly attributed to the depth of being able to show the characteristics of the underlying distribution; obtaining different depth values when computed with respect to different probability distributions.

P-5. *Receptivity to convex hull width across the domain.*

This property is only considered in functional spaces. This is due to the fact that is typical to find in functional spaces data that exhibit little variability in part of the domain and the general fact that real data contain noise. Thus, if this property were not satisfied the errors in the part of the image of the domain that exhibit little variability would have a high influence in the result of the depth.

P-6. *Continuity in P.*

This property is important for different reasons. First, it regards that the empirical depth converges to the population depth almost surely. As functional data are observed only on a finite grid of time points, it is not possible to infer the empirical depth but a reconstruction of it. However, by this property, the depth computed with respect to the reconstructed empirical depth also converges to the population depth almost surely, when the reconstructed empirical depth is close enough in Prohorov metric to the theoretical distribution.

Acknowledgements Alicia Nieto-Reyes would like to thank the Scientific Committee of the IWFOS 2017 for being invited to present her work at the workshop.

References

[1] Chakraborty, A., Chaudhuri, P.: The spatial distribution in infinite dimensional spaces and related quantiles and depths. Ann. Statist. **42**, 1203–1231 (2014)
[2] Claeskens, G., Hubert, M., Slaets, L., Vakili, K.: Multivariate Functional Halfspace Depth. J. Amer. Statist. Assoc. **109**, 411-423 (2014)
[3] Cuesta-Albertos, J. A., Nieto-Reyes, A.: The random Tukey depth. Comput. Statist. Data Anal. **52**, 4979–4988 (2008)
[4] Cuesta-Albertos, J. A., Nieto-Reyes, A.: Functional Classification and the Random Tukey Depth. Practical Issues. Combining Soft Computing and Statistical Methods in Data Analysis. Advances in Intelligent and Soft Computing. Editors: Borgelt, C. et. al. Springer Berlin / Heidelberg. **77**, 123–130 (2010)

[5] Cuevas, A., Febrero, M., Fraiman, R.: Robust estimation and classification for functional data via projection-based depth notions. Comput. Statist. **22**, 481–496 (2007)
[6] Cuevas, A., Fraiman, R.: On depth measures and dual statistics. A methodology for dealing with general data. J. Multivariate Anal. **100** (2009)
[7] Einmahl, J. H. J., Li, J., Liu, R. Y.: Bridging centrality and extremity: Refining empirical data depth using extreme value statistics. Ann. Statist. **43**, 2738–2765 (2015)
[8] Fraiman, R., Muniz, G.: Trimmed means for functional data. Test. **10**, 419–440 (2001)
[9] Genton, M. G., Hall, P.: A tilting approach to ranking influence. J. R. Stat. Soc. Ser. B Stat. Methodol.. **78**, 77–97 (2014)
[10] Liu, R. Y.: On a notion of data depth based on random simplices. Ann. Statist. **18**, 405–414 (1990)
[11] López-Pintado, S., Romo, J.: On the concept of depth for functional data. J. Amer. Statist. Assoc. **104**, 718–734 (2009)
[12] López-Pintado, S., Romo, J.: A half-region depth for functional data. Comput. Statist. Data Anal. **55**, 1679–1695 (2011)
[13] Nieto-Reyes, A.: On the Properties of Functional Depth. Recent Advances in Functional Data Analysis and Related Topics: Selected papers from IWFOS'2011. Editor: Ferraty, F. Physica-Verlag/Springer, Heidelberg. 239–244 (2011)
[14] Nieto-Reyes, A., Battey, H.: A topologically valid definition of depth for functional data. Statist. Sci. **31**, 61–79 (2016)
[15] Tukey, J. W.: Mathematics and the picturing of data. Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974). **2**, 523–531. Canad. Math. Congress, Montreal, Que. (1975)
[16] Zuo, Y., Serfling, R.: General notions of statistical depth function. Ann. Statist. **28**, 461–482 (2000)

Chapter 27
Differential interval-wise testing for local inference in Sobolev spaces

Alessia Pini, Lorenzo Spreafico, Simone Vantini and Alessandro Vietti

Abstract We present a local non-parametric inferential technique - namely, the differential interval-wise testing, or D-IWT - able to test the distributional equality of two samples of functional data embedded in Sobolev spaces. D-IWT can impute differences between the two samples to specific parts of the domain and to specific orders of differentiation. The proposed technique is applied to the functional data analysis of a data set of tongue profiles.

27.1 Introduction

Inference is a lively area in the field of functional data analysis [12, 5, 7]. The literature dealing with inference of functional data has pursued different approaches. One one hand, literature has focused both on parametric methods - relying on parametric distributional models to compute the distribution of the test statistic under the null hypothesis - and on non parametric methods - relying on computationally intensive re-sampling techniques (e.g., bootstrapping or permuting). On the other hand, the testing methods proposed in the literature can be divides into global methods - providing the analyst with a "simple" rejection or non-rejection of the null hypothesis - and local methods - providing the analyst with portions of the domain where the null hypothesis is rejected or not rejected.

The majority of works dealing with inference for functional data rely on global parametric methods [4, 7, and references therein], but there is a consistent literature pertaining also to global non-parametric methods (e.g., [2, 3, 6]). Recently, some

Alessia Pini (✉) and Simone Vantini
MOX - Department of Mathematics, Politecnico di Milano, Milano, Italy e-mail: alessia.pini@polimi.it, simone.vantini@polimi.it

Lorenzo Spreafico and Alessandro Vietti
ALPs - Alpine Laboratory of Phonetics and Phonology Free University of Bozen-Bolzano, Bolzano, Italy e-mail: lorenzo.spreafico@unibz.it, alessandro.vietti@unibz.it

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_27

works have been proposed in the framework of local parametric techniques (e.g., [1]) and local non-parametric techniques (e.g., [10, 11, 15]). In this work we focus on a non-parametric local framework, and we consider as a starting point the interval-wise testing procedure (IWT) proposed by [11]. The IWT is an inferential procedure for functional data that selects the specific parts of the domain imputable for the rejection of a functional null hypothesis.

The technique that we discuss in this paper - namely the differential interval-wise testing (D-IWT) - is an extension of the IWT that jointly exploits the information of the curves and of their derivatives, and can select the specific orders of differentiation and parts of the domain imputable for the rejection of the null hypothesis. When comparing two or more functional populations it is indeed natural – on the one hand – to compute derivatives of the curves, which can convey a deep insight on the data themselves and – on the other one – to localize along the domain the possible differences between the functional populations under inspection.

27.2 Methodology

We embed the null hypothesis testing problem in the Sobolev space $H^d(T)$ of all real-valued squared-integrable functions on the domain T with squared-integrable derivatives up to order d (where T is an open interval of $\mathbb{R}$).

Assume to observe two independent samples of functional data ξ_{ji}, $j = 1,2$, $i = 1,\dots,n_j$ taking values in the Sobolev space $H^d(T)$, $d \geq 1$. Assume $\{\xi_{1i}\}_{i=1,\dots,n_1} \sim$ iid $\boldsymbol{\xi}_1$ and $\{\xi_{2i}\}_{i=1,\dots,n_2} \sim$ iid $\boldsymbol{\xi}_2$, where $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are two independent random functions. We aim at performing the following test:

$$H_0 : \boldsymbol{\xi}_1 \stackrel{\mathrm{d}}{=} \boldsymbol{\xi}_2 \text{ against } H_1 : \boldsymbol{\xi}_1 \stackrel{\mathrm{d}}{\neq} \boldsymbol{\xi}_2. \tag{27.1}$$

In case of rejection of the null hypothesis, we aim at imputing the rejection to: (*i*) specific parts of the domain of functional data, and (*ii*) specific characteristics of functional data that can be naturally conveyed by one or more derivatives.

The D-IWT - which is fully described in detail in [12] - addresses the problem of testing null hypothesis of equality in distribution of two functional populations taking derivatives into account. To perform such a test, we replace the original test (27.1) by the following family of tests, each focusing on a specific order of differentiation $k = 0,\dots,d$:

$$H_0 : \boldsymbol{\xi}_1 \stackrel{\mathrm{d}}{=} \boldsymbol{\xi}_2 \text{ against } H_1^k : \mathbb{E}[D^k \boldsymbol{\xi}_1] \neq \mathbb{E}[D^k \boldsymbol{\xi}_2]. \tag{27.2}$$

The outputs of the D-IWT procedure are:

- $d+1$ **partial adjusted** p**-value functions** $\widetilde{p}_{D^k} : T \to [0,1]$, one for each order of differentiation, for testing separately the partial hypotheses (27.2);
- $d+1$ **multi-derivative adjusted** p**-value functions** $\widetilde{\widetilde{p}}_{D^k} : T \to [0,1]$, one for each order of differentiation, for testing jointly the partial hypotheses (27.2).

The adjusted partial p-value functions are computed by applying the interval-wise testing [11] to every test of the family (27.2). In details, for every k, test (27.2) is performed by means of a non-parametric permutation test [9] on every interval of the domain $\mathscr{I} \subseteq T$. Let $p_{D^k}^{\mathscr{I}}$ denote the p-value of such test. The adjusted p-value $\widetilde{p}_{D^k}(t)$ of order $k = 0, \ldots, d$ and point $t \in T$ is computed as the supremum of all p-values of tests on intervals containing t:

$$\widetilde{p}_{D^k}(t) = \sup_{\mathscr{I} \ni t} p_{D^k}^{\mathscr{I}}. \tag{27.3}$$

The adjusted multi-derivative p-value functions of the $d+1$ orders of differentiation are computed by adjusting the $d+1$ partial p-value functions $\widetilde{p}_{D^k}(t)$ by means of a closed testing procedure [8]. In detail, for all possible combinations of differentiation orders indexed by $\mathbf{k} = \{k_1, k_2, \ldots, k_Q\}$ with $\forall q : k_q \in \{0, \ldots, d\}$ and $Q \in \{2, \ldots, d+1\}$, a Q-variate IWT is performed to test hypotheses:

$$H_0 : \boldsymbol{\xi}_1 \overset{\mathrm{d}}{=} \boldsymbol{\xi}_2 \text{ against } H_1^{\mathbf{k}} : \bigcup_{k \in \mathbf{k}} \mathbb{E}[D^k \boldsymbol{\xi}_1] \neq \mathbb{E}[D^k \boldsymbol{\xi}_2]. \tag{27.4}$$

The tests (27.4) are performed by means of permutation tests based on Sobolev norms (or semi-norms) on the corresponding orders of differentiation. The adjusted p-value functions $\widetilde{p}_{D^{\mathbf{k}}}(t)$ are computed according to formula (27.3) based on the obtained p-values of tests (27.4). Finally, the $d+1$ adjusted multi-aspect p-value functions $\widetilde{\widetilde{p}}_{D^k}(t)$ are calculated by taking for each order of differentiation the pointwise maximum of all adjusted p-value functions $\widetilde{p}_{D^{\mathbf{k}}}(t)$ involving that order:

$$\widetilde{\widetilde{p}}_{D^k}(t) = \sup_{\mathbf{k} \ni k} \widetilde{p}_{D^{\mathbf{k}}}(t) \tag{27.5}$$

The p-value functions $\widetilde{\widetilde{p}}_{D^k}(t)$ can be thresholded at level α to select the intervals of the domain presenting significant differences between the two populations on the corresponding order of differentiation.

The D-IWT is provided by a joint control of the family-wise error rate (FWER) on each sub-interval of the domain T over the $d+1$ orders of differentiation. For every interval of the domain where the null hypothesis is not violated, the selection procedure allows to control the probability that the interval is wrongly selected as significant for at least one of the tested derivatives. Specifically, $\forall \alpha \in (0,1)$:

$$\forall \mathscr{I} \subseteq T : H_0 \text{ true on } \mathscr{I} \Rightarrow \mathbb{P}\left(\exists t \in \mathscr{I}, \exists k \in \{0,1,\ldots,d\} \text{ s.t. } \widetilde{\widetilde{p}}_{D^k}(t) \leq \alpha\right) \leq \alpha.$$

The D-IWT consistency is also proven. Proofs pertaining to both FWER control and consistency, and extentions to more complex testing problems are shown in [12].

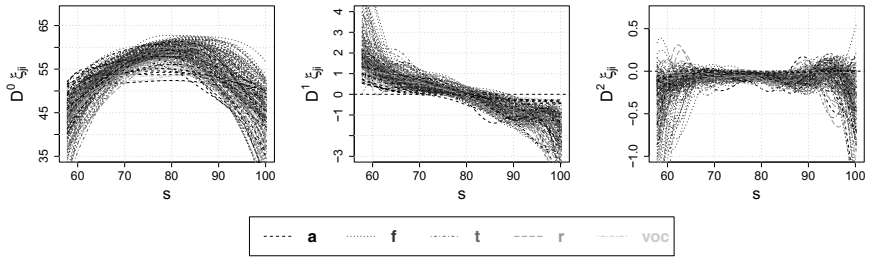


Fig. 27.1: Left: functional data of sagittal tongue profiles corresponding to the five groups of /R/ articulation: approximant (*a*), fricative (*f*), tap (*t*), trill (*r*), and vocalized (*voc*). Center: first derivatives of the functional data. Right: second derivatives of functional data.

27.3 Data Analysis

To better understand the potential of the D-IWT in the practice, we report in this section the analysis of a data set of tongue profiles. The aim of the analysis is to test the differences between five different manners of articulating the uvular /R/ in terms of vertical position of the tongue, tongue slope, and tongue concavity. The five groups that compose the data set can be described as follows:

- *f* FRICATIVE: it is produced by constricting airflow through a narrow channel at the place of articulation, causing turbulence. The fricative variant is produced with a contact between the tongue and the palate.
- *a* APPROXIMANT: it shares with *f* the way of transmission of the sound through the air, although there is no contact between tongue and palate.
- *r* TRILL: it is produced by directing the air over the tongue so that it vibrates. There is a contact between the tongue and the palate.
- *t* TAP: it is produced with a single contraction of the muscles so that the tongue is thrown against the palate. There is a contact between the tongue and the palate.
- *voc* VOCALIZATION: the airstream proceeds along the sides of the tongue but it is blocked by the tongue from going through the middle of the mouth. There is no contact between the tongue and the palate.

Data are composed of 117 tongue profiles of five variants of uvular /R/ recorded from one native speaker of Tyrolean, collected by ultrasound imaging techniques at the Alpine Laboratory of Phonetic Sciences and Phonology of the Free University of Bozen - Bolzano (Italy). For a detailed description of the data set, see [14]. The functional data have been obtained by a penalized B-spline smoothing of order six. The penalization parameter was computed via generalized cross-validation criterion (see [12] for details). The obtained functional data of the five groups and their first and second derivatives are displayed in Figure 27.1.

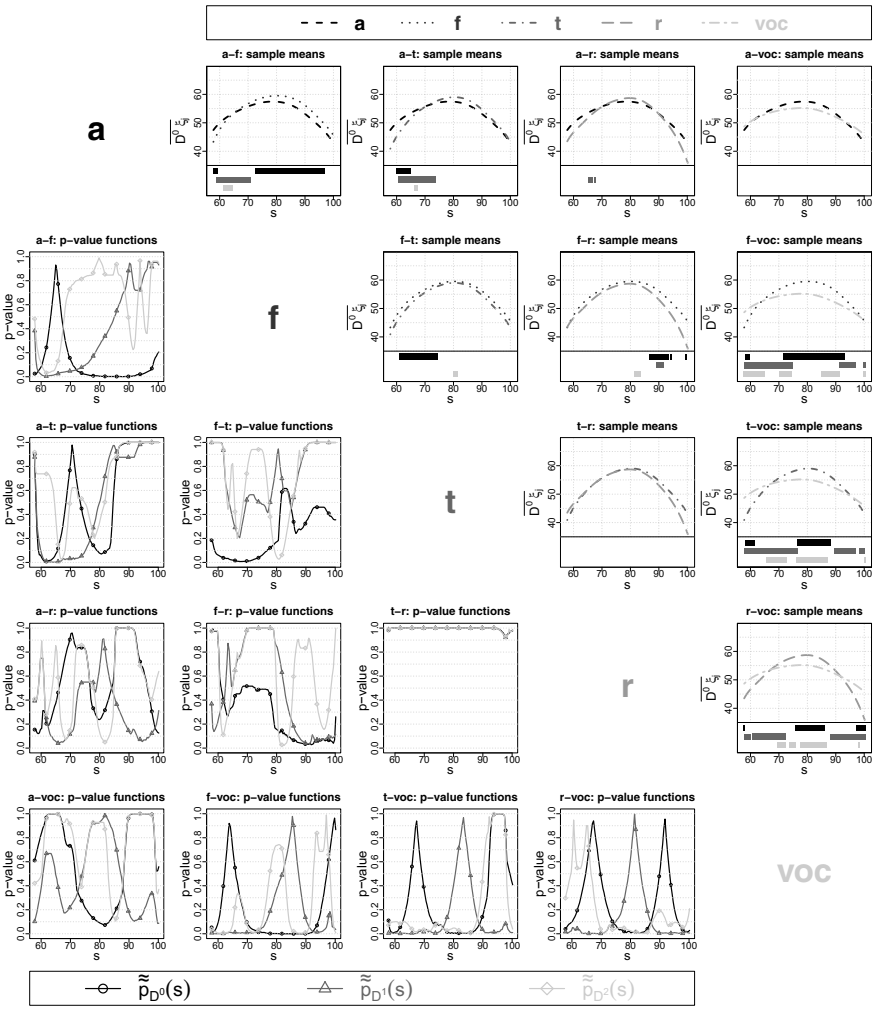


Fig. 27.2: Scatter matrix of pairwise differences between the five groups. Groups are identified in the diagonal. For each couple of groups: the upper-diagonal box indicates the two sample means (upper part) and the significant intervals at 5% level (lower part); the lower diagonal box indicates the three multi-aspect adjusted p-value functions $\tilde{\tilde{p}}_{D^0}(t)$ (black), $\tilde{\tilde{p}}_{D^1}(t)$ (dark gray), and $\tilde{\tilde{p}}_{D^2}(t)$ (light gray).

We perform a D-IWT-based inferential analysis of tongue profiles, in order to identify the possible pairwise differences between the five variants in the curves and the first two orders of differentiation ($d = 2$). Figure 27.2 displays the results of the tests. In detail - for each pairwise comparison - the upper diagonal plots show the two sample mean curves, and the lower diagonal panel shows the three multi-aspect

adjusted p-value functions $\widetilde{\widetilde{p}}_{D^0}(t)$ (black), $\widetilde{\widetilde{p}}_{D^1}(t)$ (dark gray), and $\widetilde{\widetilde{p}}_{D^2}(t)$ (light gray). The three bars in the lower part of each upper diagonal plot indicate the intervals with associated adjusted p-value lower than 5%. The color of the bars is consistent with the one of the three adjusted p-value functions. Inference in terms of multi-derivative adjusted p-value functions provides a highly informative and detailed representation of the regions of the tongue where a significant difference is located. Looking at Figure 27.2, we obtain indeed a clear representation of data differences.

As expected, we observe more pronounced differences when comparing *a* or *voc* (produced without touching the palate) with *f*, *t*, or *r* (produced by touching the palate), while there are less pronounced differences when comparing *a* with *voc* and when comparing two variants of the group *f*, *t*, and *r*. For instance, we notice that there are no significant difference between trill /R/ (*r*) and tap /R/ (*t*) in all orders of differentiation. Conversely, we notice that at $\alpha = 1\%$, approximant /R/ (*a*) and fricative /R/ (*f*) (second panel of the first row) are pointed out as not identically distributed. They show significantly different vertical positions in the front part of the tongue, different slopes in the very back, and no difference in the concavities, and these findings can be interpreted in terms of the differences in articulation between the two variants. Indeed, fricative /R/ is a consonant, and it is produced by touching the palate, while approximant /R/ is produced without touching the palate. Coherently, we observed significant differences in vertical position between the two variants, with *f* reaching higher vertical positions than *a*. In addition, having a lower degree of constriction, approximant /R/ has a lower slope in the back part of the tongue. A more detailed analysis of the results, as well as a simulation study assessing the performances of the D-IWT can be found in [12].

27.4 Conclusions

We presented a local testing procedure for testing differences between two functional samples and selecting the specific intervals of the domain and the specific orders of differentiation where the two functional populations are significantly different. The proposed method, i.e., the D-IWT, is a fully non-parametric procedure based on permutation tests. Hence, it is not based on parametric distributional assumptions on the functional data, and it is provided with a control of the family-wise error rate on each sub-interval of the domain for any (even very small) sample size. The D-IWT hereby described can be readily extended to more complex models, such as the comparison of variances between two populations, or the comparison of several functional populations.

We applied the D-IWT to perform the pairwise comparison between five groups of tongue profiles associated with five different variants of uvular /R/. The comparison between the variants is performed in terms of vertical position, slope, and concavity of the tongue. Inference in terms of multi-derivative adjusted p-value functions provides a highly informative and detailed representation of the regions of the tongue where a significant difference is located.

The data analysis here reported pertains to the comparison of tongue profiles on a specific instant during the articulation of the /R/. Our future work will focus on extending the analysis in order to consider the whole time-varying articulation. Such an extension can be performed by considering tongue profiles as functions of both space and time, and hence defined on a bi-dimensional domain. In this case, differential quantities include slope and concavity of the tongue on one hand and velocity and acceleration of the tongue on the other one.

References

[1] Abramovich, F., Heller, R.: Local functional hypothesis testing. Mathematical Methods of Statistics, **14** (3), 253 (2005)
[2] Cardot, H., Goia, A., Sarda, P.: Testing for no effect in functional linear regression models, some computational approaches. Comm. Statist. Simulation Comput., **33** (1), 179–199 (2004)
[3] Cuesta-Albertos, J. A., Febrero-Bande, M.: A simple multiway ANOVA for functional data. TEST, **19** (3), 537–557 (2010)
[4] Cuevas, A., Febrero-Bande, M., Fraiman, R.: An ANOVA test for functional data. Comput. Statist. Data Anal., **47** (1): 111–122 (2004)
[5] Ferraty, F., Vieu, P.: Nonparametric functional data analysis: theory and practice. Springer (2006).
[6] Hall, P., Tajvidi, N.: Permutation tests for equality of distributions in high-dimensional settings. Biometrika, **89** (2), 359–374 (2002).
[7] Horváth, L., Kokoszka, P.: Inference for functional data with applications. Springer (2012)
[8] Marcus, R., Peritz, E., Gabriel, K. R.: On closed testing procedures with special reference to ordered analysis of variance. Biometrika, **63** (3), 655–660 (1976)
[9] Pesarin, F., Salmaso, L.: Permutation tests for complex data: theory, applications and software. John Wiley & Sons Inc (2010)
[10] Pini, A., Vantini, S.: The interval testing procedure: a general framework for inference in functional data analysis. Biometrics, **72**, 835–845 (2016)
[11] Pini, A., Vantini, S.: Interval-Wise Testing for Functional Data. J. Nonparametr. Stat.. To appear, (2017)
[12] Pini, A., Vantini, S.: Multi-aspect local inference for functional data: analysis of ultrasound tongue profiles. Tech. Rep. 2017 MOX, Politecnico di Milano (2017)
[13] Ramsay, J. O., Silverman, B. W.: Functional data analysis. Springer, New York (2005)
[14] Vietti, A., Spreafico, L., Galatà, V.: An ultrasound study of the phonetic allophony of Tyrolean /r/. ICPhS 2015 Proceedings (2015)
[15] Vsevolozhskaya, O., Greenwood, M., Holodov, D.: Pairwise comparison of treatment levels in functional analysis of variance with application to erythrocyte hemolysis. Ann. Appl. Stat., **8** (2), 905–925 (2014)

[16] Zhang, J., Liang, X.: One-way ANOVA for functional data via globalizing the pointwise F-test. Scand. J. Stat., **41** (1), 51–71 (2014)

Chapter 28
Hotelling in Wonderland

Alessia Pini, Aymeric Stamm and Simone Vantini

Abstract While Hotelling's T^2 statistic is traditionally defined as the Mahalanobis distance between the sample mean and the true mean induced by the inverse of the sample covariance matrix, we hereby propose an alternative definition which allows a unifying and coherent definition of Hotelling's T^2 statistic in any Hilbert space independently from its dimensionality and sample size. In details, we introduce the definition of random variables in Hilbert spaces, the concept of mean and covariance in such spaces and the relevant operators for formulating a proper definition of Hotelling's T^2 statistic relying on the concept of Bochner integral.

28.1 Introduction

Statisticians are more and more confronted with the analysis of *complex* data, where *complexity* can take on various forms. For example, the advent and development of technologies able to capture real-time and/or geolocalized information has provided the statistician with data that can be viewed as functions with a certain degree of smoothness which are the foundations of functional data analysis (FDA) [22, 5]. More recently, the literature has moved in the direction of object-oriented data analysis (OODA) [16, 18] which pertains to analyzing data that are represented with abstract mathematical constructs, often belonging to some space on which a Hilbert structure is assumed. While FDA and OODA are expending rapidly (see [1, 11] for excellent recent reviews on contributions in these fields), the theoretical study of statistical tools for making inference in such spaces is still a vibrant area of methodological investigation [8, 25, 6, 24, 2, 12, 4, 5, 13, 3, 10, 17, 20].

In this work, we will focus on the inferential problem of constructing a statistical test for the means of random variables belonging to Hilbert spaces of possibly infinite dimension. Starting with Hotelling's T^2 statistic widely used in multivariate data

Alessia Pini, Aymeric Stamm (✉) and Simone Vantini,
MOX - Department of Mathematics, Politecnico di Milano, Italy, e-mail: aymeric.stamm@polimi.it

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_28

analysis (MDA) for testing the mean, we will first get the reader to realize that not only it can be defined for high-dimensional data in which the dimensionality p exceeds the sample size n but Hotelling's T^2 statistic can be coherently defined in any Hilbert space independently from its dimensionality and sample size. In details, Section 28.2 reviews the traditional definition of the statistic and proposes an alternative, more general, definition from which the generalization to Hilbert spaces can be straightforwardly foreseen. In Section 28.3 we introduce the definition of random variables in Hilbert spaces, the concept of mean and covariance in such spaces and the relevant operators for formulating, at the end, our proposed general definition of Hotelling's T^2 statistic in Hilbert spaces.

28.2 Hotelling's T^2 revisited

Given a sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ of n *i.i.d* $\mathbb{R}^p$-valued random variables with mean $\mathbf{m} \in \mathbb{R}^p$ and variance-covariance matrix Σ, one can open any textbook on multivariate statistical analysis and reads that, if $n > p$, inference on the mean vector of this population can be carried out using Hotelling's T^2 statistic defined as:

$$T^2 := n\,(\mathbf{m}_n - \mathbf{m})^\top \Sigma_n^{-1} (\mathbf{m}_n - \mathbf{m})\,, \tag{28.1}$$

where $\mathbf{m}_n$ and Σ_n are the sample counterparts of $\mathbf{m}$ and Σ respectively:

$$\mathbf{m}_n := \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \quad \text{and} \quad \Sigma_n := \frac{1}{n-1}\sum_{i=1}^{n} (\mathbf{X}_i - \mathbf{m}_n)(\mathbf{X}_i - \mathbf{m}_n)^\top. \tag{28.2}$$

What is less known is the connection between Hotelling's T^2 and the univariate Student t statistic. In effect, one can make inference on the mean vector $\mathbf{m}$ by computing a Student t statistic for any direction $\mathbf{a} \in \mathbb{R}^p$ as follows:

$$t_{\mathbf{a}} = \sqrt{n}\frac{\mathbf{a}^\top (\mathbf{m}_n - \mathbf{m})}{\sqrt{\mathbf{a}^\top \Sigma_n \mathbf{a}}}.$$

It can then be shown that Hotelling's T^2 defined in eq. (28.1) reads:

$$T^2 := \max_{\mathbf{a} \in \mathbb{R}^p \setminus \{\mathbf{0}\}} t_{\mathbf{a}}^2 = n \max_{\mathbf{a} \in \mathbb{R}^p \setminus \{\mathbf{0}\}} \frac{\left[\mathbf{a}^\top (\mathbf{m}_n - \mathbf{m})\right]^2}{\mathbf{a}^\top \Sigma_n \mathbf{a}}. \tag{28.3}$$

At this point, we can make four important remarks:

1. The sample variance-covariance matrix Σ_n is a.s. positive definite since $n > p$. Hence, the maximum in eq. (28.3) can be searched over the space spanned by the columns of Σ_n which coincides with $\mathbb{R}^p$;
2. The scalar product between two elements of $\mathbb{R}^p$, say $\mathbf{a}$ and $\mathbf{b}$, is denoted by $\mathbf{a}^\top \mathbf{b}$ in eq. (28.3). It is actually nothing but the inner product $\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbb{R}^p}$;

3. The dot product between two elements $\mathbf{a}$ and $\mathbf{b}$ of $\mathbb{R}^p$ is a rank-1 matrix denoted by $\mathbf{a}\mathbf{b}^\top$ in the above equations. It is nothing but the tensor product $\mathbf{a} \otimes_{\mathbb{R}^p} \mathbf{b}$;
4. We can define a rank-1 matrix $D_n = (\mathbf{m}_n - \mathbf{m}) \otimes_{\mathbb{R}^p} (\mathbf{m}_n - \mathbf{m})$ so that the numerator in the maximization problem involved in eq. (28.3) reads $\langle \mathbf{a}, D_n \mathbf{a} \rangle_{\mathbb{R}^p}$.

In definitive, the definition of Hotelling's T^2 statistic provided by eq. (28.3) can be equivalently stated as follows. Given a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of n i.i.d. $\mathbb{R}^p$-valued random variables with mean $\mathbf{m}$ and covariance matrix Σ, Hotelling's T^2 statistic is defined as:

$$T^2 := n \max_{\mathbf{a} \in \mathrm{Im}(\Sigma_n) \setminus \{\mathbf{0}\}} \frac{\langle \mathbf{a}, D_n \mathbf{a} \rangle_{\mathbb{R}^p}}{\langle \mathbf{a}, \Sigma_n \mathbf{a} \rangle_{\mathbb{R}^p}}, \tag{28.4}$$

where

$$D_n = (\mathbf{m}_n - \mathbf{m}) \otimes_{\mathbb{R}^p} (\mathbf{m}_n - \mathbf{m}),$$
$$\Sigma_n = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{X}_i - \mathbf{m}_n) \otimes_{\mathbb{R}^p} (\mathbf{X}_i - \mathbf{m}_n).$$

This definition is more general and provides indeed a definition of Hotelling's T^2 statistic that holds also in the case $p \geq n$ ([23]). Moreover, it naturally opens to the possibility of extending the definition beyond the Euclidean framework.

28.3 Hotelling's T^2 in Hilbert spaces

The aim of this section is to show that eq. (28.4) is actually not restricted to the Hilbert space $\mathbb{R}^p$ endowed with its natural inner product but can actually serve as the definition of Hotelling's T^2 statistic in any Hilbert space. To accomplish this goal, we shall begin with introducing the notions of random variable on a generic Hilbert space as well as the concepts of mean and covariance in such a space.

Let $(\mathbb{H}, \oplus_{\mathbb{H}}, \odot_{\mathbb{H}}, \langle \cdot, \cdot \rangle_{\mathbb{H}})$ be a generic vector space endowed with a Hilbert structure, where $\oplus_{\mathbb{H}}$, $\odot_{\mathbb{H}}$ and $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ are the addition, scalar multiplication and inner product operations, respectively. For ease of notation, let also $\ominus_{\mathbb{H}} := \oplus_{\mathbb{H}}(-1) \odot_{\mathbb{H}}$ be the subtraction operation. We further assume that $\mathbb{H}$ is separable. This last assumption is key from a practical point of view: it guarantees indeed the possibility of computing an accurate finite dimensional approximation of Hotelling's T^2 statistic in real applications [21].

An $\mathbb{H}$-valued random variable χ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is a mapping $\chi : \Omega \to \mathbb{H}$ such that $\langle \chi, h \rangle_{\mathbb{H}}$ is measurable for all $h \in \mathbb{H}$. In order to define the expected value and covariance operator of χ, we need a proper definition of integration on $\mathbb{H}$ w.r.t. the probability measure $\mathbb{P}$. In the following, we will use the Bochner integral ([15]) which can be viewed as the natural extension of the traditional Lebesgue integration on $\mathbb{R}^p$ to the integration on Hilbert spaces. Its proper general definition for any measurable function taking value in a separable Hilbert space is outside of the scope of this work. We hereby rather give a simpler definition of

the Bochner integral for $\mathbb{H}$-valued random variables. Let χ be an $\mathbb{H}$-valued random variable on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The Bochner integral of χ with respect to the probability measure $\mathbb{P}$ is the unique element $\mathbb{E}_{\mathbb{H}}[\chi]$ of $\mathbb{H}$ such that

$$\langle \mathbb{E}_{\mathbb{H}}[\chi], f \rangle_{\mathbb{H}} = \mathbb{E}_{\mathbb{R}}\left[\langle \chi, f \rangle_{\mathbb{H}}\right], \quad \text{for all } f \in \mathbb{H}, \tag{28.5}$$

with $\mathbb{E}_{\mathbb{R}}$ being the expectation operator for $\mathbb{R}$-valued random variables.

Thanks to the Bochner integral [15], we can now introduce the definition of mean and covariance for an $\mathbb{H}$-valued random variable χ. It can be shown that these entities exist and are finite if and only if $\mathbb{E}_{\mathbb{R}}[\|\chi\|_{\mathbb{H}}^2] < \infty$. Under this assumption, we define the mean m of χ as its Bochner integral, i.e. $m := \mathbb{E}_{\mathbb{H}}[\chi] \in \mathbb{H}$ and we define the covariance operator $\mathscr{K}$ of χ as the Bochner integral of the Hilbert-Schmidt operator $(\chi \ominus_{\mathbb{H}} m) \otimes_{\mathbb{H}} (\chi \ominus_{\mathbb{H}} m)$, i.e. $\mathscr{K} := \mathbb{E}_{\mathscr{B}_{\mathrm{HS}}(\mathbb{H})}[(\chi \ominus_{\mathbb{H}} m) \otimes_{\mathbb{H}} (\chi \ominus_{\mathbb{H}} m)] \in \mathscr{B}_{\mathrm{HS}}(\mathbb{H})$, where $\mathscr{B}_{\mathrm{HS}}(\mathbb{H})$ is the space of Hilbert-Schmidt operators on $\mathbb{H}$. Note that the covariance operator is perfectly well defined because the space of Hilbert-Schmidt operators $\mathscr{B}_{\mathrm{HS}}(\mathbb{H})$ on a separable Hilbert space $\mathbb{H}$ is itself a separable Hilbert space.

Assume now that we collected a sample $\chi_1, \ldots, \chi_n$ of n i.i.d. $\mathbb{H}$-valued random variables with mean m and covariance operator $\mathscr{K}$, such that $\mathbb{E}_{\mathbb{R}}[\|\chi_i\|_{\mathbb{H}}^2] < \infty$, for all $i = 1, \ldots, n$. Using standard algebraic operations, it can be shown that the following estimators m_n and $\mathscr{K}_n$ are unbiased estimators for the mean and covariance operator respectively:

$$m_n := \frac{1}{n} \odot_{\mathbb{H}} \bigoplus_{i=1}^{n} \chi_i \quad \text{and} \quad \mathscr{K}_n := \frac{1}{n-1} \odot_{\mathbb{H}} \bigoplus_{i=1}^{n} (\chi_i \ominus_{\mathbb{H}} m_n) \otimes_{\mathbb{H}} (\chi_i \ominus_{\mathbb{H}} m_n). \tag{28.6}$$

While m_n and $\mathscr{K}_n$ naturally play the roles of $\mathbf{m}_n$ and Σ_n in eq. (28.2) respectively, looking back at eq. (28.4), the definition of Hotelling's T^2 statistic requires the introduction of a novel operator that captures the squared error loss of the sample mean m_n in estimating the true mean m for replacing D_n. We therefore define the *sample mean squared-error loss operator* $\mathscr{D}_n$ as:

$$\mathscr{D}_n := (m_n \ominus_{\mathbb{H}} m) \otimes_{\mathbb{H}} (m_n \ominus_{\mathbb{H}} m).$$

The sample mean squared-error loss operator is a $\mathscr{B}_{\mathrm{HS}}(\mathbb{H})$-valued random variable.

We now have all the components required for defining Hotelling's T^2 statistic in a generic separable Hilbert space. In details, given a sample $\chi_1, \ldots, \chi_n$ of n i.i.d. $\mathbb{H}$-valued random variables with mean m and covariance operator $\mathscr{K}$, Hotelling's T^2 statistic is defined as:

$$T^2 := n \max_{f \in \mathrm{Im}(\mathscr{K}_n) \setminus \{0\}} \frac{\langle f, \mathscr{D}_n f \rangle_{\mathbb{H}}}{\langle f, \mathscr{K}_n f \rangle_{\mathbb{H}}}, \tag{28.7}$$

where

$$\mathscr{D}_n := (m_n \ominus_{\mathbb{H}} m) \otimes_{\mathbb{H}} (m_n \ominus_{\mathbb{H}} m),$$
$$\mathscr{K}_n := \frac{1}{n-1} \odot_{\mathbb{H}} \bigoplus_{i=1}^{n} (\chi_i \ominus_{\mathbb{H}} m_n) \otimes_{\mathbb{H}} (\chi_i \ominus_{\mathbb{H}} m_n).$$

All the theoretical properties pertaining to Hotelling's T^2 statistic as defined above are discussed in details in [21]. Equation (28.7), contrary to common belief, demonstrates that Hotelling's T^2 statistic is actually well defined in separable Hilbert spaces of infinite dimension. Famous examples of such Hilbert spaces are showcased in [21]. An important application that follows from Equation 28.7 is the development of new null hypothesis significance testing (NHST) procedures for making inference on the mean element in Hilbert spaces. In [21], one- and two-population NHST procedures based on Hotelling's T^2 statistic in separable Hilbert spaces are proposed.

Acknowledgements This work was partly supported by the 2014 Polimi International Fellowship program.

References

[1] Bongiorno, E. G., Salinelli, E., Goia, A., Vieu, P.: Contributions in infinite-dimensional statistics and related topics. Società Editrice Esculapio (2014)
[2] Cardot, H., Prchal, L., Sarda, P.: No effect and lack-of-fit permutation tests for functional regression. Comput. Statist. **22**, 371–390 (2007)
[3] Corain, L., Melas, V. B., Pepelyshev, A., Salmaso, L.: New insights on permutation approach for hypothesis testing on functional data. Adv. Data Anal. Classif. **8**, 339–356 (2014)
[4] Cox, D. D., Lee, J. S.: Pointwise testing with functional data using the Westfall–Young randomization method. Biometrika. **95**, 621–634 (2008)
[5] Cuesta-Albertos, J. A., Febrero-Bande, M.: A simple multiway ANOVA for functional data. TEST. **19**, 537–557 (2010)
[6] Cuevas, A., Febrero, M., Fraiman, R.: An ANOVA test for functional data. Comput. Statist. Data Anal. **47**, 111–122 (2004)
[7] Egozcue, J. J., Díaz-Barrero, J. L., Pawlowsky-Glahn, V.: Hilbert space of probability density functions based on Aitchison geometry. Acta Math. Sin. **22**, 1175–1182 (2006)
[8] Fan, J., Lin, S. K.: Test of significance when data are curves. J. Amer. Statist. Assoc. **93**, 1007–1021 (1998)
[9] Ferraty, F., Vieu, P.: Nonparametric functional data analysis: theory and practice. Springer, New York, (2006)
[10] Galeano, P., Esdras, J., Lillo, R.E.: The Mahalanobis distance for functional data with applications to classification. Technometrics. **57**(2), 281–291 (2015)

[11] Goia, A., Vieu, P.: An introduction to recent advances in high/infinite dimensional statistics. J. Multivariate Anal. **146**, 1–6, Special Issue on Statistical Models and Methods for High or Infinite Dimensional Spaces (2016)
[12] Hall, P., Van Keilegom, I.: Two-sample tests in functional data analysis starting from discrete data. Statist. Sinica **17**, 1511 (2007)
[13] Horváth, L., Kokoszka, P.: Inference for functional data with applications. Springer, New York (2012)
[14] Hron, K., Menafoglio, A., Templ, M., Hrzová, K., Filzmoser, P.: Simplicial principal component analysis for density functions in Bayes spaces. Comput. Statist. Data Anal. **94**, 330–350 (2016)
[15] Hsing, T., Eubank, R.: Theoretical foundations of functional data analysis, with an introduction to linear operators. John Wiley & Sons (2015)
[16] Marron, J. S., Alonso, A. M.: Overview of object oriented data analysis. Biom. J. (2014)
[17] Menafoglio, A., Petris, G.: Kriging for Hilbert-space valued random fields: The operatorial point of view. J. Multivariate Anal. **146**, 84–94 (2016)
[18] Menafoglio, A., Secchi, P.: Statistical analysis of complex and spatially dependent data: A review of object oriented spatial statistics. European J. Oper. Res. (2016)
[19] Pesarin, F., Salmaso, L.: Permutation tests for complex data: theory, applications and software. John Wiley & Sons Inc, Chichester (2010)
[20] Pini, A., Vantini, S.: The interval testing procedure: A general framework for inference in functional data analysis. Biometrics (2016)
[21] Pini, A., Stamm, A., Vantini, S.: Hotellings T^2 statistic and test in separable Hilbert spaces. MOX Technical report, Politecnico di Milano (2016+)
[22] Ramsay, J. O., Silverman, B. W.: Functional data analysis. Springer, New York (2005)
[23] Secchi, P., Stamm, A., Vantini, S.: Inference for the mean of large p small n data: a finite-sample high-dimensional generalization of Hotellings theorem. Electron. J. Stat. **7**, 2005–2031 (2013)
[24] Shen, Q., Faraway, J.: An F test for linear models with functional responses. Statist. Sinica 1239–1257 (2004)
[25] Spitzner, D. J., Marron, J. S., Essick, G. K.: Mixed-model functional ANOVA for studying human tactile perception. J. Amer. Statist. Assoc. **98**, 263–272 (2003)
[26] Van Den Boogaart, K. G., Egozcue, J. J., Pawlowsky-Glahn, V.: Bayes Hilbert Spaces. Aust. N. Z. J. Stat. **56**, 171–194 (2014)

Chapter 29
Confidence and prediction intervals in semi-functional partial linear regression

Paula Raña, Germán Aneiros, Philippe Vieu and Juan Vilar

Abstract Semi-functional partial linear regression model allows to deal with a nonparametric and a linear component within the functional regression. Naïve and wild bootstrap procedures are proposed to approximate the distribution of the estimators for each component in the model, and their asymptotic validities are obtained in the context of dependence data, under α-mixing conditions. Based on that bootstrap procedures, confidence intervals can be obtained for each component in the model, which can be also extended to deal with prediction intervals and prediction densities.

29.1 Introduction

Functional Data Analysis (FDA) is a relative recent field in Statistics, which has been increasing its presence over the last years. In spite of its novelty, most statistical techniques have been generalized to the functional context. One can find in [5] a nice monograph on FDA, from the nonparametric point of view, and in [7], focused on inference, or, more recently, in [8].

Paula Raña (✉)
Departamento de Matemáticas, Universidade da Coruña, A Coruña, Spain,
e-mail: paula.rana@udc.es
Germán Aneiros
Departamento de Matemáticas, Universidade da Coruña, A Coruña, Spain,
e-mail: ganeiros@udc.es
Philippe Vieu
Institut de Mathématiques, Université Paul Sabatier, Toulouse, France
e-mail: philippe.vieu@math.univ-toulouse.fr
Juan Vilar
Departamento de Matemáticas, Universidade da Coruña, A Coruña, Spain,
e-mail: juan.vilar@udc.es

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_29

Our interest is focused on functional regression. Specifically, this contribution deals with the Semi-Functional Partial Linear Regression model (SFPLR), which has been introduced in [1] in a context of independent data and then, it was analysed in [2] under dependence conditions.

The objective of this study is to build confidence and prediction intervals for the SFPLR model when dealing with functional time series, that is, for functional data under dependence conditions. For that purpose, bootstrap procedures will be used in order to approximate the distribution of the estimators in the SFPLR model, and their asymptotically validity will be obtained. This kind of bootstrap procedures have been proposed for functional linear regression in [6] and for Functional Nonparametric Regression model (FNP) in [4], in a context of dependence data, which has been extended in [11] to deal with functional time series. However, in the context of partial linear models, one can only find in [9] and [12] some approximations to this kind of procedures, but in a context of scalar models and not in functional data.

The rest of the paper is organized as follows. Section 29.2 presents the SFPLR model and their estimators, while the bootstrap procedures, together with their asymptotic validity, are developed in Section 29.3. Section 29.4 contains the procedures to build the confidence intervals, which can be extended to build also prediction intervals (some general ideas are given in Section 29.5). Finally, Section 29.6 includes some conclusions.

29.2 Semi-functional partial linear regression

The SFPLR model considered in this paper can be written as

$$Y = \mathbf{X}^T \beta + m(\chi) + \varepsilon, \tag{29.1}$$

where $\beta = (\beta_1, \ldots, \beta_p)^T$ is a vector of unknown real parameters and m is an unknown smooth real-valued operator. The explanatory random variables $\mathbf{X}$ and χ are valued in $\mathbb{R}^p$ and some infinite-dimensional space $\mathscr{H}$ endowed with a semi-metric $d(\cdot,\cdot)$, respectively, while the random error ε verifies $\mathbb{E}(\varepsilon|(\mathbf{X},\chi)) = 0$ and $\mathbb{E}(\varepsilon^2|(\mathbf{X},\chi)) < \infty$.

Throughout the paper we assume that the sample we have at hand,

$$\mathscr{S} = \{(\mathbf{X}_1, \chi_1, Y_1), \ldots, (\mathbf{X}_n, \chi_n, Y_n)\},$$

comes from a sequence of random vectors that are α-mixing and identically distributed as $(\mathbf{X}, \chi, Y)$, while the corresponding random errors $\{\varepsilon_i\}_{i=1}^n$ are independent. Note that these assumptions especially fit to the particular case where $Y_i = G(\chi_{i+1})$ ($G(\cdot)$ denotes a known operator), which plays a main role in prediction of functional time series (see, for instance, [3] for the SFPLR context, or [11] for the pure nonparametric setting).

29.2.1 Estimators

Let us denote

$$\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^T,\ \mathbf{Y} = (Y_1, \ldots, Y_n)^T \text{ and } \mathbf{W}_h = (w_h(\chi_i, \chi_j))_{1 \le i,j \le n}$$

with

$$w_h(\cdot, \chi_j) = \frac{K(d(\cdot, \chi_j)/h)}{\sum_{i=1}^n K(d(\cdot, \chi_i)/h)},$$

where $K(\cdot)$ is a kernel function and $h > 0$ is a smoothing parameter. In addition, for any $(n \times q)$ matrix $\mathbf{A}$ $(q \ge 1)$, we denote

$$\widetilde{\mathbf{A}}_h = (\mathbf{I} - \mathbf{W}_h)\mathbf{A}.$$

In this contribution, the vector parameter β and the functional operator $m(\cdot)$ in the SFPLR model (29.1) will be estimated by means of

$$\widehat{\beta}_b = (\widetilde{\mathbf{X}}_b^T \widetilde{\mathbf{X}}_b)^{-1} \widetilde{\mathbf{X}}_b^T \widetilde{\mathbf{Y}}_b \quad \text{and} \quad \widehat{m}_h(\cdot) = \sum_{j=1}^n w_h(\cdot, \chi_j)(Y_j - \mathbf{X}_j^T \widehat{\beta}_h),$$

respectively. Both estimators are based on least squares estimation and kernel smoothing, and they were studied, under α-mixing conditions, in [2].

29.3 Bootstrap in the SFPL model

Two bootstrap procedures are proposed in this section. The first one, called "naïve bootstrap", is designed for homoscedastic models ($\sigma_\varepsilon^2(\mathbf{X}, \chi) := \mathbb{E}(\varepsilon^2 | (\mathbf{X}, \chi)) = \sigma_\varepsilon^2$). The second one, called "wild bootstrap", works under heterocedasticity.

The bootstrap procedures follow the following algorithms:

Naïve bootstrap.
Step 1: Construct the residuals $\widehat{\varepsilon}_{i,b} = Y_i - \mathbf{X}_i^T \widehat{\beta}_b - \widehat{m}_b(\chi_i)$, $i = 1, \ldots, n$.
Step 2: Draw n i.i.d. random variables, $\varepsilon_1^*, \ldots, \varepsilon_n^*$, from the empirical distribution function of $(\widehat{\varepsilon}_{1,b} - \overline{\widehat{\varepsilon}}_b, \ldots, \widehat{\varepsilon}_{n,b} - \overline{\widehat{\varepsilon}}_b)$, where $\overline{\widehat{\varepsilon}}_b = n^{-1} \sum_{i=1}^n \widehat{\varepsilon}_{i,b}$.
Step 3: Obtain $Y_i^* = \mathbf{X}_i^T \widehat{\beta}_b + \widehat{m}_b(\chi_i) + \varepsilon_i^*$, $i = 1, \ldots, n$.
Step 4: Define

$$\widehat{\beta}_b^* = (\widetilde{\mathbf{X}}_b^T \widetilde{\mathbf{X}}_b)^{-1} \widetilde{\mathbf{X}}_b^T \widetilde{\mathbf{Y}}_b^* \text{ and } \widehat{m}_{hb}^*(\cdot) = \sum_{j=1}^n w_h(\cdot, \chi_j)(Y_j^* - \mathbf{X}_j^T \widehat{\beta}_b^*),$$

Wild bootstrap.

Change Step 2 in the naïve bootstrap: define $\varepsilon_i^* = \widehat{\varepsilon}_{i,b} V_i$ $(i = 1, \ldots, n)$, where $V_1, \ldots, V_n$ are i.i.d. random variables that are independent of the sample $\mathscr{S}$ and that satisfy $\mathbb{E}(V_1) = 0$, $\mathbb{E}(V_1^2) = 1$ and $\mathbb{E}(V_1^r) \leq \mathrm{C} < \infty$ for some $r > 6$. Maintain the other three steps.

29.3.1 Validity of the bootstrap procedures

Several assumptions must be imposed in order to warranty the validity of the proposed bootstrap procedures. Those assumptions have been proposed previously in [2], related to the asymptotic distribution of the estimators in the SFPLR model, and in [11], related to the bootstrap procedures applied to the FNP model. One can see in those references, and also in [10], a detailed analysis and justification of the considered assumptions, which are common in that context.

Let $P^{\mathscr{S}}$ denote the probability conditionally on the sample $\mathscr{S}$ while we use $\mathbf{a}$ to denote a constant vector in $\mathbb{R}^p$. Assume that χ is valued in some given compact subset $\mathscr{C}$ of $\mathscr{H}$ and denote by χ_0 a fixed element of $\mathscr{C}$. Denote $F_{\chi_0}(l) = P(\chi \in B(\chi_0, l))$ for $l > 0$, where $B(\chi_0, l) = \{\gamma \in \mathscr{H} \text{ such that } d(\gamma, \chi_0) \leq l\}$.

Theorem 29.1. *Under the assumptions indicated above, for the wild bootstrap, one has:*

$$\sup_{y \in \mathbb{R}} \left| P^{\mathscr{S}} \left(\sqrt{n} \mathbf{a}^T (\widehat{\beta}_b^* - \widehat{\beta}_b) \leq y \right) - P \left(\sqrt{n} \mathbf{a}^T (\widehat{\beta}_b - \beta) \leq y \right) \right| \to_P 0. \tag{29.2}$$

In addition, if the model is homoscedastic, the result holds for the naïve bootstrap.

Theorem 29.2. *Under the assumptions indicated above, for the wild bootstrap, one has:*

$$\sup_{y \in \mathbb{R}} \left| P^{\mathscr{S}} \left(\sqrt{n F_{\chi_0}(h)} (\widehat{m}_{hb}^*(\chi_0) - \widehat{m}_b(\chi_0)) \leq y \right) - P \left(\sqrt{n F_{\chi_0}(h)} (\widehat{m}_h(\chi_0) - m(\chi_0)) \leq y \right) \right| \to_P 0. \tag{29.3}$$

In addition, if the model is homoscedastic, the result holds for the naïve bootstrap.

Theorems 29.1 and 29.2 establish the validity of the bootstrap procedures proposed for the linear and the nonparametric component of the SFPLR model, respectively. From the practical point of view, those results allow to perform statistical inference on β and m using the bootstrap distribution. This can be useful, for instance, to build confidence and prediction intervals, as in the following sections.

29.4 Building the confidence intervals

Confidence intervals can be computed for each component in the SFPLR model. The procedure to build the corresponding bootstrap confidence intervals is presented in this section. Because of its generality, the wild bootstrap procedure will be considered.

Given the SFPLR model

$$Y = \mathbf{X}^T\beta + m(\chi) + \varepsilon, \tag{29.4}$$

an observation (X, χ_0) from $(\mathbf{X}, \chi)$ and a sample

$$\mathscr{S} = \{(\mathbf{X}_1, \chi_1, Y_1), \ldots, (\mathbf{X}_n, \chi_n, Y_n)\}$$

from model (29.4) satisfying the assumptions in this paper, bootstrap $(1-\alpha)$-confidence intervals for β, $m(\chi_0)$ and the regression function $r(X, \chi_0) := X^T\beta + m(\chi_0)$ can be built as follows:

- Bootstrap confidence interval for β (assuming, for simplicity, that $\beta \in \mathbb{R}$):

$$I^{\beta,*}_{1-\alpha} = (\widehat{\beta}_b + q^{\beta,*}_{\alpha/2}, \widehat{\beta}_b + q^{\beta,*}_{1-\alpha/2}).$$

- Bootstrap confidence interval for $m(\chi_0)$:

$$I^{m,*}_{\chi_0,1-\alpha} = (\widehat{m}_h(\chi_0) + q^{m,*}_{\alpha/2}(\chi_0), \widehat{m}_h(\chi_0) + q^{m,*}_{1-\alpha/2}(\chi_0)).$$

- Bootstrap confidence interval for $r(X, \chi_0)$:

$$I^{r,*}_{X,\chi_0,1-\alpha} = (X^T\widehat{\beta}_b + \widehat{m}_h(\chi_0) + q^{r,*}_{\alpha/2}(X, \chi_0), X^T\widehat{\beta}_b + \widehat{m}_h(\chi_0) + q^{r,*}_{1-\alpha/2}(X, \chi_0)),$$

where the quantiles $(q^{u,*}_p)$ involved in the confidence intervals above can be computed in the following way:

Bootstrap quantiles: $q^{\beta,*}_p$, $q^{m,*}_p(\chi_0)$ and $q^{r,*}_p(X, \chi_0)$.

1. Repeat B times the wild bootstrap algorithm (see Section 29.3) over $\mathscr{S}$ by using iid random variables V_i drawn from the mixture of the two Dirac distributions

$$0.1(5+\sqrt{5})\delta_{(1-\sqrt{5})/2} + 0.1(5-\sqrt{5})\delta_{(1+\sqrt{5})/2},$$

   giving the B bootstrap estimates $\{\widehat{\beta}^{*,t}_b\}^B_{t=1}$ and $\{\widehat{m}^{*,t}_{hb}(\chi_0)\}^B_{t=1}$.
2. Compute the sets of bootstrap errors (BE):
   a. BE.$\beta = \{\widehat{\beta}_b - \widehat{\beta}^{*,t}_b\}^B_{t=1}$.
   b. BE.$m = \{\widehat{m}_b(\chi_0) - \widehat{m}^{*,t}_{hb}(\chi_0)\}^B_{t=1}$.
   c. BE.$r = \{X^T(\widehat{\beta}_b - \widehat{\beta}^{*,t}_b) + (\widehat{m}_b(\chi_0) - \widehat{m}^{*,t}_{hb}(\chi_0))\}^B_{t=1}$.

3. Compute the bootstrap quantiles:

   a. $q_p^{\beta,*}$: is the quantile of order p of BE.β.
   b. $q_p^{m,*}(\chi_0)$: is the quantile of order p of BE.m.
   c. $q_p^{r,*}(X,\chi_0)$: is the quantile of order p of BE.r.

Finally, the estimates $\widehat{\beta}_b$, $\widehat{m}_b(\chi_0)$ and $\widehat{m}_h(\chi_0)$ used in the algorithms above are obtained from the sample $\mathscr{S}$.

This procedure can be seen as an extension of the algorithms presented in [11], from the FNP to the SFPLR model, and taking into account the three possibilities of considering β, $m(\chi_0)$ or $r(X,\chi_0)$, separately.

29.5 Building the prediction intervals

From the practical point of view, if the objective relies on functional time series forecasting (see, for instance [3] applied to electricity data), one may be interested in prediction intervals instead of confidence intervals. It is of main importance to distinguish between confidence intervals and prediction intervals and so, their differences will be analysed in the following paragraphs.

Consider the SFPLR model in (29.1), together with the estimators for β and $m(\cdot)$:

$$\widehat{\beta}_b = (\widetilde{\mathbf{X}}_b^T \widetilde{\mathbf{X}}_b)^{-1} \widetilde{\mathbf{X}}_b^T \widetilde{\mathbf{Y}}_b \quad \text{and} \quad \widehat{m}_h(\cdot) = \sum_{j=1}^{n} w_h(\cdot,\chi_j)(Y_j - \mathbf{X}_j^T \widehat{\beta}_h).$$

In the case of the confidence intervals, one may deal with the conditional expectation

$$\mathbb{E}(Y|\{X,\chi_0\}) = X^T\beta + m(\chi_0).$$

Then, one can obtain a confidence interval (level $1-\alpha$) for this conditional expectation based on the estimators $\widehat{\beta}_b$ and $\widehat{m}_h(\cdot)$. This interval will be built as:

$$(X^T\widehat{\beta}_b + \widehat{m}_h(\chi_0) + q_{\alpha/2}(X,\chi_0), X^T\widehat{\beta}_b + \widehat{m}_h(\chi_0) + q_{1-\alpha/2}(X,\chi_0)),$$

where $q_{\alpha/2}(X,\chi_0)$ and $q_{1-\alpha/2}(X,\chi_0)$ are the quantiles from the distribution of $X^T(\beta - \widehat{\beta}_b) + (m(\chi_0) - \widehat{m}_h(\chi_0))$, which can be approximated, as in the previous section, by the bootstrap procedures.

This confidence interval is devoted to cover the true value of the regression operator, $X^T\beta + m(\chi_0)$, and it deals only with the variability due to the estimations.

The procedure to obtain a prediction interval is similar, but changing its "philosophy". Prediction intervals are devoted to cover the response and not the regression

operator and so, they include, not only the variability due to the estimation, but also to the error in the model. That is, instead of working with the expectation of Y conditionally on $\{X, \chi_0\}$, one deals with the distribution of Y conditionally on $\{X, \chi_0\}$. One may consider that, in this context, the predictor for $Y/\{X, \chi_0\}$ is $X^T\widehat{\beta}_b + \widehat{m}_h(\chi_0)$.

The procedure to build a prediction interval follows the same idea as the confidence interval, but now this variability due to the model error has to be added. The starting point is that now, one looks for an interval (a,b) such that

$$P(Y|\{X, \chi_0\} \in (a,b)) = 1 - \alpha,$$

which allows to obtain an interval for the conditional response of the regression model, being of main importance from the practical usefulness of those procedures if one deals with real data prediction problems.

The extension of the bootstrap procedures, developed to build confidence intervals in Section 29.4, to deal now with the prediction intervals allows to estimate also prediction densities, which are a nice complement for any prediction study.

29.6 Concluding remarks

This study proposes two bootstrap procedures to approximate the asymptotic distribution of the estimators in both components (linear and nonparametric) of the SFPLR model, considering scalar response and functional predictor and in which one adds linear effect of scalar covariates. The validity of these two procedures has been established in the setting of dependent data, assuming α–mixing conditions on the sample, and can be also applied to the setting of independent data as a particular case.

By means of these two bootstrap procedures one can construct pointwise confidence and prediction intervals for the SFPLR model. Procedures to build confidence intervals have been developed in detail, while the extension to build not only prediction intervals but also prediction densities has been proposed.

Acknowledgements This research is partially supported by Grant MTM2014-52876-R, from Spanish Ministerio de Economía y Competitividad and Grants CN2012/130 and ED431C 2016-015, from Xunta de Galicia.

References

[1] Aneiros-Pérez, G., Vieu, P.: Semi-functional partial linear regression. Statist. Probab. Lett. **76**, 1102–1110 (2006)
[2] Aneiros-Pérez, G., Vieu, P.: Nonparametric time series prediction: A semi-functional partial linear modeling. J. Multivariate Anal. **99**, 834–857 (2008)
[3] Aneiros, G., Vilar, J., Raña, P.: Short-term forecast of daily curves of electricity demand and price. Electr. Power. Energy. Syst. **80**, 96–108 (2016)
[4] Ferraty, F., Van Keilegom, I., Vieu, P.: On the validity of the bootstrap in non-parametric functional regression. Scand. J. Statist. **37**, 286–306 (2010)
[5] Ferraty, F., Vieu, P.: Nonparametric functional data analysis. Springer-Verlag, New York (2006)
[6] González-Manteiga, W., Martínez-Calvo, A.: Bootstrap in functional linear regression. J. Statist. Plann. Inference **141**, 453–461 (2011)
[7] Horváth, L., Kokoszka, P.: Inference for functional data with applications. Springer Series in Statistics. Springer, New York (2012)
[8] Hsing, T., Eubank, R.: Theoretical foundations of functional data analysis, with an introduction to linear operators. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester (2015)
[9] Liang, H., Härdle, W., Sommerfeld, V.: Bootstrap approximation in a partially linear regression model. J. Statist. Plann. Inference **91**, 413–426 (2000)
[10] Raña, P.: Pointwise forecast, confidence and prediction intervals in electricity demand and price. Doctoral Thesis. Universidade da Coruña (2016)
[11] Raña, P., Aneiros, G., Vilar, J., Vieu, P.: Bootstrap confidence intervals in functional nonparametric regression under dependence. Electron. J. Statist. **10(2)**, 1973–1999 (2016)
[12] You, J., Zhou, X.: Bootstrap approximation of a semiparametric partially linear model with autoregressive errors. Statist. Sinica **15**, 117–133 (2005)

Chapter 30
Linear causality in the sense of Granger with stationary functional time series

Matthieu Saumard

Abstract In this paper, we investigate the causality in the sense of Granger for functional time series. The concept of causality for functional time series is defined and a statistical procedure of testing the hypothesis of non-causality is proposed. This procedure is based on a test of equality of covariance operators for dependent processes.

30.1 Introduction

Functional data analysis has become an important field of modern statistics, and now there exists an abundant literature over this topic. Nevertheless, the case of dependent observations between the functional objects is less studied. We extend the classical notion of causality in the sense of Ganger (see [3]) to the case of functional time series. This extension is important and useful to the statistical community. To the best of our knowledge, there exists only one article dealing with the causality in the sense of Granger adapted to the functional context, see [1]. In their proceedings, they define one definition of the causality for signals. A test procedure is proposed but a theoretical background is needed. We propose another procedure of testing the causality in the sense of Granger adapted to the functional time series context.

Since the pioneering work of Wiener (1956) [8] and Granger (1969) [3], an abundant literature is now available on causality of classical time series. There is nowadays different tools to analyse dependency for functional data: Mixing, linear process Bosq [2], Hörmann and Kokoszka [4] have proposed a definition of dependency for functional data that generalize the m-dependency, they call it $L^p - m-$approximable

Matthieu Saumard (✉)
Conservatoire National des Arts et Métiers, 292 Rue Saint-Martin, 75003 Paris, e-mail: matthieu.saumard@lecnam.net

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_30

sequences.

In Section 2, we introduce the definition of causality for functional time series and an example in the autoregressive functional processes. In Section 3, we recall one procedure of testing the equality of covariance operators for dependent functional processes and one procedure of estimation of the autocorrelation operator. With these procedures, we conclude this section by a procedure of testing the causality in the model defined in Section 2.

30.2 Linear causality with two functional time series

30.2.1 Definition

Let H be a separable Hilbert space, with norm $\|\cdot\|$ and the inner product associated $\langle\cdot,\cdot\rangle$. In the sequel, for every functional time series $\{Z_t\}_{t\in\mathbb{N}}$ valued in H, we make the assumptions that $\mathbb{E}\left[\|Z_t\|^2\right] < +\infty$ and $\mathbb{E}[Z_t] = 0$, $\forall t \in \mathbb{N}$, moreover we assume that $\{Z_t\}_{t\in\mathbb{N}}$ is stationary. Define the operator of covariance Γ_Z of the stationary time series Z_t by:

$$\Gamma_Z(U) = \mathbb{E}\left[\langle Z,U\rangle Z\right], \quad \forall U \in H, \tag{30.1}$$

where we omit the index t of time, due to stationarity. It is well known that this operator is a self-adjoint, positive, nuclear operator.

Let X_t and Y_t be two stationary functional time series valued in H. Let U_t be the information accumulated since time $t-1$, and $U_t - X_t$ denote all this information without the series X_t. Define the predictive error series by

$$\varepsilon_t(Y|U) = Y_t - P_t(Y|U),$$

where $P_t(Y|U)$ is the best linear predictor of Y_t using the information U_t. Then, we have the following definition.

Definition 30.1 (Causality). We say that X is causing Y if $\Gamma_{\varepsilon(Y|U-X)} - \Gamma_{\varepsilon(Y|U)}$ is a positive definite operator.

30.2.2 An example

Let $\mathbb{X}_t = (Y_t, X_t)'$ be an autoregressive process valued in $\mathscr{H} := L^2[0,1] \times L^2[0,1]$ associated with (ρ, ε) where ρ is a linear operator of $\mathscr{H}$ and ε is a $\mathscr{H}$-strong white noise. For more information on the definition of autoregressive process valued in separable Hilbert space, see [2]. We have

$$\mathbb{X}_t = \rho(\mathbb{X}_{t-1}) + \varepsilon_t. \tag{30.2}$$

We can rewrite the model as

$$\begin{cases} Y_t = \rho_{11}(Y_{t-1}) + \rho_{12}(X_{t-1}) + \varepsilon_{1t}, \\ X_t = \rho_{21}(Y_{t-1}) + \rho_{22}(X_{t-1}) + \varepsilon_{2t}, \end{cases}$$

with $\rho_{11} := \left(\rho \restriction_{L^2[0,1]\times 0}\right)^{(1)}$, $\rho_{12} := \left(\rho \restriction_{0\times L^2[0,1]}\right)^{(1)}$ and the same idea for ρ_{21} and ρ_{22}. According to the definition 30.1, we say that (X_t) does not linear cause (Y_t) if and only if $\rho_{12} = 0$. In fact, if $\rho_{12} = 0$, the two covariance operators of the definition are equals and then the operator can not be positive definite.

30.2.3 Extensions

Other functional time series models can be studied, like autoregressive processes of order p, functional moving average processes of order q and FARMA(p,q) processes. Let $\mathbb{X}_t = (Y_t, X_t)'$ be a FARMA(p,q) process valued in $\mathscr{H}$ such that

$$\sum_{j=0}^{p} \rho_j(\mathbb{X}_{t-j}) = \sum_{j=0}^{q} \theta_j(\varepsilon_{t-j}), \tag{30.3}$$

for $\rho_0 = \theta_0 = I$, ρ_j, θ_j bounded linear operators and ε a strong white noise. Here, we do not discuss about the existence, uniqueness of a stationary solution, we refer to [2]. Now, we can rewrite the model as

$$\begin{cases} \sum_{j=0}^{p} \rho_{11,j}(Y_{t-j}) + \sum_{j=1}^{p} \rho_{12,j}(X_{t-j}) = \sum_{j=0}^{q} \theta_{11,j}(\varepsilon_{1,t-j}) + \sum_{j=1}^{q} \theta_{12,j}(\varepsilon_{2,t-j}), \\ \sum_{j=1}^{p} \rho_{21,j}(Y_{t-j}) + \sum_{j=0}^{p} \rho_{22,j}(X_{t-j}) = \sum_{j=1}^{q} \theta_{21,j}(\varepsilon_{1,t-j}) + \sum_{j=0}^{q} \theta_{22,j}(\varepsilon_{2,t-j}), \end{cases}$$

where we use the same idea as in the previous example. According to the definition, we say that (X_t) does not linear cause (Y_t) if and only if $\forall j$, $\rho_{12,j} = 0$ and $\forall j$, $\theta_{12,j} = 0$.

30.3 Testing linear causality in the autoregressive model

30.3.1 Weak dependence

The notion of dependence, we will use, is defined in the article of Hörmann and Kokoszka (2010) [4]. Denoting by L_H^p, the space of H valued random variables X such that $\nu_p(X) = (\mathbb{E}[\|X\|^p])^{1/p} < +\infty$, we can state the definition, see [4] for the properties of this notion.

Definition 30.2 ($L^p - m-$**approximable sequence, see [4]).** A sequence $\{X_t\} \in L_H^p$ is called

$L^p - m-$approximable if each X_t admits the representation,

$$X_t = f(\varepsilon_t, \varepsilon_{t-1}, \ldots),$$

where the ε_t are S valued i.i.d. sequence and f is a mesurable function of S^∞ to H. Defining

$$X_t^{(m)} = f(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-m+1}, \varepsilon'_{t-m}, \varepsilon'_{t-m-1}, \ldots),$$

where $\{\varepsilon'_t\}$ is an independent copy of $\{\varepsilon_t\}$, we assume that

$$\sum_{m=1}^{\infty} \nu_p(X_m - X_m^{(m)}) < \infty.$$

30.3.2 Test of equality of the covariance operators

Zhang and Shao (2015) [9] recently have proposed a test procedure to compare the covariance operators of two mean zero stationary functional time series. We refer to this article for the general procedure. Let us make a summary of this procedure in the special case of equal number of the two samples. The null hypothesis is

$$H_0 : \Gamma_X = \Gamma_Y, \tag{30.4}$$

against the alternative

$$H_1 : \Gamma_X \neq \Gamma_Y,$$

where X and Y are two mean zero stationary functional time series, $\{X_i\}_{i=1}^N$, $\{Y_i\}_{i=1}^N$. Define $\{\hat{\lambda}_{XY}^j\}$ and $\{\hat{\phi}_{XY}^j\}$ the eigenvalues and eigenfunctions of

$$\hat{\Gamma}_{XY} = \frac{1}{2N}\left(\sum_{i=1}^{N} X_i \otimes X_i + Y_i \otimes Y_i\right).$$

Let $\hat{\Gamma}_{X,m} = 1/m \sum_{i=1}^m X_i \otimes X_i$. Let $\{\hat{\lambda}_{X,m}^j\}$ and $\{\hat{\phi}_{X,m}^j\}$ be the eigenvalues and eigenfunctions of $\hat{\Gamma}_{X,m}$. Similar quantities are defined for the second sample. Let K be a fixed user-chosen number and

$$c_k^{i,j} = \langle (\hat{\Gamma}_{X,\lfloor k/2 \rfloor} - \hat{\Gamma}_{Y,\lfloor k/2 \rfloor})(\hat{\phi}_{XY}^i), \hat{\phi}_{XY}^j \rangle, \quad 2 \leq k \leq 2N,\ 1 \leq i, j \leq K.$$

Denote by $\hat{\alpha}_k = \text{vech}(C_k)$, with $C_k = (c_k^{i,j})_i^K, j = 1$. To take the dependence into account, they introduce a self-normalized matrix:

$$V = \frac{1}{4N^2} \sum_{k=1}^{2N} k^2 (\hat{\alpha}_k - \hat{\alpha}_{2N})(\hat{\alpha}_k - \hat{\alpha}_{2N})'.$$

The test statistic is then

$$G = 2N\hat{\alpha}'_{2N}V^{-1}\hat{\alpha}_{2N}.$$

Define $B_q(r)$ as a q-dimensional vector of independent Brownian motion, $W_q = B'_q(1)J_q^{-1}B_q(1)$, where

$$J_q = \int_0^1 (B_q(r) - rB_q(1))(B_q(r) - rB_q(1))'dr.$$

The critical values of W_q have been tabulated by Lobato (2001) [5]. [9] have demonstrated that the test statistic converges to $W_{K(K+1)/2}$ under the null and diverges under the alternative hypothesis.

Assumption 1
- *Assume* $\{(X_i(t), Y_i(t))\}_{i=1}^{+\infty} \in L^4_{H\times H}$ *is an* $L^4 - m-$*approximable sequence.*
- *Assume* $\lambda_X^1 > \lambda_X^2 > > \lambda_X^{m_0+1}$ *and* $\lambda_Y^1 > \lambda_Y^2 > > \lambda_Y^{m_0+1}$, *for some positive integer* $m_0 \geq 2$.

Zhang and Shao (2015) [9] consider the local alternative $H_{1,N} : \Gamma_X - \Gamma_Y = L\Gamma/\sqrt{N}$, with Γ being a Hilbert-Schmidt operator, and L a constant different of 0. Let $\widetilde{\phi}^i$ be the eigenfunctions of the operator $(\Gamma_X + \Gamma_Y)/2$, whose empirical counterpart are the $\hat{\phi}^i_{XY}$.

Assumption 2 *Define* $\Delta = (\langle \Gamma\widetilde{\phi}^i, \widetilde{\phi}^j\rangle)_{i,j=1}^K$. *Assume* $vech(\Delta) \neq 0 \in \mathbb{R}^{K(K+1)/2}$.

Theorem 30.1. *Suppose Assumptions* 1 *and* 2 *hold with* $m_0 \geq K$. *Further assume that the asymptotic covariance matrice is positive definite, see Lemma 6.3 of Zhang and Shao (2015) [9]. Then under* H_0, $G \longrightarrow W_{K(K+1)/2}$ *and under* $H_{1,N}$, $\lim_{|L|\to+\infty} lim_{N\to+\infty} G = +\infty$.

30.3.3 Estimation of ρ

First of all, we must have an estimate of ρ. We follow the general description of Bosq (2000) [2]. For this task, define the covariance operator

$$\Gamma(x) = \mathbb{E}[\langle \mathbb{X}_0, x\rangle \mathbb{X}_0], \quad x \in \mathscr{H}.$$

If $\mathbb{X}_1, \ldots, \mathbb{X}_n$ are observed, a natural estimator of Γ is the empirical covariance operator, defined as

$$\hat{\Gamma}(x) = \frac{1}{n}\sum_{t=1}^{n} \langle \mathbb{X}_t, x\rangle \mathbb{X}_t, x \in \mathscr{H}.$$

Let us denote the eigenelements of Γ and $\hat{\Gamma}$ by $(\lambda_k, \phi_k)_{k=1,\ldots,\infty}$ and $(\widehat{\lambda}_k, \widehat{\phi}_k)_{k=1,\ldots,n}$ respectively. Now, define the lag-1 empirical operator by

$$\hat{D}(x) = \frac{1}{n-1}\sum_{t=1}^{n-1} \langle \mathbb{X}_t, x\rangle \mathbb{X}_{t+1}.$$

Noting π^{p_n} the orthogonal projection on the space spanned by the $(\widehat{\phi}_1,\dots,\widehat{\phi}_{p_n})$, the estimator of ρ can be

$$\widehat{\rho} = \pi^{p_n}\hat{D}\hat{\Gamma}^{-1}\pi^{p_n}.$$

We can write this estimator in another form

$$\widehat{\rho}(x) = \sum_{l=1}^{p_n} \widehat{\rho}_l(x)\widehat{\phi}_l,$$

with

$$\widehat{\rho}_l(x) = \frac{1}{n-1}\sum_{t=1}^{n-1}\sum_{j=1}^{p_n} \widehat{\lambda}_j^{-1}\langle x,\widehat{\phi}_j\rangle\langle \mathbb{X}_t,\widehat{\phi}_j\rangle\langle \mathbb{X}_{t+1},\widehat{\phi}_l\rangle.$$

30.3.4 The test procedure

In the previous sections, we have recalled one procedure of testing the equality of covariance operators and one procedure of estimation of the parameters. We are now ready to state our result.

Testing procedure for the autoregressive model of order 1:

1. (Data) We dispose of $(X_1,\dots,X_N)$ and $(Y_1,\dots,Y_N)$
2. (Estimation of the parameters) We perform the estimation of ρ according to the previous section. We separate the two different models, one estimation of ρ without the Y_i (nested model) and one which include them (pooled model).
3. (Estimation of the errors) ε_1 is then estimated in the different models for $t = 2,\dots,N$, by

$$\begin{cases} \widehat{\varepsilon}_t^1 = Y_t - \widehat{\rho}_{11}(Y_{t-1}) + \widehat{\rho}_{12}(X_{t-1}), \text{ (pooled model)},\\ \widehat{\varepsilon}_t^2 = Y_t - \widehat{\rho}_{11}(Y_{t-1}), \text{ (nested model)}. \end{cases}$$

4. (Test the equality of operators based on the errors) We perform the proposed test of Zhang and Shao on the estimated errors.
   There are straigthforward extensions of this procedure to test the non-causality in the other models mentioned in section 2. Possible alternatives to the test of Zhang and Shao would be to perform a global F-test like in [7], or to test the effect of the variable X on errors like [6]. Unfortunately these tests are studied in the independent case.

30.4 Conclusion

The notion of linear causality has been studied and we have proposed a definition and a statistical procedure to test this concept in a special case i.e the autoregressive model of order 1. An empirical analysis of the numerical performance of the test will be the object of another article.

References

[1] Amblard, P.O., Michel, O.: Causalité de granger pour des signaux à valeurs fonctionnelles. In: Actes Colloque GRESTI 2013 (2013)
[2] Bosq, D.: Linear processes in function spaces, *Lecture Notes in Statistics*, vol. 149. Springer-Verlag, New York (2000).
[3] Granger, C.W.J.: Investigating causal relations by econometric models and cross-spectral methods. Econometrica **37**(3), pp. 424–438 (1969).
[4] Hörmann, S., Kokoszka, P.: Weakly dependent functional data. Ann. Statist. **38**(3), 1845–1884 (2010).
[5] Lobato, I.N.: Testing that a dependent process is uncorrelated. J. Amer. Statist. Assoc. **96**(455), 1066–1076 (2001)
[6] Patilea, V., Sánchez-Sellero, C., Saumard, M.: Testing the predictor effect on a functional response. J. Amer. Statist. Assoc. **111**(516), 1684–1695 (2016)
[7] Shen, Q., Faraway, J.: An f test for linear models with functional responses. Statist. Sinica pp. 1239–1257 (2004)
[8] Wiener, N.: The theory of prediction. Modern mathematics for engineers **1**, 125–139 (1956)
[9] Zhang, X., Shao, X.: Two sample inference for the second-order property of temporally dependent functional data. Bernoulli **21**(2), 909–929 (2015)

Chapter 31
Grouped multivariate functional time series method: An application to mortality forecasting

Han Lin Shang and Yang Yang

Abstract Age-specific mortality rates are often disaggregated by different attributes, such as sex and state. Forecasting age-specific mortality rates at the sub-national levels may not add up to the forecasts at the national level. Further, the independent forecasts may not utilize correlation among sub-populations to improve forecast accuracy. Using Japanese mortality data, we extend the grouped univariate functional time series methods to grouped multivariate functional time series forecasting methods.

31.1 Introduction

It is common to have multivariate functional time series that are ordered by a grouped structure. For instance, we consider age-specific mortality rates observed annually as an example of a functional time series, where the continuum is the age variable. These age-specific mortality rates can be observed at the national level, and can be disaggregated by various attributes such as sex or state. In practice, forecasts are often required for national mortality, as well as sub-national mortality disaggregated by different attributes. When a functional time series forecasting method is applied to each series, the sum of the forecasts will not generally add up to the forecasts obtained by applying the method to the national data. This is known as the issue of forecast reconciliation (see, e.g., [5]).

In the multivariate functional time series literature, there has been few forecasting methods that take account of aggregation constraints. Two noticeable exceptions are [4] and [3], where the optimal combination method is proposed to reconcile point forecasts of age-specific mortality, and potentially improve the point forecast accuracy at different levels of a hierarchy. In these two works, each series in a group is independently forecast by a univariate functional time series (FTS) forecasting

Han Lin Shang (✉) and Yang Yang
Research School of Finance, Actuarial Studies and Statistics, Australian National University, Acton ACT 2601, Australia; e-mail: hanlin.shang@anu.edu.au; yang.yang@anu.edu.au

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_31

method and then the forecasts are reconciled with respect to the group structure. To further improve forecast accuracy, we consider a multivariate functional time series (MFTS) forecasting method to jointly model the correlation among sub-populations and forecast multivariate functional time series at each level.

Using the national and sub-national Japanese age-specific mortality rates from 1975 to 2014 in Sect. 31.2, we introduce the multivariate functional principal component regression in Sect. 31.3 for producing point forecasts, and revisit grouped functional time series forecasting methods in Sect. 31.4. We evaluate the point forecast accuracy between independent and grouped functional time series forecasting methods in Sect. 31.5.

31.2 Japanese data set

We study the Japanese age-specific mortality rates from 1975 to 2014, obtained from the [2]. We consider age groups from 0 to 99 in single years of age, while the last age group contains all ages at and over 100. The data structure is described as follows: At the top level, we have total age-specific mortality rates for Japan. We can split these total mortality rates by sex, by region, or by prefecture. Japan consists of 8 regions which contain 47 prefectures as lower level administrative divisions. The most disaggregated data arise when we consider the mortality rates for each combination of prefecture and sex, giving a total of $47 \times 2 = 94$ series. In total, across all levels of disaggregation, there are 1+2+8+16+47+94=168 series.

31.3 Multivariate functional principal component analysis

31.3.1 Nonparametric smoothing

Let $y_t^j(x_i)$ be the log central mortality rates observed at the beginning of each year for year $t = 1, 2, \ldots, n$ at observed ages $x_1, x_2, \ldots, x_p$ where x is a continuous variable, p denotes the number of ages, and superscript j represents either male or female in the case of two populations.

We assume there is an underlying continuous and smooth function $f_t^j(x)$ observed at discrete data points with errors. That is

$$y_t^j(x_i) = f_t^j(x_i) + \delta_t^j(x_i)\varepsilon_{t,i}^j, \tag{31.1}$$

where x_i represents the center of each age group for $i = 1, \ldots, p$, $\varepsilon_{t,i}^j$ is an independent and identically distributed (i.i.d.) standard normal random variable for each age in year t, and $\delta_t^j(x_i)$ measures the variability in mortality at each age in year t for the j^{th} population. Jointly, $\delta_t^j(x_i)\varepsilon_{t,i}^j$ represents the smoothing error.

31.3.2 Multivariate functional principal component analysis

Given our application has the same domain, we consider data where each observation consists of $\omega \geq 2$ functions, $[f^{(1)}(x),\ldots,f^{(\omega)}(x)]^\top \in R^\omega$.

The multivariate functional time series are combined in a vector with $\boldsymbol{f}(x) = \left[f^{(1)}(x),\ldots,f^{(\omega)}(x)\right] \in R^\omega$. We assume that $\mu(x) := \left\{\mathrm{E}\left[f^{(1)}(x)\right],\ldots,\mathrm{E}\left[f^{(\omega)}(x)\right]\right\} = 0$. For $s,t \in \Gamma$, the covariance function is defined with elements

$$C_{lj}(s,t) := \mathrm{Cov}\left[f^{(l)}(s),f^{(j)}(t)\right] = \mathrm{E}\left\{\left[f^{(l)}(s)-\mu^{(l)}(s)\right]\left[f^{(j)}(t)-\mu^{(j)}(t)\right]\right\}, \tag{31.2}$$

where $\mu^{(l)}(t)$ denotes the mean function for l^{th} sub-population. By assuming f is a continuous and square-integrable covariance function, the function K induces the kernel operator given by

$$(Kf)^{(j)}(t) = \sum_{l=1}^{\omega}\int_\Gamma C_{lj}(s,t)f^{(l)}(s)ds, \qquad f^{(j)} \in L^2(\Gamma). \tag{31.3}$$

Via Mercer's lemma, there exists an orthonormal sequence (ϕ_k) of continuous function in $L^2(\Gamma)$ and a non-increasing sequence λ_k of positive number, such that

$$C_{lj}(s,t) = \sum_{k=1}^{\infty}\lambda_k\phi_k^l(s)\phi_k^j(t). \tag{31.4}$$

By the separability of Hilbert spaces, the Karhunen-Loève expansion of a stochastic process $f^{(l)}(x)$ can be expressed as

$$f_t^{(l)}(x) = \sum_{k=1}^{K_l}\beta_{t,k}^{(l)}\phi_k^{(l)}(x), \tag{31.5}$$

where K_l denotes the number of retained functional principal components in the l^{th} population, and its matrix formulation is

$$\boldsymbol{f}_t(x) = \boldsymbol{\Phi}(x)\boldsymbol{\beta}_t^\top, \tag{31.6}$$

where $\boldsymbol{f}_t(x) = \left[f_t^{(1)}(x),\ldots,f_t^{(\omega)}(x)\right]$, $\boldsymbol{\beta}_t = \left(\beta_{t1}^{(1)},\ldots,\beta_{tK_1}^{(1)},\ \beta_{t1}^{(2)},\ldots,\beta_{tK_2}^{(2)},\ \ldots,\ \beta_{t1}^{(\omega)},\ldots,\beta_{tK_\omega}^{(\omega)}\right)$ being the vector of the basis expansion coefficients, and

$$\boldsymbol{\Phi}(x) = \begin{pmatrix} \phi_1^{(1)}(x) & \cdots & \phi_{K_1}^{(1)}(x) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \phi_1^{(2)}(x) & \cdots & \phi_{K_2}^{(2)}(x) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \phi_1^{(\omega)}(x) & \cdots & \phi_{K_\omega}^{(\omega)}(x) \end{pmatrix}_{\omega \times \sum_{\ell=1}^{\omega} K_\ell}. \tag{31.7}$$

31.3.3 Functional principal component regression

By using MFPCA, a time series of smoothed function $\boldsymbol{f}^{(l)}(x) = \{f_1(x), \ldots, f_n(x)\}$ is decomposed into orthogonal functional principal components and their associated principal component scores, given by

$$f_t^{(l)}(x) = \widehat{\mu}^{(l)}(x) + \sum_{k=1}^{\infty} \widehat{\beta}_{t,k}^{(l)} \widehat{\phi}_k^{(l)}(x) = \widehat{\mu}^{(l)}(x) + \sum_{k=1}^{K_l} \widehat{\beta}_{t,k}^{(l)} \widehat{\phi}_k^{(l)}(x) + e_t^{(l)}(x), \tag{31.8}$$

where $\widehat{\mu}^{(l)}(x) = \frac{1}{n}\sum_{t=1}^{n} f_t^{(l)}(x)$ denotes the estimated mean function for the l^{th} sub-population; $\left\{\widehat{\phi}_1^{(l)}(x), \ldots, \widehat{\phi}_{K_l}^{(l)}(x)\right\}$ is a set of the first K_l estimated functional principal components; $\widehat{\boldsymbol{\beta}}_1^{(l)} = \left(\widehat{\beta}_{1,1}^{(l)}, \ldots, \widehat{\beta}_{1,n}^{(l)}\right)^{\top}$ and $\{\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_{K_l}\}$ denote a set of estimated principal component scores, $\widehat{\beta}_k^{(l)} \sim N(0, \widehat{\lambda}_k^{(l)})$ where $\widehat{\lambda}_k^{(l)}$ is the k^{th} estimated eigenvalue of the covariance function for the l^{th} sub-population in (31.2); $e_t^{(l)}(x)$ denotes the model truncation error function with mean zero and finite variance.

Conditioning on the estimated functional principal components and historical data, the forecast curves can be obtained by $\widehat{f}_{n+h|n}^{(l)}(x) = \widehat{\mu}^{(l)}(x) + \sum_{k=1}^{K_l} \widehat{\beta}_{n+h|n,k}^{(l)} \widehat{\phi}_k^{(l)}(x)$, where $\widehat{\beta}_{n+h|n,k}^{(l)}$ represents the forecast k^{th} principal component scores, which can be obtained from the autoregressive integrated moving average model.

31.4 Grouped functional time series forecasting techniques

31.4.1 Notation

The Japanese data follow a multi-level geographical hierarchy coupled with a sex grouping variable. The geographical hierarchy is displayed in Figure 31.1, where Japan is split into eight regions which can be further split into 47 prefectures.

The data can also be split by sex. Thus, each of the nodes in the geographical hierarchy can be separated into one male series and one female series. We denote a particular disaggregated series using the notation $G \times S$ meaning the geographical

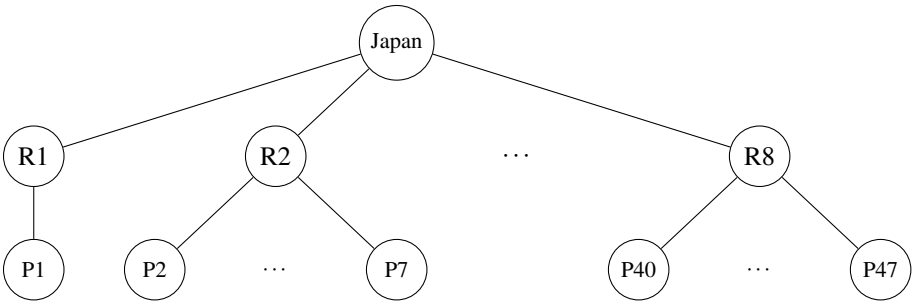
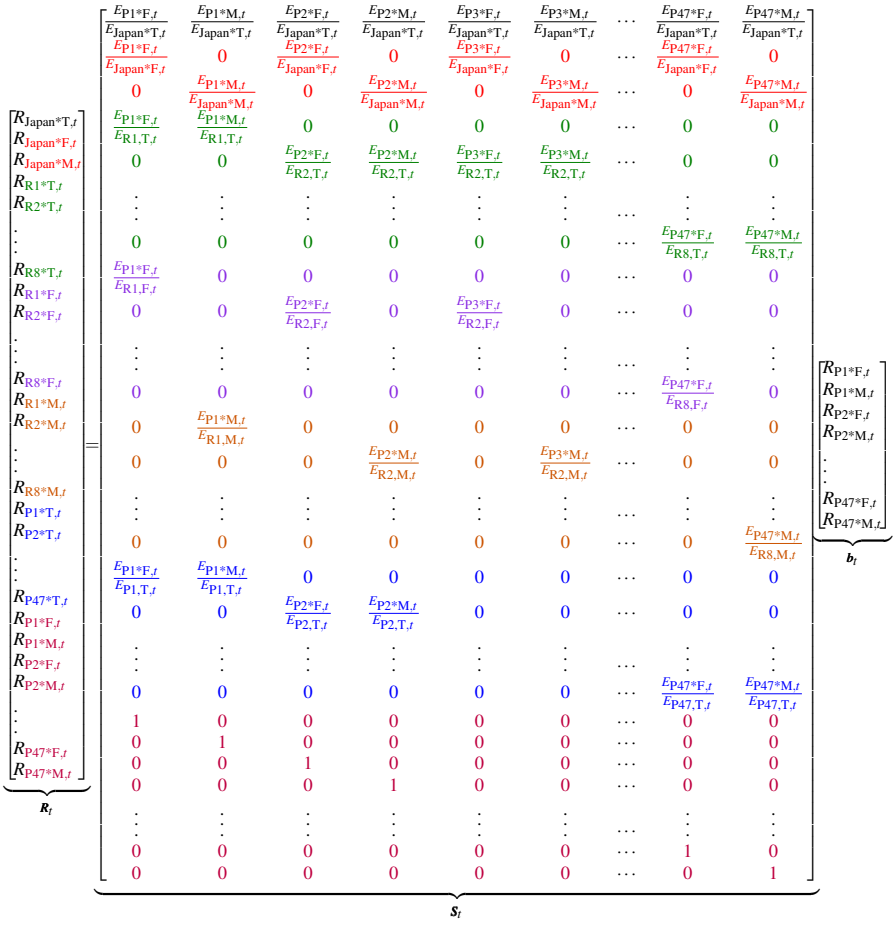


Fig. 31.1: The Japanese geographical hierarchy tree diagram, with eight regions and 47 prefectures.

area G and the sex S, where G can take the values shown in any node of Figure 31.1 and S can take values M (males), F (females) or T (total). For instance, R1$\times$ F denotes females in Region 1; P1$\times$ T denotes females and males in Prefecture 1; Japan$\times$ M denotes males in Japan; and so on.

Let $E_{G\times S,t}(z)$ denote the exposure-at-risk for series $G\times S$ in year t and age z, and let $D_{G\times S,t}(z)$ be the number of deaths for series $G\times S$ in year t and age z. Then age-specific mortality rate is given by $R_{G\times S,t}(z)=D_{G\times S,t}(z)/E_{G\times S,t}(z)$. To simplify expressions, we will drop the age argument (z). For a given age, we can express



or $\boldsymbol{R}_t = \boldsymbol{S}_t \boldsymbol{b}_t$, where $\boldsymbol{R}_t$ is a vector containing all series at all levels of disaggregation, $\boldsymbol{b}_t$ is a vector of the most disaggregated series, and $\boldsymbol{S}_t$ shows how the two are related.

31.4.2 Optimal combination method

[1] proposed a method in which independent (also known as base) forecasts for all aggregated and disaggregated series are computed independently, and then the resulting forecasts are reconciled so that they satisfy the aggregation constraints. As the base forecasts are independently generated, these forecasts will not generally be "aggregate consistent". The optimal combination methods combine the base forecasts through linear regression by generating a series of revised forecasts that are not only close to the base forecasts but also aggregate consistently within the group. The method is derived by writing the base forecasts as the response variable of the linear regression

$$\widehat{\boldsymbol{R}}_{n+h} = \boldsymbol{S}_{n+h}\boldsymbol{\beta}_{n+h} + \boldsymbol{\varepsilon}_{n+h}, \tag{31.9}$$

where $\widehat{\boldsymbol{R}}_{n+h}$ is a matrix of h-step-ahead base forecasts for all series, stacked in the same order as for raw data; $\boldsymbol{\beta}_{n+h}$ is the unknown mean of the forecast distributions of the most disaggregated series; $\boldsymbol{\varepsilon}_{n+h}$ represents the reconciliation errors.

To estimate the regression coefficients, [1] proposed a weighted least squares solution which we adapt to our problem as follows:

$$\widehat{\boldsymbol{\beta}}_{n+h} = \left(\boldsymbol{S}_{n+h}^{\top}\boldsymbol{W}^{-1}\boldsymbol{S}_{n+h}\right)^{-1}\boldsymbol{S}_{n+h}^{\top}\boldsymbol{W}^{-1}\widehat{\boldsymbol{R}}_{n+h}, \tag{31.10}$$

where $\boldsymbol{W}$ is assumed to be a diagonal matrix containing the one-step-ahead forecast variances for each series. The revised forecasts can be given by

$$\overline{\boldsymbol{R}}_{n+h} = \boldsymbol{S}_{n+h}\widehat{\boldsymbol{\beta}}_{n+h} = \boldsymbol{S}_{n+h}\left(\boldsymbol{S}_{n+h}^{\top}\boldsymbol{S}_{n+h}\right)^{-1}\boldsymbol{S}_{n+h}^{\top}\widehat{\boldsymbol{R}}_{n+h}. \tag{31.11}$$

31.5 Results of the point forecasts

31.5.1 Forecast error criterion

Using the first 25 observations from 1975 to 1999 in the Japanese age-specific mortality rates, we produce one- to 15-step-ahead point forecasts. Through an expanding window approach, we re-estimate the parameters using the first 26 observations from 1975 to 2000. Forecasts from the estimated models are then produced for one to 14-step-ahead. We iterate this process by increasing the sample size by one year until reaching the end of data period in 2014. This process produces 15 one-step-ahead forecasts, 14 two-step-ahead forecasts, …, and one 15-step-ahead forecast.

To evaluate the point forecast accuracy, we use the root mean squared forecast error (RMSFE). For each series k, it can be written as

$$\text{RMSFE}_k(h) = \sqrt{\frac{1}{101\times(16-h)}\sum_{\varsigma=h}^{15}\sum_{j=1}^{101}\left[f_{n+\varsigma}^{k}(x_j) - \widehat{f}_{n+\varsigma}^{k}(x_j)\right]^2}, \tag{31.12}$$

where $f_{n+\varsigma}^{k}(x_j)$ denotes the holdout sample for the j^{th} age and ς^{th} curve of the forecasting period in the k^{th} series, while $\widehat{f}_{n+\varsigma}^{k}(x_j)$ denotes the point forecasts.

By averaging $\text{RMSFE}_k(h)$ across the number of series within each level of disaggregation, we obtain

$$\text{RMSFE}(h) = \frac{1}{m_k}\sum_{k=1}^{m_k}\text{RMSFE}_k(h), \qquad k = 1,\ldots,K, \tag{31.13}$$

where m_k denotes the number of series at the k^{th} level of disaggregation.

31.5.2 Point forecast accuracy

Averaging over all series at each level of the Japanese data hierarchy, Table 31.1 presents RMSFE(h) values between the univariate and multivariate functional time series forecasting methods, without forecast reconciliation.

Table 31.1: RMSFE ($\times 100$) in the holdout sample between the univariate and multivariate functional time series forecasting methods applied to the Japanese age-specific mortality rates, without forecast reconciliation. The bold entries highlight the method that performs best for each level of the hierarchy and each forecast horizon, as well as summary statistics.

	Total		Sex		Region		Region + Sex		Prefecture		Prefecture + Sex	
h	FTS	MFTS	FTS	MFTS	FTS	MFTS	FTS	MFTS	FTS	MFTS	FTS	MFTS
1	0.29	0.29	0.38	**0.35**	0.57	**0.53**	0.77	0.77	0.89	**0.80**	1.45	**1.40**
2	0.42	0.42	0.50	**0.39**	0.65	**0.55**	0.82	**0.80**	0.90	**0.82**	1.48	**1.38**
3	0.52	0.52	0.63	**0.42**	0.71	**0.56**	0.86	**0.83**	0.92	**0.84**	1.67	**1.37**
4	0.60	0.60	0.73	**0.45**	0.76	**0.59**	0.90	**0.80**	0.89	**0.82**	1.34	1.34
5	0.68	0.68	0.81	**0.49**	0.80	**0.63**	0.94	**0.83**	0.90	**0.85**	**1.32**	1.34
6	0.75	0.75	0.88	**0.52**	0.83	**0.67**	0.98	**0.85**	0.90	**0.87**	**1.32**	1.34
7	0.83	0.83	0.96	**0.62**	0.91	**0.76**	1.05	**0.90**	0.95	0.95	**1.35**	1.36
8	0.90	0.90	1.04	**0.72**	0.96	**0.85**	1.13	**0.96**	**0.98**	1.00	**1.37**	1.38
9	0.96	0.96	1.12	**0.76**	1.01	**0.91**	1.20	**0.99**	**1.01**	1.04	1.39	**1.37**
10	1.00	1.00	1.18	**0.81**	1.07	**1.01**	1.26	**1.00**	**1.04**	1.12	1.43	**1.39**
11	1.01	1.01	1.20	**0.77**	1.06	**1.01**	1.27	**0.95**	**1.03**	1.09	1.41	**1.32**
12	0.97	0.97	1.16	**0.72**	1.01	**0.99**	1.28	**0.91**	**1.01**	1.07	1.40	**1.28**
13	0.91	0.91	1.06	**0.65**	**0.94**	0.97	1.18	**0.86**	**0.97**	1.05	1.34	**1.28**
14	0.78	0.78	0.83	**0.61**	**0.84**	0.89	1.07	**0.87**	0.93	0.93	**1.27**	1.33
15	0.61	0.61	**0.57**	0.58	0.87	**0.75**	1.07	**0.95**	0.86	**0.80**	**1.21**	1.45
Mean	0.75	0.75	0.87	**0.59**	0.87	**0.78**	1.05	**0.88**	0.95	**0.94**	1.38	**1.36**

In Table 31.2, we demonstrate the superior forecast accuracy obtained from the optimal combination method between the univariate and multivariate functional time series forecasting methods.

Table 31.2: RMSFE ($\times 100$) in the holdout sample between the univariate and multivariate functional time series forecasting methods applied to the Japanese age-specific mortality rates, with the optimal combination method. The bold entries highlight the method that performs best for each level of the hierarchy and each forecast horizon, as well as summary statistics.

	Total		Sex		Region		Region + Sex		Prefecture		Prefecture + Sex	
h	FTS	MFTS	FTS	MFTS	FTS	MFTS	FTS	MFTS	FTS	MFTS	FTS	MFTS
1	0.36	**0.30**	0.41	**0.37**	0.53	**0.44**	0.69	**0.63**	0.88	**0.80**	1.40	**1.34**
2	0.40	**0.34**	0.46	**0.41**	0.59	**0.47**	0.74	**0.65**	0.93	**0.80**	1.41	**1.31**
3	0.47	**0.37**	0.54	**0.44**	0.77	**0.49**	0.88	**0.66**	1.12	**0.81**	1.55	**1.31**
4	0.49	**0.39**	0.58	**0.47**	0.58	**0.50**	0.77	**0.68**	0.84	**0.79**	1.34	**1.28**
5	0.50	**0.40**	0.63	**0.49**	0.59	**0.51**	0.81	**0.70**	0.84	**0.80**	1.34	**1.29**
6	0.52	**0.40**	0.69	**0.51**	0.60	**0.51**	0.85	**0.71**	0.84	**0.79**	1.36	**1.29**
7	0.58	**0.47**	0.77	**0.59**	0.66	**0.57**	0.92	**0.77**	0.88	**0.84**	1.39	**1.31**
8	0.61	**0.51**	0.84	**0.64**	0.68	**0.59**	0.97	**0.81**	0.90	**0.86**	1.43	**1.33**
9	0.65	**0.51**	0.92	**0.67**	0.74	**0.61**	1.07	**0.85**	0.92	**0.86**	1.47	**1.33**
10	0.68	**0.53**	0.97	**0.69**	0.77	**0.64**	1.12	**0.88**	0.94	**0.89**	1.51	**1.35**
11	0.65	**0.43**	1.00	**0.62**	0.75	**0.56**	1.15	**0.82**	0.92	**0.83**	1.52	**1.29**
12	0.60	**0.37**	0.97	**0.52**	0.70	**0.49**	1.13	**0.72**	0.91	**0.81**	1.51	**1.24**
13	0.50	**0.38**	0.89	**0.44**	0.63	**0.53**	1.07	**0.68**	0.85	**0.81**	1.45	**1.18**
14	**0.35**	0.54	0.69	**0.39**	**0.52**	0.63	0.89	**0.60**	**0.80**	0.88	1.36	**1.11**
15	**0.28**	0.91	**0.45**	0.58	**0.52**	0.99	0.78	**0.75**	**0.77**	1.20	1.25	**1.19**
Mean	0.51	**0.46**	0.72	**0.52**	0.64	**0.57**	0.92	**0.73**	0.89	**0.85**	1.42	**1.28**

References

[1] Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., Shang, H. L.: Optimal combination forecasts for hierarchical time series. Comput. Statist. Data Anal., **55**, 2579-2589 (2011)

[2] Japanese Mortality Database: National Institute of Population and Social Security Research (2016) Available via http://www.ipss.go.jp/p-toukei/JMD/index-en.html. Cited 18 July 2016

[3] Shang, H. L., Haberman, S.: Grouped multivariate and functional time series forecasting: An application to annuity pricing. Living to 100: Insight on the challenges and opportunities of longevity, Orlando, USA (2017)

[4] Shang, H. L., Hyndman, R. J.: Grouped functional time series forecasting: An application to age-specific mortality rates. J. Comput. Graph. Statist., forthcoming (2017)

[5] Stone, R., Champernowne, D. G., Meade, J. E.: The pension of national income estimates. Rev. Econ. Stud., 9(2), 111-125 (1942)

Chapter 32
Tests for separability in nonparametric covariance operators of random surfaces

Shahin Tavakoli, Davide Pigoli and John A. D. Aston

Abstract We consider the problem of testing for separability in nonparametric covariance operators of multidimensional functional data is considered. We cast the problem in a tensor product of Hilbert space framework, where the role of the partial trace operator is emphasized, and the tests proposed are computationally tractable. An applications to acoustic phonetic data is also presented.

32.1 Introduction

Many applications involve hypersurface data, data that is both functional data [as in functional data analysis (FDA), see e.g. 11] and multidimensional, such as images from medical devices such as MRI [8] or PET [17], spectrograms derived from audio signals [10, and as in the application we consider in Section 32.3] or geolocalized data [see, e.g., 13]. In these problems, the number of available observations (hypersurfaces) is often small relative to the high-dimensional nature of the individual observation, and not usually large enough to estimate a full multivariate covariance function.

A way of circumventing this problem is by assuming that the covariance is *separable*. This assumption greatly simplifies both the estimation and the computational

Shahin Tavakoli (✉)
Statistical Laboratory, Department of Pure Maths and Mathematical Statistics, University of Cambridge, Cambridge, UK, and PRDA, Darwin College, Cambridge e-mail: s.tavakoli@statslab.cam.ac.uk

Davide Pigoli
Statistical Laboratory, Department of Pure Maths and Mathematical Statistics, University of Cambridge, Cambridge, UK, e-mail: d.pigoli@statslab.cam.ac.uk

John A. D. Aston
Statistical Laboratory, Department of Pure Maths and Mathematical Statistics, University of Cambridge, Cambridge, UK, e-mail: j.aston@statslab.cam.ac.uk

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_32

cost in dealing with multivariate covariance functions, and allows for a positive definite covariance to be specified.

In the context of space-time data $X(s,t)$, for instance, where $s \in [-S,S]^d$, $S > 0$, denotes the location in space, and $t \in [0,T]$, $T > 0$, is the time index, the assumption of separability translates into

$$c(s,t,s',t') = c_1(s,s')c_2(t,t'), \quad s,s' \in [-S,S]^d; t,t' \in [0,T], \tag{32.1}$$

where c, c_1, and c_2, are respectively the full covariance function, the space covariance function and the time covariance function. In words, this means that the full covariance function factorises as a product of the spatial covariance function with the time covariance function. The separability assumption [see e.g. 5] is a simplification of the covariance structure of the process, and makes estimation of the covariance easier. It can be argued that it introduces some bias in the estimator, for the benefit of reduced variance.

The separability assumption can be very useful in practice, and is indeed often implicitly made in many higher dimensional applications when using isotropic smoothing [17, 8]. However, very little has been done to develop tools to assess its validity on a case by case basis. In the classical multivariate setup, some tests for the separability assumption are available [see 14, and references therein].

When the covariance is being non-parametrically specified, as will be the case in this paper, estimation of the full covariance is at best computationally complex with large estimation errors, and in many cases simply computationally infeasible. Indeed, we highlight that, while the focus of this work is on checking the viability of a separable structure for the covariance, this is done without any parametric assumption on the form of $c_1(s,s')$ and $c_2(t,t')$, thus allowing for the maximum flexibility. The methods we develop here are aimed to applications typical of FDA, where replicates from the underlying random process are available [see 16, for a recent review of functional data analysis]. This is different from the spatio-temporal setting, where usually only one realization of the process is observed. See also [3] for a likelihood based approach to testing separability for functional data, [2] for an example of how separability can be used for modeling function-valued stochastic processes, and also [6] which illustrates how the concept of separability can be applied to matrix-variate data arising from functional MRI imaging data.

In this paper—which is an extended abstract of [1]—we propose a class of tests to verify if the data at hand are in agreement with a separability assumption. Our test does not require the estimation of the full covariance structure, but only the estimation of the separable structure (32.1), thus avoiding both the computational issues and the diminished accuracy involved in the former. To do this, we rely on a strategy from Functional Data Analysis, which consists in projecting the observations onto a carefully chosen low-dimensional subspace. The key fact for the success of our approach is that, under the null hypothesis, it is possible to determine this subspace using only the marginal covariance functions. In the end, the proposed test checks the separability in the chosen subspace, which will often be the focus of following analyses. In Section 32.2, we define covariance separability in an abstract Hilbert

space framework, introduce tests for separability and study their asymptotic behavior. In Section 32.3 we show an application to acoustic phonetic data.

32.2 Separable Covariances: definitions, estimators and asymptotic results

We start by introducing some definitions and notation about operators in a Hilbert space [see e.g. 12]. Let H be a real separable Hilbert space (that is, a Hilbert space with a countable orthonormal basis), whose inner product and norm are denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. The space of bounded (linear) operators on H is denoted by $\mathscr{S}_\infty(H)$. The space of Hilbert–Schmidt operators on H is denoted by $\mathscr{S}_2(H)$, and is a Hilbert space with the inner-product $\langle S, T \rangle_{\mathscr{S}_2} = \sum_{i \geq 1} \langle Se_i, Te_i \rangle$ and induced norm $|||\cdot|||_2$, where $(e_i)_{i \geq 1} \subset H$ is an orthonormal basis of H. The space of trace-class operator on H is denoted by $\mathscr{S}_1(H)$, and consists of all compact operators T with finite trace-norm, i.e. $|||T|||_1 = \sum_{n \geq 1} s_n(T) < \infty$, where $s_n(T) \geq 0$ denotes the n-th singular value of T. For any trace-class operator $T \in \mathscr{S}_1(H)$, we define its trace by $\mathrm{Tr}(T) = \sum_{i \geq 1} \langle Te_i, e_i \rangle$, where $(e_i)_{i \geq 1} \subset H$ is an orthonormal basis, and the sum is independent of the choice of the orthonormal basis.

If H_1, H_2 are real separable Hilbert spaces, we denote by $H = H_1 \otimes H_2$ their tensor product Hilbert space, which is obtained by the completion of all finite sums $\sum_{i,j=1}^N u_i \otimes v_j$, $u_i \in H_1, v_j \in H_2$, under the inner-product $\langle u \otimes v, z \otimes w \rangle = \langle u, z \rangle \langle v, w \rangle$ for $u, z \in H_1, z, w \in H_2$ [see e.g. 7]. If $C_1 \in \mathscr{S}_\infty(H_1)$, $C_2 \in \mathscr{S}_\infty(H_2)$, we denote by $C_1 \mathbin{\widetilde{\otimes}} C_2$ the unique linear operator on $H_1 \otimes H_2$ satisfying

$$\left(C_1 \mathbin{\widetilde{\otimes}} C_2\right)(u \otimes v) = C_1 u \otimes C_2 v, \quad \text{for all } u \in H_1, v \in H_2. \tag{32.2}$$

It is a bounded operator on H. Furthermore, if $C_1 \in \mathscr{S}_1(H_1)$ and $C_2 \in \mathscr{S}_1(H_2)$, then $C_1 \mathbin{\widetilde{\otimes}} C_2 \in \mathscr{S}_1(H_1 \otimes H_2)$ and $\left|\left|\left|C_1 \mathbin{\widetilde{\otimes}} C_2\right|\right|\right|_1 = |||C_1|||_1 |||C_2|||_1$. We denote by $\mathrm{Tr}_1 : \mathscr{S}_1(H_1 \otimes H_2) \to \mathscr{S}_1(H_2)$ the *partial trace with respect to* H_1. It is the unique bounded linear operator satisfying $\mathrm{Tr}_1(A \mathbin{\widetilde{\otimes}} B) = \mathrm{Tr}(A)B$, for all $A \in \mathscr{S}_1(H_1), B \in \mathscr{S}_1(H_2)$. $\mathrm{Tr}_2 : \mathscr{S}_1(H_1 \otimes H_2) \to \mathscr{S}_1(H_1)$ is defined symmetrically.

If $X \in H$ is a random element with $\mathbb{E}\|X\| < \infty$, then $\mu = \mathbb{E}X \in H$, the mean of X, is well defined. Furthermore, if $\mathbb{E}\|X\|^2 < \infty$, then $C = \mathbb{E}[(X - \mu) \otimes_2 (X - \mu)]$ defines the *covariance operator* of X, where $f \otimes_2 g$ is the operator on H defined by $(f \otimes_2 g)h = \langle h, g \rangle f$, for $f, g, h \in H$. The covariance operator C is a trace-class hermitian operator on H, and encodes all the second-order fluctuations of X around its mean.

32.2.1 Separability

We recall now that we want to define separability so that the covariance function can be written as $c(s,t,s',t') = c_1(s,s')c_2(t,t')$, for some $c_1 \in L^2\left([-S,S]^d \times [-S,S]^d, \mathbb{R}\right)$ and $c_2 \in L^2\left([0,T]\times[0,T],\mathbb{R}\right)$. This can be extended to the covariance operator of a random elements $X \in H = H_1 \otimes H_2$, where H_1, H_2 are arbitrary separable real Hilbert spaces. We call its covariance operator C *separable* if

$$C = C_1 \,\widetilde{\otimes}\, C_2, \tag{32.3}$$

where C_1, respectively C_2, are trace-class operators on H_1, respectively on H_2, and $C_1 \,\widetilde{\otimes}\, C_2$ is defined in (32.2). Notice that though the decomposition (32.3) is not unique, since $C_1 \,\widetilde{\otimes}\, C_2 = (\alpha C_1) \,\widetilde{\otimes}\, (\alpha^{-1} C_2)$ for any $\alpha \neq 0$, this will not cause any problem at a later stage since we will ultimately be dealing with the product $C_1 \,\widetilde{\otimes}\, C_2$, which is identifiable.

In practice, neither C nor $C_1 \,\widetilde{\otimes}\, C_2$ are known. If $X_1, \ldots, X_N \overset{\text{i.i.d.}}{\sim} X$ and (32.3) holds, the sample covariance operator $\widehat{C}_N$ is not necessarily separable in finite samples. However, we can estimate a separable approximation of it by

$$\widehat{C}_{1,N} \,\widetilde{\otimes}\, \widehat{C}_{2,N}, \tag{32.4}$$

where $\widehat{C}_{1,N} = \mathrm{Tr}_2(\widehat{C}_N)/\sqrt{\mathrm{Tr}(\widehat{C}_N)}$, $\widehat{C}_{2,N} = \mathrm{Tr}_1(\widehat{C}_N)/\sqrt{\mathrm{Tr}(\widehat{C}_N)}$. The intuition behind (32.4) is that

$$\mathrm{Tr}(T)T = \mathrm{Tr}_2(T) \,\widetilde{\otimes}\, \mathrm{Tr}_1(T),$$

for all $T \in \mathscr{S}_1(H_1 \otimes H_2)$ of the form $T = A \,\widetilde{\otimes}\, B$, $A \in \mathscr{S}_1(H_1), B \in \mathscr{S}_1(H_2)$, with $\mathrm{Tr}(T) \neq 0$.

We stress again that we aim to develop a test statistic that solely relies on the estimation of the separable components C_1 and C_2, and does not require the estimation of the full covariance C. We can expect that under the null hypothesis $H_0 : C = C_1 \,\widetilde{\otimes}\, C_2$, the difference $D_N = \widehat{C}_N - \widehat{C}_{1,N} \,\widetilde{\otimes}\, \widehat{C}_{2,N}$ between the sample covariance operator and its separable approximation should take small values. We propose therefore to construct our test statistic by projecting D_N onto the first eigenfunctions of C, since these encode the directions along which X has the most variability. If we denote by $C_1 = \sum_{i\geq 1} \lambda_i u_i \otimes_2 u_i$ and $C_2 = \sum_{j\geq 1} \gamma_j v_j \otimes_2 v_j$ the Mercer decompositions of C_1 and C_2, we have

$$C = C_1 \,\widetilde{\otimes}\, C_2 = \sum_{i,j\geq 1} \lambda_i \gamma_j (u_i \otimes v_j) \bigotimes\nolimits_2 (u_i \otimes v_j),$$

The eigenfunctions of C are therefore of the form $u_r \otimes v_s$, where $u_r \in H_1$ is the r-th eigenfunction of C_1 and $v_s \in H_2$ is the s-th eigenfunction of C_2. We define a test statistic based on the projection

$$T_N(r,s) = \sqrt{N}\,\langle D_N(\hat{u}_r \otimes \hat{v}_s), \hat{u}_r \otimes \hat{v}_s \rangle\,, \quad r,s \geq 1 \text{ fixed}, \tag{32.5}$$

where we have replaced the eigenfunctions of C_1 and C_2 by their empirical counterpart, i.e. the Mercer decompositions of $\widehat{C}_{1,N}$, respectively $\widehat{C}_{2,N}$, are given by $\widehat{C}_{1,N} = \sum_{i\geq 1} \hat{\lambda}_i \hat{u}_i \otimes \hat{u}_i$, respectively $\widehat{C}_{2,N} = \sum_{j\geq 1} \hat{\gamma}_j \hat{v}_j \otimes \hat{v}_j$. Notice that though the eigenfunctions of $\widehat{C}_{1,N}$ and $\widehat{C}_{2,N}$ are defined up to a multiplicative constant $\alpha = \pm 1$, our test statistic is well defined. The key fact for the practical implementation of the method is that $T_N(r,s)$ can be computed without the need to estimate (and store in memory) the operator D_N, since $T_N(r,s) = \sqrt{N}\left(\frac{1}{N}\sum_{k=1}^N \left\langle X_k - \overline{X}_N, \widehat{v}_i \otimes \widehat{u}_j \right\rangle^2 - \hat{\lambda}_r \hat{\gamma}_s\right)$. In particular, the computation of $T_N(r,s)$ does *not* require an estimation of the full covariance operator C, but only the estimation of the marginal covariance operators C_1 and C_2, and their eigenstructure.

32.2.2 Asymptotics

The theoretical justification for using a projection of D_N to define a test procedure is that, under the null hypothesis $H_0 : C = C_1 \widetilde{\otimes} C_2$, we have $\|\|D_N\|\|_1 \xrightarrow{p} 0$ as $N \to \infty$, i.e. D_N convergences in probability to zero with respect to the trace norm. In fact, we will show in Theorem 32.1 that $\sqrt{N} D_N$ is asymptotically Gaussian under the following regularity conditions:

Condition 1 *X is a random element of the real Hilbert space H satisfying*

$$\sum_{j=1}^{\infty} \left(\mathbb{E}\left[\left\langle X, e_j \right\rangle^4 \right] \right)^{1/4} < \infty, \tag{32.6}$$

for some orthonormal basis $(e_j)_{j\geq 1}$ of H.

Recall that $\widehat{C}_N = \frac{1}{N}\sum_{j=1}^N (X_i - \overline{X}) \otimes_2 (X_i - \overline{X})$, where $\overline{X} = N^{-1}\sum_{k=1}^N X_k$. The following result establishes the asymptotic distribution of $D_N = \widehat{C}_N - \frac{\mathrm{Tr}_2(\widehat{C}_N) \,\widetilde{\otimes}\, \mathrm{Tr}_1(\widehat{C}_N)}{\mathrm{Tr}(\widehat{C}_N)}$:

Theorem 32.1. *Let H_1, H_2 be separable real Hilbert spaces, $X_1, \ldots, X_N \sim X$ be i.i.d. random elements on $H_1 \otimes H_2$ with covariance operator C, and $\mathrm{Tr}C \neq 0$.*

If X satisfies Condition 1 (with $H = H_1 \otimes H_2$), then, under the null hypothesis

$$H_0 : C = C_1 \widetilde{\otimes} C_2, \qquad C_1 \in \mathscr{S}_1(H_1), C_2 \in \mathscr{S}_1(H_2),$$

we have

$$\sqrt{N}\left(\widehat{C}_N - \frac{\mathrm{Tr}_2(\widehat{C}_N) \,\widetilde{\otimes}\, \mathrm{Tr}_1(\widehat{C}_N)}{\mathrm{Tr}(\widehat{C}_N)} \right) \xrightarrow{d} Z, \quad \text{as } N \to \infty, \tag{32.7}$$

where Z is a Gaussian random element of $\mathscr{S}_1(H_1 \otimes H_2)$ with mean zero, whose covariance structure is given in [1].

Condition 1 is used here because we need $\sqrt{N}(\widehat{C}_N - C)$ to converge in distribution in the topology of the space $\mathscr{S}_1(H_1 \otimes H_2)$; it could be replaced by any (weaker) condition ensuring such convergence. The assumption $\mathrm{Tr}C \neq 0$ is equivalent to assuming

that X is not almost surely constant. We can now give the asymptotic distribution of $T_N(r,s)$ under Gaussian assumptions on X. Recall that $T_N(r,s)$ is defined in (32.5) as the (scaled) projection of D_N in a direction given by the tensor product of the empirical eigenfunctions $\hat{u}_r$ and $\hat{v}_s$. Note that the asymptotic variance-covariance of $(T_N(r,s))_{(r,s)\in\mathscr{I}}$ can be entirely expressed in terms of the covariance operator C. The asymptotic distribution is also derived without the Gaussian assumptions in [1], but there the expression of the asymptotic covariance is more involved.

Corollary 32.1. *Assume the conditions of Theorem 32.1 hold, and that X is Gaussian. If $\mathscr{I} \subset \{(i,j) : i,j \geq 1\}$ is a* finite *set of indices such that $\lambda_r\gamma_s > 0$ for each $(r,s) \in \mathscr{I}$, then*

$$(T_N(r,s))_{(r,s)\in\mathscr{I}} \xrightarrow{d} N(0,\Sigma), \quad \text{as } N \to \infty.$$

where

$$\Sigma_{(r,s),(r',s')} = \frac{2\lambda_r\lambda_{r'}\gamma_s\gamma_{s'}}{\operatorname{Tr}(C)^2}\left(\delta_{rr'}\operatorname{Tr}(C_1)^2 + |||C_1|||_2^2 - (\lambda_r+\lambda_{r'})\operatorname{Tr}(C_1)\right) \times \left(\delta_{ss'}\operatorname{Tr}(C_2)^2 + |||C_2|||_2^2 - (\gamma_s+\gamma_{s'})\operatorname{Tr}(C_2)\right),$$

and $\delta_{ij} = 1$ if $i = j$, and zero otherwise. In particular, notice that Σ itself is separable.

The results presented in this Section allow one to construct asymptotically chi-squared tests for covariance separability, and to approximate the finite sample distribution of the test by means of bootstrap techniques, which have very good performance in simulations [1].

32.3 Application to acoustic phonetic data

An interesting case where the proposed methods can be useful are phonetic spectrograms. These data arise in the analysis of speech records, since relevant features of recorded sounds can be better explored in a two dimensional time-frequency domain.

In particular, we consider here the dataset of 23 speakers from five different Romance languages that has been first described in [9]. The speakers were recorded while pronouncing the words corresponding to the numbers from one to ten in their language and the recordings are converted to a sampling rate of 16000 samples per second. Since not all these words are available for all the speakers, we have a total of 219 speech records. We focus on the spectrum that speakers produce in each speech recording $x_{ik}^L(t)$, where L is the language, $i = 1,\ldots,10$ the pronounced word and $k = 1,\ldots,n_L$ the speaker, n_L being the number of speakers available for language L. We then use a short-time Fourier transform to obtain a two dimensional log-spectrogram: we use a Gaussian window function $w(\cdot)$ with a window size of 10 milliseconds and we compute the short-time Fourier transform as $X_{ik}^L(\omega,t) = \int_{-\infty}^{+\infty} x_{ik}^L(\tau)w(\tau-t)e^{-j\omega\tau}\mathrm{d}\tau$. The spectrogram is defined as the magnitude of the Fourier transform and the log-spectrogram (in decibel)

is therefore $\mathfrak{S}_{ik}^{L}(\omega,t) = 10\log_{10}(|X_{ik}^{L}(\omega,t)|^2)$. The raw log-spectrograms are then smoothed [with the robust spline smoothing method proposed in 4] and aligned in time using an adaptation to 2-D of the procedure in [15]. The alignment is needed because a phase distortion can be present in acoustic signals, due to difference in speech velocity between speakers. Since the different words of each language have different mean log-spectrograms, the focus of the linguistic analysis—which is the study cross-linguistics changes—is on the residual log-spectrograms $R_{ik}^{L}(\omega,t) = S_{ik}^{L}(\omega,t) - (1/n_i)\sum_{k=1}^{n_i} S_{ik}^{L}(\omega,t)$. Assuming that all the words within the language have the same covariance structure, we disregard hereafter the information about the pronounced words that generated the residual log-spectrogram, and use the surface data $R_j^{L}(\omega,t), j = 1,\ldots,N_L$, i.e. the set of observations for the language L including all speakers and words, for the separability test.

We thus apply the Studentized version of the empirical bootstrap test for separability [1] to the residual log-spectrograms for each language individually. Here, we take into consideration different choices for set of eigendirections to be used in the definition of the test statistic $\widetilde{G}_N(\mathscr{I})$, namely $\mathscr{I} = \mathscr{I}_1 = \{(1,1)\}$, $\mathscr{I} = \mathscr{I}_2 = \{(r,s) : 1 \le r \le 2, 1 \le s \le 3\}$, $\mathscr{I} = \mathscr{I}_3 = \{(r,s) : 1 \le r \le 8, 1 \le s \le 10\}$. For all cases we use $B = 1000$ bootstrap replicates.

Table 32.1: P-values for the test for the separability of the covariance operators of the residual log-spectrograms of the five Romance languages, using the Studentized version of the empirical bootstrap.

$\mathscr{I}$	French	Italian	Portuguese	American Spanish	Iberian Spanish
$\mathscr{I}_1$	0.65	< 0.001	< 0.001	< 0.001	< 0.001
$\mathscr{I}_2$	0.078	0.197	0.022	0.36	0.013
$\mathscr{I}_3$	0.001	0.002	0.001	0.001	< 0.001

The resulting p-values for each language and for each set of indices can be found in Table 32.1. Taking into account the multiple testings with a Bonferroni correction, we can conclude that the separability assumption does not appear to hold. We can also see that the departure from separability is caught mainly on the first component for the two Spanish varieties. In conclusion, a separable covariance structure is not a good fit for these languages and thus, when practitioners use this approximation for computational or modeling reasons, they should bear in mind that relevant aspects of the covariance structure may be missed in the analysis.

References

[1] Aston, J. A. D., Pigoli, D., Tavakoli, S.: Tests for separability in nonparametric covariance operators of random surfaces. Ann. Statist., in press, (2017)

[2] Chen, K., Delicado, P., Müller, H.-G.: Modelling function-valued stochastic processes, with applications to fertility dynamics. J. R. Stat. Soc. Ser. B Stat. Methodol., **79**(1):177–196, (2017)
[3] Constantinou, P., Kokoszka, P., Reimherr, M.: Testing separability of space–time functional processes. ArXiv e-prints, September (2015)
[4] Garcia, D.: Robust smoothing of gridded data in one and higher dimensions with missing values. Comput. Statist. Data Anal., **54**(4):1167–1178, (2010)
[5] Gneiting, T., Genton, M. G., Guttorp, P.: Geostatistical space-time models, stationarity, separability, full symmetry. Monogr. Statist. Appl. Probab., **107**:151, (2007)
[6] Huang, L., Reiss, P. T., Xiao, L., Zipunnikov, V., Lindquist, M. A., Crainiceanu, C. M.: Two-way principal component analysis for matrix-variate data, with an application to functional magnetic resonance imaging data. Biostatistics, kxw040, (2016)
[7] Kadison, R.V., Ringrose, J.R.: Fundamentals of the Theory of Operator Algebras: Elementary theory. Number v. 1 in Fundamentals of the Theory of Operator Algebras. American Mathematical Society, (1997)
[8] Lindquist, M. A.: The statistical analysis of fMRI data. Statist. Sci., **23**(4):439–464, (2008)
[9] Pigoli, D., J. A. D. Aston, Dryden, I. L., Secchi, P.: Distances and inference for covariance operators. Biometrika, **101**(2):409–422, (2014)
[10] Rabiner, L. R., Schafer, R. W.: Digital processing of speech signals, volume 100. Prentice-hall Englewood Cliffs, (1978)
[11] Ramsay, J. O., Silverman, B. W.: Functional data analysis. Springer Series in Statistics. Springer, New York, second edition, (2005)
[12] Ringrose, J. R.: Compact non-self-adjoint operators. Van Nostrand Reinhold Co., London, (1971)
[13] Secchi, P., Vantini, S., Vitelli, V.: Analysis of spatio-temporal mobile phone data: a case study in the metropolitan area of Milan. Stat. Methods Appl., (2015) in press.
[14] Simpson, S. L., Edwards, L. J., Styner, M. A., Muller, K. E.: Separability tests for high-dimensional, low-sample size multivariate repeated measures data. J. Appl. Stat., **41**(11):2450–2461, (2014)
[15] Tang, R., Müller, H.-G.: Pairwise curve synchronization for functional data. Biometrika, **95**(4):875–889, (2008)
[16] Wang, J.-L., Chiou, J.-M., Müller, H.-G.: Review of Functional Data Analysis. Annual Review of Statistics and Its Application, (2016)
[17] Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., Evans, A. C.: A unified statistical approach for determining significant signals in images of cerebral activation. Human brain mapping, **4**(1):58–73, (1996)

Chapter 33
Contribution of functional approach to the classification and the identification of acoustic emission source mechanisms

Oumar I. Traore, Paul Cristini, Nathalie Favretto-Cristini, Laurent Pantera, Philippe Vieu and Sylvie Viguier-Pla

Abstract In a context of nuclear Reactivity Initiated Accident, we describe acoustic emission signals, for which a problem of classification is open. As classical approaches with a reduced number of variables do not give satisfactory discrimination, we propose to use the envelopes of the received signals. We perform a k-means clustering and discuss the first results of this approach.

Introduction

Several non-destructive methods are used for the monitoring of nuclear safety experiment reactors. Among them the acoustic emission (AE) technique is of major interest. It has the advantage of being simple to adapt to nuclear-oriented purposes and allows a quasi-real-time monitoring of experiments. One goal of the AE testing is to process the AE signals recorded from the reactor in such a way that they can be associated with specific physical source mechanisms occurring in the tested structure or material. In general, this characterization is done by computing classical AE variables from the received signal like the energy, the rise time, the duration...

Oumar. I. Traore (✉), Paul Cristini and Nathalie Favretto-Cristini
Aix-Marseille Univ., CNRS, Centrale Marseille, L.M.A., France, e-mail: toumarissiaka@gmail.com, cristini@lma.cnrs-mrs.fr, favretto@lma.cnrs-mrs.fr

Laurent Pantera
CEA, DEN, DER/SRES, Cadarache,F13108 Saint-Paul-Lez-Durance, France, e-mail: laurent.pantera@cea.fr

Philippe Vieu and Sylvie Viguier-Pla
Equipe de Stat. et Proba., Institut de Mathématiques, UMR5219, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France, e-mail: vieu@math.univ-toulouse.fr

Sylvie Viguier-Pla
Université de Perpignan via Domitia, LAMPS, 52 av. Paul Alduy, 66860 Perpignan Cedex 9, France, e-mail: viguier@univ-perp.fr

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_33

When the number of recorded signals to process becomes very important, classical data-mining methods based on these variables are used to classify them and then the physical source mechanism associated with each class is identified. However, depending on the diversity of source mechanisms (cracks, fractures, delaminations ...) and the type of material (nuclear fuel, zircaloy, inox...) very different types of variables constructed from the AE signals can be discriminant [4, 3, 2, 9]. In the case of nuclear safety experiment which is of interest in this article, the test device is composed of several types of materials and interact with a very complex environment, leading to a difficulty to get enough discriminant variables for a very heterogeneous sample of source mechanisms.

Unsupervised classification based on functional approaches has been, in literature, widely of interest, as the problem on the overall curves shape, their origins and regularity. Some authors use a basis expansion, and process a clustering based on the coefficients of expansions (see for example Abraham et al. [1] for the use of a B-spline basis or Giacofci et al.[6], for the use of wavelet basis). Other use the scores of the functional principal components analysis (as for example, Peng and Muller [12]).

In this work, we present a functional approach based on the envelopes of the received signals. This choice rather than the use of the raw received signals has been motivated by their more effectiveness and robustness to estimate the time-delay between signals recorded from the same source mechanism at two different sensors, a crucial variable for source mechanisms identification.

33.1 Context of the study and Raw data processing

Reactivity Iniatiated Accident (RIA) is a nuclear safety experiment which involves an unexpected and very fast increase in fission rate and reactor power due to the ejection of a control rod. The power increases may damage the fuel clad and the fuel pellets of the reactor. The French Alternative Energies and Atomic Energy Commission (CEA) operates a pool-type reactor dedicated to fuel behavior study in RIA conditions. During these RIA experiments, the test device is equipped with two AE sensors (microphones) allowing to record information about the fuel behavior.

Since 1993, fourteen RIA experiments have been operated by the laboratory in charge of the preparation and the realization of the experiments, after each of them, the raw microphone signals are processed in order to give to the experimentalists the first estimations about the fuel behavior. This experiment result analysis process is composed of several steps [11]. In the first one, the physical measurements performed by the two sensors are converted into numerical signals (Figure 33.1).

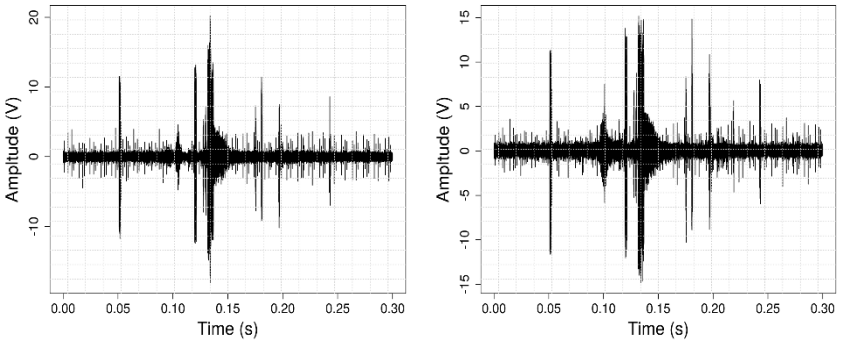


Fig. 33.1: Microphone numerical signals recorded during an experiment realized in 1994. Microphone 1 (left) and microphone 2 (right)

Then, a hits (events) detection function is used in order to isolate segments of signal which are potentially associated with physical source mechanism of interest. This hits detection strategy based on a moving variance consists, for each microphone, in:

1. defining a learning sample of signal corresponding to the background noise of the reactor and in fixing a threshold equal to the variance of this noise sample
2. computing the moving variance vector associated with the microphone signal
3. identifying the segments of the microphone signal corresponding to a threshold violation.

After the realization of the hits detection for the two microphones, their results are merged in order to get the same hits starts and ends times. At the end of this first treatment, a certain number of hits, so of potential physical source mechanisms of interest, are associated with the experiment (Figure 33.2).

In order to implement a statistical classification algorithm able to perform an automatic source mechanism identification for the future experiments, the results of the hits detection process for the fourteen first experiments have been gathered to form a sample of 168 hits.

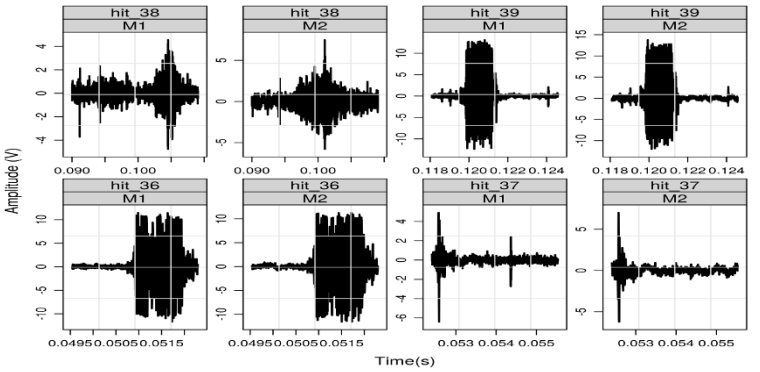


Fig. 33.2: Representation of the first four hits detected after the application of the moving variance strategy on the experiment signal presented in Figure 33.1

Functional dataset creation

The 168 hits resulting from the application of the hits detection strategy on the fourteen experiments constituted a dataset of signal of very different duration (length). In order to get a dataset with discretized curves of the same length and to cope with some limitations due to changes on signal sampling ratio and differences of period between the two microphone signals, we have chosen to perform the classification on the re-sampled envelopes of the hit signals. The envelope of a signal has got the advantage of being characteristic of a signal and smoother than the raw data. Here is the method for the computation of this envelope [10].

Let $x(t)$ be the real-valued signal associated with a given hit. The analytic signal $z(t)$ of $x(t)$ is defined as follows:

$$z(t) = x(t) + iy(t),$$

where $y(t)$ is the Hilbert transform of $x(t)$, that is

$$y(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(t)}{t-\tau} \mathrm{d}\tau,$$

Then the signal $z(t)$ can be written as follows:

$$z(t) = A(t)\mathrm{e}^{i\phi(t)}$$

The function $A(t)$ is named the envelope of the signal $x(t)$. It can also be written $A(t) = \sqrt{x^2(t) + y^2(t)}$.

Then, the envelope based on the Hilbert transform of each hit is computed and its length is coerced to 500 observation points by resampling (Figure 33.3).

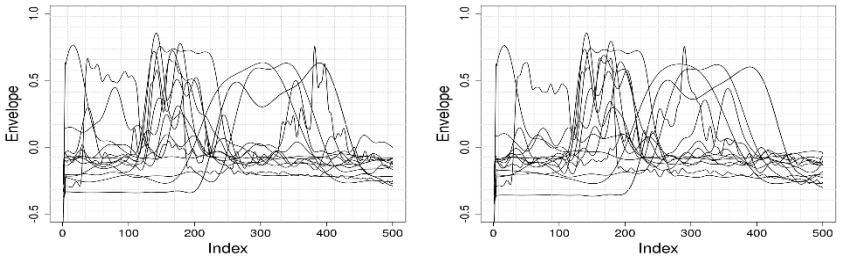


Fig. 33.3: Illustration of hits envelopes dataset associated with the experiment signals presented in Figure 33.1 and hits sample presented in Figure 33.2. Microphone 1 (left) and microphone 2 (right)

At the end of this raw data processing step, each hit can be considered as the realization of a functional random variable written as:

$$X = \{X(t), t \in \mathscr{T}\}$$

where $\mathscr{T}$ is the interval of hit observation time.

Of course, this formulation is a misnomer as in practice we have a finite set of observation points $\in \{1,\ldots,500\}$. Then, in practice, X_i the i^{th} realization of X should be written as:

$$X_i = \{X_{ij}, j = 1,\ldots,500\}$$

In order to transform the envelopes (discretized curves) into functional data, several methods exist. According to Jacques and Preda [8], the most common solution to this problem is to consider that sample paths belong to a finite dimensional space spanned by some basis of functions. In our case, as the sampling rate of signal where about $2.5\mu s$, the observation grid is very fine, then we can consider each envelope as a continuous curve [5].

33.2 Unsupervised classification method and parameters settings

Based on the classification of clustering methods for functional data proposed by Jacques and Preda [8], we distinguish four principal approaches. Among them, the non parametric one is a priori very well adapted to the envelopes. Let us consider

the usual non parametric k-means clustering method for which a description can be found in several articles (see for example [14, 7]). The key point when applying such algorithm in non parametric functional context is of course the choice of a function to measure the proximity between two curves. As the envelops are smooth curves, semi-metrics are good candidates to do so [5].

Let X be the functional random variable associated with the envelopes taking its values in an infinite dimensional semi-metric space (E,d). The semi-metric considered in order to compute the proximity between two envelopes x_i and $x_{i'}$ is defined as follows:

$$d_l(x_i,x_{i'}) = \left(\int_{\mathscr{T}} (x_i^{(l)}(t) - x_{i'}^{(l)})^2 \mathrm{d}t\right)^{1/2},$$

where $\mathscr{T}$ is the interval where X is defined, and $x^{(l)}$ is the l derivative of x.

Recalling that in our case the observation grid is fine enough to consider the discretized envelopes as good approximation of functional ones, $d_l(x_i,x_{i'})$ is estimated by

$$\widehat{d}_l(x_i,x_{i'}) = \left(\sum_{t=1}^{500} (x_i^{(l)}(t) - x_{i'}^{(l)}(t))^2\right)^{1/2}$$

i indexing the n realizations of X, and t indexing the $T = 500$ points of observation of each realization.

Then performing a successful clustering is equivalent to choose the best value of the order of derivative l of the curves, this is discussed by Ferraty and Vieu [5]. An over crucial parameter being of course the number of clusters. To do so, some strategies are proposed in the literature among which the use of silhouette values [13].

Four our purpose we have given priority to physical considerations. Indeed, except for non physical source mechanisms taking place out of the test device, the two microphones are supposed to give the same signals. Then, the derivative order has been chosen in order to get the closest possible classification result between the two microphones. Furthermore, as experimentalists have a good a priori about the characteristics of many source mechanisms, the number of clusters have been set according to their appreciation and give the priority to physical interpretability of each cluster.

33.3 Some analysis results

Figure 33.4 shows the result of the clustering in 6 clusters of the 168 envelopes of the dataset by using derivative of order 2 of the curves. It highlights a very important

closeness between envelopes belonging to the same cluster and a good concordance between the two microphone results.

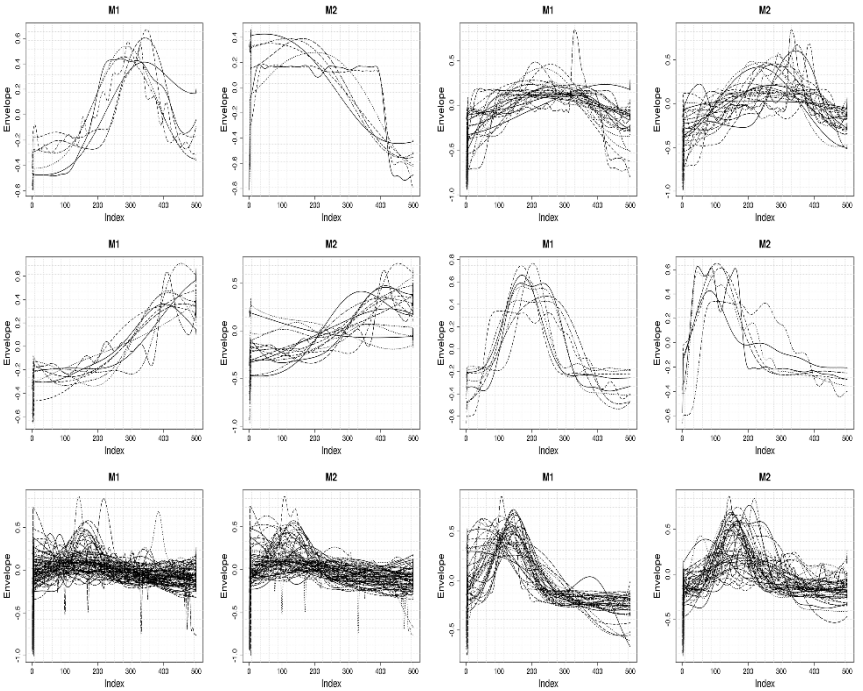


Fig. 33.4: Results of *k*-means clustering of the datasets of the two microphones. Clusters from 1 to 6 to be read from top to bottom and left to right.

Back to physical source mechanisms

The natural continuation of the process is the identification of source mechanisms associated with each cluster. For some clusters, this is very simple after a return to the raw detected signals. For example, hits associated with cluster 4 are supposed to correspond to fuel clad failure source mechanism as those of cluster 5 correspond to noise (not associated with source mechanism of interest).

33.4 conclusion

In this work, we have study the opportunity of applying functional data clustering to classify acoustic emission signals recorded from two microphones during nuclear

safety experiments. A functional dataset construction process in several steps has been presented. We propose a hit detection strategy based on moving variance. The detected hits for two microphones are then merged and their resampled envelopes are computed. A k-means classification algorithm has been performed, the number of classes being chosen in order to get the best physical interpretabilty, we have chosen the semi-metric taking into account the non regular character of the envelopes and the necessity to get close the classification results for each microphone. The results confirm the potential of the functional approaches for this kind of data. A natural further work could be a multivariate classification process for the treatment of the two microphones together.

References

[1] Abraham, C., Cornillon, P., Matzner-Løber, E., Molinari, N.: Unsupervised curve clustering using b-splines. Scand. J. Stat. **30**(3), 581–595 (2003)

[2] Ai, Q., Liu, C., Chen, X., He, P., Wang, Y.: Acoustic emission of fatigue crack in pressure pipe under cyclic pressure. *Nuclear Engineering and Design* **240**(10), 3616–3620 (2010)

[3] Anastassopoulos, A.A., Philippidis, T.P.: Clustering methodology for the evaluation of acoustic emission from composites. *Journal of Acoustic Emission* **13**(1-2), 11–22 (1995)

[4] Favretto-Cristini, N., Hégron, L., Sornay, P.: Identification of the fragmentation of brittle particles during compaction process by the acoustic emission technique. Ultrasonics **67**, 178–189 (2016)

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

[6] Giacofci, M., Lambert-Lacroix, S., Marot, G., Picard, F.: Wavelet-based clustering for mixed-effects functional models in high dimension. Biometrics **69**(1), 31–40 (2013)

[7] Ieva, F., Paganoni, A., Pigoli, D., Vitelli, V.: Multivariate functional clustering for the morphological analysis of electrocardiograph curves. J. R. Stat. Soc. Ser. C. Appl. Stat. **62**(3), 401–418 (2013)

[8] Jacques, J., Preda, C.: Functional data clustering: a survey. Adv. Data Anal. Classif. **8**(3), 231–255 (2014)

[9] Keyvan, S., Nagaraj, J.: Pattern recognition of acoustic signatures using art2: A neural network. Journal of Acoustic Emission **14**(2), 97–102 (1996)

[10] Mbu Nyamsi, R., Aubin, T., Bremond, J.: On the extraction of some time dependent parameters of an acoustic signal by means of the analytic signal concept. its application to animal sound study. Bioacoustics **5**(3), 187–203 (1994)

[11] Pantera, L., Traore, O.: Reproducible data processing research for the CABRI RIA experiments acoustic emission signal analysis. In: Advancements in

Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA), 2015 4th International Conference on, pp. 1–8. IEEE (2015)

[12] Peng, J., Müller, H.: Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. Ann. Appl. Stat. pp. 1056–1077 (2008)

[13] Struyf, A., Hubert, M., Rousseeuw, P., et al.: Clustering in an object-oriented environment. Journal of Statistical Software **1**(4), 1–30 (1997)

[14] Tarpey, T., Kinateder, K.: Clustering functional data. J. Classification **20**(1), 093–114 (2003)

Chapter 34
Parameter regimes in partially functional linear regression for panel data

Fabian Walders and Dominik Liebl

Abstract We introduce a novel semiparametric partially functional linear regression model for panel data. The parametric model part is completely time varying, whereas the functional non-parametric component is allowed to vary over a set of different (functional) parameter regimes. These parameter regimes are assumed latent and need to be estimated from the data additionally to the unknown model parameters. We develop asymptotic theory for the suggested estimators including rates of convergence as $n, T \to \infty$. Our statistical model is motivated from economic theory on asset pricing. It allows to identify different risk regimes, governing the pricing of idiosyncratic risk in stock markets. For our application we develop necessary theoretical ground and offer a vast empirical study based on high-frequency stock-level data for the S&P 500 Index.

34.1 Introduction

This work contributes a novel semiparametric regression model for panel data. The suggested approach allows a scalar response to be affected by a random function as well as by real-valued predictors in a time varying manner. The nonparametric functional parameter changes over time by switching between different parameter regimes which have to be estimated from the data. The real-valued parametric parameters are allowed to vary over time independently from the regimes. Estimation relies on estimators from the functional data literature and a recent nonparametric classification strategy that allows to identify the different parameter regimes. We

Fabian Walders
University of Bonn, BGSE & Institute for Financial Economics and Statistics, Bonn, Germany. fwalders@uni-bonn.de

Dominik Liebl (✉)
University of Bonn, Institute for Financial Economics and Statistics, Bonn, Germany. dliebl@uni-bonn.de

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_34

develop asymptotic theory for the estimators as n and T diverge simultaneously. Given a set of standard assumptions, we prove classification consistency and derive rates of convergence.

Our approach relates to the recent literature on limited parameter instabilities in panel data models. For instance, [3] consider regime varying parameters in a linear panel model with unknown regime classes. Further important works are those of [11] and [10] which postulate group structures governing dissimilarities between regression functions in the cross section. Another important reference is the contribution of [9] who introduce the partially functional regression model in a cross section context. Our work can be understood as an extension of that work to a panel data context with additional unknown parameter regimes. A second important reference in the functional data literature is the work of [6], introducing parameter instabilities in classical functional linear regression.

The specific form of our suggested statistical model is motivated by relevant economic theory. It is particularly well suited to identify pricing regimes of idiosyncratic risk in stock markets. In our application we offer a combination of economic and statistical theory. Beyond that we provide a vast empirical study for the US stock market examining the idiosyncratic volatility puzzle (see, e.g., [2]).

The remainder of this paper is structured as follows. In Sections 2 and 3 we introduce the model and present the estimation procedure. In Section 4 we present the asymptotic theory, including rates of convergence. Subsequently, we provide economic theory for the application as well as a glance at the corresponding empirical work in Section 5. A last Section 6 briefly concludes.

34.2 Model

We suggest a linear regression model which is formally obtained as a partially functional regression in a panel data context. Generically, the task is to model the effect of a square integrable random function $X_{it} \in L^2[0,1]$ on a scalar response y_{it} in the presence of a real-valued random variable z_{it}. Indexing the cross section and time dimensions $1 \leq i \leq N$ and $1 \leq t \leq T$ respectively, the ultimate statistical model reads as

$$y_{it} = \alpha_{0,t} + \int_0^1 \alpha_t(s) X_{it}(s) \mathrm{d}s + \beta_t z_{it} + \varepsilon_{it}, \tag{34.1}$$

with $\alpha_{0,t}$ being a t-specific intercept parameter and ε_{it} being random disturbance. The unknown functional parameters $\alpha_t \in L^2[0,1]$ differ across K different time regimes $G_1, \ldots, G_K$ which form a latent partition of the index set $\{1, \ldots, T\}$. These time regimes are mutually exclusive, i.e., $G_k \cap G_l = \emptyset$ for all $k, l \in \{1, \ldots, K\}$, form a complete partition, i.e., $\bigcup_{k=1}^K G_k = \{1, \ldots, T\}$, and may consist of non-adjacent time points t. Each regime G_k is associated with a square integrable parameter function $A_k \in L^2[0,1]$ governing the effect of X_{it} on y_{it}, i.e.,

$$\alpha_t = A_k \text{ if } t \in G_k.$$

Model (34.1) compromises two extreme specifications. On the one hand it might be the case that $K = 1$ and hence $G_1 = \{1, \ldots, T\}$. In this situation the effect of the random functions X_{it} on the response y_{it} is time invariant. On the other hand the model nests complete heterogeneity if $K = T$ and all regimes G_k are singletons which turns (34.1) into a collection of T different cross section models. Further purely functional or purely parametric specifications are possible in case $\beta_t = 0$ or $\alpha_t = 0$ for all $1 \leq t \leq T$ respectively.

34.3 Estimation

The objective is to estimate the parameters A_k, β_t and $\alpha_{0,t}$ as well as the regimes $G_1, \ldots, G_K$ from realizations of the random variables $\{(y_{it}, X_{it}, z_{it}) : 1 \leq i \leq n, 1 \leq t \leq T\}$. Without loss of generality all variables are assumed to be centered (see Assumption 1). We suggest a four step estimation procedure which comprises an initial auxiliary estimation (Step 1) that serves as a basis for the subsequent classification (Step 2) of the regression functions. The two final estimation steps (Step 3 and 4) re-estimate the model parameters borrowing strength from the identified regime structure. In the following we give a detailed description:

Step 1 Estimate the parameters α_t, β_t and $\alpha_{0,t}$ separately for each $1 \leq t \leq T$ using the (modified) estimators in [9] as well as within-t-averages. For the intercepts simply employ the estimators $\hat{\alpha}_{0,t} = n^{-1} \sum_{i=1}^{n} y_{it}$ and use these in turn to form centered responses $\breve{y}_{it} = y_{it} - \hat{\alpha}_{0,t}$. The estimators for α_t and β_t are based on cross section estimates of the relevant covariance objects. Consider for a fixed t the cross section $\{(y_{it}, X_{it}, z_{it}) : 1 \leq i \leq n\}$. For each t, the empirical covariance operator of the $\{X_{it} : 1 \leq i \leq n\}$ is the integral operator $\hat{\Gamma}_t : L^2[0,1] \to L^2[0,1]$ defined via its kernel, namely, the empirical covariance function $\hat{K}_{X,t}(u,v) := n^{-1} \sum_{i=1}^{n} X_{it}(u) X_{it}(v)$. Let $(\hat{\lambda}_{1,t}, \hat{\phi}_{1,t}), \ldots, (\hat{\lambda}_{n,t}, \hat{\phi}_{n,t})$ denote the eigenvalue-eigenfunction pairs of $\hat{\Gamma}_t$ ordered according to $\hat{\lambda}_{1,t} \geq \cdots \geq \hat{\lambda}_{n,t}$ and $\langle \cdot, \cdot \rangle$ the inner product in $L^2[0,1]$. In a similar t-wise fashion we define the following estimators:

$$\hat{K}_{zx,t}(s) := n^{-1} \sum_{i=1}^{n} z_{it} X_{it}(s), \qquad \hat{K}_{yx,t}(s) := n^{-1} \sum_{i=1}^{n} \breve{y}_{it} X_{it}(s),$$

$$\hat{K}_{z,t}(s) := n^{-1} \sum_{i=1}^{n} z_{it}^2, \qquad \hat{\Phi}_t(g) := \sum_{j=1}^{m} \frac{\langle \hat{K}_{zx,t}, \hat{\phi}_{j,t} \rangle \langle \hat{\phi}_{j,t}, g \rangle}{\hat{\lambda}_{j,t}} \text{ for } g \in L^2[0,1].$$

Given a truncation parameter m, with $1 \leq m < n$, one obtains least squares estimators for (a_{jt}^*) $1 \leq j \leq m$ and β_t in the approximate empirical model $\breve{y}_{it} \approx \sum_{j=1}^{m} \langle X_{it}, \hat{\phi}_{j,t} \rangle a_{jt}^* + z_{it} \beta_t + \varepsilon_{it}$. These estimators read as

$$\hat{\beta}_t = \left[\hat{K}_{z,t} - \hat{\Phi}_t\left(\hat{K}_{zx,t}\right)\right]^{-1}\left[\hat{K}_{zy,t} - \hat{\Phi}_t\left(\hat{K}_{xy,t}\right)\right]$$
$$\hat{a}_{j,t} = \hat{\lambda}_{j,t}^{-1}\frac{1}{n}\sum_{i=1}^{n}\langle X_{it}, \hat{\phi}_{j,t}\rangle(\check{y}_{it} - \hat{\beta}_t z_{it}), \quad 1 \le j \le m,$$

where the final estimate for α_t is obtained as $\hat{\alpha}_t = \sum_{j=1}^{m}\hat{a}_{j,t}\hat{\phi}_{j,t}$. For the next step we use the scaled estimator $\hat{\alpha}_t^{(\Delta)} := \sum_{j=1}^{m}\frac{\hat{\lambda}_{j,t}^{1/2}}{\hat{\sigma}_{t,\varepsilon}}\hat{a}_{j,t}\hat{\phi}_{j,t}$, where $\hat{\sigma}^2_{t,\varepsilon}$ is the empirical residual variance from the above regression for time t.

Step 2 Classify the t-specific regression functions into K regimes using a thresholding procedure along the lines of [11]. To do so, select a subset $S \subset \{1,\dots,T\}$ and some $s \in S$. Suppose s is in some regime G_k. The squared L^2 distances of the modified estimators $\hat{\alpha}_t^{(\Delta)}$ from other periods $t \in S$ to $\hat{\alpha}_s^{(\Delta)}$ are denoted as $\hat{\Delta}_{ts} := ||\hat{\alpha}_t^{(\Delta)} - \hat{\alpha}_s^{(\Delta)}||_2^2$. The corresponding order statistics $\hat{\Delta}_{t(1)} \le \hat{\Delta}_{t(2)} \le \cdots \le \hat{\Delta}_{t(|S|)}$ now form the basis for classification.

Given a pre-selected threshold τ_{nT}, see discussion below, define the regime estimate $\hat{G}_k$ as $\{(1),\dots,(\hat{p})\}$, where $(\hat{p})$ is defined according to $\hat{\Delta}_{t(\hat{p})} \le \tau_{nT} < \hat{\Delta}_{t(\hat{p}+1)}$. Iteratively proceed with this procedure for the remaining points in time, i.e., for $t \in \{1,\dots,\mathrm{T}\}\setminus\hat{G}_k$. This naturally provides an estimate $\hat{K}$.

Step 3 Re-estimate the t-specific parameters β_t by borrowing strength from the identified regime structure. For each $1 \le k \le \hat{K}$, the regime specific empirical covariance operator $\hat{\Gamma}_k$ of the $\{X_{it} : 1 \le i \le n,\ t \in \hat{G}_k\}$ is defined via its integral kernel $\tilde{K}_{X,k}(u,v) := \left(n|\hat{G}_k|\right)^{-1}\sum_{i=1}^{n}\sum_{t\in\hat{G}_k}X_{it}(u)X_{it}(v)$. The eigenfunction-eigenvalue-pairs of this operator are denoted by $(\tilde{\lambda}_{j,k}, \tilde{\phi}_{j,k})$, $1 \le j \le n|\hat{G}_k|$ ordered according to $\tilde{\lambda}_{1,k} \ge \cdots \ge \tilde{\lambda}_{n|\hat{G}_k|,k} \ge 0$. Further, necessary regime specific quantities are obtained according to

$$\tilde{K}_{zx,k}(s) := \left(n|\hat{G}_k|\right)^{-1}\sum_{i=1}^{n}\sum_{t\in\hat{G}_k} z_{it}X_{it}(s), \qquad \tilde{K}_{z,k}(s) := \left(n|\hat{G}_k|\right)^{-1}\sum_{i=1}^{n}\sum_{t\in\hat{G}_k} z_{it}^2,$$
$$\tilde{\Phi}_k(g) := \sum_{j=1}^{\tilde{m}}\frac{\langle\tilde{K}_{zx,k},\tilde{\phi}_{j,k}\rangle\langle\tilde{\phi}_{j,k},g\rangle}{\tilde{\lambda}_{j,k}} \quad \text{for } g \in L^2[0,1],$$

where $\tilde{m}$ denotes a regime specific truncation parameter with $1 \le \tilde{m} < n|\hat{G}|$. From these objects the final slope estimate is calculated according to

$$\tilde{\beta}_t := \left[\tilde{K}_{z,k} - \tilde{\Phi}_k\left(\tilde{K}_{zx,k}\right)\right]^{-1}\left[\hat{K}_{zy,t} - \tilde{\Phi}_k\left(\hat{K}_{xy,t}\right)\right] \quad \text{for all } t \in \hat{G}_k.$$

Step 4 Estimate the regime specific parameter functions A_k. To do so, the response variables need to be transformed for each $t \in \hat{G}_k$ according to $\tilde{y}_{it} = \check{y}_{it} - \tilde{\beta}_t z_{it}$. For any $1 \le k \le \hat{K}$ the pairs $\{(\tilde{y}_{it}, X_{it}) : 1 \le i \le n,\ t \in \hat{G}_k\}$ are then pooled in order to estimate A_k by the following regression:$\tilde{y}_{it} = \langle X_{it}, A_k\rangle + \varepsilon_{it}^*$ using the procedure in [5]. Here, the error term is composed according to $\varepsilon_{it}^* := \varepsilon + (\beta_t - \tilde{\beta}_t)z_{it}$. The resulting estimator $\tilde{A}_k$ obtains as

$$\tilde{A}_k(s) = \sum_{j=1}^{\tilde{m}} \tilde{a}_{j,k}\tilde{\phi}_{j,k}(s), \quad \text{where} \quad \tilde{a}_{j,k} = \tilde{\lambda}_{j,k}^{-1}\left(n|\hat{G}_k|\right)^{-1}\sum_{i=1}^{n}\sum_{t\in\hat{G}_k}\langle X_{it},\tilde{\phi}_{j,k}\rangle\tilde{y}_{it}.$$

In summary, the final estimates are $\hat{G}_1,\ldots,\hat{G}_k$ for the regimes as well as $\tilde{\beta}_t$, $\tilde{A}_k$ and $\hat{\alpha}_{0,t}$ for the regression parameters.

Practical choice of τ_{nt}. For large n and given m, it can be shown that the law of $2^{-1}\cdot n\cdot\Delta_{ts}$ can be reasonably well approximated by a χ^2_m-distribution. For the actual implementation we thus suggest to set the threshold to $\hat{\tau}_{nT} := 2\cdot n^{-1}\cdot q_{0.99}(\chi^2_m)$, with $q_{0.99}(\chi^2_m)$ being the 99% quantile of a χ^2_m-distribution.

34.4 Asymptotic Theory

Two types of problems add to the well understood estimation in functional linear regression. The first one is the additional classification error contaminating the estimation of A_k. The second one is that estimation of A_k suffers from the presence of nuissance parameters $\alpha_{0,t}$ and β_t. For our asymptotic analysis we rely on a set of standard assumptions present in the literature[1]. In the following we present a detailed list of our assumptions.

Assumption 1. Suppose that

1. $\{(\varepsilon_{it},X_{it},z_{it}):\ 1\le i\le n,\ 1\le t\le T\}$ are centered and iid over i and t,
2. $E\left[||X||_2^4\right]<\infty$, $E\left[z_{it}^4\right]<\infty$, $E\left[\varepsilon_{it}^4\right]<\infty$
3. ε_{it} is independent from X_{js} and z_{js} for any $1\le i,j\le n$ and $1\le t,s\le T$.

These conditions are even standard for analyzing classical regression models.

The next assumption contains the regularity conditions on the covariance structure of the functional regressor, on the complexity of the corresponding parameter functions, and on the covariance between functional and scalar regressors.

Assumption 2. Suppose there exist constants $0<C_\lambda,C'_\lambda,C_\theta,C_a,C_{zX},C_\beta<\infty$, such that

1. $C_\lambda^{-1}j^{-\mu}\le\lambda_j\le C_\lambda j^{-\mu}$ and $\lambda_j-\lambda_{j+1}\ge C'_\lambda j^{-(\mu+1)}$, $j\ge 1$ for the eigenvalues $\lambda_1>\lambda_2>\ldots$ of the covariance operator Γ of X_{it} and a $\mu>1$,
2. $E\left[\langle X_{it},\phi_j\rangle^4\right]<C_\theta\lambda_j^2$ for the eigenfunction ϕ_j of Γ corresponding to the j-th eigenvalue,
3. $|\langle A_k,\phi_j\rangle|\le C_a j^{-\nu}$ for all $1\le k\le K$,
4. $|\langle K_{zX},\phi_j\rangle|\le C_{zX}j^{-(\mu+\nu)}$, where $K_{zX}:=E[X_{it}z_{it}]$, and
5. $\sup_{1\le t\le T}\beta_t^2<C_\beta$.

Statements 1 and 3 are traditional in the literature (see, e.g., [5] or [7] among others) while [9] introduces (variant of) statements 2 and 3.

[1] See, for instance, [9], [11], [6] and [5]

The following assumption specifies the panel asymptotics we are considering.

Assumption 3. Suppose that

1. $|G_k| \propto T$ and
2. $T \propto n^{\delta}$ for some $\delta > 0$.

By writing $n, T \to \infty$ it is meant that n and T diverge simultaneously on the path specified by the second point in Assumption 3. The only restriction on δ is indirectly formulated in terms of ν and μ.

Assumption 4. Suppose that $\nu > \max\{r_1, r_2, r_3\}$, where $r_1 := 3(1+\mu)$, $r_2 := (\delta)^{-1}(1+\delta+\mu/2)$ and $r_3 := ((1+\delta)+(1+2\delta)\mu)/2$.

Assumption 4 can be understood as setting an upper bound on the complexity of the parameter functions A_k in terms of the complexity of the regressor X and the rate at which nuisance parameters β_t are added. Another issue which is generic to many nonparametric problems is the specification of truncation, i.e., smoothing parameters, addressed in the following Assumption.

Assumption 5. Suppose for the truncation parameters $m = m(n)$ and $\tilde{m} = \tilde{m}(n,T)$ that $m \propto n^{\frac{1}{\mu+2\nu}}$ and $\tilde{m} \propto (n|G_k|)^{\frac{1}{\mu+2\nu}}$. While m is the standard truncation parameter as in the related literature on cross sections, $\tilde{m}$ is a logical extension to within-regime observations.

The following assumption is borrowed from [9] and serves as the central regularity condition in the estimation of $\hat{\beta}_t$ in step 1.

Assumption 6. The random variables $q_{it} := z_{it} - \int_0^1 X(u)\left(\sum_{j=1}^{\infty} \frac{\langle K_{zX}, \phi_j\rangle}{\lambda_j}\phi_j(u)\right) \mathrm{d}u$ are iid and $E[q_{it}|X_{1t},\ldots,X_{nt}] = 0$ as well as $E[q_{it}^2|X_{1t},\ldots,X_{nt}] > 0$.

A last assumption is designed to ensure identification of different the regimes given the thresholding procedure.

Assumption 7.

1. The threshold parameter $\tau_{nT} \to 0$ satisfies $\mathbb{P}\left(\max_{t,s\in G_k} \hat{\Delta}_{ts} \leq \tau_{nT}\right) \to 1$ for all $1 \leq k \leq K$.
2. There exists some $C_\Delta > 0$ such that for any $1 \leq k \leq K$ and any $t \in G_k$

$$||\alpha_t^{(\Delta)} - \alpha_s^{(\Delta)}||_2^2 =: \Delta_{ts} \begin{cases} \geq C_\Delta & \text{if } s \notin G_k \\ = 0 & \text{if } s \in G_k, \end{cases}$$

where $\alpha_t^{(\Delta)} := \sigma_\varepsilon^{-1} \sum_{j=1}^{\infty} \lambda_j^{1/2} \langle \alpha_t, \phi_j\rangle \phi_j$ and $\sigma_\varepsilon^2 := E[\varepsilon_{it}^2]$.

Assumption 7 consist of the same ingredients as the corresponding requirements in [11]. It ensures that different regimes have different parameter functions in the relevant L^2-metric and that the threshold tends to zero sufficiently slowly.

Based on the above assumptions we conclude with the following results starting with a Lemma extending Theorems 3.1 and 3.2 in [9].

Lemma 34.1. *Given Assumptions 1,2,4,5 and 6 it holds for all* $1 \leq t \leq T$ *as* $n \to \infty$ *that*

$$(\hat{\alpha}_{0,t} - \alpha_{0,t})^2 = O_p\left(n^{-1}\right), \tag{34.2}$$

$$\left(\hat{\beta}_t - \beta_t\right)^2 = O_p\left(n^{-1}\right), \tag{34.3}$$

$$||\hat{\alpha}_t - \alpha_t||_2^2 = O_p\left(n^{\frac{1-2\nu}{\mu+2\nu}}\right). \tag{34.4}$$

Based on the consistency of the $\hat{\alpha}_t$ for a fixed t our estimation procedure achieves classification consistency in the sense of the following theorem.

Theorem 34.1. *Given Assumptions 1–7 it holds that*

$$\mathbb{P}\left(\{\hat{G}_1, \ldots, \hat{G}_{\hat{K}}\} \neq \{G_1, \ldots, G_K\}\right) = o(1) \quad \text{as } n, T \to \infty. \tag{34.5}$$

The following result establishes rates of convergence for the suggested estimators.

Theorem 34.2. *Given Assumptions 1–7 it holds for all* $1 \leq k \leq \hat{K}$ *that*

$$\left|\hat{G}_k\right|^{-1} \sum_{t \in \hat{G}_k} (\hat{\alpha}_{0,t} - \alpha_{0,t})^2 = O_p\left(n^{-1}\right), \tag{34.6}$$

$$\left|\hat{G}_k\right|^{-1} \sum_{t \in \hat{G}_k} \left(\tilde{\beta}_t - \beta_t\right)^2 = O_p\left(n^{-1}\right), \tag{34.7}$$

$$\left|\left|\tilde{A}_k - A_k\right|\right|_2^2 = O_p\left(n^{-1}\right) \quad \text{as } n, T \to \infty. \tag{34.8}$$

Here, root-n consistency of $\tilde{A}_k$ is mainly a consequence of Assumption 4.

34.5 Regime Dependent Pricing of Idiosyncratic Risk

The specific form of our suggested statistical model is strongly motivated by economic theory. According to a standard asset pricing approach, underdiversified investors ask for a premium compensating for the idiosyncratic risk of an asset. Proxying idiosyncratic risk with idiosyncratic volatility, our statistical model offers a tailor-made tool to uncover dynamics in such premiums from price data[2]. In Section 34.5.1, we motivate our statistical model in (34.1) from a theoretical viewpoint and argue how to construct suitable functional regressors from discrete data. We provide a brief outline of the ongoing empirical study in Section 34.5.2.

[2] See, e.g., [1] for the notion of time varying risk premiums.

34.5.1 Economic Modeling and Volatility-Curve Estimation

Our approach is motivated by the stochastic volatility framework of [4]. Let us denote the underlying probability space as $(\Omega, \mathscr{A}, \mathbb{P})$, where randomness is emphasized by making the dependence on an $\omega \in \Omega$ explicit. For a point in time $s \geq 0$ we suggest considering the following differential equation

$$\mathrm{d}\log(P(s,\omega)) = \mu(s,\omega)\mathrm{d}s + \alpha(s)\sigma^2(s,\omega)\mathrm{d}s + \sigma(s,\omega)\mathrm{d}W_s \tag{34.9}$$

to model the log returns corresponding to the asset price P, where μ and W_s denote a stochastic drift and a Wiener process. The object $\sigma(s,\omega)$ is the stochastic instantaneous volatility and α a time varying risk premium. Model (34.9) generalizes the log-price process in [4].

For our purpose it is of interest to deal with the returns between the beginning and the end of a certain period such as a trading day, a week or a month. Choosing without loss of generality a unit interval, the start and end points of the tth period are given by $t-1$ and t. It follows then immediately from (34.9) that

$$y_t := \log\left(\frac{P(t,\omega)}{P(t-1,\omega)}\right) = \gamma_t + \int_0^1 \alpha_t(v)\sigma_t^2(v)\mathrm{d}v + \varepsilon_t. \tag{34.10}$$

with $\gamma_t := \int_{t-1}^t \mu(v,\omega)\mathrm{d}v$ and $\varepsilon_t := \int_{t-1}^t \sigma(v,\omega)\mathrm{d}W_v$. Further $\alpha_t(v)$ and $\sigma_t^2(v)$ are defined as $\alpha_t(v) := \alpha(t-1+v)$ and $\sigma_t^2(v) := \sigma^2(t-1+v)$ to emphasize the t-dependence of the two objects.

Beyond that we postulate that the drift term is at each time $s > 0$ proportional to a stochastic process $Z(s)$, describing pricing relevant market frictions. The proportionality factor is a deterministic step function $b(s)$ taking values β_t in the interval $[t-1,t]$. That is, we assume that $\mu(s) \propto b(s)Z(s)$ which implies that $\gamma_t = \beta_t z_t$ with $z_t := \int_{t-1}^t Z(v)\mathrm{d}v$. As a consequence (34.10) reads as

$$y_t = \beta_t z_t + \int_0^1 \alpha_t(v)\sigma_t^2(v)\mathrm{d}v + \varepsilon_t. \tag{34.11}$$

Above, we implicitly operated on two time scales: the discrete time periods indexed by t and the intra-period times over the generic interval $[0,1]$. The notation in (34.11) indicates that the parameters (α_t, β_t) might change on the first time scale. We postulate that the parameter of interest, α_t, is *regime*-specific with different possible parameter values $A_1, \ldots, A_K$, where each parameter A_k can be interpreted as the regime specific idiosyncratic risk premium. Such regimes are then collections of periods with similar investor's perception of the asset-specific risk relevance.

As the Log-Volatility (LV)-trajectories σ_t^2 are latent, they need to be recovered from observed discrete log-returns. Along the lines of [8] this proceeds as follows. Without loss of generality, intraday trading time is indexed in the unit interval $[0,1]$. Denote the incremental log-return over an interval of length $0 < \Delta \ll 1$ for some $s \geq 0$ as

$$Y_\Delta(s) := \Delta^{-\frac{1}{2}} \log\left(\frac{P(s+\Delta)}{P(s)}\right) = \beta_t \Delta^{\frac{1}{2}} Z(s) + \Delta^{\frac{1}{2}} \alpha(s)\sigma^2(s) + \sigma(s) W_\Delta(s) + \sum_{j=1}^{3} R_{j,\Delta},$$

where $W_\Delta(s) := \Delta^{-\frac{1}{2}}(W(s) - W(s-\Delta))$ and $R_{1,\Delta}, R_{2,\Delta}, R_{3,\Delta}$ are discretization errors as described in the appendix . The latter are negligible in the following sense.

Theorem 34.3. *Given Assumption 8 in the appendix,* $\sum_{j=1}^{3} R_{j,\Delta} = O_p(\Delta^{1/2})$ *as* $\Delta \to 0$.

This justifies the central small-Δ approximation $Y_\Delta(s) \approx \sigma(s) W_\Delta(s)$. From this it can be concluded that $\log\left(Y_\Delta(s)^2\right) + c_0 \approx X(s) + e_s$, where $c_0 \approx 1.27$ and $e_s := \log\left(W_\Delta(s)^2\right) - c_0$ denotes an error term (for details see [8] and the references therein). Most importantly the LV process is defined according to $X(s) := \log(\sigma(s)^2)$. Given returns are observed on an intraday grid $\mathbf{D} := \{0, \Delta, 2\Delta, \ldots, 1-\Delta, 1\}$, it is reasonable to understand $\log\left(Y_\Delta(s)^2\right)$ as discrete noisy observations of the LV-process, provided one is willing to assume a suitable dependence structure for $e_s,\ s \in \mathbf{D}$. As a consequence the method from [8] to estimate $X(s)$ from the discrete prices also applies to our setup without further adjustment.

34.5.2 Empirical Study: Risk Regimes in the US Stock Market

Using the presented framework we examine risk pricing in the US stock market. Data for the S&P 500 constituents is recorded over 100 trading days in 2016. Beyond asset prices, which are sampled every 10 minutes, market frictions, e.g. illiquidity, are proxied by a daily bid-ask spread. Using the estimation procedure indicated in the previous section we are able to construct daily LV-trajectories for each (i,t)-combination. These LV-trajectories are employed as functional regressors X_{it}, while y_{it} is the end-of-day price and z_{it} the maximum bid-ask spread reflecting frictions relevant for asset i at day t.

First estimation results indicate that the regimes and risk premiums distinguish times of financial turmoil from tranquil days.

34.6 Conclusion

In this paper we present a novel regression framework, allowing to examine regime specific effects of a random function on a scalar response in the presence of parametric nuissance terms. Our estimation procedure is designed for a panel data context. We prove consistency and derive rates of convergence. The relevance of our semiparametric model is underlined by an application to idiosyncratic risk pricing in the presence of frictions.

References

[1] Anderson, R.M.: Time-varying risk premia? J. Math. Econom., **47** (3): 253–259, (2011)
[2] Ang, A., Hodrick, A.R., Xing, Y., Zhang, X.: Have we solved the idiosyncratic volatility puzzle? Journal of Financial Economics, **121** (1): 167–194, (2016)
[3] Bada O., Gualtieri J., Kneip A., Sickles R.C.: Panel data models with multiple jump discontinuities in the parameters. Working Paper, (2015)
[4] Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realized volatility and its use in estimating stochastic volatility models. J. R. Stat. Soc. Ser. B Stat. Methodol., **64** (2): 253–280 (2002)
[5] Hall, P., Horowitz, J.L.: Methodology and convergence rates for functional linear regression. Ann. Statist., **35** (1): 70–91, (2007)
[6] Horváth, L., Reeder, R.: Detecting changes in functional linear models. J. Multivariate Anal., **111**: 310–334, (2012)
[7] Kneip, A., Poss, D., Sarda, P.: Functional linear regression with points of impact. Ann. Statist., **44** (1): 1–30, (2016)
[8] Müller, H.G., Sen, R., Stadtmüller, U.: Functional data analysis for volatility. J. Econometrics, 165 (2): 233–245, (2011)
[9] Shin, H.: Partial functional linear regression. J. Statist. Plann. Inference, **139** (10): 3405–3418, (2009)
[10] Su, L., Shi, Z., Phillips, P.C.B.: Identifying latent structures in panel data. Econometrica, **84** (6): 2215–2264, (2016)
[11] Vogt, M., Linton, O.: Classification of nonparametric regression functions in longitudinal data models. Forthcomming in: J. R. Stat. Soc. Ser. B Stat. Methodol. (2016)

Chapter 35
Registration for exponential family functional data

Julia Wrobel and Jeff Goldsmith

Abstract We consider the problem of aligning curves from exponential family distributions. The approach is based on the combination of alignment and functional principal components analysis, and is facilitated by recent extensions of FPCA to non-Gaussian settings. Our work is motivated by the study of physical activity using accelerometers, wearable devices that provide around-the-clock monitoring of activity and produce non-Gaussian measurements. We apply the proposed methods to activity counts using a Poisson distribution, and to a binary "active" vs "inactive" indicator using a binomial distribution. After alignment, the trajectories show clear peaks of activity in the morning and afternoon with a dip in the middle of the day.

35.1 Introduction

Functional data are often observed with variation in both phase and amplitude. As a result, registration or alignment of curves to remove phase variation is a basic problem in functional data analysis (FDA) with a rich associated literature, and serves either as a preprocessing step necessary to understand amplitude variation or as a important modeling step in it's own right.

An excellent overview of the importance of registration and the conceptual and practical challenges in addressing phase variation in functional data is given in [11]. Early approaches to registration include dynamic time warping [13] and landmark registration [3]. Later efforts focused on choosing smooth warping functions by minimizing a loss or distance. The most natural loss function, the $\mathbb{L}^2$ distance

Julia Wrobel
Columbia University Department of Biostatistics, New York, USA, e-mail: jw3134@cumc.columbia.edu

Jeff Goldsmith (✉)
Columbia University Department of Biostatistics, New York, USA, e-mail: ajg2202@cumc.columbia.edu

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_35

between the observed function and the target, was considered in [12]; citing practical concerns when combining this loss function with a flexible class of warping functions, other losses were considered in [14, 18, 21], among others. In parallel, model-based approaches that pose observed curves as warped versions of a common smooth mean function have been developed by e.g. [2] and [19].

The FDA literature, including the preceding work for registration, has traditionally focused on real-valued curves $y_i(t) \in \mathbb{R}$ for subject $1 \leq i \leq I$ at time $t \in [0,T]$. Recently this literature has expanded to include exponential family curves, for which $y_i(t)$ is the realization of a random variable with an exponential family distribution. Examples of such data include sparse longitudinal measurements of the presence or absence of hepatomegaly in patients with primary biliary cirrhosis [7]; high frequency measurements on the feeding behavior of pigs [5]; and binary indicators of abstinence in studies of drug abuse and treatment [8]. In this work, we are motivated by accelerometer data, which provides around-the-clock quantification of physical activity through activity count measurements.

Methods for exponential-family curves has grown in parallel with the rise of this data class. Models now exist for function-on-scalar regression [6, 15] and clustering [8]. Importantly, methods for functional principal components analysis have been extended to the exponential-family setting; these efforts take two distinct approaches. Building on the covariance decomposition framework to functional principal components analysis (FPCA) [22], [7] consider cross-sectional binary curves and [17] consider multilevel spatial curves. Several papers have used probabilistic or Bayesian approaches to FPCA, similar in spirit to the reduced rank approach of [9]. A variational Bayes approach for cross-sectional exponential-family functional data was developed in [20], while a sampling-based approach to multilevel data was presented in [6]. Comments on the interpretation of FPCA for exponential-family curves, emphasizing issues that arise in the direct adaptation of covariance decomposition methods to this setting, appear in [4].

Despite this growing literature for exponential-family functional data, there is a gap for the registration of such data. The purpose of the present work is to address this gap. Our approach is to extend FPCA-based registration methods, originally developed exclusively for the Gaussian case [10], to the current setting. We describe our methodology in Section 35.2 and apply the techniques to accelerometer data in Section 35.3.

35.2 Methods

We begin by establishing our notation and goals. Suppose we observe functional data $y_i(t^*)$ for subjects $1 \leq i \leq I$ over time $t^* \in [0,1]$. The argument t^* represents the originally observed domain, sometimes referred to as "clock time"; our goal is to estimate warping functions $h_i(t^*) = t$, where $t \in [0,1]$ is the domain after alignment, sometimes referred to as "system time". Although t^* and t are observed on the same domains and curves $y_i(\cdot)$ can be evaluated on both, they are conceptually different

and our notation stresses that. The warping functions $h_i(t^*)$ characterize the phase shift in the observed data and are the main target to be estimated.

After alignment, curves $y_i(t)$ follow an exponential family distribution with density

$$p[y_i(t)|\eta_i(t)] = \exp\{(y_i(t)\eta_i(t) - b[\eta_i(t)])/\phi + c[y_i(t),\phi]\}$$

for each $t \in [0,1]$, where $E[y_i(t)|\eta_{ij}(t)] = \mu_i(t) = b'[\eta_i(t)]$ and $Var[y_i(t)|\eta_i(t)] = b''[\eta_i(t)]\phi$. For the binary and count response curves that are primarily discussed in the paper, the dispersion parameter ϕ is known; for other distributions (or to allow overdispersion) it may be necessary to model this parameter. The mean is related to a linear predictor by a known link function $g[\mu_{ij}(t)]$ through

$$\begin{aligned} \mathrm{E}[y_i(t)|\boldsymbol{c}_i] &= \mu_i(t) \\ g[\mu_i(t)] &= \alpha(t) + \sum_{k=1}^{K} c_{ik}\psi_k(t). \end{aligned} \tag{35.1}$$

Model (35.1) generalizes FPCA to exponential family curves and includes the mean $\alpha(t)$, basis functions $\psi_k(t)$, and subject-specific scores c_{ik} which have mean zero and variance λ_k. We generally use the canonical link $\alpha_{ij}(t) = g[\mu_{ij}(t)]$, although this is not necessary for our methodology.

Given the mean $\alpha(t)$, basis functions $\psi_k(t)$, and scores c_{ik}, we register the observed curve $y_i(t^*)$ to the fitted value $\mu_i(t)$ given by model (35.1). The exponential family distribution provides the mechanism through which warping functions $h_i(t^*)$ can be estimated. Our estimate $\hat{h}_i(t^*)$ is obtained my maximizing the log likelihood of the observed data over candidate warping functions:

$$\hat{h}_i(t^*) = \arg\max_{h_i} -[\ell(h_i; y_i(h_i(t^*)), \mu_i(h_i(t^*))] \tag{35.2}$$

As an example, in the case of Bernoulli functions $y_i(t^*)$ our log likelihood is

$$\begin{aligned} \ell(h_i; y_i(h_i(t^*)), \mu_i(h_i(t^*)) &= y_i(h_i(t^*)) \cdot \log(\hat{\mu}_i(h_i(t^*)) \\ &\quad + [1 - y_i(h_i(t^*))] \cdot \log(1 - \hat{\mu}_i(h_i(t^*)) \end{aligned} \tag{35.3}$$

In the Gaussian case, of course, maximizing the log likelihood in (35.2) is equivalent the minimizing the usual $\mathbb{L}^2$ loss; in this sense our proposed approach generalizes existing methods. In practice, estimation through (35.2) is augmented by constraints that $h_i(0) = 0$, $h_i(1) = 1$, and $h_i(t^*)$ is monotone increasing.

The preceding steps suggest an iterative algorithm for the joint estimation of the FPCA model and the warping functions. After obtaining initial estimates of the $h_i(t^*)$, one iterates between the following steps:

- Given the current estimates $\hat{h}_i(t^*)$, estimate the parameters in model (35.1).
- Given the current estimates $\hat{\mu}_i(t)$, estimate the warping functions through (35.2).

These steps are alternated until the change in successive iterations is negligible.

Because the approach to estimate the parameters in model (35.1) differs from the usual covariance decomposition, we now provide a brief description. The mean

function $\alpha(t)$ and the basis functions $\psi_k(t)$ are expanded using a cubic B-spline basis $\Theta(t) = \{\theta_1(t), \ldots, \theta_{K_\theta}(t)\}$. To ensure smoothness, the spline coefficients for the mean and each basis function are assumed to be multivariate Normal with a covariance structure that penalizes magnitude of the second derivative. Scores are treated as random effects. The resulting parameters of this specification can be estimated using an EM algorithm, sampling-based Bayesian approaches, or variational Bayes; see [6] for details. Meanwhile, the warping functions $h_i(t^*)$ are also expanded using a spline basis with coefficients estimated through constrained optimization.

35.3 Data Analysis

The motivation for this manuscript is to align patterns of physical activity measured using accelerometers in a sample of elderly subjects enrolled in the Baltimore Longitudinal Study on Aging (BLSA) [16]. Continuous monitoring of activity using accelerometers has revolutionized the measurement of physical activity by providing objective, unbiased, and detailed observation around the clock. Accelerometers generally measure physical activity through "activity counts", which are devised by summarizing the voltage signals across a monitoring period, and exemplify exponential-family functional data. In the BLSA, a study of normative human aging, participants wore the Actiheart, a combined heart rate and physical activity monitor adhesively placed on the chest [1]. Subjects were asked to wear the device at all times other than bathing or swimming.

For 500 subjects, we observe activity count trajectories measured over 24 hours. Binary "activity" vs "inactivity" trajectories are derived from these by thresholding activity counts at 20. We apply the methods described in Section 35.2 to obtain estimated warping functions for both binary and count trajectories; for the former we assume a binomial distribution and logit link, and for the latter we assume a Poisson distribution and log link. To estimate the parameters in model (35.1) we use a B-spline basis of dimension $K_\theta = 10$ and truncate the expansion to use $K = 1$ basis function. Warping functions $h_i(t^*)$ are estimated using a B-spline basis with 5 functions. To reduce the computational burden of the analysis, data were thinned to one data point for every 10 minutes, giving 144 observations per subject.

Figure 35.1 illustrates our results for binary activity trajectories. In both panels, we plot Gaussian kernel smooths with a narrow bandwidth of the observed binary data, which can be loosely interpreted as trajectories showing the probability of being active. The observed data show a diurnal pattern of binary activity, with lower probability of activity in the nighttime and higher probability of activity during the day. The right panel shows aligned data; after alignment, there are clear morning and afternoon active periods, and a mid-day dip that was much less pronounced in the observed data. Several subjects are highlighted in both panels to give a sense of the phase shift in the observed data and the effect of registration.

Warping functions $h_i(t^*)$ obtained in the registration of binary activity trajectories are shown in Figure 35.2. These suggest a non-trivial degree of misalignment in the

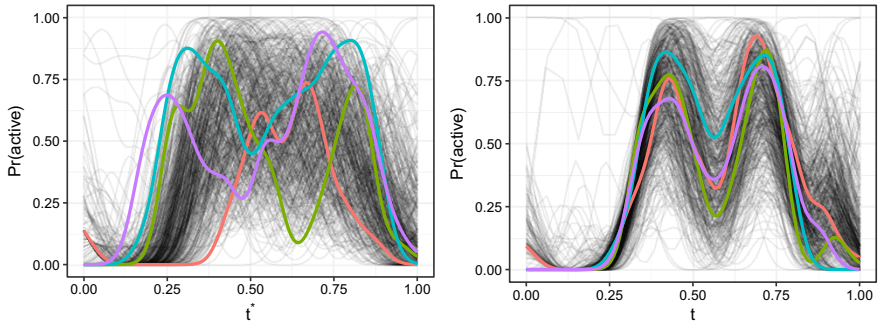


Fig. 35.1: Analysis results for binary activity trajectories. The left panel shows kernel smooths of binary activity against observed time; the right panel shows kernel smooths of binary activity against aligned time.

observed data. Additional analyses to examine this source of variability may be of interest. For reference, the subjects highlighted in Figure 35.1 are highlighted in Figure 35.2

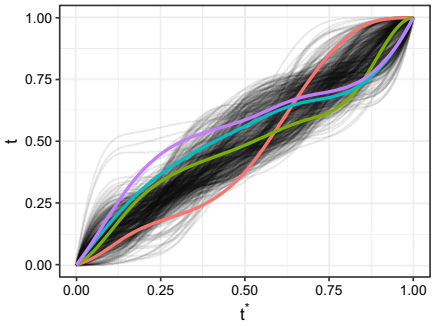


Fig. 35.2: Warping functions $h_i(t^*)$ obtained in the registration of binary activity trajectories.

To demonstrate that our methods are applicable for a variety of data distributions, we briefly present the results for activity count trajectories. Figure 35.3 shows the observed activity count trajectories in the left panel. The data show a diurnal pattern of activity similar to that for binary activity trajectories. A careful examination of the observed data suggests somewhat higher activity in the morning than in the evening. The results after alignment also suggest periods of increased activity intensity in the morning and in the afternoon, and an obvious mid-day dip.

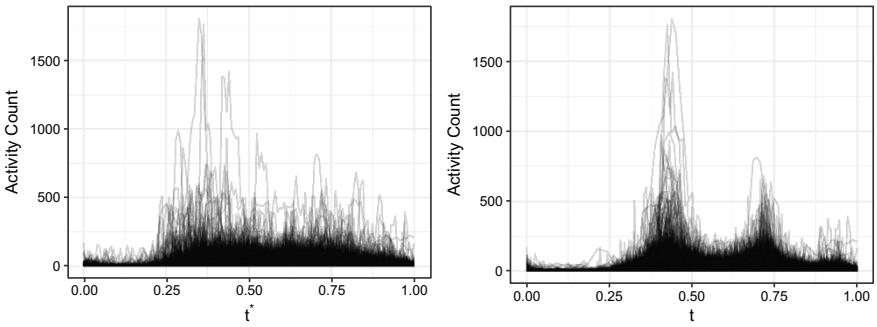


Fig. 35.3: Analysis results for activity count trajectories. The left panel shows activity counts against observed time; the right panel shows activity counts against aligned time.

Acknowledgements The authors' research was supported in part by Award R01NS097423-01 from the National Institute of Neurological Disorders and Stroke. The second author's research was additionally supported in part by Award R21EB018917 from the National Institute of Biomedical Imaging and Bioengineering.

References

[1] Brage, S., Brage, N., Ekelund, U., Luan, J., Franks, P. W., Froberg, K., Wareham, N. J.: Effect of combined movement and heart rate monitor placement on physical activity estimates during treadmill locomotion and free-living. European Journal of Applied Physiology, **96**:517–524, (2006)

[2] Brumback, L. C., Lindstrom, M. J.: Self modeling with flexible, random time transformations. Biometrics, **60**:461–470, (2004)

[3] Gasser, T., Kneip, A.: Searching for structure in curve samples. J. Amer. Statist. Assoc., **90**:1179–1188, (1995)

[4] Gertheiss, J., Goldsmith, J., Staicu, A.-M.: A note on modeling sparse exponential-family functional response curves. Comput. Statist. Data Anal., **105**:46–52, (2017)

[5] Gertheiss, J., Maier, V., Hessel, E., Staicu, A.-M.: Marginal functional regression models for analyzing the feeding behavior of pigs. J. Agric. Biol. Environ. Stat., **20**:353–370, (2015)

[6] Goldsmith, J., Zipunnikov, V., Schrack, J.: Generalized multilevel function-on-scalar regression and principal component analysis. Biometrics, **71**:344–353, (2015)

[7] Hall, P., Müller, H.-G., Yao, F.: Modelling sparse generalized longitudinal observations with latent gaussian processes. J. R. Stat. Soc. Ser. B Stat. Methodol., **70**:703–723, (2008)

[8] Huang, H., Li, Y., Guan, Y.: Joint modeling and clustering paired generalized longitudinal trajectories with application to cocaine abuse treatment data. J. Amer. Statist. Assoc., **109**:1412–1424, (2014)
[9] James, G. M., Hastie, T. J., Sugar, C. A.: Principal component models for sparse functional data. Biometrika, **87**:587–602, (2000)
[10] Kneip, A., Ramsay, J. O.: Combining registration and fitting for functional models. J. Amer. Statist. Assoc., **103**:1155–1165, (2008)
[11] Marron, J. S., Ramsay, J. O., Sangalli, L. M., Srivastava, A.: Functional data analysis of amplitude and phase variation. Statist. Sci., **30**:468–484, (2015)
[12] Ramsay, J. O., Silverman, B. W.: Functional Data Analysis. New York: Springer, (2005)
[13] Sakoe, H. and Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE transactions on acoustics, speech, signal processing, **26**:43–49, (1978)
[14] Sangalli, L. M., Secchi, P., Vantini, S., Veneziani, A.: A case study in exploratory functional data analysis: geometrical features of the internal carotid artery. J. Amer. Statist. Assoc., **104**:37–48, (2009)
[15] Scheipl, F., Gertheiss, J., Greven, S.: Generalized functional additive mixed models. Electron. J. Stat., **10**:1455–1492, (2015)
[16] Schrack, J. A., Zipunnikov, V., Goldsmith, J., Bai, J., Simonshick, E. M., Crainiceanu, C. M., Ferrucci, L.: Assessing the "physical cliff": Detailed quantification of aging and physical activity. Journal of Gerontology: Medical Sciences, (2014)
[17] Serban, N., Staicu, A.-M., Carrol, R. J.: Multilevel cross-dependent binary longitudinal data. Biometrics, **69**:903–913, (2013)
[18] Srivastava, A., Wu, W., Kurtek, S., Klassen, E., Marron, J. S.: Registration of functional data using Fisher-Rao metric. arXiv preprint arXiv:1103.3817, (2011)
[19] Telesca, D., Inoue, L. Y. T.: Bayesian hierarchical curve registration. J. Amer. Statist. Assoc., **103**:328–339, (2008)
[20] van der Linde, A.: A Bayesian latent variable approach to functional principal components analysis with binary and count. Adv. Stat. Anal., **93**:307–333, (2009)
[21] Vantini, S.: On the definition of phase and amplitude variability in functional data analysis. TEST, **21**:676–696, (2012)
[22] Yao, F., Müller, H.-G., Wang, J.: Functional data analysis for sparse longitudinal data. J. Amer. Statist. Assoc., **100**(470):577–590, (2005)

Chapter 36
Random functional variable and Fourier series

Jiří Zelinka

Abstract This paper presents how a functional random variable can be expressed in the form of Fourier series. This expansion can be used for the definition of components of the functional random variable and for the approximation of the random curves as the partial sum of the Fourier series. Thus we can estimate the distribution of the components, simulate the functional random variable with given components and compute some characteristics of the distribution of its norm.

36.1 Introduction

Statistical methods are often based on the properties of distribution of random variables or random vectors. We try to estimate their unknown distribution to obtain important knowledge about the data. In the case of functional data analysis (FDA) we do not work with random observation containing a finite random vector but the whole function is one observation. We call it *the functional random variable*. It is clear that the standard approach that used for instance cumulative distribution functions is inapplicable in this case. Non-parametric methods, namely kernel smoothing, proved far more useful (see [2], [1] and many citations inside).

Simulations are also very important for testing statistical methods. In FDA there are infinitely many parameters that affect the observation. This fact makes it difficult to simulate an observation of the functional random variable and the authors mostly use only several parameters in simulations (see [4], for instance). Application of the Fourier series in simulations allows to use of any number of parameters. So any observation of the functional random variable can be approximated with arbitrary accuracy.

Jiří Zelinka (✉)
Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Kotlářská 2, Brno, Czech Republic, e-mail: zelinka@emath.muni.cz

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_36

We will name the functional random variable as *random function* or *random curve* for brevity. Let us suppose that random function χ is a measurable mapping of a probability space $(\Omega, \mathscr{A}, P)$ into a real separable Hilbert space H. In this case there exists the complete orthonormal system $\Psi = (\psi_n)_{n=0}^{\infty}$ in the Hilbert space H and for every $\omega \in \Omega$ can be $\chi(\omega)$ expressed by the Fourier series

$$\chi(\omega) = \sum_{n=0}^{\infty} X_n(\omega)\psi_n, \tag{36.1}$$

where

$$X_n(\omega) = (\chi(\omega), \psi_n) \tag{36.2}$$

and

$$\|\chi(\omega)\|^2 = \sum_{n=0}^{\infty} X_n(\omega)^2 < \infty. \tag{36.3}$$

We can call the one-dimensional random variables X_n, $n = 0, 1, 2, \ldots$ given by the relationship (36.2) as *components of the random function* χ *with respect to the orthonormal system* Ψ.

Stochastic independence of components of a random vector is very important. So we can say that *the components* X_n, $n = 0, 1, 2, \ldots$ *of the random function* χ *with respect to the orthonormal system* Ψ *are (mutually) independent if every finite subset of the components is (mutually) independent.*

In functional data analysis the probability of small balls $B(\chi, h)$ is also an important aim:

$$B(\chi, h) = \{\eta \in H; \|\chi - \eta\| < h\}.$$

If

$$\chi = \sum_{n=0}^{\infty} x_n \psi_n$$

then the probability of the ball $B(\chi, h)$ is given by the probability of the set of all $\omega \in \Omega$ for which

$$\sum_{n=0}^{\infty} (x_n - X_n(\omega))^2 < h^2. \tag{36.4}$$

It is clear that the information about the distributions of the components X_n can help us to evaluate the probability of small balls. On the other had we can made simulations using components of the random function with predetermined distributions and obtain the exact values of some estimated quantities by this way.

36.2 Decomposition into components in L^2

Let's have the random sample of the random function χ in the sense that we have observations of independent random functions $\chi_1, \ldots, \chi_N$. We can also take this random sample as the values of χ in (unknown) points $\omega_1, \ldots \omega_N$, i.e. $\chi_j = \chi(\omega_j)$.

Typically, the random function χ_j is given only by its values in some discrete set of nodes $x_1,\ldots,x_M$.

The spectrometric dataset cited in many publications (see for instance [2], [3]) can serve as the basic example of the random sample of random function. This data are given as the sample of 210 random curves, each of them consists of values of absorbance measured at 100 wavelengths from 850 to 1050 nm. First ten curves from this dataset are displayed in Figure 36.1.

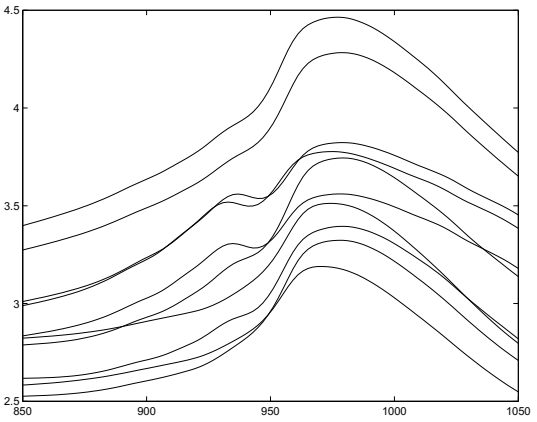


Fig. 36.1: Spectrometric data: first 10 curves

The scalar product $X_n(\omega) = (\chi(\omega), \psi_n)$ can be approximated for $\chi(\omega) \in L^2(a,b)$ by the value $\tilde{X}_n(\omega)$ using some quadrature formula. Composite trapezoidal or Simpson's rule are appropriate numerical methods for this purpose. Thus we evaluate approximation of the Fourier series of the observation of the random function χ_j

$$\chi_j = \chi(\omega_j) \approx \sum_{n=0}^{\infty} \tilde{X}_n(\omega_j)\psi_n.$$

Figure 36.2 presents the partial sum of 4 or 11 members of the cosine Fourier series (solid line) for the first random curve from the spectrometric data set (dashed line), respectively. We can see that the second approximation of the random curve is good enough.

The distributions of the components can be tested by many statistical tools. In the Figure 36.3 we can see the histograms of components X_0 and X_1 for all 215 curves from spectrometric dataset. The approximations $\tilde{X}_0$ and $\tilde{X}_1$ were used, of course.

A good view into the distributions of the components of the random function is given by boxplot. The boxplots for random variables $X_0,\ldots,X_{10}$ are displayed in Figure 36.4.

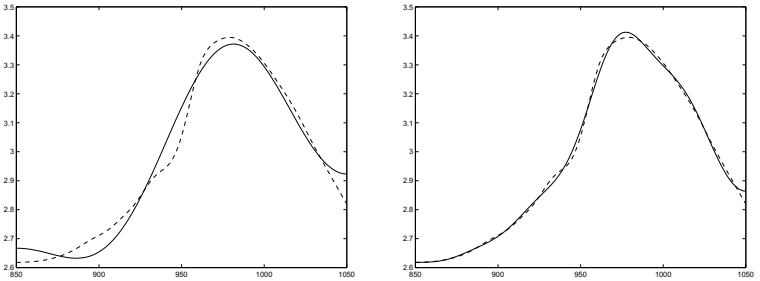


Fig. 36.2: Approximations of random curve – sum of 4 and 11 members of cosine series

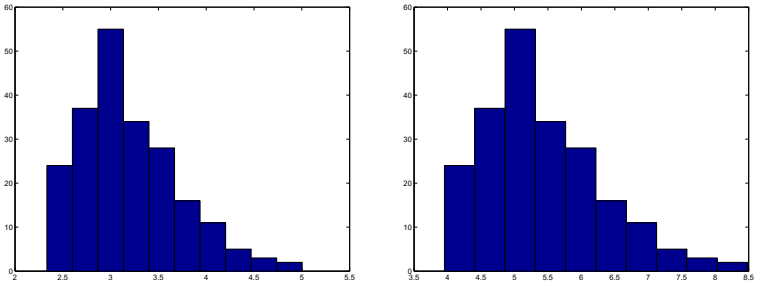


Fig. 36.3: Histograms for random variables X_0 and X_1.

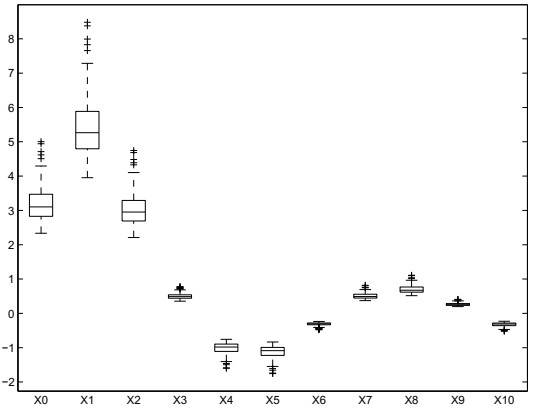


Fig. 36.4: Boxplots for $X_0,\ldots,X_{10}$

36.3 Simulations

The first idea how to simulate random function is simulate its components X_n, $n = 0,1,2,dots$ and then create the random function χ in the form

$$\chi = \sum_{n=0}^{\infty} X_n \psi_n$$

for given orthonormal system $\Psi = (\psi_n)_{n=0}^{\infty}$. If we simulate the independent random samples $X_{n,1}, \ldots, X_{n,N}$, $n = 0, 1, 2, \ldots$, we obtain arbitrarily accurate approximation of the random sample $\chi_1, \ldots, \chi_N$ as the partial sum of the Fourier series

$$\chi_j = \sum_{n=0}^{K} X_{n,j} \psi_n$$

for some $K \in \mathbf{N}$.

We mustn't forget the condition $\sum_{n=0}^{\infty} X_n(\omega)^2 < \infty$. This condition may not be met for every $\omega \in \Omega$, but just when the set where it is not fulfilled has probability equal to zero. Or equivalently

$$P\left(\sum_{n=0}^{\infty} X_n^2 < \infty\right) = 1. \tag{36.5}$$

The validity of this condition have to be ensured by selecting the appropriate distributions of the components $X_0, X_1, \ldots$. Sufficient condition can be derived using Markov's inequality

$$P(X \geq a) \leq \frac{E(X)}{a}$$

for non-negative random value X and $a > 0$. Denoting

$$Y_N = \sum_{n=0}^{N} X_n^2$$

we get

$$P\left(\sum_{n=0}^{N} X_n^2 \geq N\right) = P(Y_N \geq N) \leq \frac{E(Y_N)}{N} = \frac{E\left(\sum_{n=0}^{N} X_n^2\right)}{N} = \frac{\sum_{n=0}^{N} E(X_n^2)}{N}$$

for independent components $X_1, X_2 \ldots$.

Now, it can be easily seen that if $\sum_{n=0}^{\infty} E(X_n^2) = \sigma^2 < \infty$ then the probability $P\left(\sum_{n=0}^{N} X_n^2 \geq N\right)$ is arbitrarily small for N sufficiently large. So we have sufficient condition for (36.5) in the form

$$\sum_{n=0}^{\infty} E(X_n^2) < \infty. \tag{36.6}$$

36.4 Normally distributed components

Now let us suppose independent and normally distributed components of random function χ. For simplicity let us assume that the components X_n are of zero mean, i.e. $X_n \sim N(0, \sigma_n^2)$. In the case $E(X_n) = \mu_n \neq 0$ the following considerations should be applied to random values $X_n - \mu_n$.

As it follows from the relationships in the section 36.1 the probability $P(\|\chi\| < h)$ is equal to the probability $P\left(\sum_{n=0}^{\infty} X_n^2 < h^2\right)$. Due to this fact we focus on properties of the distribution of $\sum_{n=0}^{\infty} X_n^2 =: Y$.

The random value X_n^2 has chi-squared distribution with 1 degree of freedom. Its probability density function is

$$\frac{1}{\sigma_n \sqrt{1\pi x}} e^{-\frac{x}{2\sigma_n^2}}$$

and $E(X_n^2) = \sigma_n^2$, $\sum_{n=0}^{\infty} \sigma_n^2 < \infty$.

The characteristic function of X_n^2 takes the form

$$\varphi_{X_n^2}(t) = \frac{1}{\sqrt{1 - i2\sigma_n^2 t}}.$$

Now it is seen that the characteristic function of $\sum_{n=0}^{\infty} X_n^2$ is

$$\varphi_Y(t) = \frac{1}{\sqrt{\prod_{n=0}^{\infty} (1 - i2\sigma_n^2 t)}}. \tag{36.7}$$

We have to answer the question of convergence of the infinite product in (36.7). The following theorem from the 15-th chapter of [5] solves this problem:

Theorem 36.1. *Suppose $\{u_n\}$ is a sequence of bounded complex functions on a set S, such that $\sum_{n=0}^{\infty} |u_n(t)|$ converges uniformly on S. Then the product $\prod_{n=0}^{\infty} (1 + u_n(t))$ converges uniformly on S.*

The functions $u_n(t)$ are equal to $-i2\sigma_n^2 t$ and $\sum_{n=0}^{\infty} |u_n(t)|$ converges uniformly on any compact set S, because $\sum_{n=0}^{\infty} |u_n(t)| = 2t \sum_{n=0}^{\infty} \sigma_n^2$. So the infinite product in (36.7) converges uniformly on any compact set S. It is also clear that

$$\varphi_Y(t) = \frac{1}{\prod_{n=0}^{\infty} \sqrt{1 - i2\sigma_n^2 t}}.$$

Theoretically, the cumulative distribution function of Y can be evaluated using the formula (see [6])

$$F_Y(x) = \frac{1}{2} + \frac{i}{2\pi} p.v. \int_{-\infty}^{\infty} e^{-itx} \frac{\varphi_Y(t)}{t} dt$$

where $p.v.$ denotes the Cauchy principal value. The question of the practical application of this relationship remains open but the characteristic function φ_Y can be well used for the calculation of the moments of Y^k because of the equality

$$E(Y^k) = (-i)^k \varphi_Y^{(k)}(0).$$

Differentiation of φ_Y gives

$$\varphi_Y'(t) = \sum_{n=0}^{\infty} \frac{1}{\prod_{m \neq n} \sqrt{1 - i2\sigma_m^2 t}} \cdot \left(-\frac{1}{2} \frac{-2i\sigma_n^2}{(1 - i2\sigma_n^2 t)^{3/2}} \right) = \varphi_Y(t) \sum_{n=0}^{\infty} \frac{i\sigma_n^2}{1 - i2\sigma_n^2 t}.$$

This formula gives to us usable tool for computation of higher derivatives, e.g. for the second derivation of φ_Y we obtain

$$\varphi_Y''(t) = -\varphi_Y(t) \left[\left(\sum_{n=0}^{\infty} \frac{\sigma_n^2}{1 - i2\sigma_n^2 t} \right)^2 + \sum_{n=0}^{\infty} \frac{2\sigma_n^4}{(1 - i2\sigma_n^2 t)^2} \right].$$

As $\varphi_Y(0) = 1$ we have

$$E(Y) = (-i)\varphi_Y'(0) = \sum_{n=0}^{\infty} \sigma_n^2,$$

$$E(Y^2) = -\varphi_Y''(0) = \left(\sum_{n=0}^{\infty} \sigma_n^2 \right)^2 + 2 \sum_{n=0}^{\infty} \sigma_n^4$$

and the higher moments can be easily evaluated, too.

Acknowledgements This work was supported by Czech Science Foundation: GA15 - 06991S.

References

[1] Ferraty, F. ed.:Recent Advances in Functional Data Analysis and Related Topics. Springer Science & Business Media (2011)

[2] Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics, Springer, New York (2006)
[3] NPFDA web page: https://www.math.univ-toulouse.fr/staph/npfda/
[4] Rachdi, M., Vieu P.: Nonparametric regression for functional data: Automatic smoothing parameter selection. J. Statist. Plann. Inference **137**(9), 2784-2801, (2007)
[5] Rudin, W.: Real and Complex Analysis. McGraw Hill (1970)
[6] Ushakov, N.G.: Selected Topics in Characteristic Functions, De Gruyter (1999)

Authors Index

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2