

Frédéric Ferraty and Philippe Vieu

# Nonparametric Functional Data Analysis

Theory and Practice

April 18, 2006

Springer

Berlin Heidelberg New York

Hong Kong London

Milan Paris Tokyo

---

## Preface

This work is the fruit of recent advances concerning both nonparametric statistical modelling and functional variables and is based on various publications in international statistical reviews, several post-graduate courses and international conferences, which are the result of several years of research. In addition, all these developments have their roots in the recent infatuation for functional statistics. In particular, the synergy around the activities of the working group STAPH is a permanent source of inspiration in these statistical functional fields.

This book presents in a original way new nonparametric statistical methods for functional data analysis. Because we support the idea that statistics should take advantage of interactions between applied and theoretical aspects, we deliberately decided not to privilege one over the other. So, this work proposes two levels of reading. Recent theoretical advances, as far as possible, are presented in self-contained sections while statistical methodology, practical aspects, and elementary mathematics are accessible to a very large public. But, in any case, each part of this book starts with the presentation of general ideas concerning theoretical as well as applied issues.

This book could be useful as well for practitioners as for researchers and students. Non expert researchers and students will find detailed proofs and mathematical tools for theoretical advances presented in this book. For experienced researchers, these advances have been selected to balance the trade-off between comprehensive reading and up-to-date results. Because nonparametric functional statistics is a recent field of research, we discuss the existing bibliography by emphasizing open problems. This could be the starting point for further statistical developments. Practitioners will find short descriptions on how to implement the proposed methods while the companion website (<http://www.lsp.u-ps-tlse.fr/staph/npfda>) includes large details for codes, guidelines, and examples of use. So, the use of such nonparametric functional procedures will be easy for any users. In this way, we can say that this book is really intended for a large public: practitioners, theoreticians and anybody else who is interested in both aspects.

The novelty of nonparametric functional statistics obliges us to start by clarifying the terminology, by presenting the various statistical problems and by describing the kinds of data (mainly curves). Part I is devoted to these generalities. The remaining parts consist in describing the nonparametric statistical methods for functional data, each of them being basically split into theoretical, applied, and bibliographical issues. Part II focuses on prediction problems involving functional explanatory variables and scalar response. We study regression, conditional mode and conditional quantiles and their kernel nonparametric estimates. Part III concerns the classification of functional data. We focus successively on curve discrimination (prediction of a categorical response corresponding to the class membership) and unsupervised classification (i.e., the class membership is unobserved). Because time series can be viewed as a particular case of functional dataset, we propose in Part IV to extend most of the previous developments to dependent samples of functional data. The dependence structure will be taken into account through some mixing notion. In order to keep the main body of the text clear, theoretical tools are put at the end of this monograph in the appendix.

All the routines are implemented in the R and S+ languages and are available on the companion website (<http://www.lsp.ups-tlse.fr/staph/npfda>). S+ is an object-oriented language intensively used in engineering and applied mathematical sciences. Many universities, institutions and firms use such a software which proposes just as well a very large number of standard statistical methods as a programming language for implementing and popularizing new ones. In addition, all subroutines are translated into R because many other people work with such software, which is a free-version of S+ developed by academic researchers.

Science finds its source in the collective knowledge which is based on exchanges, collaborations and communications. So, as with any scientific production, this book has taken many benefits from contacts we had along the last few years. We had the opportunity to collaborate with various people including A. Ait-Saidi, G. Aneiros, J. Bougaran, C. Camlong, H. Cardot, V. Couallier, S. Dabo-Niang, G. Estévez, W. Gonzalez-Manteiga, L. Györfi, A. Goia, W. Härdle, J. Hart, I. Horova, R. Kassa, A. Laksaci, A. Mas, S. Montcaup, V. Nuñez-Antón, L. Pélégrina, A. Quintela del Río, M. Rachdi, J. Rodriguez-Poo, P. Sarda, S. Sperlicht and E. Youndjé, and all of them have in some sense indirectly participated to this work. Many other statisticians including J. Antoch, D. Bosq, A. Cuevas, A. Kneip, E. Kontoghiorghes, E. Mammen, J.S. Marron, J. Ramsay and D. Tjostheim have also been useful and fruitful supports for us.

Of course, this book would not have become reality without the permanent encouragements of our colleagues in the working group STAPH in Toulouse. This group acting on functional and operatorial statistics is a source of inspiration and in this sense, A. Boudou, H. Cardot, Y. Romain, P. Sarda and S. Viguier-Pla are also indirectly involved in this monograph. We would also like to express our gratitude to the numerous participants in the activities of

STAPH, with special thanks to J. Barrientos-Marin and L. Delsol for their previous reading of this manuscript and their constructive comments.

Gérard Collomb (1950-1985) was a precursor on nonparametric statistics. His international contribution has been determinant for the development of this discipline, and this is particularly true in Toulouse. Undoubtly, his stamp is on this book and we wish to take this opportunity for honoring his memory.

*Frédéric Ferraty*  
*Philippe Vieu*  
Toulouse, France  
January, 2006

---

# Contents

Preface	VII
List of Abbreviations and Symbols	XVII
List of Figures	XIX

---

## Part I Statistical Background for Nonparametric Statistics and Functional Data

---

<b>1</b>	<b>Introduction to Functional Nonparametric Statistics</b>	5
1.1	What is a Functional Variable?	5
1.2	What are Functional Datasets?	6
1.3	What are Nonparametric Statistics for Functional Data	7
1.4	Some Notation	9
1.5	Scope of the Book	10
<b>2</b>	<b>Some Functional Datasets and Associated Statistical Problematics</b>	11
2.1	Functional Chemometric Data	11
2.1.1	Description of Spectrometric Data	12
2.1.2	First Study and Statistical Problems	13
2.2	Speech Recognition Data	15
2.2.1	What are Speech Recognition Data?	15
2.2.2	First Study and Problematics	15
2.3	Electricity Consumption Data	17
2.3.1	The Data	17
2.3.2	The Forecasting Problematic	18
<b>3</b>	<b>What is a Well-Adapted Space for Functional Data?</b>	21
3.1	Closeness Notions	21

3.2	Semi-Metrics as Explanatory Tool . . . . .	22
3.3	What about the Curse of Dimensionality? . . . . .	25
3.4	Semi-Metrics in Practice . . . . .	28
3.4.1	Functional PCA: a Tool to Build Semi-Metrics . . . . .	28
3.4.2	PLS: a New Way to Build Semi-Metrics . . . . .	30
3.4.3	Semi-metrics Based on Derivatives . . . . .	32
3.5	R and S+ Implementations . . . . .	33
3.6	What About Unbalanced Functional Data? . . . . .	33
3.7	Semi-Metric Space: a Well-Adapted Framework . . . . .	35
<b>4</b>	<b>Local Weighting of Functional Variables . . . . .</b>	<b>37</b>
4.1	Why Use Kernel Methods for Functional Data? . . . . .	37
4.1.1	Real Case . . . . .	38
4.1.2	Multivariate Case . . . . .	39
4.1.3	Functional Case . . . . .	41
4.2	Local Weighting and Small Ball Probabilities . . . . .	42
4.3	A Few Basic Theoretical Advances . . . . .	43

---

## Part II Nonparametric Prediction from Functional Data

---

<b>5</b>	<b>Functional Nonparametric Prediction Methodologies . . . . .</b>	<b>49</b>
5.1	Introduction . . . . .	49
5.2	Various Approaches to the Prediction Problem . . . . .	50
5.3	Functional Nonparametric Modelling for Prediction . . . . .	52
5.4	Kernel Estimators . . . . .	55
<b>6</b>	<b>Some Selected Asymptotics . . . . .</b>	<b>61</b>
6.1	Introduction . . . . .	61
6.2	Almost Complete Convergence . . . . .	62
6.2.1	Regression Estimation . . . . .	62
6.2.2	Conditional Median Estimation . . . . .	66
6.2.3	Conditional Mode Estimation . . . . .	70
6.2.4	Conditional Quantile Estimation . . . . .	76
6.2.5	Complements . . . . .	76
6.3	Rates of Convergence . . . . .	79
6.3.1	Regression Estimation . . . . .	79
6.3.2	Conditional Median Estimation . . . . .	80
6.3.3	Conditional Mode Estimation . . . . .	87
6.3.4	Conditional Quantile Estimation . . . . .	90
6.3.5	Complements . . . . .	92
6.4	Discussion, Bibliography and Open Problems . . . . .	93
6.4.1	Bibliography . . . . .	93
6.4.2	Going Back to Finite Dimensional Setting . . . . .	94
6.4.3	Some Tracks for the Future . . . . .	95

**7 Computational Issues** . . . . . 99

7.1 Computing Estimators . . . . . 99

7.1.1 Prediction via Regression . . . . . 100

7.1.2 Prediction via Functional Conditional Quantiles . . . . . 103

7.1.3 Prediction via Functional Conditional Mode . . . . . 104

7.2 Predicting Fat Content From Spectrometric Curves . . . . . 105

7.2.1 Chemometric Data and the Aim of the Problem . . . . . 105

7.2.2 Functional Prediction in Action . . . . . 106

7.3 Conclusion . . . . . 107

**Part III Nonparametric Classification of Functional Data**

**8 Functional Nonparametric Supervised Classification** . . . . . 113

8.1 Introduction and Problematic . . . . . 113

8.2 Method . . . . . 114

8.3 Computational Issues . . . . . 116

8.3.1 kNN Estimator . . . . . 116

8.3.2 Automatic Selection of the kNN Parameter . . . . . 117

8.3.3 Implementation: R/S+ Routines . . . . . 118

8.4 Functional Nonparametric Discrimination in Action . . . . . 119

8.4.1 Speech Recognition Problem . . . . . 119

8.4.2 Chemometric Data . . . . . 122

8.5 Asymptotic Advances . . . . . 122

8.6 Additional Bibliography and Comments . . . . . 123

**9 Functional Nonparametric Unsupervised Classification** . . . . . 125

9.1 Introduction and Problematic . . . . . 125

9.2 Centrality Notions for Functional Variables . . . . . 127

9.2.1 Mean . . . . . 127

9.2.2 Median . . . . . 129

9.2.3 Mode . . . . . 130

9.3 Measuring Heterogeneity . . . . . 131

9.4 A General Descending Hierarchical Method . . . . . 131

9.4.1 How to Build a Partitioning Heterogeneity Index? . . . . . 132

9.4.2 How to Build a Partition? . . . . . 132

9.4.3 Classification Algorithm . . . . . 134

9.4.4 Implementation: R/S+ Routines . . . . . 135

9.5 Nonparametric Unsupervised Classification in Action . . . . . 135

9.6 Theoretical Advances on the Functional Mode . . . . . 137

9.6.1 Hypotheses on the Distribution . . . . . 138

9.7 The Kernel Functional Mode Estimator . . . . . 140

9.7.1 Construction of the Estimates . . . . . 140

9.7.2 Density Pseudo-Estimator: a.co. Convergence . . . . . 141

9.7.3 Mode Estimator: a.co. Convergence . . . . . 144

9.7.4	Comments and Bibliography . . . . .	145
9.8	Conclusions . . . . .	146

---

## Part IV Nonparametric Methods for Dependent Functional Data

---

<b>10</b>	<b>Mixing, Nonparametric and Functional Statistics . . . . .</b>	<b>153</b>
10.1	Mixing: a Short Introduction . . . . .	153
10.2	The Finite-Dimensional Setting: a Short Overview . . . . .	154
10.3	Mixing in Functional Context . . . . .	155
10.4	Mixing and Nonparametric Functional Statistics . . . . .	156
<b>11</b>	<b>Some Selected Asymptotics . . . . .</b>	<b>159</b>
11.1	Introduction . . . . .	159
11.2	Prediction with Kernel Regression Estimator . . . . .	160
11.2.1	Introduction and Notation . . . . .	160
11.2.2	Complete Convergence Properties . . . . .	161
11.2.3	An Application to the Geometrically Mixing Case . . . . .	163
11.2.4	An Application to the Arithmetically Mixing Case . . . . .	166
11.3	Prediction with Functional Conditional Quantiles . . . . .	167
11.3.1	Introduction and Notation . . . . .	167
11.3.2	Complete Convergence Properties . . . . .	168
11.3.3	Application to the Geometrically Mixing Case . . . . .	171
11.3.4	Application to the Arithmetically Mixing Case . . . . .	175
11.4	Prediction with Conditional Mode . . . . .	177
11.4.1	Introduction and Notation . . . . .	177
11.4.2	Complete Convergence Properties . . . . .	178
11.4.3	Application to the Geometrically Mixing Case . . . . .	183
11.4.4	Application to the Arithmetically Mixing Case . . . . .	184
11.5	Complements on Conditional Distribution Estimation . . . . .	185
11.5.1	Convergence Results . . . . .	185
11.5.2	Rates of Convergence . . . . .	187
11.6	Nonparametric Discrimination of Dependent Curves . . . . .	189
11.6.1	Introduction and Notation . . . . .	189
11.6.2	Complete Convergence Properties . . . . .	190
11.7	Discussion . . . . .	192
11.7.1	Bibliography . . . . .	192
11.7.2	Back to Finite Dimensional Setting . . . . .	192
11.7.3	Some Open Problems . . . . .	193
<b>12</b>	<b>Application to Continuous Time Processes Prediction . . . . .</b>	<b>195</b>
12.1	Time Series and Nonparametric Statistics . . . . .	195
12.2	Functional Approach to Time Series Prediction . . . . .	197
12.3	Computational Issues . . . . .	198
12.4	Forecasting Electricity Consumption . . . . .	198



12.4.1 Presentation of the Study ..... 198  
 12.4.2 The Forecasted Electrical Consumption ..... 200  
 12.4.3 Conclusions..... 201

---

**Part V Conclusions**

---

**13 Small Ball Probabilities and Semi-metrics** ..... 205  
 13.1 Introduction ..... 205  
 13.2 The Role of Small Ball Probabilities ..... 206  
 13.3 Some Special Infinite Dimensional Processes ..... 207  
     13.3.1 Fractal-type Processes ..... 207  
     13.3.2 Exponential-type Processes ..... 209  
     13.3.3 Links with Semi-metric Choice ..... 212  
 13.4 Back to the One-dimensional Setting ..... 214  
 13.5 Back to the Multi- (but Finite) -Dimensional Setting..... 219  
 13.6 The Semi-metric: a Crucial Parameter ..... 223

**14 Some Perspectives** ..... 225

**Appendix: Some Probabilistic Tools**..... 227  
 A.1 Almost Complete Convergence ..... 228  
 A.2 Exponential Inequalities for Independent r.r.v. .... 233  
 A.3 Inequalities for Mixing r.r.v. .... 235

**References** ..... 239

**Index** ..... 255