Title
Prospect of Nonparametric Modeling

Author
Jianqing Fan

Publication Date
2011-10-25

Prospects of nonparametric modeling *

Jianqing Fan

Department of Statistics, University of California, Los Angeles, CA 90095

Abstract

This article briefly summarizes key developments in nonparametric function estimation over the past three decades. It highlights some potentially fruitful areas of future research on nonparametric modeling and its applications.

1 Introduction

Modern computing facilities allow statisticians to explore fine data structures that were unimaginable two decades ago. Driven by many sophisticated applications, demanded by the need of nonlinear modeling and fueled by modern computing power, many computationally intensive data-analytic modeling techniques have been invented to exploit possible hidden structures and to reduce modeling biases of traditional parametric methods. These data-analytic approaches are also referred to as nonparametric techniques. For an introduction to these nonparametric techniques, see the books by Devroye and Györfi (1985), Silverman (1986), Eubank (1988), Müller (1988), Györfi, Härdle, Sarda and View (1989), Hastie and Tibshirani (1990), Wahba (1990), Scott (1992), Green and Silverman (1994), Wand and Jones (1995), Fan and Gijbels (1996), Simonoff (1996), Bowman and Azzalini (1997), Hart (1997), Ramsay and Silverman (1997), Ogden (1997), Bosq (1998), Efromovich (1999), Vidakovic (1999), among others.

An aim of nonparametric techniques is to reduce possible modeling biases of parametric models. Such parametric models are simple and convenient linear models to facilitate computational expediency before 1980s'. They are typically not derived from physical laws and can not be expected to fit all data well. An erroneous parametric model can create excessive modeling biases and leads to wrong conclusions. Nonparametric techniques intend to fit a much larger class of models to reduce modeling biases. They allow data to search appropriate nonlinear forms that best describe the available data. They also provide useful tools for parametric nonlinear modeling and for model diagnostics.

Over the past three decades, intensive efforts have been devoted to nonparametric function estimation. Many new nonparametric models have been introduced and a vast array of new techniques have been invented. Many new phenomena have been unveiled and deep insights have been gained. The field of nonparametric modeling has progressed steadily and dynamically. This trend will continue for decades to come. With the advance of information and technology, more and more complicated data mining problems emerge. The research and applications of data-analytic techniques will be proven to be even more fruitful in the next millennium.

The field of nonparametric modeling is vast. It has taken many books to describe a part of the art. Indeed, most of parametric models have their nonparametric counterpart. It is impossible to

*Fan's research was partially supported by NSF grant DMS-9804414 and a grant from University of California at Los Angeles.

give a complete survey of this wide field. Rather, the aim of this article is to highlight some of the important achievements and outline some potentially fruitful topics of research. For a more complete review of the literature, see the aforementioned books and the references therein.

2 Overview of developments

2.1 Density estimation and nonparametric regression

Density estimation summarizes data distributions via estimating underlying densities and nonparametric regression smoothes scatterplots via estimating regression functions. They provide the simplest setup for understanding nonparametric modeling techniques and serve as useful building blocks for high-dimensional modeling. They are relatively well developed and understood.

Many useful techniques have been proposed for univariate smoothing. Among those, kernel methods (Rosenblatt 1956; Gasser and Müller 1979; Müller 1988; Wand and Jones, 1995), local polynomial methods (Stone 1977; Cleveland 1979; Fan, 1993; Fan and Gijbels, 1996), spline methods (Wahba 1977; Eubank 1988; Nychka 1988; Wahba, 1990; Green and Silverman, 1994; Stone, *et al.* 1997), Fourier methods (Efromovich and Pinsker, 1982, Efromovich, 1999) and wavelet methods (Donoho and Johnstone 1994; Donoho, Johnstone, Kerkyacharian and Picard 1995; Hall and Patil 1995; Ogden, 1997, Antoniadis, 1999; Vidakovic, 1999). Different techniques have their own merits. Chapter 2 of Fan and Gijbels (1996) gives an overview of these techniques.

Each nonparametric technique involves selection of smoothing parameters. Several data-driven methods have been developed. Cross-validation (Allen 1974; Stone 1974; Rudemo 1982) and generalized cross-validation (Wahba, 1977) are generally applicable methods. Yet, their resulting bandwidths can vary substantially (Hall and Johnstone, 1991). Plug-in methods are more stable. In addition to the methods surveyed in Jones, Marron and Sheather (1996), the pre-asymptotic substitution method by Fan and Gijbels (1995) and the empirical-bias method by Ruppert (1997) provide useful alternatives. See also Marron and Padgett (1987).

2.2 Multivariate nonparametric modeling

Univariate smoothing techniques can be extended in a straightforward manner to multivariate settings. But such extensions are not useful due to so-called "curse-of-dimensionality". Many powerful models have been proposed to avoid using "saturated" nonparametric models and hence attenuate the problems of the "curse-of-dimensionality". Different models incorporate different knowledge into data analyses and explore different aspects of data. Examples include additive models (Breiman and Friedman, 1985; Hastie and Tibshirani 1990; Stone, 1994), varying coefficient models (Cleveland, *et al*, 1991, Hastie and Tibshirani, 1993), low-dimensional interaction models (Friedman 1991; Gu and Wahba, 1993; Stone, *et al* 1997), multiple-index models (Härdle and Stoker 1989, Li 1991), and partially linear models (Speckman 1988; Green and Silverman 1994), and their hybrids (Carroll, *et al.* 1997; Fan *et al* 1998), among others. See also semiparametric models in Bickel *et al.* (1994). They together form useful tool kits for processing data that arise from many scientific disciplines and for checking adequacy of commonly-used parametric models.

The area of multivariate data-analytic modeling is very dynamic. A vast array of innovative ideas have been proposed. Each method relies on certain univariate smoothing techniques as building blocks. To name a few, backfitting methods (Hastie and Tibshirani, 1990) and average regression surface methods (Tjöstheim, 1991, Linton and Nielsen, 1995) for additive modeling, sliced inverse regression method (Duan and Li, 1991 and Li, 1991) and average derivative methods (Härdle and Stoker 1990 and Samarov, 1993) for multiple-index models, and combinations of these methods (Carroll, *et al.*, 1997), among others. Polynomial splines and smoothing splines can be directly applied to low-dimensional interaction models.

Tree-based regression models (Friedman, *et al*, 1993; Zhang and Signer 1999) are based on different ideas. They are also powerful tools for nonparametric multivariate regression and classification.

Nonparametric regression problems arise often from other statistical contexts such as generalized linear models and the Cox proportional hazards model. While they is some theory and methods available, nonparametric techniques are relatively underdeveloped for likelihood and pseudo-likelihood models.

2.3 Theoretical developments

Apart from creative technological inventions, many foundational insights have been gained, and many new phenomena in infinite dimensional spaces have been discovered. It is now well known that many nonparametric functions can not be estimated at a root-n rate (Farrell 1972; Donoho and Liu 1991), while some functionals such as integrated square densities (Bickel and Ritov, 1988; Fan, 1991) can be estimated at a root-n rate. These optimal rates of convergence depend on the smoothness of the function classes. Adaptive procedures have been constructed so that they are nearly optimal for each given smoothness of a class of functions. See Lepski (1991, 1992), Donoho, Johnstone, Kerkyacharian and Picard (1995), Brown and Low (1996), among others. Adaptive estimation based on penalized least-squares can be found in Andrew, Birgé and Massart (1999). Optimal rates for hypothesis testing have also been developed. See Ingster (1993) and Spokoiny (1996).

Optimal rates for multivariate ANOVA types of nonparametric models offer valuable theoretical insights into high-dimensional function estimation problems (Stone, 1994, Huang, 1999). It was shown that asymptotically estimating a component in additive separable models is just as hard as the case when the other components are known (Fan, Mammen and Härdle, 1998). This property is not shared by parametric models.

3 Future research

With increasing complexity of statistical applications and the need for refinements of traditional techniques, the future of nonparametric modeling and its applications is bright and prosperous. Cross-fertilization of parametric and nonparametric techniques will be fruitful and powerful. Applications of nonparametric techniques to other scientific and engineering disciplines are increasingly demanding. While some areas of research are outlined here, they are far from exhaustive.

3.1 Nonparametric inferences

Maximum likelihood estimation, likelihood ratio statistics and the bootstrap offer generally applicable tools for parametric analysis. Yet, there are no generally applicable principles available for nonparametric inference. Consider the example of additive models. How can one construct simultaneous confidence bands for estimated functions? Are a set of variables significant in the models? Does a given nonlinear parametric model adequately fit the data or is a given nonparametric model over parametrized? While there are many collective efforts and much progress, the area still requires intensive research and widely applicable methods should be sought. Recently, Fan *et al* (1999) made a start in this direction via proposing a sieve likelihood ratio method and demonstrated that it possesses various good statistical properties..

3.2 High-dimensional nonparametric modeling

Many interesting statistical problems are multivariate and high-dimensional, with a mix of discrete and continuous variables. While they are a number of creative nonparametric models, they can not be expected to handle all of these statistical problems. A lack of inference tools and availability of software have hampered their applications. High-dimensional classification problems are in increasing demand. Applications of nonparametric modeling techniques to other statistical contexts need further developing.

3.3 Functional data analysis

Massive data sets can nowadays easily be collected for each individual in the form of curves or images. In ophthalmology, for example, images of a patient's cornea maps are recorded along with other demographic and ophthalmic variables. Interesting questions include studying associations between cornea shapes and demographic and ophthalmic variables, testing whether there are any differences among two or more clinical groups or treatment methods, and monitoring regression/progression of clinical surgery. More examples and problems can be found in Kneip and Gasser (1992), Capra and Müller (1997), and Ramsay and Silverman (1997). Feature extractions have been extensively studied (Ramsay and Silverman, 1997). Yet, predictions, modeling and inferences based on functional data need substantial developments.

3.4 Information engineering and signal processing

Modern telecommunications and information engineering create many challenging statistical problems and offer statisticians many golden opportunities. Data (signals or images) compression requires substantial dimensionality reduction techniques and estimation of high-dimensional conditional densities. Nonparametric techniques are vital to image analysis and modeling. Pattern (speech and character) recognition demands sophisticated nonlinear classification rules. Monitoring network traffic needs creative visualization tools. Dynamical prediction of an event based on observed signals such as designing airbag deployments is another challenge to statistics. Research in these complicated and computationally intensive engineering problems can be fruitful. Innovative data-analytic tools and interdisciplinary collaborations are needed.

3.5 Nonlinear time series and finance modeling

Interest in nonlinear time series has surged during the last decade (Tong, 1990). New features have been discovered and better predictions can be made (Yao and Tong, 1994). From linear modeling to nonlinear prediction, there are infinitely many possibilities. This offers nonparametric techniques a tremendous opportunity to reduce modeling biases and gain prediction power. Nonparametric multivariate regression techniques can be extended readily to time series setup. They can provide useful tools for discovery of nonlinear phenomena, understanding underlying dynamics, better forecasting and model diagnostics.

Modern asset pricing theory allows one to value and hedge contingent claims once a model for the dynamics of an underlying state variable is given (Duffie, 1996). Many such models have been developed, such as the geometric Brownian motion model and the interest-rate models. Most of these asset pricing models are simple and convenient parametric models. They are not derived from any economic theory and can not be expected to fit well for all financial data. Thus, while the pricing theory gives spectacularly beautiful formulas when the underlying dynamics is correctly modeled, it offers little guidance neither in choosing a correct model nor in validating a specific model. Hence there is always a danger that misspecification of a model will lead to erroneous valuation and hedging. Various extensions and relaxing of restrictive assumptions have been made. Nonparametric approaches have recently been introduced to estimate state price densities, instantaneous return and volatility (Aït-Sahalia and Lo 1996 and Stanton 1997). These have immediate applications to value bond price and stock options. Nonparametric applications to modeling financial data and to testing existing financial models are also fruitful.

3.6 Nonparametric modeling in biostatistics

Nonparametric smoothing techniques have been applied to estimate hazard functions and to the Cox proportional hazards model. Yet, there are many possibilities of extending linear models to multivariate nonparametric models (see §2.2) and there are many other biostatistical models that require nonparametric ameriolating. Further, there are many different types of incomplete data collected in epidemiological studies. Inference tools and model diagnostic techniques need further developments. Simple and powerful diagnostic tools for checking survival time models are useful.

Longitudinal data arise often from biostatistical studies. To monitor disease progression and to examine time effect, various parametric and nonparametric models (Hoover, *et al*, 1998) have been developed. Semiparametric modeling of covariance matrices and efficient estimation of time-varying coefficient functions are needed. We still lack inference tools to answer clinically important questions such as detecting if coefficients are really time varying or if certain covariate effects become more pronounced over time. Tools for predicting future events based on an individual's history are needed.

There are many other statistical problems arising from biostatistical and epidemiological applications. They pose new challenges and offer new opportunities for nonparametric modeling.

3.7 Applications to other statistical problems

There are many other statistical problems that require data-analytic tools. They offer statisticians enormous opportunities for interdisciplinary collaboration. Context-based applications of nonparametric techniques will be fruitful.

3.8 Software developments

Applications of nonparametric techniques have been hampered by availability of software. While many nonparametric techniques have been programmed by individual researchers, they were written in many computer languages and were only tested for "in-house" use. Many modern nonparametric techniques are not available in commonly-used statistical software packages. Research into fast and robust implementations of nonparametric techniques and their software developments are needed.

References

Aït-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica*, **64**, 527-560.

Allen, D.M. (1974). The relationship between variable and data augmentation and a method of prediction. *Technometrics*, **16**, 125–127.

Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, **113**, 301–413.

Antoniadis, A. (1999). Wavelets in Statistics: A Review (with discussion). *Italian Jour. Statist.*, to appear.

Bickel, P.J., Klaassen, A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Inference in Semi-parametric Models.* Johns Hopkins University Press, Baltimore.

Bickel, P.J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp order of convergence estimates. *Sankhyā Ser. A*, **50**, 381–393.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, 2nd ed., Lecture Notes in Statistics, **110**. Springer-Verlag, Berlin.

Breiman, L. and Friedman, J.H. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). *J. Amer. Statist. Assoc.*, **80**, 580–619.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1993). *CART: Classification and Regression Trees* (first edition, 1984). Wadsworth, Belmont.

Brown, L.D. and Low, M. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.*, **24**, 2524-2535.

Bowman, A.W. & Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis.* Oxford University Press, Oxford.

Capra, W.B. and Müller, H.G. (1997). An accelerated time model for response curves. *J. Amer. Statist. Assoc.*, **92**, 72–83.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, **74**, 829–836.

Cleveland, W.S., Grosse, E. and Shyu, W.M. (1991). Local regression models. In *Statistical Models in S* (Chambers, J.M. and Hastie, T.J., eds), 309–376. Wadsworth & Brooks, Pacific Grove.

Carroll, R.J., Fan, J., Gijbels, I, and Wand, M.P. (1997). Generalized partially linear single-index models. *Jour. Ameri. Satist. Assoc.*, **92**, 477-489

Devroye, L.P. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View.* Wiley, New York.

Donoho, D.L. and Johnstone, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? *J. Royal Statist. Soc. B*, **57**, 301–369.

Donoho, D.L. and Liu, R.C. (1991). Geometrizing rate of convergence III. *Ann. Statist.*, **19**, 668–701.

Duan, N. and Li, K.-C. (1991). Slicing regression: a link-free regression method. *Ann. Statist.*, **19**, 505–530.

Duffie, D. (1996). *Dynamic Asset Pricing Theory* (Second Edition). Princeton University Press, Princeton, New Jersey.

Efromovich, S. (1999). *Nonparametric Curve Estimation: Methods, Theory and Applications.* Springer-Verlag, New York.

Efromovich, S.Y. and Pinsker, M.S. (1982). Estimation of square-integrable probability density of a random variable. *Problems of Information Transmission*, **18**, 175–189.

Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression.* Marcel Dekker, New York.

Fan, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.*, **19**, 1273–1294.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiency. *Ann. Statist.*, **21**, 196–216.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. Royal Statist. Soc. B*, **57**, 371–394.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Fan, J., Härdle, W. and Mammen, E. (1998). Direct estimation of additive and linear components for high dimensional data (with W. Härdle and E. Mammen). *The Annals of Statistics*, **26**, 943-971.

Fan, J., Zhang, C. and Zhang, J. (1999). Sieve likelihood ratio statistics and Wilks phenomenon. *Technical report*, Department of Statistics, University of California at Los Angeles.

Farrell, R.H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.*, **43**, 170–180.

Friedman, J.H. (1991). Multivariate adaptive regression splines (with discussion). *Ann. Statist.*, **19**, 1–141.

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics, **757**, 23–68. Springer-Verlag, New York.

Green, P.J. and Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach.* Chapman and Hall, London.

Gu, C. and Wahba, G. (1993). Smoothing spline ANOVA with component-wise Bayesian "confidence intervals". J. Comput. Graph. Statist. 2 (1993), 97–117.

Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series.* Lecture Notes in Statistics, **60**. Springer-Verlag, Berlin.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Boston.

Härdle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.*, **84**, 986–995.

Hall, P. and Johnstone, I. (1992). Empirical functionals and efficient smoothing parameter selection (with discussion). *J. Royal. Statist. Soc. B*, **54**, 475–530.

Hall, P. and Patil, P. (1995). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.* **23**, 905–928.

Hart, J.D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* New York: Springer.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika, **85**, 809-822.

Hastie, T.J. and Tibshirani, R. (1990). *Generalized Additive Models.* Chapman and Hall, London.

Hastie, T.J. and Tibshirani, R. J. (1993). Varying-coefficient models. *Jour. Roy. Statist. Soc. B.*, **55**, 757-796.

Huang, J. (1999). Projection estimation in multiple regression with application to functional ANOVA models. *Annals of Statistics*, Vol 26, 242–272.

Ingster, Yu. I. (1993). Asymptotic minimax hypothesis testing for nonparametric alternatives I-III. *Math. Methods Statist*, **2**, 85-114; **3**, 171-189; **4**, 249-268.

Jones, M.C., Marron, J.S. and Sheather, S.J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.*, **91**, 401-407.

Lepski, O.V. (1991) Asymptotically minimax adaptive estimation I. *Theory Probab. Appl.*, 36, 4, 682-697.

Lepski, O.V. (1992) Asymptotically minimax adaptive estimation II. *Theory Probab. Appl.*, **37**, 3, 433-448.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.*, **86**, 316–342.

Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93-101.

Kneip, A. and Gasser, T. (1992). "Statistical tools to analyze data representing a sample of curves", *The Annals of Statistics*, 20, 1266-1305.

Marron, J.S. and Padgett, W.J. (1987). Asymptotically optimal bandwidth selection from randomly right-censored samples. *Ann. Statist.*, **15**, 1520–1535.

Müller, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data.* Lecture Notes in Statistics,

Nychka, D. (1995). Splines as local smoothers. *Ann. Statist.*, **23**, 1175–1197.

Ogden, T. (1997). *Essential wavelets for statistical applications and data analysis.* Birkhuser Boston, Boston.

Ramsay, J.O. and Silverman, B.W. (1997), "The Analysis of Functional Data", Springer-Verlag, Berlin.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832–837.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.*, **9**, 65–78.

Ruppert, D. (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *Jour. Ameri. Statist. Assoc.*, **92**, 1049–1062.

Samarov, A.M. (1993). Exploring regression structure using nonparametric functional estimation. *J. Amer. Statist. Assoc.*, **88**, 836–847.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Simonoff, J. S. (1996). *Smoothing Methods in Statistics.* Springer, New York.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Royal Statist. Soc. B*, **50**, 413–436.

Spokoiny, V.G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.*, **24**, 2477-2498.

Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley, New York.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *J. Finance*, **5**, 1973-2002.

Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5**, 595–645.

Stone, C.J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.*, **22**, 118–184.

Stone, C.J., Hansen, M., Kooperberg, C. and Truong, Y.K. (1997). Polynomial splines and their tensor products in extended linear modeling. With discussion and a rejoinder by the authors and Jianhua Z. Huang. *Ann. Statist.*, **25**, 1371–1470.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *J. Royal Statist. Soc. B*, **36**, 111–147.

Tjøstheim, D. and Auestad, B.H. (1994). Nonparametric identification of nonlinear time series: projections. *J. Amer. Statist. Assoc.* **89** 1398 - 1409.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach.* Oxford University Press, Oxford.

Vidakovic, B. (1999) *Statistical Modeling by Wavelets.* Wiley, New York.

Wahba, G. (1977). A survey of some smoothing problems and the method of generalized cross-validation for solving them. In *Applications of Statistics* (P.R. Krisnaiah, ed.), 507–523. North Holland, Amsterdam.

Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* Chapman and Hall, London.

Yao, Q. and Tong, H. (1994). On prediction and chaos in stochastic systems. *Phil. Tran. Roy. Soc. Lond. A*, **348**, 357–369.

Zhang, H.P. and Singer, B. (1999) *Recursive Partitioning in the Health Sciences.* Springer-Verlag: New York.