PROCEEDINGS OF

The First International Conference on Mathematics and Statistics (ICOMS-1)

June 19-21, 2006
Bandung, West Java - Indonesia

Enhancing Contribution of Mathematicians and Statisticians Toward World Peace and Prosperity

ISBN: 978-979-16363-0-8

Layout by : Aceng Komarudin Mutaqin & Riki Kurniawan
Universitas Islam Bandung, Indonesia

Published by : Moslem Statisticians and Mathematicians Society in South East Asia (MSMSSEA)

Cover Design : Denny Rahmat Ramdhani
Universitas Islam Bandung, Indonesia

PREFACE

Since November 2004 in South-East Asia region had stood organization of profession which called Moslem Statisticians and Mathematicians Society in South East Asia (MSMSSEA) and is centered in town Bandung. This organization open to mathematicians and statisticians from whole world which wish to congregate gives contribution in increasing peace and prosperity in South-East Asia region especially and world generally through various mathematics scientific and statistics.

In the effort realizing purpose of the MSMSSEA, cooperates with Institute for Mathematical Research - Universiti Putra Malaysia, Malaysian Mathematical Sciences Society, Indonesian Mathematics Society, and UNISBA has carried out "The First International Conference on Mathematics and Statistics (ICOMS-1)" on Junes 19 - 21, 2006, in Hotel Jayakarta Bandung. This conference attended by one hundred mathematicians and statisticians from various country, like from Australia, India, Canada, Malaysia, Pakistan, Iran, and Indonesia its self.

Some good articles in mathematics study, mathematics education and also statistics presented in this proceedings.

Hopefully is of benefit to all readers.

Yours faithfully,

President MSMSSEA,

Prof. Dr. Maman A Djauhari

Proceedings of ICoMS-1 (2006), pp. 151-157

Estimation of Spatio-Temporal Additive Model Using Mixed Model Approach with Application to Ozone Data in Surabaya City

Anik Djuraidah and Aunuddin
PhD student at Bogor Agriculture Institute
Lecturer at Satistics Departement of Bogor Agriculture Institute

Abstract

Spatio-temporal models arise when spatial data are collected over continuous time, so that one must take account of spatial correlations as well as temporal correlations. This research is focused in modelling spatio-temporal data using additive models. Spatio-temporal additive models are combination of time series additive models and spatial additive models. Functional form of predictors and response are modeled using P-spline. Since P-spline has connection with linear mixed models, the estimation of spatio-temporal additive model can be approached by linear mixed models. The models are applied to air pollution Ozone from AQMS in Surabaya. In modelling Ozone we add meteorological factors as covariates. This resulting model is used for spatial interpolation in unmonitored location. The study show that addition meteorological factors in spatio-temporal models improved the accuracy of models, although spatial interpolation are similar with models without meteorological factors.

Key words: *additive model, time series additive model, spatial additive model, P-spline, linear mixed model, smoothing parameter.*

1. Introduction

Spatio-temporal data are spatial data which are observed over continuous time. Spatio-temporal data consists of temporal correlations and spatial correlations which should be accounted in modelling. There are a lot of spatio-temporal models, expanded from the parametric method until nonparametric method.

The goal of nonparametric smoothing techniques is to estimate the regression function s in a relationships between a response variable y and a covariate x, expressed in the model

$$y_i = s(x_i) + \varepsilon_i \tag{1}$$

where ε_i denotes independent random error. When several covariates are present, Stone (1985) proposed to extend the idea of multiple regression into a flexible form known as an additive model. For data $\{(y_i, x_{i1}, \cdots, x_{ip}),\ i = 1, \cdots, n\}$, the models can be represented as

$$y_i = s_0 + \sum_{j=1}^{p} s_j(x_{ij}) + \varepsilon_i \tag{2}$$

Where the covariate $\mathbf{x}_j$ has its own associated component s_j and the regression function is constructed from the combination of the components. The errors $\boldsymbol{\varepsilon}$ are assumed to be independent. Hastie and Tibshirani (1990) proposed to extend the additive model to a wide range of distribution families known as GAM (generalized additive models). However, whilst it is flexible and efficient, the GAM framework based on backfitting with linear smoothers presents some difficulties when it comes to model selection and inference.

Eilers and Marx (1996), Ruppert and Carroll (1997) proposed low rank dimension of smoothing spline known as P-spline or penalized spline regression. P-spline have the advantage that they require only a small set of spline basis functions for each covariate with a penalty to avoid undersmoothing and can be represented as mixed models which have been described in Wang (1998), Fan dan Zhang (1998), Brumback *et al* (1999), Vebyla *et al* (1999), French *et al* (2001), Kamman dan Wand (2003), dan Wand (2003).

Mixed-effects models provide flexibility of fitting models with various fixed and random elements. The application of mixed-effects models to practical data analysis has greatly expanded with the subsequent development of theory and computer software. The parameters in these models are typically estimated by maximum likelihood (ML) or restricted maximum likelihood (REML). Because of a simple mathematical connection between P-spline models and mixed models, mixed model software can be used to fit P-splines. In the linear mixed models framework the fitted penalized splines are BLUP (best linear unbiased predictors) and the smoothing parameters are ratios of variance components which can be estimated by ML or REML. (Wand, 2003).

The aim of the research is modelling spatio-temporal additive models for Ozone air pollution in Surabaya. The data are hourly ozone concentration measured in $\mu g/m^3$ from the 5 monitoring stations that observed from January 2002 up to December 2002. Meteorological factors play an important role in studies of air pollution because of its role as a possible confounding factor, then we include them as covariate in the models. Functional form of covariates and response are modeled using P-spline. The estimation of spatio-temporal additive model using linear mixed models approach. The resulting models represent a fusion of time-series additive models and spatial additive models with meteorological factors as covariates.

2. Spatio-temporal Additive Models

Spatio-temporal additive models are combination of time series additive models and spatial additive models. Time series additive model is an extention of time series model that allow us to modelling nonlinear of lagged variables (Fan dan Yao, 2003; Huang dan Shen, 2004). Additive AR(2) denoted by AAR(2) is AR(2) with functional form of lagged variables can be modeled by parametric, nonparametric, or both. Suppose $y_{t,i}$ is Ozone concentration at time-t and location-i. AAR(2) is given as

$$y_{t,i} = f_1(y_{t-1,i}) + f_2(y_{t-2,i}) + \varepsilon_{t,i} \qquad (3)$$

where $f_1(y_{t-1,i})$ and $f_2(y_{t-2,i})$ are smooth function of lag 1 and lag 2 of y.

Spatial additive models is an extention of kriging and thin-plate spline. Kriging is surface estimation that used for spatial interpolation at an arbitrary location. Thin-plate spline is smoothing surface estimation (Green dan Silverman, 1994). Kriging and thin plate spline can be stated as linear combination of radial basis functions (Kamman dan Wand, 2003). Suppose y_i is Ozone concentration at location $\mathbf{x}_i = (x_{1i}, x_{2i}) \in \Re^2$, spatial additive models of kriging dan spline-2 is given as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + S(\mathbf{x}_i) + \varepsilon_i \qquad (4)$$

Where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ is parameter vector, $S(\mathbf{x}_i)$ in kriging is stochastic process, while in thin plate spline is radial basis function.

By combined time series additive models (4) and spatial additive models (3), we get spatio-temporal additive models

$$y_{t,i} = f_1(y_{t-1,i}) + f_2(y_{t-2,i}) + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + S(\mathbf{x}_i) + \varepsilon_{t,i} \qquad (5)$$

In model (5), we can add other covariates like meteorological facors. Suppose $g(m)$ is functional form of y with temperature m, aa is indicator variable of wind directions, ch is indicator variable of rainfall, so that we get spatio-temporal models

$$y_{t,i} = f_1(y_{t-1,i}) + f_2(y_{t-2,i}) + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + S(\mathbf{x}_i) + g(m_{t,i}) + aa_{t,i} + ch_{t,i} + \varepsilon_{t,i} \qquad (6)$$

3. Estimation of Spatio-temporal Additive Models

Suppose smooth function f_1 and f_2 in equation (3) are modeled by linear spline

$$y_{t,i} = \alpha_0 + \alpha_1 y_{t-1,i} + \sum_{k=1}^{K_1} u_k^{y1}(y_{t-1,i} - \kappa_k^{y1})_+ + \alpha_2 y_{t-2,i} + \sum_{k=1}^{K_2} u_k^{y2}(y_{t-2,i} - \kappa_k^{y2})_+ + \varepsilon_{t,i} \qquad (7)$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, u_1^{y1}, \cdots, u_{K_1}^{y1}, u_1^{y2}, \cdots, u_{K_2}^{y2})$ is parameter vector, $(w)_+^p = w^p \mathbf{I}(w \geq 0)$ is truncated power function with $\mathbf{I}$ is indicator function and $p \geq 1$, $\kappa_1^{y1} < \ldots < \kappa_{K_1}^{y1}$ and $\kappa_1^{y2} < \ldots < \kappa_{K_2}^{y2}$ are fixed knot for lag-1 and lag-2 respectively.

Component $S(\mathbf{x}_i)$ in equation (4) can be stated as combination linear of radial basis function, so that

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sum_{k=1}^{K_x} u_k^x \, C_0(\|\mathbf{x}_i - \boldsymbol{\kappa}_k^x\|) + \varepsilon_i \qquad (8)$$

Where $C_0(r) = \begin{cases} \text{correlation function for kriging} \\ r^2 \log r \text{ for thin-plate spline} \end{cases}$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, u_1^x, \ldots, u_{K_x}^x)$ is parameter vector, $\boldsymbol{\kappa}_1^x < \ldots < \boldsymbol{\kappa}_{K_x}^x$ is knot vector of location variables, and $\|\mathbf{x}_i - \boldsymbol{\kappa}_k^x\|$ is distance from location-i to knot $\boldsymbol{\kappa}_k^x$ (Djuraidah and Aunuddin, 2006b). For simplicity $C_0(\|\mathbf{x}_i - \boldsymbol{\kappa}_k^x\|)$ in (8), we denotes as z_{ik}, so that (8) can rewrite as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sum_{k=1}^{K_x} u_k^x \, z_{ik} + \varepsilon_i \qquad (9)$$

Combination between spline time series additive models and spline additive models, we get spline spatio-temporal models, is given as

$$y_{t,i} = \alpha_0 + \alpha_1 y_{t-1,i} + \sum_{k=1}^{K_1} u_k^{y1}(y_{t-1,i} - \kappa_k^{y1})_+ + \alpha_2 y_{t-2,i} + \sum_{k=1}^{K_2} u_k^{y2}(y_{t-2,i} - \kappa_k^{y2})_+ +$$

$$\beta_1^x x_{1,t,i} + \beta_2^x x_{2,t,i} + \sum_{k=1}^{K_x} u_k^x z_{t,i,k} + \varepsilon_{t,i} \qquad (10)$$

where $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \alpha_2, \beta_1^x, \beta_2^x, u_1^{y1}, \cdots, u_{K_1}^{y1}, u_1^{y2}, \cdots, u_{K_2}^{y2}, u_1^x, \cdots, u_{K_x}^x)$ is parameter vector.

The estimation for additive model with more than one covariates can be generalized from additive model with one covariate (Djuraidah and Aunuddin, 2005, 2006a). The estimator $\hat{\boldsymbol{\beta}}$ is solution of penalized spline regression criterion

$$J(s) = \|\mathbf{y} - \mathbf{C}\boldsymbol{\beta}\|^2 + \boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta} \qquad (11)$$

where C is design matrix of equation (10)

$$\mathbf{C} = \left(1 \;\; y_{t-1,i} \;\; y_{t-2,i} \;\; x_{1i} \;\; x_{2i} \;\; (y_{t-1,i}-\kappa_1^{y1})_+ \; \cdots \; (y_{t-1,i}-\kappa_{K_1}^{y1})_+, \; (y_{t-2,i}-\kappa_1^{y2})_+, \; \cdots \; (y_{t-2,i}-\kappa_{K_2}^{y2})_+, \; z_{t,i,1}, \; \cdots \; z_{t,i,K_x}\right)_{1\le t\le n, 1\le i\le s}$$ and

$\Lambda = \mathrm{diag}\left(0,\ 0,\ 0,\ 0,\ 0,\ \lambda_{y_{t-1,i}}\mathbf{1}_{K_1\times 1},\ \lambda_{y_{t-2,i}}\mathbf{1}_{K_2\times 1}\ \lambda_{x_i}\mathbf{1}_{K_x\times 1}\right)$, where $\lambda_{y_{t-1}}, \lambda_{y_{t-2}}, \lambda_{x_i}$ are smoothing parameter of lag-1, lag-2, spatial respectively. The estimator of penalized spline regression is $\hat{\mathbf{y}} = \mathbf{C}(\mathbf{C}'\mathbf{C} + \boldsymbol{\Lambda})^{-1}\mathbf{C}'\mathbf{y}$ (12)

Penalized regression spline in equation (10) can be represented in linear mixed models by treated coefficient of truncated spline basis u_k^{y1}, u_k^{y2}, and u_k^{x} as random effect in linear mixed models. Set up

$$\boldsymbol{\gamma} = (\alpha_0, \alpha_1, \alpha_2, \beta_1^x, \beta_2^x)', \quad \mathbf{b} = \left(u_1^{y1},\ \cdots\ , u_{K_1}^{y1},\ u_1^{y2},\ \cdots\ , u_{K_2}^{y2},\ u_1^{x},\ \cdots\ ,\ u_{K_x}^{x}\right)'$$

$$\mathbf{X} = \left(1,\ y_{t-1,i},\ y_{t-2,i},\ x_{1,t,i},\ x_{2,t,i}\right)_{1\le t\le n, 1\le i\le s},$$

$$\mathbf{Z} = \left((y_{t-1,i}-\kappa_1^{y1})_+,\ \cdots;\ (y_{t-1,i}-\kappa_{K_1}^{y1})_+,\ (y_{t-2,i}-\kappa_1^{y2})_+,\ \cdots\ (y_{t-2,i}-\kappa_{K_2}^{y2})_+,\ z_{t,i,1},\ \cdots;\ z_{t,i,K_x}\right)_{1\le t\le n, 1\le i\le s}$$

the models (10) can expressed in standard form of linear mixed models

$$\mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \qquad (13)$$

where $E\begin{bmatrix}\mathbf{b}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}0\\0\\0\\0\end{bmatrix}$ and $\mathbf{cov}\begin{bmatrix}\mathbf{b}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\sigma_{y1}^2\mathbf{I} & 0 & 0 & 0\\ 0 & \sigma_{y2}^2\mathbf{I} & 0 & 0\\ 0 & 0 & \sigma_x^2\mathbf{I} & 0\\ 0 & 0 & 0 & \sigma_\varepsilon^2\mathbf{I}\end{bmatrix}$

The spatio-temporal additive models in (10) can be obtained by applying REML to γ, σ_ε^2, σ_{y1}^2, σ_{y2}^2, σ_x^2, and best prediction of **b**. Smoothing parameter of spatio-temporal model is ratio of two component variance, is given as

$$\lambda_{y1} = \frac{\sigma_\varepsilon^2}{\sigma_{y1}^2},\ \lambda_{y2} = \frac{\sigma_\varepsilon^2}{\sigma_{y2}^2},\ \text{dan}\ \lambda_x = \frac{\sigma_\varepsilon^2}{\sigma_x^2} \qquad (14)$$

4. Application to Air Pollution Ozon

Regional prediction of tropospheric Ozone concentration is an important problem in environtmental monitoring. The data are hourly concentrations of Ozone (µg/m³) monitored in Surabaya from January 2002 up to December 2002. The data have been monitored by AQMS (ambient air quality system) network for monitoring and evaluation of air pollutans in Indonesia. Surabaya has 5 fixed monitoring station that placed in the background area, to describe the real air pollution condition in Surabaya. Meteorological data included in modelling are temperature, wind direction, and rainfall.

Building spatio-temporal model consists of two steps. The first step is to find the best time series additive models, the second steps is to find the best spatial additive model from residual model in step 1. After that, we combine model in step 1 and model in step 2. The criterions of choosing the best model are AIC (*Akaike Information Criteria*) of models, and ACF (*autocorrelation*) function, PACF (*partial autocorrelation*) function, spatial correlation function of residual.

The result summary of some spatio-temporal model is displayed in Table 1. Spatio-temporal models with meteorological factors (Model 3 and Model 4) have smaller AIC than models without meteorological factors. AIC and smoothing parameter of Model 3 and Model 4 are equal, except for smoothing parameter of spatial. The differences values caused by the difference trend of radial basis function's.

Table1. Summary of AIC and smoothing parameter of spatio-temporal models

Model	Component model	AIC	Smoothing Parameter				
			Lag-1	Lag-2	Hour	Temperature	Spatial
1	AAR(2)1) + Spline-2 (K=5)	20203.2	2.456	2.318	2.112	-	18.930
2	AAR(2) + Kriging(K=5)	20203.1	2.455	2.318	2.112	-	4.353
3	AAR(2) + Spline-2(K=5) +f(M)2)	19563.7	2.996	2.498	2.148	19.413	17.061
4	AAR(2) + Kriging (K=5)+ f(M)	19563.6	2.996	2.498	2.148	19.413	3.929

1) AAR(2) = Lag-1 Ozon(K=5) + Lag-2 Ozon(K=5) + hour (K=23)
2) f(M) = SH(K=5) + AA(I) + H(I), K= number of knot, where SH = Temperature ,AA = wind direction,
H = rainfall, I = indicator variable

The spatial correlation, ACF, and PACF plot of residual Model 3 and Model 4 are equal, shown in Figure 1a, 1b, and 1c respectively. The spatial correlation plot shows that the values are small, although still has exponential trend but the R^2 of model is small. ACF and PACF plot of residual have small value and random pattern.

The smoothing parameter is ratio of two component variances, has been explained in equation (14). Thin-plate spline has big value, it means that spatial variance's is small, so that thin-palte spline has smooth response curve. The opposite of thin-plate spline, kriging has rough response curve. Thin-plate spline and kriging are used to spatial interpolation, so that we should choose the best one. We compare the standard error of the two methods, shown in Figure 2a and Figure 2b. The standard error pattern of kriging can described the variation in an arbitrary location. However the standard error of thin-plate spline can not described it. Based on this results, the best spatio-temporal model for Ozone is Model 4.

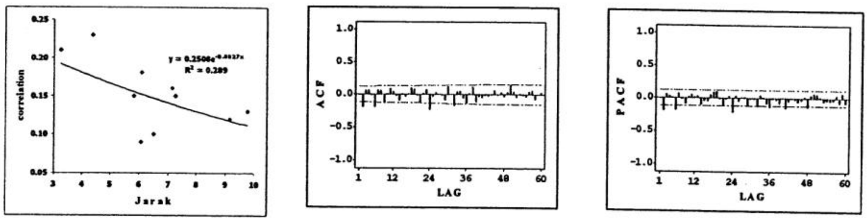
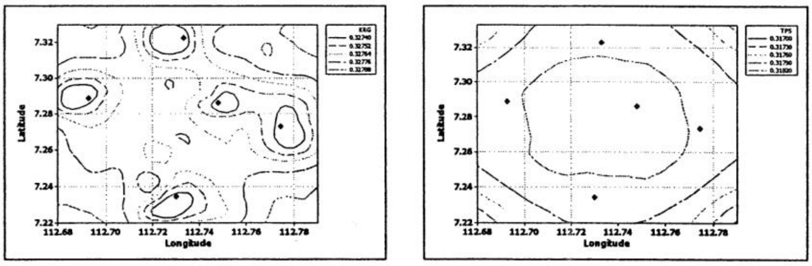


Figure 1. (a) Spatial correlation plot, (b) ACF plot, (c) PACF plot of residual Model 4



(a) (b)

Figure 2. (a) Standard error of kriging (b) standard error of thin-plate spline.

Contour of Model 2 and Model 4 for August 31, 2002 at 8 am, 9 am, and 6 pm, are shown in Figure 3a up to 3f. The figure show that spatial prediction of Model 2 and Model 4 are similar.

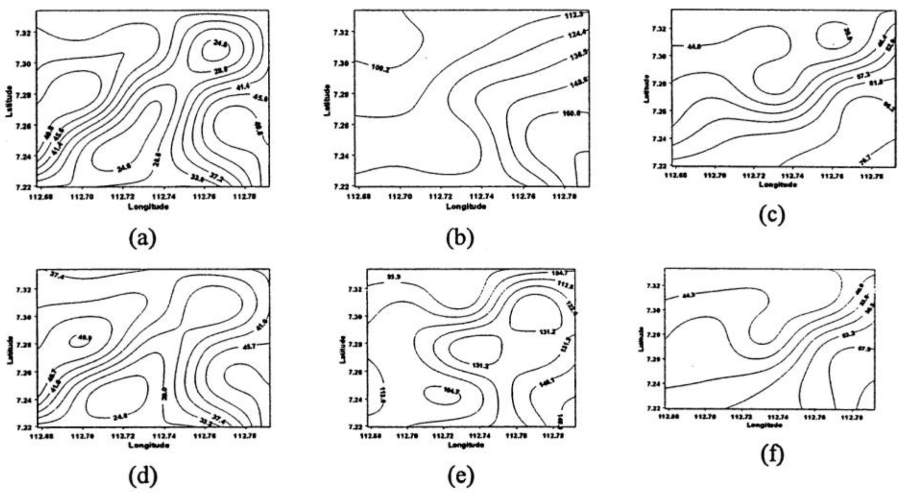


Figure 3. Contour of Model 2 at (a) 8 am, (b) 9 am, (c) 6 pm; Contour of Model 4 at (d) 8 am, (e) 9 am, (f) 6 pm

5. Conclusion

Spatio-temporal additive models can be used for modelling data that contains spatial and temporal correlation. Estimation of additive models using linear mixed models approach give simplification in computation and inference. Addition meteorological factors covariate to spatio-temporal model can reduce AIC, although spatial prediction is similar with model without meteorolgical factors. The response curve of spatial prediction using thin-plate spline is smooth, but using kriging is rough.

6. References

Brumback, B.A., Ruppert, D., and Wand, M.P. (1999). Comment on Variable selection and function estimation in additive nonparametric regression using a data-based prior by Shively, Kohn and Wood, *Journal of the American Statistical Association*, Vol 94, pp. 794-797.

Djuraidah, A. and Aunuddin. (2005). Regresi Spline Terpenalisasi dengan Pendekatan Model Campuran, Prosiding Seminar Nasional dalam Rangka Dies ke-50 Fakultas MIPA UGM, Yogyakarta, September 17, 2005.

Djuraidah, A. and Aunuddin. (2006a). Model Aditif Deret Waktu dengan Pendekatan Model campuran, *Inferensia* Jurnal Statistika FMIPA-ITS (will appear).

Djuraidah, A. and Aunuddin. (2006b). Kriging dan *Thin-Plate* Spline dengan Pendekatan Model Campuran, *Matematika Integratif* Jurnal Matematika FMIPA-UNPAD (will appear).

Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties (with discussion), *Statistical Science,* Vol 11, pp. 89-121.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series, Nonparametric and Parametric Methods,* Springer-Verlag, New York

Fan, J. and Zhang, J.T. (1998). Comment on Smoothing spline models for the analysis of nested and crossed samples of curves by Brumback and Rice, *Journal of the American Statistical Association,* Vol 93, pp. 961-994.

French, J.L., Kammann, E.E., and Wand, M.P. (2001). Comment on Semiparametric nonlinear mixed-effects models and their applications by Ke and Wang, *Journal of the American Statistical Association,* Vol 96, pp.1285-1288.

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

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

Huang, J.Z. and Shen, H. (2004). Functional Coefficient Regression Models for Nonlinear Time series: A Polynomial Spline Approach, *Scandinavian Journal of Statistics,* Vol 31, pp. 515-534

Kammann, E.E. and Wand, M.P. (2003). Geoadditive models, *Applied Statistics,* Vol 52, pp. 1-18.

Ruppert, D. and Carroll, R.J. (1997). Penalized regression splines, Unpublished manuscript. http://www.orie.cornell.edu/~davidr/papers/index/index.html

Stone, C.J. (1985). Additive Regression and Other Nonparametric Models, *The Annals of Statistics,* Vol 13, pp. 689–705.

Verbyla, A.P., Cullis, B.R., Kenward, M.G., and Welham, S.J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion), *Journal of the Royal Statistics Society,* Series C, Vol 48, pp. 269-312.

Wand, M. (2003). Smoothing and mixed models, *Computational Statistics,* Vol 18, pp. 223–249.

Wang, Y. (1998). Mixed effects smoothing spline analysis of variance, *Journal of the Royal Statistical Society,* Series B, Vol 60, pp. 159-174.