Bayesian surface smoothing under anisotropy

Subhashish Chakravarty

University of Iowa

Follow this and additional works at: https://ir.uiowa.edu/etd

Part of the Statistics and Probability Commons

Copyright © 2007 Subhashish Chakravarty

This dissertation is available at Iowa Research Online: https://ir.uiowa.edu/etd/169

Recommended Citation

https://doi.org/10.17077/etd.wj8x9tak

Follow this and additional works at: https://ir.uiowa.edu/etd

Part of the Statistics and Probability Commons
BAYESIAN SURFACE SMOOTHING UNDER ANISOTROPY

by

Subhashish Chakravarty

An Abstract

Of a thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics in the Graduate College of The University of Iowa

July 2007

Thesis Supervisors: Professor George Woodworth Lecturer Matthew Bognar
ABSTRACT

Bayesian surface smoothing using splines usually proceeds by choosing the smoothness parameter through the use of data driven methods like generalized cross validation. In this methodology, knots of the splines are assumed to lie at the data locations. When anisotropy is present in the data, modeling is done via parametric functions.

In the present thesis, we have proposed a non-parametric approach to Bayesian surface smoothing in the presence of anisotropy. We use eigenfunctions generated by thin-plate splines as our basis functions. Using eigenfunctions does away with having to place knots arbitrarily, as is done customarily. The smoothing parameter, the anisotropy matrix, and other parameters are simultaneously updated by a Reversible Jump Markov Chain Monte Carlo (RJMCMC) sampler. Unique in our implementation is model selection, which is again done concurrently with the parameter updates.

Since the posterior distribution of the coefficients of the basis functions for any given model order is available in closed form, we are able to simplify the sampling algorithm in the model selection step. This also helps us in isolating the parameters which influence the model selection step.

We investigate the relationship between the number of basis functions used in the model and the smoothness parameter and find that there is a delicate balance which exists between the two. Higher values of values of the smoothness parameter correspond to more number of basis functions being selected.

Use of a non-parametric approach to Bayesian surface smoothing provides for more modeling flexibility. We are not constrained by the shape defined by a parametric shape of the covariance as used by earlier methods. A Bayesian approach also
allows us to include the results obtained from previous analysis of the same data, if any, as prior information. It also allows us to evaluate pointwise estimates of variability of the fitted surface. We believe that our research also poses many questions for future research.

Abstract Approved: __________________________
Thesis Supervisor

Title and Department __________________________

Date ________________________________________

Thesis Supervisor __________________________

Title and Department __________________________

Date ________________________________________
BAYESIAN SURFACE SMOOTHING UNDER ANISOTROPY

by

Subhashish Chakravarty

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics in the Graduate College of The University of Iowa

July 2007

Thesis Supervisors: Professor George Woodworth
Lecturer Matthew Bognar
CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. thesis of

Subhashish Chakravarty

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Statistics at the July 2007 graduation.

Thesis Committee:

George Woodworth, Thesis Supervisor

Matthew Bognar, Thesis Supervisor

Kate Cowles

Richard Dyskstra

Brian Smith
To my parents and grandparents
I would like to thank my advisor, George Woodworth, my co-advisor, Matthew Bognar, for their guidance, encouragement, motivation and help. This thesis would not have been possible without their help. Many thanks to my dissertation committee; Kate Cowles, Richard Dykstra and Brian Smith for their support. I would also like to thank Chetan Tiwari and Aditya Sehgal for their help with computer related issues. Special thanks to Tammy Siegel for her unwavering support and encouragement.

I am truly blessed to have a family which has always supported me: Baba, Ma, Didi and Nevil. Diduma, Dadu and Dimma, I have been really lucky to have grandparents like you. Without you all I would not have even dared to dream big and set my goals high.
ABSTRACT

Bayesian surface smoothing using splines usually proceeds by choosing the smoothness parameter through the use of data driven methods like generalized cross validation. In this methodology, knots of the splines are assumed to lie at the data locations. When anisotropy is present in the data, modeling is done via parametric functions.

In the present thesis, we have proposed a non-parametric approach to Bayesian surface smoothing in the presence of anisotropy. We use eigenfunctions generated by thin-plate splines as our basis functions. Using eigenfunctions does away with having to place knots arbitrarily, as is done customarily. The smoothing parameter, the anisotropy matrix, and other parameters are simultaneously updated by a Reversible Jump Markov Chain Monte Carlo (RJMCMC) sampler. Unique in our implementation is model selection, which is again done concurrently with the parameter updates.

Since the posterior distribution of the coefficients of the basis functions for any given model order is available in closed form, we are able to simplify the sampling algorithm in the model selection step. This also helps us in isolating the parameters which influence the model selection step.

We investigate the relationship between the number of basis functions used in the model and the smoothness parameter and find that there is a delicate balance which exists between the two. Higher values of the smoothness parameter correspond to more number of basis functions being selected.

Use of a non-parametric approach to Bayesian surface smoothing provides for more modeling flexibility. We are not constrained by the shape defined by a parametric shape of the covariance as used by earlier methods. A Bayesian approach also
allows us to include the results obtained from previous analysis of the same data, if any, as prior information. It also allows us to evaluate pointwise estimates of variability of the fitted surface. We believe that our research also poses many questions for future research.
## TABLE OF CONTENTS

LIST OF TABLES .......................................................... vii

LIST OF FIGURES ........................................................ viii

CHAPTER

1 INTRODUCTION ...................................................... 1

2 SMOOTHING GAUSSIAN DATA ..................................... 7
   2.1 Model ........................................................... 8
   2.2 Bayesian formulation of the model ....................... 8
      2.2.1 Prior distributions ................................... 9
      2.2.2 Joint Distribution .................................... 12
      2.2.3 Posterior Sampling .................................. 13
      2.2.4 Convergence of the Sampler ....................... 20

3 ANALYSIS OF THE COAL ASH DATA ......................... 22
   3.1 Data .......................................................... 22
   3.2 Modified coal ash data ................................... 23

4 CONCLUSIONS AND FUTURE RESEARCH .................... 32

APPENDIX

A GRAPHS FOR NORMAL DATA .......................... 35

REFERENCES ......................................................... 72
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Posterior mean, 95% credible intervals and monte carlo error for $\theta_1$</td>
<td>25</td>
</tr>
<tr>
<td>3.2</td>
<td>Posterior mean, 95% credible intervals and monte carlo error for $\theta_2$</td>
<td>25</td>
</tr>
<tr>
<td>3.3</td>
<td>Posterior mean, 95% credible intervals and monte carlo error for $\theta_3$</td>
<td>25</td>
</tr>
<tr>
<td>3.4</td>
<td>Posterior mean, 95% credible intervals and monte carlo error for $A$</td>
<td>26</td>
</tr>
<tr>
<td>3.5</td>
<td>Posterior mean, 95% credible intervals and monte carlo error for $\angle$</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>of anisotropy in degrees.</td>
<td></td>
</tr>
<tr>
<td>3.6</td>
<td>Posterior mean, 95% credible intervals and monte carlo error for $\tau_e$</td>
<td>27</td>
</tr>
<tr>
<td>3.7</td>
<td>Posterior mean, 95% credible intervals and monte carlo error for $\tau$</td>
<td>27</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure

3.1 Scatter plot of the core values in %coal ash. ................................. 22
3.2 Surface plot of the observed values. ............................................. 23
3.3 Surface plot of the fitted values from sample 1 ............................. 24
A.1 Surface plot of the observed values. ............................................. 35
A.2 Surface plot of the fitted values from chain 1. .............................. 36
A.3 Surface plot of the fitted values from chain 2. .............................. 37
A.4 Surface plot of the fitted values from chain 3. .............................. 38
A.5 Histogram of the Model indicator (K) from chains 1, 2 and 3 respectively. 39
A.6 Trace plots of K for chains 1 (red) 2 (blue) and 3 (brown). ......... 40
A.7 Trace plot of the deviance $-2\log(\pi(y|\omega))$ for chains 1 (red) 2 (blue) and 3 (brown). ......................................................... 41
A.8 Trace plots of the $[1,1]^{th}$ element of the anisotropy matrix A for chains 1 (red) 2 (blue) and 3 (brown). ......................................................... 42
A.9 Trace plots of the $[1,2]^{th}$ element of the anisotropy matrix A for chains 1 (red) 2 (blue) and 3 (brown). ......................................................... 43
A.10 Trace plots of the $[2,2]^{th}$ element of the anisotropy matrix A for chains 1 (red) 2 (blue) and 3 (brown). ......................................................... 44
A.11 Trace plots of $\theta_1$ for chains 1 (red) 2 (blue) and 3 (brown). .... 45
A.12 Trace plots of $\theta_2$ for chains 1 (red) 2 (blue) and 3 (brown). .... 46
A.13 Trace plots of $\theta_3$ for chains 1 (red) 2 (blue) and 3 (brown). .... 47
A.14 Trace plots of $\tau_e$ for chains 1 (red) 2 (blue) and 3 (brown). .... 48
A.15 Trace plots of $\tau$ for chains 1 (red) 2 (blue) and 3 (brown). .... 49
A.16 Trace plots of $\alpha_e$ for chains 1 (red) 2 (blue) and 3 (brown). .... 50
A.17 Trace plots of $\beta_e$ for chains 1 (red) 2 (blue) and 3 (brown). .... 51
A.18 ACF plots of the $[1,1]^{th}$ element of the anisotropy matrix A. ...... 52
A.19 ACF plots of the $[1,2]^{th}$ element of the anisotropy matrix A. ...... 53
A.20 ACF plots of the $[2,2]^{th}$ element of the anisotropy matrix A. ...... 54
<table>
<thead>
<tr>
<th>Section Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.21</td>
<td>ACF plots of the $\tau_e$, the data precision.</td>
<td>55</td>
</tr>
<tr>
<td>A.22</td>
<td>ACF plots of the $\tau$, the smoothness parameter.</td>
<td>56</td>
</tr>
<tr>
<td>A.23</td>
<td>ACF plots of the $\theta_1$.</td>
<td>57</td>
</tr>
<tr>
<td>A.24</td>
<td>ACF plots of the $\theta_2$.</td>
<td>58</td>
</tr>
<tr>
<td>A.25</td>
<td>ACF plots of the $\theta_3$.</td>
<td>59</td>
</tr>
<tr>
<td>A.26</td>
<td>ACF plots of the $\alpha_e$.</td>
<td>60</td>
</tr>
<tr>
<td>A.27</td>
<td>ACF plots of the $\beta_e$.</td>
<td>61</td>
</tr>
<tr>
<td>A.28</td>
<td>Scatter plot showing posterior correlation between $\tau_e$ and $\tau$.</td>
<td>62</td>
</tr>
<tr>
<td>A.29</td>
<td>Scatter plot showing posterior correlation between $\theta_1$ and $\theta_2$.</td>
<td>63</td>
</tr>
<tr>
<td>A.30</td>
<td>Scatter plot showing posterior correlation between $\theta_1$ and $\theta_3$.</td>
<td>64</td>
</tr>
<tr>
<td>A.31</td>
<td>Scatter plot showing posterior correlation between $\theta_2$ and $\theta_3$.</td>
<td>65</td>
</tr>
<tr>
<td>A.32</td>
<td>Scatter plot showing posterior correlation between $\tau$ and $A[1,1]$.</td>
<td>66</td>
</tr>
<tr>
<td>A.33</td>
<td>Scatter plot showing posterior correlation between $\tau$ and $A[1,2]$.</td>
<td>67</td>
</tr>
<tr>
<td>A.34</td>
<td>Scatter plot showing posterior correlation between $\tau$ and $A[2,2]$.</td>
<td>68</td>
</tr>
<tr>
<td>A.35</td>
<td>Scatter plot showing posterior correlation between angle of anisotropy and $K$.</td>
<td>69</td>
</tr>
<tr>
<td>A.36</td>
<td>Scatter plot showing posterior correlation between angle of anisotropy and $K$.</td>
<td>70</td>
</tr>
<tr>
<td>A.37</td>
<td>Plot of the fitted values using PROC TPSPLINE in SAS.</td>
<td>71</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Splines are an alternative to parametric models for noisy data. For example, Wood et al. [25] model the variation in global temperature from a historical standard against latitude and longitude using splines. Another example could be taken from mine engineering. It might be of interest to estimate the quantity of ore in a deposit, and thus core samples are taken at various points and the resultant surface is estimated. In fact we will study a similar example in the next chapter. Cowles [4] provides another example of such data. It deals with the amount of ammonium deposition through rainfall measured at specific locations across 46 states within the United States of America.

Splines can be used for both smoothing and interpolation. Smoothing splines are functions which pass close to the data, without the constraint of interpolating the data, but are smooth in some sense, as their name implies. Usually, smoothness is defined in terms of the derivatives of the function, thus relating it to the curvature of the function. A method of fitting these smoothing splines then immediately appeals to the intuition: minimizing a criterion which has a least squares like component and a penalty component which penalizes rough functions. This method is called the method of penalized least squares. For a detailed discussion about different types of splines and methods of fitting them to data can be found in Wegman and Wright [24].

Suppose the data model is

\[ y_i = f(x_i) + \epsilon_i, \quad i = 1, 2, \ldots, n. \]  (1.1)

where \( f(.) \) is a smooth real-valued function to be estimated (defined on \( \mathbb{R}^d \)), \( x_i = (x_1(i), x_2(i), \ldots, x_d(i))^t \) is a point in \( \mathbb{R}^d \), and \( \epsilon_i \) \( \sim N(0, \frac{1}{\tau_e}) \) is the error term where \( \tau_e \) denotes the data precision. We assume that \( f \in \mathcal{X} \) where \( \mathcal{X} \) is a space of functions whose partial derivatives of total order \( d \) are in \( L_2(\mathbb{R}^d) \). This inclusion
property defines our notion of *smoothness* of \( f \). Let \( \mathbf{y} = (y_1, y_2, \ldots, y_n)^t \) denote the vector of observations. Let \( \mathbf{f} = (f(x_1), f(x_2), \ldots, f(x_n))^t \) denote the values of \( f \) at the *locations* \( \mathbf{x}_i = (x_{1(i)}, x_{2(i)}, \ldots, x_{d(i)})^t, \ i = 1, 2, \ldots, n \). Then the problem of estimating the smooth function \( f \) can be reformulated as minimizing

\[
\frac{1}{n} \| \mathbf{y} - \mathbf{f} \|^2 + \tau J_{m,d}(f)
\]

where \( \| \cdot \| \) is the Euclidean norm in \( d \)-space, \( \tau > 0 \) is the smoothness parameter, balancing fit to the data, with a penalty on the measure of *wiggliness* of \( f \) [Wegman and Wright [24]] defined by \( J_{m,d}(f) \). The function \( J_{m,d}(f) \) is appropriately called a *smoothness penalty functional*. In this thesis, we shall use the *thin-plate* spline penalty functional, defined for general \( m \) and \( d \) by

\[
J_{m,d}(f) = \sum_{\alpha_1 + \alpha_2 + \ldots + \alpha_d = m} \frac{m!}{\alpha_1! \ldots \alpha_d!} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left( \frac{\partial^m f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \right)^2 \prod_{j=1}^d dx_j
\]

where the summation is over all \( \alpha_1 + \alpha_2 + \ldots + \alpha_d = m \), \( \alpha_j \)'s being non-negative integers. In the particular case of \( d = 2 \) and \( m = 2 \), the smoothness penalty becomes

\[
J_{2,2}(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{\partial^2 f}{\partial x_1^2} + 2 \frac{\partial^2 f}{\partial x_1 \partial x_2} + \frac{\partial^2 f}{\partial x_2^2} \right)^2 dx_1 dx_2,
\]

which corresponds to the bending energy of a thin plate, a generalization in two-dimensions of a thin elastic beam [Wegman & Wright [24]]. For this reason, the family of solutions to (1.2) is called the set of thin plate smoothing splines. We shall restrict attention to \( d = 2 \) and \( m = 2 \) in this thesis.

Duchon [8] has shown that for a fixed value of \( \tau \), the minimizer of (1.2) has the representation

\[
f_\tau(\mathbf{x}) = \sum_{j=1}^M \theta_j \phi_j(\mathbf{x}) + \sum_{i=1}^n c_i E(\mathbf{x}, \mathbf{x}_i)
\]

(1.5)
where the constants $\theta_1, \theta_2, \ldots, \theta_M$ and $c_1, c_2, \ldots, c_n$ are to be determined. The $\phi_j$'s span the space of all polynomials of degree less than $m$. $E(\cdot, \cdot)$ is defined in equation (1.8) below. The $c_i$'s are generalized divided differences (g.d.d.), i.e. they annihilate all polynomials of total degree less than $m$. For more details on g.d.d. the reader can refer to Wahba [23]. At this point we would like to remind the reader that since $d = 2$ and $m = 2$, the sum in the first term of (1.5) will have 3 terms (the first term corresponds to $\phi_1 = 1$). Thus, if we write $V = E(x_i, x_j), i, j = 1, 2, \ldots, n$, $X = ((1, (x_i)^t))_{n \times 3}$, $\Theta = (\theta_1, \theta_2, \theta_3)^t$, then the minimization problem (1.2) can be rewritten as, minimizing

$$
\frac{1}{n} ||y - X\Theta - Vc||^2 + \tau c^t Vc \tag{1.6}
$$

subject to the condition $X^t c = 0$. Notice that the penalty function in (1.2) has been replaced by a quadratic form in (1.6). This can be done since $J_{m,d}(f)$ defines a seminorm on $\mathcal{X}$ and its null space is the $M$ dimensional space spanned by polynomials in $d$ variables of total degree less than $m$. For more details, the reader is referred to Wahba [23]. An explicit solution for $\Theta$ and $c$ can be found in Wahba [23] (pg. 33).

The minimization problem (1.2) can also be solved under the Bayesian paradigm. This is achieved by Wahba [22]. Using the methodology of reproducing kernel Hilbert spaces, Wahba [22] shows that the minimizer of (1.2), for a given value of $\tau$, is the posterior mean of the sample path of a stochastic process whose prior distribution is the same as the distribution of the stochastic process

$$
X_a(s) = \sum_{i=1}^{M} \theta_i \phi_i(s) + \frac{1}{\sqrt{\tau}} Z(s) \tag{1.7}
$$

where $s \in \mathbb{R}^2$, $\Theta = (\theta_1, \theta_2, \theta_3)^t$ has a diffuse Normal prior with variance $1/\alpha I$ as $\alpha \to 0$, $\phi_j$'s retain the same interpretation as in (1.5), $\tau$ has a proper prior distribution, and $Z(\cdot)$ is a zero mean Gaussian random field (see Adler [1] for details on random fields).
with covariance given by $P_X E P_X$ where $E$ is given by

$$
E(x_i, x_j) = \frac{1}{16\pi} \|x_i - x_j\|^2 \log(\|x_i - x_j\|^2)
$$

(1.8)

and $P_X$ is a projection to the null space of span $\{\phi_j\}$. Note that $\phi_1(s) = 1$, $\phi_2(s) = s_1$, $\phi_3(s) = s_2$ for the case $d = 2$, $m = 2$. By smoothness prior we will mean the distribution of the stochastic process $X_a(\cdot)$ in this thesis. The RHS of (1.7) gives us a representation of the smooth function $f$ that we are trying to fit in (1.1). Other approaches to smoothness priors can be found in Gersch [11] and the references cited therein.

We express $f$ using the dominant principal components of the prior covariance of the Gaussian random field $Z(s)$. This is done as follows. We first rewrite (1.1) using matrix notation as

$$
y = X\Theta + T\Gamma + \epsilon,
$$

(1.9)

where $y = (y_1, y_2, \ldots, y_n)^t$, and $X = ((1, (x_i)^t))_{n \times 3} (i = 1, 2, \ldots, n)$ denote the matrix of locations. Note that $X$ contains the functions $\phi_1, \phi_2, \phi_3$ evaluated at the locations. $X$ is assumed to be of full column rank, i.e we shall assume that there are no tied covariate values. $\Theta$ is as before.

The dominant principle component matrix $T$ is obtained as follows. Let the covariance matrix of $Z$ be denoted by $V$, where $V(x_i, x_j) = \((P_X E P_X)_{ij}) (i, j = 1, 2, \ldots, n)$. Here

$$
P_X = I - X(X^t X)^+ X^t.
$$

(1.10)

The positive sign in the exponent indicates a generalized inverse of the matrix. In other words, $P_X$ is a projection to the null space corresponding to the column space of $X$. Then $V = U\Delta U^t$, using a spectral decomposition; the columns of $U$ are
orthonormal, arranged according to descending eigenvalues of $V$, and $\Delta$ is a diagonal matrix with the eigenvalues arranged in descending order along the diagonal. Let $\Delta^{1/2}$ denote the Cholesky decomposition of $\Delta$. We set

$$
T_{n \times K} = U_K \Delta_K^{1/2}.
$$

Here $K$ is the number of dominant eigenvalues of $V$. The coefficients $\Gamma_{K \times 1}$ has a proper Normal prior with zero mean, but with variance $(1/\tau)\mathbf{I}$. Thus the smoothness parameter $\tau$ in (1.2) turns out to be a precision parameter, on which we will place a proper prior distribution. The vector $T\Gamma$ completes the specification of the random field in terms of its dominant principal values. As before $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^t$.

The vectors $X\Theta$ and $T\Gamma$ represent the linear and non-linear parts of the model respectively.

Our next step is to introduce geometric anisotropy into the model (1.9). If the elements of $V$, as given by (1.8), are not just a function of distances ($h_{ij} = \|x_i - x_j\|$) between them, but also of direction, then $Z$ is said to be anisotropic. Furthermore, if there exists a positive definite matrix $A$, admitting a Cholesky factorization $A^{1/2}$, and the elements of $V$ are a function of the transformed lag-distances $[A^{1/2}]^t h_{ij}$, and not of direction, then $Z$ is called geometrically anisotropic. Cressie [5] and Zimmerman [28] give details about geometric and other kinds of anisotropy. We estimate $A$ simultaneously with the linear and non-linear coefficients. Ecker and Gelfand [10], Johnson [16] and Cowles [4] also incorporate geometric anisotropy in their model using a parametric function as the covariance for the observed data, but it is fundamentally different from ours, as our model uses a nonparametric form for the same. This is evident from the expression of the covariance of the Gaussian random field that we have used.
Inference from our model is based on Markov chain Monte Carlo (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller [18]) simulation techniques, also called MCMC in short. In this thesis we shall include the number of dominant principle components, $K$, as a parameter in our model. The number of dominant principal components, $K$ (also referred to as model order in the sequel), is chosen using the reversible jump Markov chain Monte Carlo (RJMCMC) method of Green [13]. This achieves variable selection in our model. We have a nested structure in our model; if $K = r$, then the first $r - 1$ basis functions are also included in the model. This structure is adopted since the columns of $T$ are eigenvectors corresponding to eigenvalues sorted in descending order and eigenvalues larger in magnitude correspond to more variation being explained (like in principal components analysis).

To the best of our knowledge, our approach of using smoothness priors in 2-dimensions to model anisotropy in spatial data and achieve variable selection through RJMCMC is new. The rest of the thesis is organized as follows. Chapter 2 presents our model for smoothing Gaussian data. We implement the model on a dataset. Finally in Chapter 3, we shall summarize and present some further research that needs to be done. The simulations were done using the C programming language, the R and SAS software packages were used for pre/post processing of the data and for generating graphics. We used ArcGIS for the values of the mean of the anisotropy matrix in its prior distribution. We have used the GNU Scientific Library (GSL) [14] for our C programs. The computer on which the simulations were done had a Intel Core 2 Duo 2.66 GHz processor, 2 GB of memory and was running Mandriva Linux 2007 and Windows XP.
CHAPTER 2
SMOOTHING GAUSSIAN DATA

As stated in the Introduction, we will be dealing with data locations in two dimensions. For such spatial data, regression has been performed by Ecker and Gelfand [9] [10], Wood et al. [25], Denison et al. [6] [7], Cowles [4] and Johnson [16], among others. Of these, anisotropy has been modeled by Ecker and Gelfand [10], Cowles [4] and Johnson [16].

Ecker and Gelfand [10] propose a Bayesian methodology which simultaneously estimates the linear transformation (anisotropy matrix) on the lags of the locations which reduce the region to isotropy and estimate other parameters in the semivariogram. They use a parametric form of the semivariogram. They recommend a method to detect departures from isotropy in the data. They do not propose a method for variable selection. Cowles [4] uses a parametric form for the correlations between the observed values in her model for spatial regression. Model selection is achieved via Bayes’ factors. Wood et al. [25] do not explicitly include anisotropy in their model, but we feel that their method of adaptively smoothing spatial data has connections to anisotropy. We leave this as a topic for future research. They use thin-plate splines in their non-parametric model of the regression surface. Variable selection is done via Bayesian information criterion.

We have dealt with the problem of smoothing spatial data in this thesis as opposed to spatial regression. We have modeled the required surface non-parametrically using thin-plate splines. We have also achieved model selection simultaneously with estimating other relevant parameters via a RJMCMC sampler. The rest of the chapter is organized as follows. We present our model, its Bayesian formulation, and implement our approach on a dataset. We also provide an estimate of the surface
using PROC TPSPLINE in SAS.

2.1 Model

We recall from the Introduction, the form of our model

\[ y = X\Theta + T_A\Gamma + \epsilon \]  

(2.1)

with the distinction of \( T \) being replaced by \( T_A \). This serves to reinforce the fact that the matrix \( T \) depends on the anisotropy matrix \( A \). As before, \( \epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^t \) represents the error in the model which might have arisen due to the non-inclusion of all the eigenvectors in \( T_A \). Furthermore, \( \epsilon \sim N(0, \tau_e I_n) \), where \( \tau_e \) denotes the data precision. \( X\Theta \) forms the linear part of the model, and \( T_A\Gamma \) forms the non-linear part of the model. In our current notation, we also recall that \( T'_A T_A = \Delta \), where \( \Delta \) is a diagonal matrix containing the eigenvalues of the matrix \( V \). Also, by construction, \( X'T_A = 0 \), since the columns of \( T_A \) are scaled eigenvectors of a matrix which belongs to the null space corresponding to the column space of \( X \). We would like to remind the reader that the eigenvalue-eigenvector decomposition of \( V \) may result in eigenvalues which are close to zero, negative, or have multiplicities. The reason for the mentioned anomalies is assigned to numerical limitations of the computer algorithm being used to perform the operations. Experience suggests that such is rarely the case with the first few eigenvalues of \( V \).

2.2 Bayesian formulation of the model

A Bayesian formulation of the model starts with the prior distributions on the parameters of interest and then proceeds to evaluate the posterior distributions of the parameters (either analytically, or via simulations) conditional on the data.
2.2.1 Prior distributions

$\Theta$ contains the coefficients for the monomials of first degree which make up the linear part of the model. A diffuse $N(0, \frac{1}{c_1}I_3)$ prior is assigned to $\Theta$, where $c_1$ is a small number representing the practitioner’s beliefs for the particular application. $\Gamma$ contains the random coefficients associated with the basis functions. The smoothness parameter $\tau$ is incorporated into the model as the prior precision of $\Gamma$. Conditional on the value of $K$ (the number of basis functions) and $\tau$, the prior distribution of $\Gamma$ is $N(0, \frac{1}{\tau}I_K)$ where $I_K$ is an identity matrix of order $K$. In the limit, as $\tau \to \infty$, the fitted surface would be a hyperplane; as $\tau \to 0$, there would be no smoothing, and the fitted surface would be an interpolant to the data. Thus the prior for $\tau$ should not allocate too much probability close to zero. In our experience with spatial data, the data by themselves do not carry much information about smoothness and it is necessary that the prior distribution of the smoothness parameter should be informative. We express the prior beliefs about $\tau$ by a $\text{Gamma}(1, \beta_\tau)$ distribution, where $\beta_\tau$ represents the rate parameter in a $\text{Gamma}$ distribution. Kadane and Woodworth [17] deal with employment decision data where the data is available quarterly. In their prior elicitation they consider changes in the odds of terminating a protected employee relative to an unprotected employee across a half-quarter expressed in rescaled time units (as a fraction of the total observation period). Since business data is usually available in quarterly time intervals, their method of prior elicitation can be used. For spatial data though, there is no standard grid over which data is observed. For example, we might be dealing with death rate data collected in all counties of a state. The data locations would probably be defined by the latitude and longitude of the centroid of the county. As a second example, we might be dealing with data collected on ore deposits in a single mine. Here the data locations could just be reference points relative to a fixed point, as in the Coal ash data example in Cressie [5] from
the Robena mine property. Also, unlike 1-dimensional data, additional consideration also needs to be made for the direction of variation amongst the observations for higher dimensional data. In our case, careful consideration has to be made while deciding on a value for \( \beta \tau \) in conjunction with \( K \). This point is further discussed below when we address the prior distribution for \( K \). Obviously results from a previous analysis could be used for determining values of \( \beta \tau \). Also, an approximate value for \( \beta \tau \) can be obtained from a preliminary analysis using software like PROC TPSPLINE in SAS. PROC TPSPLINE has the capability to evaluate the smoothness parameter using cross validation. The data precision \( \tau_e \) is given a conjugate Gamma distribution prior with hyper-paramters \( \text{shape} = \alpha_e \) and \( \text{rate} = \beta_e \). The hyper-parameters, in turn, have \( \text{Lognormal}(0,1) \) distributions.

The value of \( K \) equals the number of basis functions in the model. In some sense it can also be thought of as a model indicator. \( K = 0 \) denotes a no-basis-function model. In our experience, we have found that eigenvectors from the variance-covarinance matrix based on thin-plate splines corresponding to high values of \( K \) are visually rougher than those eigenvectors which correspond to lower values of \( K \). Thus there are two parameters in our model which influence smoothness of the required surface. A high value of \( K \) also indicates lack of parsimony in our model, in terms of many non-linear terms being included. Thus, the practitioner’s prior beliefs about the parameters in the prior distribution of the smoothness parameter and \( K \) should address this delicate balance between smoothness and parsimony.

Let \( K_{\text{max}} \) denote the maximum number of basis functions to be allowed in the model. The value of \( K_{\text{max}} \) can be decided based on how many eigenvalues of \( V \) account for 95–99% of the variation. A truncated Poisson distribution with mean \( m_K \) over \( \{0, 1, \ldots, K_{\text{max}}\} \) is chosen as the prior of \( K \). The point made earlier about the numerical limitations in calculating the eigenvalues and eigenvectors should also
be kept in mind while deciding on $K_{\text{max}}$. A preliminary analysis could be run for this purpose.

The positive definite anisotropy matrix $A$ is given a Wishart prior with degrees of freedom 2 and mean matrix $\Sigma$. The angle of anisotropy, $a^*$, is given by

$$\cot(2a^*) = \frac{a_{11} - a_{22}}{2a_{12}}$$  \hspace{1cm} (2.2)

where $a_{ij}$, $i, j = 1, 2$ are the elements of the anisotropy matrix $A$. The range of anisotropy in the direction $\eta$, $r_\eta$, is defined as that lag value at which the correlation between observations separated by that lag distance and in the direction $\eta$ is zero. In terms of the elements of the anisotropy matrix, the range of anisotropy is given by

$$r_\eta = \frac{c}{(h_\eta^t A h_\eta)^{1/2}}$$  \hspace{1cm} (2.3)

where $h_\eta = (\cos \eta, \sin \eta)$ is a unit vector in the direction $\eta$ and $c$ is a constant determined by the form of the correlation function chosen. For Gaussian, $c = \sqrt{3}$, for Cauchy, $c = \sqrt{19}$ etc. The ratio of anisotropy, $\lambda$, is defined as the ratio of the major axis of the ellipse (associated with anisotropy) to the minor axis and is given by

$$\lambda = \left\{\frac{h_{\pi-a^*}^t A h_{\pi-a^*}}{h_{a^*}^t A h_{a^*}}\right\}^{1/2}$$  \hspace{1cm} (2.4)

where $a^*$ is the angle of anisotropy and $h_{(\cdot)}$ is again a unit vector in the direction of its subscript. If estimates of the angle of anisotropy, range and ratio of anisotropy are available from historical data, equations (2.2)–(2.4) can be used for values of $\Sigma$, the prior mean of the anisotropy matrix. If historical data is not available, then we can figure out the range of anisotropy for three different angles and assuming a particular form of the correlation function, to solve for the elements of $\Sigma$ (which being symmetric, has only three distinct elements) using software like ArcGIS or SAS.

If $\pi(\cdot)$ denotes the density of its argument, then the joint prior distribution can
be written as

\[ \pi(\Theta, A, \Gamma, \tau, \tau_e, \alpha_e, \beta_e, K) = \]

\[ \pi(A)\pi(K)\pi(\alpha_e)\pi(\beta_e)\pi(\tau)\pi(\Theta) \]

\[ \cdot \pi(\tau_e|\alpha_e, \beta_e)\pi(\Gamma|K, \tau) \] (2.5)

### 2.2.2 Joint distribution

We are now in a position to write out the joint distribution of the data and the parameters. We do not include the constants of proportionality. Noticing that the data likelihood is multivariate normal, the joint distribution is written as

\[ \pi(y, \Theta, A, \Gamma, \tau, \tau_e, \alpha_e, \beta_e, K) \propto \]

\[ \tau_e^{n/2} \exp \left\{ -0.5\tau_e(y - X\Theta - TA\Gamma)'(y - X\Theta - TA\Gamma) \right\} \]

\[ \cdot \exp \left\{ -0.5(c_1)\Theta^t\Theta \right\} \]

\[ \cdot \tau^K/2 \exp \left\{ -0.5\Gamma^t\Gamma \right\} \]

\[ \cdot |A|^{-0.5} \exp \left\{ -0.5tr(A\Sigma^{-1}) \right\} \]

\[ \cdot \tau^{1-1} \exp \left\{ -\beta_e\tau \right\} \]

\[ \cdot 1/(\Gamma(\alpha_e))\beta_e^{\alpha_e}\tau_e^{\alpha_e-1} \exp \left\{ -\beta_e\tau_e \right\} \]

\[ \cdot (1/\alpha_e) \exp \left\{ -0.5[\log(\alpha_e)]^2 \right\} \]

\[ \cdot (1/\beta_e) \exp \left\{ -0.5[\log(\beta_e)]^2 \right\} \]

\[ \cdot \exp(-m_K)m_K^k/k! \] (2.6)
2.2.3 Posterior Sampling

A hybrid MCMC algorithm is used to sample from the posterior distribution of the parameters. This is in the spirit of the algorithm suggested in Denison et al. [7] on page 55. As will be seen below, we have conjugate posterior distributions for $\Theta$, $\tau$, $\tau_e$ and $\Gamma$, conditional on the other parameters. Hence we use a Gibbs sampler to sample from the posterior of $\Theta$, $\tau$ and $\tau_e$. Since $\Gamma$ changes dimension across models, we use a partially analytic reversible jump sampler (described below) to explore from its conditional posterior distribution. The hyperparameters $\alpha_e$ and $\beta_e$ have conditional posteriors which are of unknown form, and hence we use a Metropolis-Hastings algorithm to sample from their posteriors. We explain the details of the sampler below.

We introduce some further notation which will make it easier to write out the posterior distributions. Let

$$\omega = (\Theta, \Gamma, \tau, \tau_e, \alpha_e, \beta_e, A, K)^t$$

(2.7)
denote the vector of parameters of interest. In the above, $A$ represents the three distinct elements of the anistropy matrix, written as a vector. We will denote $\omega_{-(\cdot)}$ as a subvector of $\omega$ which doesn’t contain the parameter ($\cdot$).

The coefficients of the linear term, $\Theta$, is sampled from its posterior density, $\pi(\Theta|\omega_{-\Theta}, y)$, conditional on the current value of all the other parameters and the data $y$. This can be accomplished by collecting all terms in (2.6) corresponding to $\Theta$. Comparing these terms with a multivariate normal density, and using the fact that $X^tT_A = 0$, we see that $\pi(\Theta|\omega_{-\Theta}, y)$ is multivariate normal $N_3(\mu_\Theta, \Sigma_\Theta)$, where

$$\mu_\Theta = \tau_e[\tau_eX^tX + (c_1)I_3]^{-1}X^ty,$$

and

$$\Sigma_\Theta = [\tau_eX^tX + (c_1)I_3]^{-1}.$$

To sample from the posterior distribution of $A$, $\pi(A|\omega_{-A}, y)$, a Metropolis-Hastings algorithm is used where we use the prior as the proposal distribution. Hence
the probability of accepting such a move is only composed of relevant terms from the data likelihood. It is given by

\[ \alpha = \min \left\{ 1, \frac{h(A^*)}{h(A)} \right\} \]  

(2.8)

where

\[ h(A) = \exp \left\{ -0.5\tau_e (-2\Gamma^t T^t A y \Gamma + \Gamma^t T^t A T A \Gamma) \right\} \]

with \( A^* \) being the proposed value.

To update \( \alpha_e \) from \( \pi(\alpha_e | \omega^{-\alpha_e}, y) \), a random walk Metropolis Hastings proposal is used to generate a proposal for \( \log(\alpha_e) \). The variance of the proposal is tuned so that the acceptance rate is approximately 30–60%. Let the proposed value for \( \log(\alpha_e) \) be denoted by \( z_1^* \), and the current value be denoted by \( z_1 \). Then the logarithm of the acceptance probability \( \log(\alpha) = \min(0, ld) \), where \( ld \) is given by

\[ ld = -\frac{1}{2}(z_1^*)^2 - \log(\Gamma(\alpha_e)) + \alpha_e \log(\beta_e \tau_e) \]  

(2.9)

To update \( \beta_e \) from \( \pi(\beta_e | \omega^{-\beta_e}, y) \), a random walk Metropolis Hastings proposal, with variance tuned as before, is used to generate a proposal for \( \log(\beta_e) \). Let the proposed value for \( \log(\beta_e) \) be denoted by \( z_2^* \), and the current value be denoted by \( z_2 \). Then the logarithm of the acceptance probability \( \log(\alpha) = \min(0, ld) \), where \( ld \) is given by

\[ ld = -\frac{1}{2}(z_2^*)^2 + \alpha_e z_2 - \beta_e \tau_e \]  

(2.10)

The difference in the logarithm of the density evaluated at \( z_2^* \) and \( z_2 \) is calculated. The acceptance probability is the minimum of zero and this difference.

\( \tau \) is updated from its conjugate conditional posterior distribution \( \pi(\tau | \omega^{-\tau}, y) \), which is Gamma with parameters \( \text{shape} = \frac{K}{2} + 1 \) and \( \text{rate} = \frac{1}{2} \Gamma^t T + \beta \). This is
obtained by collecting relevant terms in (2.6).

Similarly, \( \tau_e \) is updated from its conjugate conditional posterior distribution \( \pi(\tau_e|\omega_{-\tau_e}, y) \), which is \textit{Gamma} with parameters \( \text{shape} = \frac{n}{2} + \alpha_e \) and \( \text{rate} = \frac{1}{2}(y - X\Theta - T_A\Gamma)^t(y - X\Theta - T_A\Gamma) + \beta_e \).

Now, we consider the posterior distribution of \( \Gamma, \pi(\Gamma|\omega_{-\Gamma}, y) \). Again, collecting terms involving \( \Gamma \) from the joint distribution (2.6) and comparing them to the multivariate normal density function, we see that the posterior distribution of \( \Gamma \) is multivariate normal \( N(\mu_\Gamma, \Sigma_\Gamma) \) where \( \mu_\Gamma = \tau_e [\tau_e T_A^t T_A + \tau I_K]^{-1} T_A^t y, \) and \( \Sigma_\Gamma = [\tau_e T_A^t T_A + \tau I_K]^{-1}. \)

We now consider the \textit{partially analytic} reversible jump MCMC sampler. Godsill [12] proposes a composite representation for model uncertainty problems. Let \( k \in K \) denote the model indicator. Let \( \omega \) denote a collection of parameters whose elements (might be vectors themselves) are parameters associated with each value of \( k \). The elements of \( \omega \) might have different dimensionalities. When parameters are not shared across models, only one element of from the collection \( \omega \) corresponds to a unique value of \( k \). To define a prior distribution over the entire parameter space means that a probability distribution needs to be defined over the entire product space of candidate models and their parameters. For a particular value of the model indicator \( k \), the likelihood only depends on the corresponding element of \( \omega \), which we denote by \( \omega_k \), as follows

\[
\pi(y|k, \omega) = \pi(y|k, \omega_k)
\]

Denoting the prior for the parameter by \( \pi(\omega_k|k) \) and the model prior by \( \pi(k) \), the full posterior distribution might be written as

\[
\pi(k, \omega_k|y) = \frac{\pi(y|k, \omega_k)\pi(\omega_k|k)\pi(\omega_{-k}|\omega_k, k)\pi(k)}{\pi(y)} (2.11)
\]
where $\omega_{-k}$ denotes the parameters not used by model $k$. The term $\pi(\omega_{-k}|\omega_k, k)$ is called the pseudo-prior. The key feature of the composite model space is that the dimension remains fixed even when the model indicator $k$ changes.

We now show how the reversible jump sampler (Green [13]) can be obtained by a special form of the Metropolis Hastings [15] (MH) proposal to the composite model. The reversible jump sampler moves through the model space using the MH algorithm constructed in such a way that detailed balance within each move type is preserved. Consider a move from model $k$ with $\omega_k$ parameters to a model $k'$ with $\omega_{k'}$ using a proposal $q(k', \omega_{k'}|k, \omega_k)$. Assuming all densities exist, the acceptance probability for such a move is given by

$$\alpha = \min\left\{1, \frac{\pi(k', \omega_{k'}|y)q(k, \omega_k|k', \omega_{k'})}{\pi(k, \omega_k|y)q(k', \omega_{k'}|k, \omega_k)}\right\}$$

(2.12)

Now consider a proposal distribution $q(\cdot|\cdot)$ from the current state of the composite model $(k, \omega)$ to a future state $(k', \omega')$ of the form

$$q(k', \omega'|k, \omega) = q_1(k'|k)q_2(\omega_{k'}'|\omega_k)p(\omega_{-k'}|\omega_{k'}, k')$$

(2.13)

This factors the proposal into three parts

1. a proposal for changing the model index from $k$ to $k'$
2. a proposal for the parameters used by model $k'$
3. a proposal for the remaining unused parameters, which are chosen equal to the pseudo-prior.

Thus we have defined a joint proposal across the whole state space of parameters and the model indicator. Since new states are proposed depending only on the current state, the Markov property of the MH method is satisfied. Also, there are no new concerns about variable dimension in the composite space. The standard MH
acceptance probability for this move is given by

\[ \alpha = \min \left\{ 1, \frac{q(k, \omega|k', \omega') \pi(k', \omega'|y)}{q(k', \omega'|k, \omega) \pi(k, \omega|y)} \right\} \]  

(2.14)

Substituting (2.13) above and using (2.11) we get

\[ \alpha = \min \left\{ 1, \frac{q_1(k|k') q_2(\omega_k|\omega_k') p(\omega_{-k} | k, k') \pi(k', \omega_{k'}|y) p(\omega_{-k'}|\omega_{k'}, k') \pi(k, \omega_k|y)}{q_1(k'|k) q_2(\omega_{k'}|\omega_k) p(\omega_{-k'}|\omega_{k'}, k') \pi(k', \omega_{k'}|y) p(\omega_{-k'}|\omega_k)} \right\} \]  

(2.15)

which, after obvious cancellation, reduces to

\[ \alpha = \min \left\{ 1, \frac{q_1(k|k') q_2(\omega_k|\omega_k') \pi(k', \omega_{k'}|y)}{q_1(k'|k) q_2(\omega_{k'}|\omega_k) \pi(k, \omega_k|y)} \right\} \]  

(2.16)

which on comparing with (2.12) is exactly the reversible jump acceptance probability of Green [13] where the proposal in (2.12) has been factored into two parts. The first part proposes a new model indicator and the second proposes the parameter corresponding to this new model indicator. The acceptance probability is independent of the value of any parameters which are unused by both models \(k\) and \(k'\).

Now, if the full conditional posterior density \( \pi(\omega_{k'}|k', y) \) is known, then we might use it as the proposal distribution \( q_2(\cdot|\cdot) \) in the description above. For the special case when the full conditional posterior density \( \pi(\omega_{k'}|y, k') \) is known, we can use it as the proposal distribution \( q_2(\cdot|\cdot) \) in (2.16) above. This gives

\[ \alpha = \min \left\{ 1, \frac{q_1(k|k') \pi(\omega_k|y) \pi(k', \omega_{k'}|y)}{q_1(k'|k) \pi(\omega_{k'}|y) \pi(k, \omega_k|y)} \right\} \]  

(2.17)

In (2.17) we substitute the identity

\[ \frac{\pi(k, \omega|y)}{\pi(\omega|k, y)} = \pi(k|y) \]  

(2.18)

then the reversible jump acceptance probability reduces to

\[ \alpha = \min \left\{ 1, \frac{\pi(k'|y) q_1(k|k')}{\pi(k|y) q_1(k'|k)} \right\} \]  

(2.19)
which is nothing but the standard MH acceptance probability where \( \pi(k|y) \) is the target distribution and \( q_1(k'|k) \) is the proposal. The acceptance probability in (2.19) is independent of the parameter values and depends only on the proposal distribution for model order and the posterior odds ratio. We can then use standard MCMC techniques to sample from the posterior distributions of the other parameter values.

Usually, the full conditional posterior is unavailable in most models of practical interest. We might have the full conditional posterior for a subset of the parameters where the other parameters are shared. Godsill [12] extends the above reasoning to include the case when parameters are shared across models. We introduce some more notation and state his result. We denote the parameters corresponding to the model indicator \( k' \) by \( \omega_{k'} \). Let \( \omega_U \) denote the subset of \( \omega \) which changes from model to model and whose full conditional is available in closed form. In our case, this is \( \Gamma \). We will use a superscript, i.e. \( \Gamma^{(k)} \), to indicate that the dimension of \( \Gamma \) is \( k \). The subvector \( \omega_{-U} \) of \( \omega \), denotes the parameters which don’t change dimension for different values of \( K \). In our case, this is the vector \((\Theta, \tau, \tau_e, \alpha_e, \beta_e, A)^t\). Suppose that we consider a move to model \( k' \). We choose a reversible jump proposal which sets \( \omega_{k'-U} = \omega_{k-U} \), and proposes the remaining parameter vector \( \omega_{k',U} \) from its full conditional, \( \pi(\omega_{k',U}|\omega_{k'-U}, k', y) \). The reverse move would set \( \omega_{k-U} = \omega_{k'-U} \) and propose the remaining parameter vector \( \omega_{k,U} \) from its full conditional, \( \pi(\omega_{k,U}|\omega_{k,-U}, k, y) \). Using the same reasoning as demonstrated for the case when we have the full conditional posterior, the reversible jump acceptance probability for such a move is then derived by using the same reasoning as above in Godsill [12] as

\[
\alpha = \min \left\{ 1, \frac{\pi(k'|\omega_{k'-U}, y) \cdot q(k|k')}{\pi(k|\omega_{k,-U}, y) \cdot q(k'|k)} \right\}
\]  

(2.20)

where \( q(\cdot|\cdot) \) denotes the probability of proposing a model changing move, and

\[
\pi(k'|\omega_{k'-U}, y) = \int \pi(k', \omega_{k',U}|\omega_{k'-U}, y) d(\omega_{k',U})
\]  

(2.21)
In our case, the integral on the right hand side of (2.21) is given by \( \int \pi(k', \Gamma | \omega_{k',-U}, y) \).

This can be obtained by collecting all terms involving \( k \) and \( \Gamma \) from the joint distribution. These terms are

\[
\exp \left\{ -0.5\tau_e (y - X\Theta - T_{A} \Gamma^{(k)})^t (y - X\Theta - T_{A} \Gamma^{(k)}) \right\}
\cdot \exp \left\{ -0.5\tau \Gamma^{(k)} \Gamma^{(k)} \right\} \cdot \tau^{k/2}
\cdot P[K = k].
\]

The first two terms are also involved in the conditional posterior distribution \( \pi(\Gamma | \omega_{- \Gamma}, y) \). Multiplying and dividing by \((2\pi)^{-k/2} |\Sigma_{\Gamma^{(k)}}|^{-1/2}\) would give the value of the integral as (using \( k \) instead of \( k' \))

\[
\pi(k | \omega_{k,-U}, y) = (2\pi)^{k/2} |\Sigma_{\Gamma^{(k)}}|^{1/2} \cdot \tau^{k/2} \cdot P[K = k]. \tag{2.22}
\]

Now recall that \( \Sigma_{\Gamma} = (\tau_e \Delta + \tau I)^{-1} \) where \( \Delta \) is the diagonal matrix of eigenvalues. Thus \( |\Sigma_{\Gamma^{(k)}}|^{1/2} = \prod_{j=1}^{k} (\tau_e \delta_j + \tau)^{-1/2} \), where \( \delta_j \) is the \( j \)th diagonal element of \( \Delta \), and the acceptance probability in (2.20) can be simplified to

\[
\alpha = \min \left\{ 1, \frac{(2\pi)^{k'/2} \cdot \prod_{j=1}^{k'} (\tau_e \delta_j + \tau)^{-1/2} \cdot \tau^{k'/2} \cdot P[K = k'] \cdot q(k'|k)}{(2\pi)^{k/2} \cdot \prod_{j=1}^{k} (\tau_e \delta_j + \tau)^{-1/2} \cdot \tau^{k/2} \cdot P[K = k] \cdot q(k|k')} \right\} \tag{2.23}
\]

We only consider three types of moves, a birth move, a death move and a non-dimension changing move which keeps the current value of \( K \). In a birth move, \( k' = k + 1 \), in a death move \( k' = k - 1 \). Thus we change the dimension of the model by only one if a birth or death move is proposed. The proposal distributions \( q(\cdot | \cdot) \) are set up as follows

1. \( q(k + 1|k) = c \cdot \min \{ P[K = k + 1]/P[K = k], 1 \} \)

2. \( q(k - 1|k) = c \cdot \min \{ P[K = k - 1]/P[K = k], 1 \} \)

3. \( q(k|k) = 1 - q(k + 1|k) - q(k - 1|k) \)

where the constant \( c \) is chosen so that \( q(k + 1|k) + q(k - 1|k) \leq 0.80 \) for all \( k \in \).
\{1, 2, \ldots, k_{\text{max}} - 1\}. \ q(1|0) = 1 = q(k_{\text{max}} - 1|k_{\text{max}}). \ The \ terms \ P[K = k + 1]/P[K = k] \ and \ P[K = k - 1]/P[K = k] \ are \ evaluated \ using \ the \ Poisson(5) \ prior \ on \ K.

We are now in a position to simplify the acceptance probability further. For a birth step, the acceptance probability is

\[
\alpha = \min \left\{ 1, (\tau \cdot 2\pi)^{1/2}[\tau_e \delta_{k+1} + \tau]^{-1/2} \cdot \frac{5}{k + 1} \cdot \frac{\min[(k + 1)/5, 1]}{\min[5/(k + 1), 1]} \right\}
\]

(2.24)

The last two terms in (2.24) cancel out, finally giving us

\[
\alpha = \min \left\{ 1, (\tau \cdot 2\pi)^{1/2}[\tau_e \delta_{k+1} + \tau]^{-1/2} \right\}
\]

(2.25)

The acceptance probability for a death move is given by

\[
\alpha = \min \left\{ 1, (\tau \cdot 2\pi)^{-1/2}[\tau_e \delta_{k} + \tau]^{1/2} \cdot \frac{k}{5} \cdot \frac{\min[5/k, 1]}{\min[k/5, 1]} \right\}
\]

(2.26)

Again the last two terms cancel out giving us

\[
\alpha = \min \left\{ 1, (\tau \cdot 2\pi)^{-1/2}[\tau_e \delta_{k} + \tau]^{1/2} \right\}
\]

(2.27)

Equations (2.25) and (2.27) enable us to identify the parameters driving the model selection step. They are \(\tau, \tau_e\) and the eigenvalue corresponding to the birth/death step. This completes the Bayesian formulation of our model.

2.2.4 Convergence of the Sampler

The issue of convergence of the sampler, i.e. when to start collecting the sample of models and how many to collect, is a topic of ongoing research. All convergence diagnostics give ideas about non-convergence rather than detect convergence. For parameter spaces with variable dimension, work in this area has been done by Brooks and Giudici [2], Castelloe and Zimmerman [3], among others. Castelloe and Zimmerman [3] consider an unbalanced ANOVA based approach, with the parameter space
being treated as a factor. They also propose to monitor multiple parameters. All methods devoted to studying convergence of varying dimensional parameters thus far, further work is needed for their application in wide ranging applications, we feel, still needs further work. We have chosen to look at non-convergence of our sampler based on a simple intuitive criterion, namely, monitoring the plot of the deviances, i.e. $-2 \log(y|\omega)$, for models traversed by the sampler. This depends on the likelihood of the data and it is easy to compute. We believe that monitoring the deviances should be undertaken in all practical applications involving sampling based exploration of the posterior distribution.
3.1 Data

We have used the data pertaining to percent coal ash present at different locations for the Robena Mine Property in Greene County, Pennsylvania (see Cressie [5]). We used ArcGIS to provide the range at three different angles and solved the equations relating the range of anisotropy and elements of $\Sigma$; the mean of the anisotropy matrix is given in (2.3). Figure 3.1 draws a scatter plot of the core measurements in %coal ash recorded at 208 points of observation.

![Scatter plot of the observations](image)

Figure 3.1: Scatter plot of the core values in %coal ash.

The top of the page faces North, and the right-hand side faces East. Figure 3.1 doesn’t show any noticeable trend in the observations. Following Kadane and Woodworth [17], where they reason that the mean of the corresponding Gamma distribution should be set away from 0, we set the mean of the smoothness parameter to be 0.01. We choose $k_{max} = 20$, thus there are a maximum of 20 basis functions in our model.
We run our MCMC sampler for 20,000 iterations with 3 chains. Convergence is evaluated by observing the plot of deviances $-2 \log(y|\omega)$, as the sampler progresses.

The surface plot for the Coal ash data is presented in Figure 3.2. The plot of the fitted posterior mean surface is plotted in Figure 3.3. The fitted surface is evaluated using the posterior mean values of the parameters.

![Surface plot of observed values](image)

**Figure 3.2:** Surface plot of the observed values.

### 3.2 Modified coal ash data

The fitted posterior mean surface appears to be a plane in 2-dimensions. Thus, to check if our procedure could fit a curvilinear surface, we added the quadratic term $0.03x^2 + 0.01y^2 + 0.005xy$ to the percent Coal ash measurements; $x$ represents the x-coordinate and $y$ represents the y-coordinate of the locations.

The sampler was run for 20,000 iterations using three chains. The run time was
6 hours. We discarded the first 10000 observations as burn in. The prior distribution of the smoothness parameter was specified to have a mean of 0.14, which corresponds to a rate $\beta_\tau = 7$. Since spatial data don’t contain much information about smoothness by themselves, and lower values of $\tau$ correspond to greater levels of smoothness (and smoothing is the objective in this thesis), we put this informative prior on $\tau$. The corresponding Gamma distribution in fact puts 30% of its mass above the mean. The prior on the precision of the linear coefficients is required to be diffuse as per the specification of Wahba [22]. We used $0.001\mathbf{I}$ as the prior precision of $\Theta$. The prior mean of the anisotropy matrix used was

$$
\Sigma = \begin{pmatrix}
4.0 & 1.2 \\
1.2 & 2.1 
\end{pmatrix} \tag{3.1}
$$

This was again obtained by solving the equations corresponding to range for three
Table 3.1: Posterior mean, 95% credible intervals and monte carlo error for $\theta_1$.

<table>
<thead>
<tr>
<th>Chain</th>
<th>Mean</th>
<th>Credible Intv.</th>
<th>MC Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.790</td>
<td>(7.210, 8.240)</td>
<td>0.002</td>
</tr>
<tr>
<td>2</td>
<td>7.797</td>
<td>(7.341, 8.254)</td>
<td>0.002</td>
</tr>
<tr>
<td>3</td>
<td>7.794</td>
<td>(7.354, 8.253)</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Table 3.2: Posterior mean, 95% credible intervals and monte carlo error for $\theta_2$.

<table>
<thead>
<tr>
<th>Chain</th>
<th>Mean</th>
<th>Credible Intv.</th>
<th>MC Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.382</td>
<td>(0.330, 0.432)</td>
<td>0.0002</td>
</tr>
<tr>
<td>2</td>
<td>0.381</td>
<td>(0.331, 0.433)</td>
<td>0.0002</td>
</tr>
<tr>
<td>3</td>
<td>0.381</td>
<td>(0.330, 0.431)</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

Table 3.3: Posterior mean, 95% credible intervals and monte carlo error for $\theta_3$.

<table>
<thead>
<tr>
<th>Chain</th>
<th>Mean</th>
<th>Credible Intv.</th>
<th>MC Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.280</td>
<td>(0.256, 0.313)</td>
<td>0.0001</td>
</tr>
<tr>
<td>2</td>
<td>0.284</td>
<td>(0.256, 0.313)</td>
<td>0.0001</td>
</tr>
<tr>
<td>3</td>
<td>0.285</td>
<td>(0.257, 0.431)</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

different angles and using ArcGIS as mentioned earlier. We utilized three chains, starting with disparate starting values, to explore the posterior distributions of the parameters. In what follows, we present results of our implementation on these transformed observations. Graphs are constructed for the parameters which do not change dimension using the post burn-in observations.

We present tables for the posterior means, 95% credible intervals and monte carlo error for some of the parameters which don’t change dimension below using the post burn-in observations.

The posterior mean, 95% credible interval and monte carlo error Tables 3.1–3.3
<table>
<thead>
<tr>
<th>A</th>
<th>Mean</th>
<th>Credible Intv.</th>
<th>MC Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain 1</td>
<td>6.040</td>
<td>(0.227, 24.950)</td>
<td>0.351</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>Chain 2</td>
<td>5.618</td>
<td>(0.002, 23.818)</td>
</tr>
<tr>
<td></td>
<td>Chain 3</td>
<td>5.799</td>
<td>(0.108, 23.121)</td>
</tr>
<tr>
<td>Chain 1</td>
<td>2.870</td>
<td>(-0.941, 12.707)</td>
<td>0.174</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>Chain 2</td>
<td>2.658</td>
<td>(0.241, 14.031)</td>
</tr>
<tr>
<td></td>
<td>Chain 3</td>
<td>2.925</td>
<td>(-0.886, 13.989)</td>
</tr>
<tr>
<td>Chain 1</td>
<td>3.560</td>
<td>(0.355, 13.240)</td>
<td>0.195</td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>Chain 2</td>
<td>3.398</td>
<td>(0.117, 12.481)</td>
</tr>
<tr>
<td></td>
<td>Chain 3</td>
<td>3.584</td>
<td>(0.234, 15.620)</td>
</tr>
</tbody>
</table>

Table 3.4: Posterior mean, 95% credible intervals and monte carlo error for $A$.

for $\Theta$ indicate that none of the coefficients corresponding to the linear term are zero.

The posterior mean, 95% credible interval and monte carlo error Table 3.4 indicates a large spread for the values of the elements of the anisotropy matrix $A$. The monte carlo error is the highest for the elements of the matrix $A$ compared to that of all the parameters studied in the tables. We feel this is the case because there is no anisotropy detected in the data by the sampler. This is evident from the similar Table 3.5 constructed for the angle of the anisotropy matrix. For all the chains, the 95% credible interval for the angle includes zero. Thus, we conclude that anisotropy is absent in the data.

The posterior mean of $K$ from the three chains was 1.99, 2.72, 2.515 respectively. Thus we included 2, 3 and 3 basis functions in the fitted surfaces constructed from chains 1, 2 and 3 respectively. We notice from the trace plot of $K$ for chain 2 (the second row in A.6) that it takes longer to settle down. We do notice that all
Table 3.5: Posterior mean, 95% credible intervals and monte carlo error for angle of anisotropy in degrees.

<table>
<thead>
<tr>
<th>Angle</th>
<th>Mean</th>
<th>Credible Intv.</th>
<th>MC Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain 1</td>
<td>0.110</td>
<td>(-0.749, 0.780)</td>
<td>0.035</td>
</tr>
<tr>
<td>Chain 2</td>
<td>0.091</td>
<td>(-0.751, 0.765)</td>
<td>0.035</td>
</tr>
<tr>
<td>Chain 3</td>
<td>0.115</td>
<td>(-0.748, 0.765)</td>
<td>0.041</td>
</tr>
</tbody>
</table>

Table 3.6: Posterior mean, 95% credible intervals and monte carlo error for $\tau_e$.

<table>
<thead>
<tr>
<th>$\tau_e$</th>
<th>Mean</th>
<th>Credible Intv.</th>
<th>MC Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain 1</td>
<td>0.758</td>
<td>(0.591, 0.932)</td>
<td>0.0025</td>
</tr>
<tr>
<td>Chain 2</td>
<td>0.760</td>
<td>(0.594, 0.933)</td>
<td>0.002</td>
</tr>
<tr>
<td>Chain 3</td>
<td>0.767</td>
<td>(0.597, 0.947)</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Table 3.7: Posterior mean, 95% credible intervals and monte carlo error for $\tau$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Mean</th>
<th>Credible Intv.</th>
<th>MC Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain 1</td>
<td>0.252</td>
<td>(0.014, 0.853)</td>
<td>0.013</td>
</tr>
<tr>
<td>Chain 2</td>
<td>0.270</td>
<td>(0.017, 0.816)</td>
<td>0.009</td>
</tr>
<tr>
<td>Chain 3</td>
<td>0.271</td>
<td>(0.017, 0.796)</td>
<td>0.008</td>
</tr>
</tbody>
</table>
the chains move rapidly away from excursions into higher valued states. The acceptance rates for $K$ were 22%, 34% and 29% for chains 1, 2 and 3 respectively. For the anisotropy matrix $A$ the acceptance rates were 19%, 11% and 16% for chains 1, 2 and 3 respectively. Along with the monte carlo error associated with each of its elements (which are considerably higher compared to the monte carlo error for other parameters), this is an indication that the sampling algorithm for $A$ is not that efficient. Johnson [16] uses a decomposition of the anisotropy matrix in his model. Since $A$ is positive definite, it can be considered as a variance covariance matrix and decomposed into three matrices. The first and the third matrices are diagonal matrices with standard deviations along the diagonal. The second matrix is a correlation matrix. Johnson [16] relates the angle of anisotropy to this correlation. We believe that the angle of anisotropy is also related to the standard deviations and not only to the correlation parameter in the decomposition. Though his approach is promising, we believe it needs further work. For $\alpha_e$ and $\beta_e$ the acceptance rates for all the three chains are approximately 35% and 38% respectively. These are in accordance the rates usually recommended for MH algorithms.

We present the graphs relating the output in Appendix A. In the plot of the observed surface, the blue color represents the underside of the surface. The fitted surface plot from the three samples presented in Figures A.2–A.4 appears to be similar. The fitted surfaces are presented in equations (3.2)–(3.4). The superscript to the matrix $T_A$ corresponds to the number of columns chosen in evaluating the fitted surface.

$$y = X \begin{pmatrix} 7.790 \\ 0.382 \\ 0.280 \end{pmatrix} + T_A^{(2)} \begin{pmatrix} 0.035 \\ -0.011 \end{pmatrix} \quad (3.2)$$
\[ y = X \begin{pmatrix} 7.797 \\ 0.381 \\ 0.284 \end{pmatrix} + T_A^{(3)} \begin{pmatrix} 0.004 \\ 0.052 \\ -0.036 \end{pmatrix} \] 

(3.3)

\[ y = X \begin{pmatrix} 7.794 \\ 0.381 \\ 0.285 \end{pmatrix} + T_A^{(3)} \begin{pmatrix} 0.006 \\ 0.074 \\ -0.067 \end{pmatrix} \] 

(3.4)

We notice from equations (3.2)–(3.4) that the values of the coefficients of the basis functions are small. We suspect that this might have to do with the numerical values of the elements of the \( T_A \) matrix obtained after the eigenvalue decomposition. This is being investigated.

The histogram plot of the model indicators, \( K \) also appears to be similar for all the three samples. This is presented in Figure A.5. For all the three samples, it appears as if states above \( K = 10 \) are not visited as frequently as lower valued states.

We expect this, since the higher order eigenvalues which drive the acceptance ratio in the partially analytic reversible jump sampler, have low magnitude and are similar in value. The trace plots of \( K \) for all the three chains are presented in Figure A.6. The graph demonstrates that the sampler tends to move swiftly away from the extreme states.

The deviance trace plots in Figure A.7 graph \(-2 \log(\pi(y|\omega))\) across the iterations for the three chains. The deviance seems appears to have settled down between 240 – 310. The trace plots for the elements of the anisotropy matrix are presented in Figures A.8–A.10. It is to be mentioned that when we analyzed the results from the original data (before adding the quadratic term), the acceptance rates for new moves for \( A \) was in the region of 60–80% for all the three chains.

The trace plots of \( \Theta \) are presented in Figures A.11–A.13. There appears to be
extremely good mixing between the three samples. Same is the case with the trace plot for $\tau_e$ in Figure A.14. The trace plot for $\tau$, $\alpha_e$ and $\beta_e$ presented in Figures A.15, A.16 and A.17 respectively, all demonstrate good mixing.

The auto-correlation function (ACF) plots of the elements of the $A$ matrix are presented in Figures A.18–A.20. as would be expected because of the low acceptance rates for the anisotropy matrix that were presented earlier. ACF plots for the $\tau_e$ and $\tau$ in Figures A.21 and A.22 respectively, show rapid decay and the autocorrelations die out towards the higher lags. Figures A.23–A.25 present the autocorrelations for the elements of $\Theta$. There are no significant autocorrelations beyond lower lags in any of these figures.

The ACF plot for $\alpha_e$ is presented in Figure A.26. The autocorrelations in this plot die out steadily. The same is true for the autocorrelation plot of $\beta_e$ in Figure A.27.

The scatter plots amongst the elements of $\Theta$, presented in Figures A.29–A.31, show negative correlation. The correlation appears to be stronger between $\theta_1$ vs. $\theta_2$ and $\theta_3$ compared to the correlation between $\theta_2$ and $\theta_3$.

In trying to understand the relationship between the smoothness parameter and the elements of the anisotropy matrix, we present the scatter plots of $\tau$ vs. $A$ in Figures A.32–A.34. It is difficult to come to any conclusion as to the relationship between $\tau$ and the individual elements of the anisotropy matrix looking at these plots. Though, we do notice that really high values for all the elements of $A$ occur only for relatively small values of $\tau$.

The scatter plot of angle vs. $K$ is plotted in Figure A.35. We don’t see any particular pattern in the graph, suggesting that the angle of anisotropy remains unaffected by the number of basis functions chosen.
From the scatter plot of \( \tau \) vs \( K \) presented in Figure A.36, their complex relationship is not very evident. The birth and death acceptance probabilities in equations (2.25) and (2.27) provide some more clues. Lower values of \( \tau \) correspond to higher smoothness. But since the first few eigenvalues will typically be much larger than \( \tau \), we would expect many death moves to be accepted. This in turn would imply lower values of \( K \). In Figure A.36, for the first chain, it appears as if higher values of \( K \) correspond to higher values of \( \tau \). We believe that this might be happening because of the role \( \tau_e \) plays in the acceptance probability. We suspect that values of \( \tau_e \) are low when this happens. The effect of the prior distribution of \( \tau_e \) heavily influencing the reversible jump sampler is similar to what was observed with the prior distribution of ladder heights in Green [13] and in Richardson and Green [19]. Explaining the relationship between \( \tau \) and \( K \) still needs more work.

The plot of fitted values using PROC TPSPLINE in SAS is presented in Figure A.37. This fitted surface has more undulations than the surface fitted in our implementation. Further work is needed to investigate this phenomenon. At the outset though, it can be said that PROC TPSPLINE uses a slightly different form for the thin-plate spline basis function. The form of the basis function used by SAS is \( r^2 \log(r) \) where \( r \) represents the distance between two points. The form used by us is \( r^2 \log(r^2) \). Also PROC TPSPLINE doesn't model anisotropy as we have done. Whether these factors explain the differences between the fitted surfaces from the two different methods is a topic for future research.
CHAPTER 4
CONCLUSIONS AND FUTURE RESEARCH

We have presented a non-parametric approach to Bayesian surface smoothing in this thesis. We have used thin-plate spline basis functions to as the non-linear part in our model. Thin-plate splines are known to work favorably only in lower dimensions as in this thesis. A topic of future research would be to use other basis functions and compare their performance with the thin-plate splines. Since we are dealing with geometric anisotropy, the choice would probably be limited to radial basis functions.

Our use of the scaled eigenvectors as the basis functions has the advantage that knots for the splines needn’t be arbitrarily placed. The disadvantage of our method is that the sampler needs to calculate the eigenvectors and eigenvalues at each iteration. Since calculating eigenvalues increases as a cubic function of the order of the matrix, specialized algorithms, suggested as for example in Wood [26], might have to be used for large datasets. Using a nonparametric approach, as compared to a parametric one, has the added advantage that we are not constrained by any particular form of the covariance. There is more modeling flexibility.

We have presented an intuitive setup for the reversible jump step in our MCMC sampler. We have chosen a Poisson prior on the number of basis functions $K$, with the belief that only a few basis functions would be sufficient in modeling the data. Preliminary analysis done using a discrete Laplace type prior has yielded similar results in smoothing Gaussian data. Future research could look into a thorough posterior sensitivity analysis on $K$ by using different priors. Our use of the partially analytic RJMCMC sampler has presented us a neat formulation of the parameters driving the model selection step. This wasn’t the case when we were using the RJMCMC sampler without the simplification of the partially analytic step during our investigations.
In spite of the fact that the birth and death acceptance probabilities are more explanatory, we have not been able to completely understand the complex relationship between the smoothness parameter $\tau$ and the anisotropy matrix $A$ and the model order $K$.

We have used a nested structure in model selection process. This is because of two reasons. First the basis functions are eigenvectors; hence the first basis function would explain more variation than the second which in turn would explain more variation than the third and so on . . . . Thus, if we are to include two basis functions in our model, we believe they should be the first two. The second reason behind this nested structure has to do with the smoothness of the basis functions. The second basis function appears more jagged than the first, and less jagged the third and so on. These two reasons have led us to use the nested structure. The nested structure has also helped us in being able to use the partially analytic reversible jump sampler.

The positive definiteness of the anisotropy matrix has been restrictive in our choice for a prior distribution for it. We have chosen a Wishart prior for the anisotropy matrix. We have used ARCGIS software to provide us with the mean of our proposal in the absence of prior data. Since we are dealing with a small dataset, we hope that when our method is applied to a large dataset, we would not have use the current dataset to come up with parameters of the prior distribution of some of the variables. In cases where historical data is available, we could use it for coming up with choices of prior parameters. We believe that the method used by Johnson [16] is promising, but needs further work. We believe that there should be additional constraints to the correlation parameter, which Johnson says has to do with the angle of anisotropy. The range of anisotropy should also be considered in the model formulation. We have been able to incorporate this range of anisotropy by using the equations stated in Ecker and Gelfand [10] to evaluate the mean of the anisotropy matrix. The relationship between
the smoothness parameter and the anisotropy matrix also needs more research.

Our only constraint on the surface has been in terms of its smoothness. Future research is needed if the surface is to be endowed with more restrictions like concavity, order restrictions etc. Extending our ideas to include anisotropy along multiple angles is another challenging research area. With these unanswered questions, we hope Bayesian nonparametric smoothing under anisotropy will be an active area of research in the future.
Figure A.1: Surface plot of the observed values.
Figure A.2: Surface plot of the fitted values from chain 1.
Figure A.3: Surface plot of the fitted values from chain 2.
Figure A.4: Surface plot of the fitted values from chain 3.
Figure A.5: Histogram of the Model indicator (K) from chains 1, 2 and 3 respectively.
Figure A.6: Trace plots of $K$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.7: Trace plot of the deviance $-2 \log(\pi(y|\omega))$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.8: Trace plots of the $[1, 1]^{th}$ element of the anisotropy matrix $A$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.9: Trace plots of the $[1, 2]^{th}$ element of the anisotropy matrix $A$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.10: Trace plots of the $[2,2]^{th}$ element of the anisotropy matrix $A$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.11: Trace plots of $\theta_1$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.12: Trace plots of $\theta_2$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.13: Trace plots of $\theta_3$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.14: Trace plots of $\tau_e$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.15: Trace plots of $\tau$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.16: Trace plots of $\alpha_e$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.17: Trace plots of $\beta_e$ for chains 1 (red) 2 (blue) and 3 (brown).
Figure A.18: ACF plots of the $[1, 1]^{th}$ element of the anisotropy matrix $A$. 
Figure A.19: ACF plots of the $[1, 2]^{th}$ element of the anisotropy matrix $A$. 
Figure A.20: ACF plots of the $[2, 2]^{th}$ element of the anisotropy matrix $A$. 
Figure A.21: ACF plots of the $\tau_e$, the data precision.
Figure A.22: ACF plots of the $\tau$, the smoothness parameter.
Figure A.23: ACF plots of the $\theta_1$. 
Figure A.24: ACF plots of the $\theta_2$. 
Figure A.25: ACF plots of the $\theta_3$. 
Figure A.26: ACF plots of the $\alpha_e$. 
Figure A.27: ACF plots of the $\beta_e$. 
Figure A.28: Scatter plot showing posterior correlation between $\tau_e$ and $\tau$. 
Figure A.29: Scatter plot showing posterior correlation between $\theta_1$ and $\theta_2$. 
Figure A.30: Scatter plot showing posterior correlation between $\theta_1$ and $\theta_3$. 
Figure A.31: Scatter plot showing posterior correlation between $\theta_2$ and $\theta_3$. 
Figure A.32: Scatter plot showing posterior correlation between $\tau$ and $A[1, 1]$. 
Figure A.33: Scatter plot showing posterior correlation between $\tau$ and $A[1, 2]$. 
Figure A.34: Scatter plot showing posterior correlation between $\tau$ and $A[2, 2]$. 
Figure A.35: Scatter plot showing posterior correlation between angle of anisotropy and $K$. 
Figure A.36: Scatter plot showing posterior correlation between *angle* of anisotropy and *K*. 
Figure A.37: Plot of the fitted values using PROC TPSPLINE in SAS.
REFERENCES


