Exploring Geometric Anisotropy for Point-Referenced Spatial Data

---

A Thesis
Presented to
Department of Statistical Science
Duke University

Angie Shen

December 2017

Approved for the
Bachelor of Science in Statistical Science

_____
Dr. Alan Gelfand

_____
Dr. Colin Rundel

_____
Dr. Katherine Heller

_____
Dr. Mine Çetinkaya-Rundel, DUS

Contents

List of Tables

List of Figures

Abstract

Isotropic covariance functions are routinely adopted in specifying models for point-referenced spatial data. Implicit in such modeling is the assumption that spatial dependence is not directional. Geometric anisotropic models offer a class of specifications which incorporate directional dependence. They have received sparse attention in the literature because their associated number of parameters increase and become difficult to identify as we increase richness of the function. Adopting a Bayesian framework, this paper attempts to illuminate when and how much such models for random effects in geostatistical settings improve predictive performance. We show that geometric anisotropy yields better predictive performance when the data significantly departs from isotropy (anisotropy ratio is much greater than one), and that the improvement is more prominent when spatial variance is greater than pure error. The improvement in predictive performance is illustrated through simulation investigation as well as modeling data on scallop catches. We implement full Bayesian inference for model parameters using Markov chain Monte Carlo with a Metropolis-Hastings algorithm, adding kriging using composition sampling.

Introduction

Researchers in diverse areas such as climatology, ecology and environmental health are increasingly interested in analyzing data that are geographically referenced. For example, epidemiologists may be interested in studying the number of lung cancer incidents by county and state. Ecologists may be interested in the location of a particular species of trees in a forest. Researchers are often interested in statistical inference tasks, such as modeling of trends and correlation structures, estimation of underlying model parameters, hypothesis testing, model comparison and selection, and prediction of observations at unobserved locations (kriging).

With the advancement of Markov chain Monte Carlo (MCMC) computing, Bayesian modeling approaches have become increasingly common, as they enable hierarchical model structures where prior belief is updated with new information, as well as natural quantification of uncertainty through sampling schemes. The (Banerjee, Carlin, & Gelfand, 2014) text presents a thorough treatment of hierarchical Bayesian approaches for a variety of complex spatial data problems.

One common type of spatial data is *point-referenced* data, often referred to as *geostatistical* data, where we observe realizations of a spatial stochastic process at a fixed set of locations. We are often interested in a geographical distribution for the realizations that accounts for spatial correlation, typically in the presence of spatially referenced covariates. The simplest choices for modeling spatial correlation are *isotropic* covariance functions, where we assume spatial correlation between locations depends only on the distance between locations. In cases where this assumption does not hold, i.e., spatial correlation varies by direction, we can consider *anisotropic* covariance functions which depend on the separation vector between locations.

In the literature we find several notions of anisotropy, e.g., sill, nugget, and geometric anisotropy (Zimmerman, 1993). From a generative modeling perspective, the most useful form of anisotropy is *Geometric Anisotropy*, where coordinates are linearly transformed, i.e., rotated and stretched, to allow for different magnitudes of correlation in different directions. (Budrikaite & Ducinskas, 2005) explored different forms of geometric anisotropic variograms. (Eriksson & Siska, 2000) provided the geometrical details for modeling various types of anisotropy (range, sill, power, slope, nugget) on an ellipse. (Allard, Senoussi, & Porcu, 2016) derived a directional representation of anisotropies to build a large class of models that include and go beyond classical anisotropies such as the geometric and zonal ones. (Porcu, Gregori, & Mateu, 2006) incorporated anisotropy into spatio-temporal covariance models. (Ecker & Gelfand, 1999) proposed a Bayesian methodology for simultaneously estimating the linear transformation of the coordinates and other variogram parameters,

which also allows full inference for any characteristic of the geometrically anisotropic model. Following (Ecker & Gelfand, 1999) who proposed to use objective, independent priors for model parameters, (Kazianka, 2013) developed default priors and studied their posterior propriety.

This paper attempts to illuminate when and how much geometric anisotropic (henceforth anisotropic) models for random effects in geostatistical settings improve predictive performance. We use a transformation matrix parametrized by a decay parameter, a rotation angle and an anisotropy ratio. In the form of a simulation study, we compare the predictive performance of isotropic and anisotropic models for data generated under anisotropy. We also test the sensitivity of the models to different parameter values of the data generating model, including different sample sizes, different choices of anisotropy ratio, and different scales and ratios of spatial variance and pure error. We find that geometric anisotropy yields better predictive performance when the data significantly departs from isotropy (anisotropy ratio is much greater than one), and the improvement is more prominent when spatial variance is greater than pure error. We use Metropolis Hastings algorithm to perform full Bayesian inference for all model parameters. We then fit isotropic and anisotropic models to data on scallop catches used in (Ecker & Gelfand, 1999) which have been shown to suggest anisotropic behavior. We show that the anisotropic model performs better in terms of empirical coverage, predictive mean squared error and continuous rank probability score.

The paper proceeds as follows. Section 2 formally defines the isotropic and geometric anisotropic models we use for point-referenced data. Section 3 details our model fitting algorithm and distribution theory for making predictions. Section 4 lays out the metrics we use for model comparison. Section 5 demonstrates the simulation exercise with associated results. Section 6 presents modeling results for scallop catches data. Finally, Section 7 discusses future work.

Chapter 1

Models

1.1 Point-Level Models

Suppose $\{Y(s) : s \in D\}$ is a stochastic process where D is a fixed subset of r-dimensional Euclidean space. When $r = 2$, we say that $Y(s)$ is a *spatial process*. Data generated by such spatial process $Y(s)$ where s varies continuously over D is *point-referenced* data. For example, $Y(s)$ may represent pollutant level over a region, and we observe pollutant measurements at a finite set of locations$\{s_1, ..., s_n\}$ where there are monitoring stations. Even though $Y(s)$ exists on an infinite dimensional function space, in reality we can only observe data at a finite number of locations. The data is a partial realization of the spatial process. The problem facing statisticians is to make inference about the spatial process $Y(s)$ and prediction at new locations based on this partial realization.

We model $Y(s)$ with a *Gaussian Process*. Suppose our spatial process has a mean $\mu(s) = E(Y(s))$ associated with it and the variance of $Y(s)$ exists for all $s \in D$. $Y(s)$ is said to be *Gaussian* if for any $n \geq 1$ and any set of n sites $\{s_1, ..., s_n\}$, $Y = (Y(s_1), ..., Y(s_n))^T$ has a multivariate normal distribution. The remarkable feature of Gaussian Process models is that, despite only observing the spatial surface at a finite number of locations, we can infer about the process at an uncountable number of locations by specifying the association between pairs of locations through *structured* dependence. Suppose we assume the random variables at two locations depend on the *distance* of the locations. If we assume spatial correlation is a function solely of the distance $d_{ii'}$ between s_i and $s_{i'}$, the covariance function is *isotropic*. One commonly used isotropic covariance specification is the exponential model, where the covariance between measurements at two locations is an exponential function decreasing in the distance between the two locations.

Suppose we have observations $\mathbf{Y} = \{Y(s_1), ..., Y(s_n)\}$. We assume a multivariate normal model where

$$\mathbf{Y} \sim N_n(\mu\mathbf{1}, \Sigma(\theta)) \tag{1.1}$$

where N_n denotes the n-dimensional normal distribution, μ is the global mean[1], and $(\Sigma(\theta))_{ii'}$ gives the covariance between $Y(s_i)$ and $Y(s_{i'})$.

[1] We assume a constant mean surface here and in the sequel since our focus is on the effects of the choice of covariance function.

If we use an exponential correlation function,

$$\Sigma(\theta)_{ii'} = \begin{cases} \sigma^2 + \tau^2 & \text{if } d_{ii'} = 0 \\ \sigma^2 \exp(-\phi d_{ii'}) & \text{if } d_{ii'} > 0 \end{cases} \tag{1.2}$$

where $d_{ii'}$ is the distance between site s_i and $s_{i'}$, $\phi > 0$ is the *decay parameter* ($1/\phi$ is the *range* parameter), $\sigma^2 > 0$ is *partial sill* or spatial variance, $\tau^2 > 0$ is *nugget* or pure error, and $\tau^2 + \sigma^2$ is *sill*. The parameters of the covariance matrix are $\theta = (\sigma^2, \tau^2, \phi)^T$.

1.2 Isotropy

Suppose we observe $\{Y(s) : s \in D \subseteq \mathbb{R}^r\}$. For any $h \in \mathbb{R}^r$, *intrinsic stationarity* assumes that $E(Y(s+h) - Y(s)) = 0$ and that Var$(Y(s+h) - Y(s))$ is a function of the separation vector h, denoted by $2\gamma(h)$, the *variogram*. The stronger assumption of *weak stationarity* asserts that $E(Y(s)) = \mu$ (the mean is constant) and $\text{Cov}(Y(s), Y(s+h)) = C(h)$, and implies *intrinsic stationarity* with $\gamma(h) = C(0) - C(h)$. Under *isotropy*, the *semivariogram* function $\gamma(h)$ depends on the separation vector only through its length $d = \|h\|$. That is, the variogram is a function of the Euclidean distance d between two sites. The *sill* is $\lim_{d\to\infty} 2\gamma(d)$. The *range* is the distance at which the sill is reached; sites separated by distances larger than the range are uncorrelated. The *effective range* is defined as the distance such that $\text{Corr}(Y(s_i), Y(s_j)) = 0.05$. The *nugget* is $\lim_{d\to 0} 2\gamma(d)$, which need not be zero, to allow for measurement error and microscale variability.

1.3 Geometric Anisotropy

Isotropic variograms are popular because of their simplicity and interpretability. However, in many cases correlation function does not simply depend on the distance between locations, but depends on separation vector between locations. As a result, association depends upon direction. Here we explore covariance functions that are stationary but not isotropic. When the variogram is a function of both length and orientation of the separation vector h, the process $Y(s)$ is said to be *anisotropic*. That is, the semivariogram is $\gamma(h)$ rather than $\gamma(\|h\|)$.

Following (Zimmerman, 1993), anisotropy can take three general forms: *sill anisotropy*, *nugget anisotropy*, and *range anisotropy*. When $\lim_{a\to\infty} \gamma(ah/\|h\|)$ depends on the separation vector h, the situation is referred to as sill anisotropy. When $\lim_{a\to 0} \gamma(ah/\|h\|)$ depends on h, the situation is referred to as nugget anisotropy. Nugget anisotropy is typically ascribed to correlated measurement errors. A third type of anisotropy is range anisotropy where the range depends upon direction. This is the form most often seen in practice. A common case of range anisotropy is geometric anisotropy where we set

$$Cov(s - s') = \sigma^2 \rho((s - s')^T B(s - s')) \tag{1.3}$$

where B is a positive semidefinite matrix and ρ is a valid correlation function in R^r. When $r = 2$, B is 2×2, and the correlation specification will have three parameters rather than one

decay parameter. $B = \phi I$ corresponds to isotropy and arises from a linear transformation of coordinates under a general B. The contour corresponding to $\rho = 0.05$ provides the range in each spatial direction, and its shape is described by an ellipse.

As detailed in (Eriksson & Siska, 2000), for sites s_i and s_j, we define the covariance function as:

$$\Sigma(\theta)_{ij} = \sigma^2 \text{exp}(-\phi(h_{ij}^T B h_{ij})^{1/2}) + \tau^2 I(i = j) \tag{1.4}$$

where

$$B = AA^T, A = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} 1/r & 0 \\ 0 & 1 \end{bmatrix} \tag{1.5}$$

\ where $h_{ij} = s_i - s_j$. Here, α is orientation of the associated ellipse and r is anisotropy ratio (ratio of the major axis to the minor axis of the ellipse). The B matrix rotates and stretches the range of the covariance. Again, when B is a constant times the identity matrix, the covariance function is isotropic.

The range c in direction θ is obtained by taking a unit vector in direction θ, h_θ, and then solving $\rho(c_\theta^2 h_\theta^T B h_\theta) = .05$ for c_θ. In the case of the exponential correlation function, we have

$$\text{exp}(-c_\theta(h_\theta^T B h_\theta)^{1/2}) = 0.05 \tag{1.6}$$

where $h_\theta = (cos(\theta), sin(\theta))$.

Figure 1.1 shows spatial ranges in every direction from 0° to 360° with different combinations of the rotation angle and anisotropy ratio. The right panel shows the range in polar coordinates, forming ellipses. We can see that when the anisotropy axis ratio is closer to 1, the range does not vary much by direction and associated ellipse is closer to a circle.

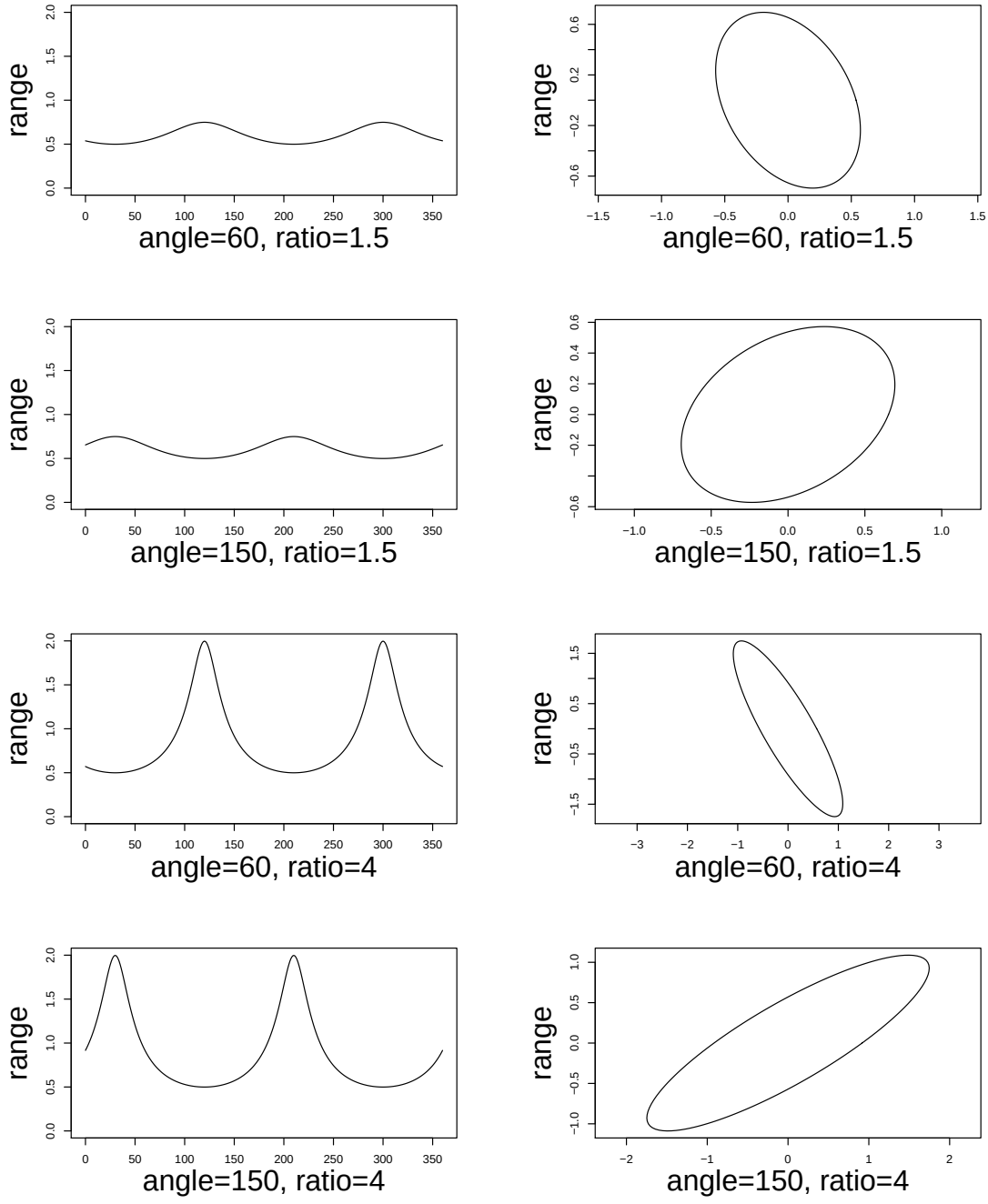

Figure 1.1: Spatial range with different specifications of rotation angle and anisotropy ratio

Chapter 2

Methodology

2.1 Model Fitting

We use the Bayesian framework to perform inference on model parameters and predictions for new locations. Not only do we simultaneously estimate for the ratio of major axis to minor axis of the ellipse, the angle of orientation of the ellipse with respect to the x-axis, the decay parameter, and the additional variogram parameters, but we provide complete inference in the form of a posterior distribution for each parameter. In addition, the Bayesian framework enables incorporation of prior information. If a priori, the process is expected to exhibit geometric anisotropy whose characteristics we can quantify, this information is easily built into the prior specifications and the model.

We model the data with a Gaussian Process

$$\mathbf{Y} \sim N_n(\mu\mathbf{1}, \Sigma(\theta)) \tag{2.1}$$

where $\Sigma(\theta)$ is specified by (4) and (5), and $\theta = (\sigma^2, \tau^2, \phi, \alpha, r)^T$. Under the likelihood implied by (7) for a set of say n locations, we complete the Bayesian formulation by assuming the prior takes the form

$$\pi(\mu, \theta) = \pi(\mu, \sigma^2, \tau^2, \phi, \alpha, r) = \pi(\mu)\pi(\sigma^2)\pi(\tau^2)\pi(\phi)\pi(\alpha)\pi(r), \tag{2.2}$$

i.e., assuming all the parameters are independent.

We use a weak $N(0, 1000)$ prior for μ. We use vague inverse gamma distributions for the variances σ^2 and τ^2 and anisotropy ratio r with shape and scale parameters both equal to one, implying no prior mean or variance. We put a uniform$(0, \pi)$ prior on rotation angle α and a uniform$(0.1D, 0.5D)$ on range $1/\phi$, where D is the maximum distance between locations.

We use a Metropolis-Hastings sampling algorithm to obtain samples from the posterior distribution of all parameters $p(\theta|y)$. The Metropolis-Hastings algorithm is a Markov chain Monte Carlo (MCMC) method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. As more and more sample values are produced, the distribution of values more closely approximates the desired posterior distribution $p(\theta|y)$.

Intuitively, assume we have a collection of $\{\theta^{(1)}, ..., \theta^{(s)}\}$. To generate a new value $\theta^{(s+1)}$, we sample a new value θ^* that is nearby $\theta^{(s)}$ and consider whether to accept the sampled θ^* into the collection of $\theta's$. We sample θ^* from a proposal distribution J centered on $\theta^{(s)}$, and accept θ^* if $p(\theta^*|y) > p(\theta^{(s)}|y)$. If $p(\theta^*|y) < p(\theta^{(s)}|y)$, we accept θ^* with some probability. In Metropolis algorithm, the proposal distribution J is symmetric. That is, $J(\theta^*|\theta^{(s)}) = J(\theta^{(s)}|\theta^*)$. In Metropolis-Hastings, J may not be symmetric.

The Metropolis-Hastings algorithm proceeds as follows.

1. Sample $\theta^* \sim J(\theta|\theta^{(s)})$

2. Compute acceptance ratio r:

$$r = \frac{p(\theta^*|y)}{p(\theta^{(s)}|y)} \times \frac{J(\theta^*|\theta^{(s)})}{J(\theta^{(s)}|\theta^*)} = \frac{p(y|\theta^*)p(\theta^*)}{p(y|\theta^{(s)})p(\theta^{(s)})} \times \frac{J(\theta^*|\theta^{(s)})}{J(\theta^{(s)}|\theta^*)}$$

   (In the case of Metropolis, $\frac{J(\theta^*|\theta^{(s)})}{J(\theta^{(s)}|\theta^*)} = 1$)

3. Sample $u \sim \text{Uniform}(0, 1)$. Set $\theta^{(s+1)} = \theta^*$ if $u < r$ and set $\theta^{(s+1)} = \theta^s$ otherwise.

We use appropriate proposal distributions for different parameters. For rotation angle α, we generate normal proposals mod π. For anisotropy ratio r, we generate proposals from truncated normal distribution from 0 to ∞ as r can only be positive. We also experimented with using log normal proposal for r, but the sampler behaved poorly as it frequently generated extremely big values. For σ^2, τ^2 and ϕ we generate proposals from log normal distribution. Finally, we generate μ from a normal distribution. Table 1 lists the proposal and prior distributions we use for all the parameters.

Because α and r are highly correlated, we update them jointly. We tune the variances of the proposal distributions so that we accept around 25% of all generated samples to achieve optimal efficiency of the sampler. If proposals vary too much, the sampler will reject too many samples which is inefficient. If proposals vary too little, the chain of samples will not move very much and might get stuck in a local mode.

	Proposal Distribution	Prior Distribution
α	Normal (mod π)	Uniform$(0, \pi)$
r	Truncated Normal $_{(0,+\infty)}$	Inverse Gamma$(1, 1)$
σ^2	Log Normal	Inverse Gamma$(1, 1)$
τ^2	Log Normal	Inverse Gamma$(1, 1)$
ϕ	Log Normal	Uniform $(3/0.5D, 3/0.2D)$
μ	Normal	Normal$(0, 1000)$

Table 2.1: Proposal and prior distributions for Metropolis algorithm

2.2 Kriging

Spatial prediction in the point-referenced data setting is often referred to as *kriging*. Given observations $Y = (Y(s_1), ..., Y(s_n))$, how do we predict $Y(s_0)$, where s_0 is a site that has not been observed? An observation at location we want to predict s_0 follows the following distribution:

$$y(s_o) = \mu + w(s_0) + \epsilon(s)$$

There are two ways of formulating the posterior predictive distribution. The first way samples spatial random effect w:

$$[Y(s_0)|Y] = \int [Y(s_0)|w(s_0), \theta][w(s_0)|w, \theta][w|\theta, Y][Y|\theta]$$
$$[w|\theta, Y] = N((\frac{1}{\tau^2}I + \Sigma^{-1})^{-1}\frac{1}{\tau^2}(Y - \mu\mathbf{1}), (\frac{1}{\tau^2}I + \Sigma^{-1})^{-1})$$
$$[w(s_0)|w] = N(r^T\Sigma^{-1}w, \sigma^2 - r^T\Sigma^{-1}r)$$
$$[Y(s_0)|w(s_0), \theta] = N(\mu + w(s_0), \tau^2)$$

where r is the $n \times 1$ covariance matrix between the new location s_0 and all other observed locations, cov$(w(s_0), w(s_i)), i = 1, 2, ...n$. Σ is the $n \times n$ covariance matrix between the random effects at the observed locations, $(w(s_1), ...w(s_n))$.\\

The second way marginalizes out spatial random effect w:

$$[Y(s_0)|Y] = \int [Y(s_0)|Y, \theta][\theta|Y]$$
$$[Y(s_0)|Y, \theta][\theta|Y] = N(\mu + r^T(\Sigma + \tau^2 I)^{-1}(Y - \mu\mathbf{1}), \sigma^2 + \tau^2 - r^T(\Sigma + \tau^2 I)^{-1}r)$$

We use the second approach in our algorithm as it has been shown to yield better sampling behavior.

Chapter 3

Model Comparison

We use the following three metrics for model comparison.

3.1 Predictive Mean Squared Error

The predictive mean squared error measures the mean squared difference between the predicted value and the observed value averaged over all hold out locations.

$$PMSE = \frac{1}{n}\sum(\hat{Y} - Y_{obs})^2$$

In the Bayesian context, $\hat{Y}$ for each location is the posterior predictive mean. We then average the squared difference between $\hat{Y}$ and the observed value over all hold out locations.

3.2 Empirical Coverage

Empirical coverage assesses how well credible intervals, derived from the posterior predictive distributions of the predictions, capture the observed values of the hold out sample. Suppose we obtain 90% credible intervals from the posterior predictive distributions for each of the hold out locations. If the credible intervals capture the observed value for $<$90% hold out locations, we have *under coverage*. If the credible intervals capture the observed value for $>$90% hold out locations, we have *over coverage*. Since the empirical coverage is random, we will criticize the adequacy of the model when the departure from the nominal coverage is consequential.

3.3 Continuous Rank Probability Score (CRPS)

To examine how concentrated the predictive distribution of $Y(s_0)$ is around the observed value, we use the Continuous Rank Probability Score (CRPS) metric, the squared integrated distance between the predictive distribution and the degenerate distribution at the observed value,

$$CRPS(F, y) = \int_{-\infty}^{\infty}(F(u) - I(u \geq y))^2 du$$

where F is the predictive distribution and y is the observed value. In our case, $Y(s_0)$ is the observation and F is the posterior predictive distribution for $Y(s_0)$. With a collection of hold out observations and associated predictive distributions, we would average the CRPS over these observations for model comparison. In our case, under MCMC model fitting, we would not have F directly, but rather a sample from F. We use the alternative form of CRPS:

$$CRPS(F, y) = \frac{1}{2} E_F|Y - Y'| + E_F|Y - y|$$

where Y and Y' are independent replicates from F. With samples from F, we can use Monte Carlo integration to compute CRPS.

Chapter 4

Simulation Study

We design a set of simulations to compare the predictive performance of isotropic and anisotropic models for data generated under anisotropy. We test the sensitivity of the models to different parameter values of the data generating model, including different sample sizes, different choices of anisotropy ratio, and different scales of spatial variance and pure error. We don't vary the choice of rotation angle, as the angle does not impact the degree of departure from isotropy.

We randomly generate locations on a unit square. We generate observations following a multivariate normal distribution with constant global mean 0 and exponential covariance function as specified in (4) and (5). Spatial range is 0.5, around half of the maximum distance between the locations. The decay parameter ϕ is $3/0.5 = 6$. We fit the model and predict for a hold out sample using the methodology outlined in section 3. We fit the isotropic model using the R package spBayes (Finley, Banerjee, & Carlin, 2007). We compare the predictive performance of the isotropic model and the anisotropic model using empirical coverage, PMSE and CRPS as defined in Section 4.

We choose three different anisotropy ratios, 1.5, 4 and 8, as well as four different combinations of spatial variance σ^2 and pure error τ^2. We anticipate that, when the anisotropy ratio is big, spatial variance of simulated data is high, and spatial variance is significantly larger compared to pure error, the predictive performance of the anisotropic model will significantly improve over the isotropic model. We use two different sample sizes 100 (with 40 additional observations for hold out) and 500 (with 100 additional observations for hold out). We further anticipate that the two models will be more distinguishable with the bigger sample size.

Table 2-7 display the comparisons of empirical coverage, PMSE, and CRPS of anisotropy and isotropy under the different data generation scenarios. All results are averaged over 10 randomly generated datasets. When sample size is 100, anisotropy yields smaller MSE than isotropy under all but one scenario when anisotropy ratio is 1.5. As we expect, the improvement is more prominent when anisotropy ratio is 8 (10%) than when it is 4 (1%), when spatial variance is greater than non-spatial variance (11%) than when they are equal (3%) and when both variances are small (7%) than when they are big (6%). The CRPS confirms these results. The empirical coverage of the anisotropic model is closer to the nominal 90% under all scenarios. These conclusions are confirmed when sample size is 500. Sample size does not seem to make a difference in distinguishing the models.

$r = 1.5$	(0.9275, 0.93)
$r = 4$	(0.92, 0.9275)
$r = 8$	(0.9125, 0.925)

	$\sigma^2/\tau^2 = 1$	$\sigma^2/\tau^2 = 5$
$\sigma^2 = 0.2$	(0.9175, 0.9175)	(0.9575, 0.96)
$\sigma^2 = 1$	(0.89, 0.8775)	(0.9125, 0.925)

Table 4.1: Model comparison for simulated data, 90% Empirical Coverage (Anisotropy, Isotropy), sample size = 100

$r = 1.5$	(0.925, 0.9)
$r = 4$	(0.916, 0.906)
$r = 8$	(0.9, 0.902)

	$\sigma^2/\tau^2 = 1$	$\sigma^2/\tau^2 = 5$
$\sigma^2 = 0.2$	(0.905, 0.904)	(0.928, 0.912)
$\sigma^2 = 1$	(0.896, 0.898)	(0.9, 0.902)

Table 4.2: Model comparison for simulated data, 90% empirical coverage (anisotropy, isotropy), sample size = 500

$r = 1.5$	(0.588, 0.528)
$r = 4$	(0.497, 0.504)
$r = 8$	(0.411, 0.457)

	$\sigma^2/\tau^2 = 1$	$\sigma^2/\tau^2 = 5$
$\sigma^2 = 0.2$	(0.265, 0.275)	(0.082, 0.093)
$\sigma^2 = 1$	(1.377, 1.407)	(0.411, 0.457)

Table 4.3: Model comparison for simulated data, predictive mean squared error (anisotropy, isotropy), sample size = 100

$r = 1.5$	(0.487, 0.399)
$r = 4$	(0.334, 0.338)
$r = 8$	(0.318, 0.345)

	$\sigma^2/\tau^2 = 1$	$\sigma^2/\tau^2 = 5$
$\sigma^2 = 0.2$	(0.226, 0.239)	(0.061, 0.068)
$\sigma^2 = 1$	(1.141, 1.197)	(0.318, 0.345)

Table 4.4: Model comparison for simulated data, predictive mean squared error (anisotropy, isotropy), sample size = 500

$r = 1.5$	(0.431, 0.409)
$r = 4$	(0.397, 0.399)
$r = 8$	(0.364, 0.385)

	$\sigma^2/\tau^2 = 1$	$\sigma^2/\tau^2 = 5$
$\sigma^2 = 0.2$	(0.292, 0.299)	(0.165, 0.176)
$\sigma^2 = 1$	(0.664, 0.672)	(0.364, 0.385)

Table 4.5: Model comparison for simulated data, continuous rank probability score (anisotropy, isotropy), sample size = 100

$r = 1.5$	(0.391, 0.355)
$r = 4$	(0.325, 0.326)
$r = 8$	(0.317, 0.329)

	$\sigma^2/\tau^2 = 1$	$\sigma^2/\tau^2 = 5$
$\sigma^2 = 0.2$	(0.268, 0.275)	(0.139, 0.147)
$\sigma^2 = 1$	(0.603, 0.617)	(0.317, 0.329)

Table 4.6: Model comparison for simulated data, continuous rank probability score (anisotropy, isotropy), sample size = 500

Chapter 5

A real Data Example: Scallop Catches

We apply the models to data on sampling of scallop abundance on the continental shelf off the coastline of the northeastern U.S. The data comes from a survey conducted by the Northeast Fisheries Science Center of the National Marine Fisheries Service. Figure 5.1 shows the sampling sites for 1993 and Figure 5.2 shows the spatial surface as well as contours of the data at the sites used for model fitting. There are a total of 148 sampling sites. We use 118 for model fitting and 30 as hold out data for kriging. We see a lot more sampling in southwest to northeast direction than in the northwest to southeast direction. Evidently it is more appropriate to follow the coastline in searching for scallops.

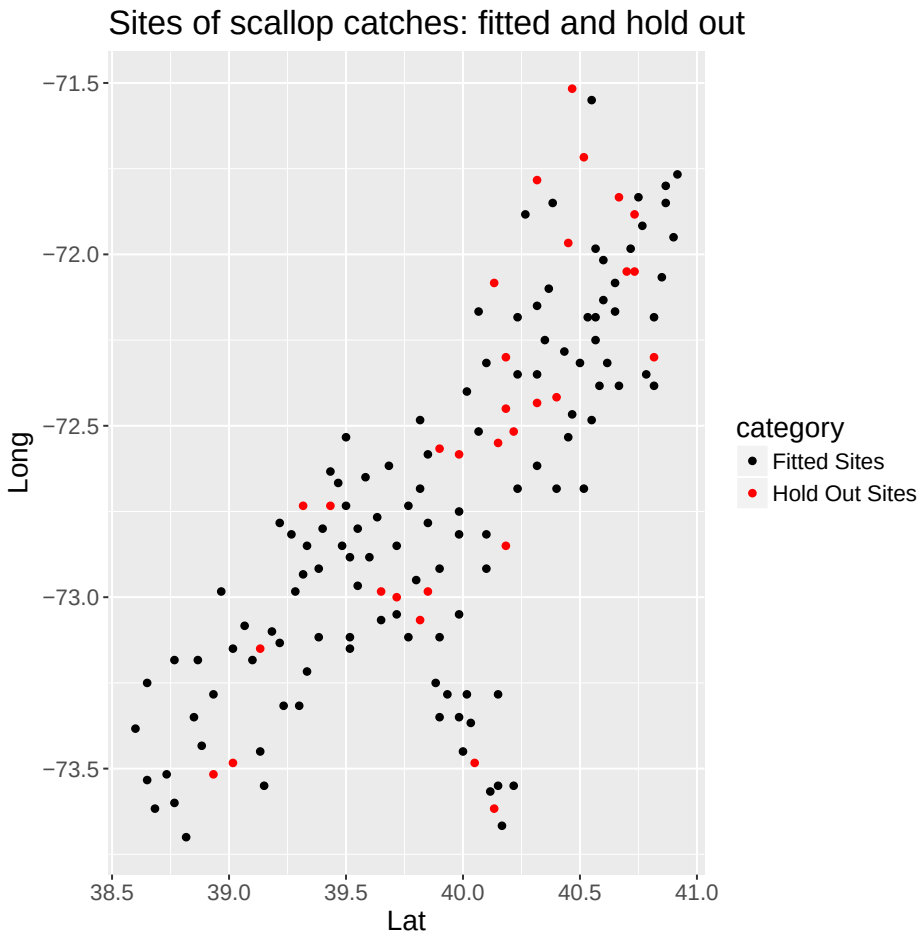

Figure 5.1: Sites sampled in the Atlantic Ocean for 1993 scallop catch data (fitted and hold out)

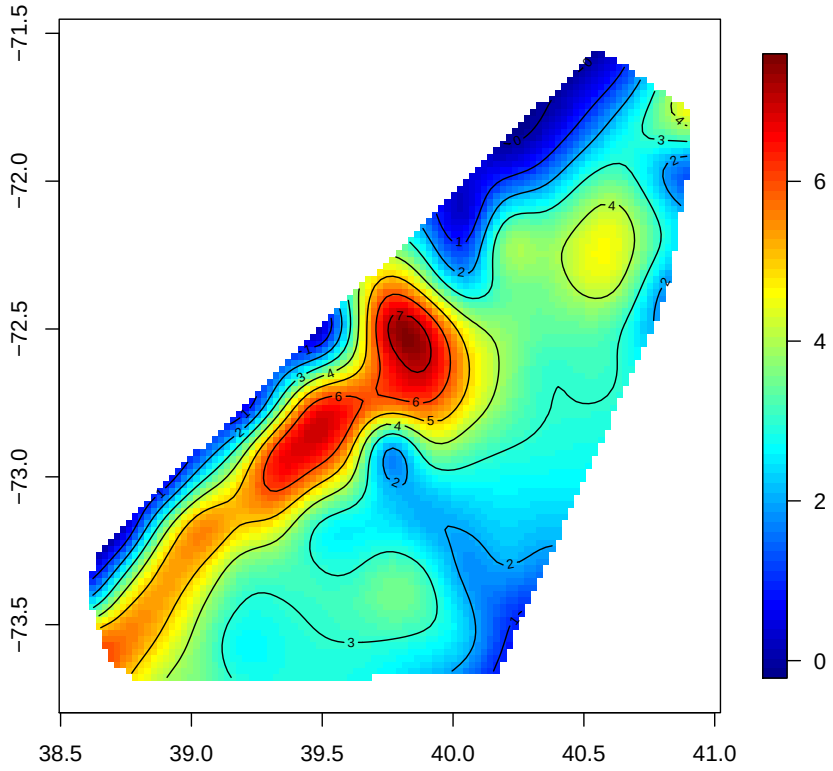

Figure 5.2: Surface and contour plot for 1993 scallop data (fitted)

Again, we fit the anisotropic model using methodology detailed in section 3, and the isotropic model using spBayes. In both cases, we run 30,000 iterations, using a burn-in of 20,000 and a thinning rate of 1/20 for the remaining 10,000 samples. Therefore, we retain 500 posterior samples of all model parameters. Figure 5.3 shows the density plot of the remaining posterior samples under anisotropy. Table 5.1 shows the posterior means and 95% credible intervals for all model parameters under isotropy and anisotropy.

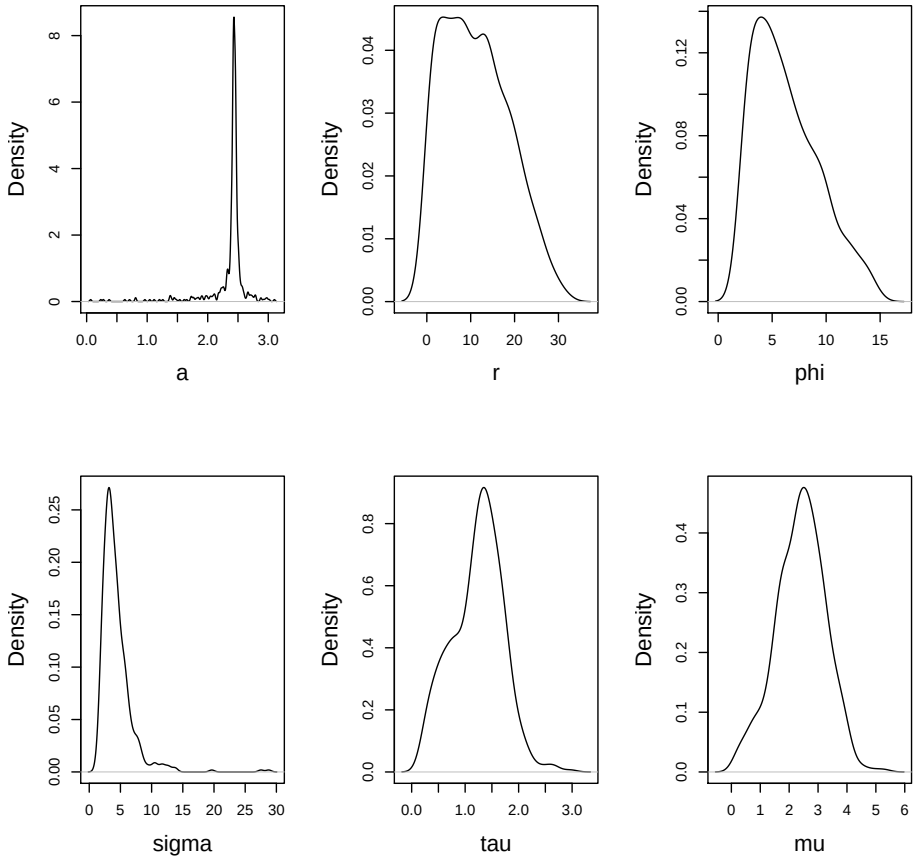

Figure 5.3: Density plots of posterior samples for all parameters under anisotropy (after burn-in and thinning)

Parameters	μ	ϕ	σ^2	τ^2	α	r
Isotropy	NA	4.92 (2.49, 8.71)	5.20 (3.10, 8.94)	0.49 (0.19, 1.15)	NA	NA
Anisotropy	2.40 (0.57, 3.96)	6.3 (2.23, 13.24)	4.36 (1.77, 10.81)	1.23 (0.30, 2.07)	2.35 (1.31, 2.74)	11.2 (0.49, 26.27)

Table 5.1: Posterior means and 95% credible intervals for all model parameters under isotropy and anisotropy

To show evidence for departure from isotropy, we obtain the posterior distribution for range in each direction. Figure 5.4 shows the mean posterior range plotted as a function of angle with associated individual 95% credible intervals. The plot on the right shows the range in polar coordinates which forms an ellipse.

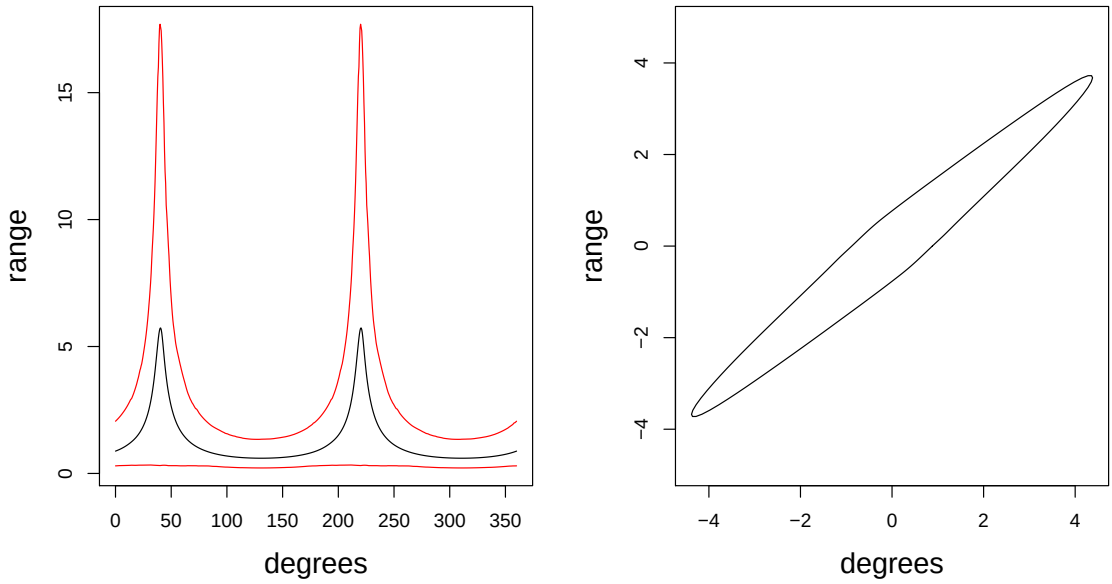

Figure 5.4: Mean posterior range plotted as a function of angle with associated individual 95 percent credible intervals. The plot on the right shows the range in polar coordinates which forms an ellipse.

We evaluate the predictive performance of the two models on the 30 hold-out sites using empirical coverage, PMSE and CRPS, as displayed in Table 5.2. The anisotropic model reduces average PMSE by 12% and reduces average CRPS by 7% over the 30 sites. Anisotropy also has higher empirical coverage than isotropy.

Model	EC	PMSE	CRPS
Isotropy	86.7%	1.792	0.781
Anisotropy	96.7%	1.581	0.725

Table 5.2: Model comparison of isotropy and anisotropy for scallops data: 90% empirical coverage, PMSE, and CRPS

Figure 5.5 shows the empirical coverage of isotropy and anisotropy, where the dots represent the observed values and the grey lines represent the 90% credible interval of the posterior predictive samples. We can see the credible intervals produced by isotropy fail to capture the observed value for 4 of the 30 sites, while anisotropy fails to capture 1 site.

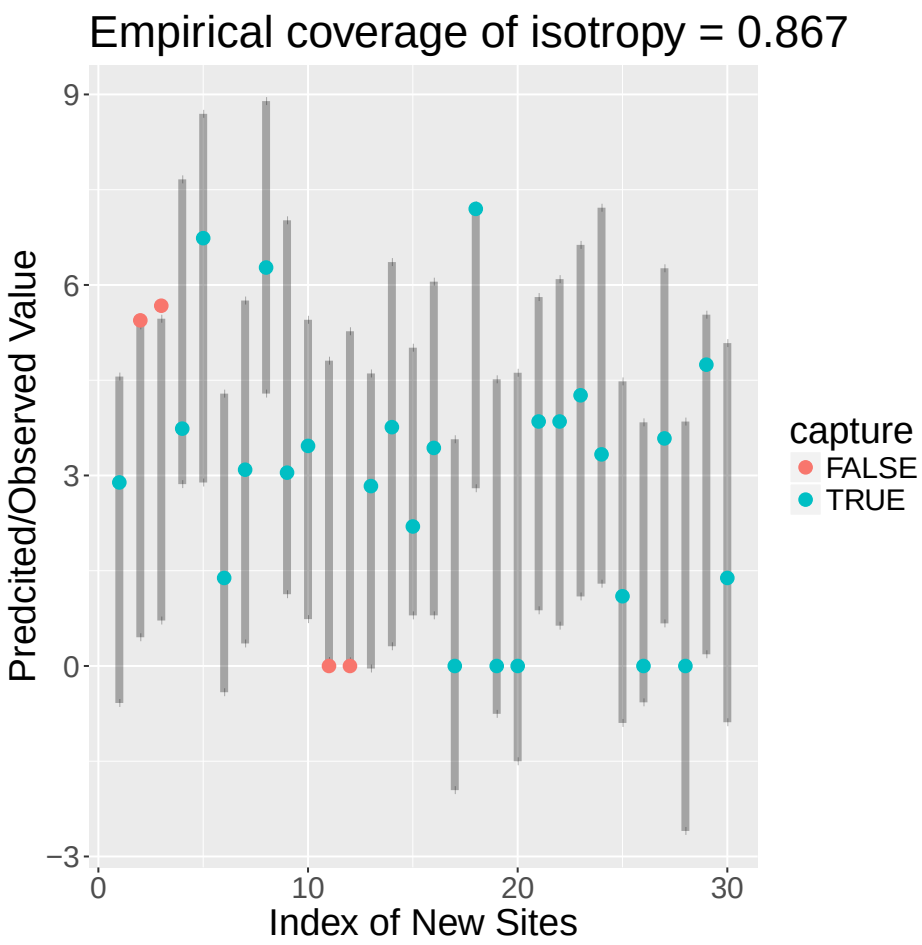

Figure 5.5: Model comparison: empirical coverage of isotropy and anisotropy

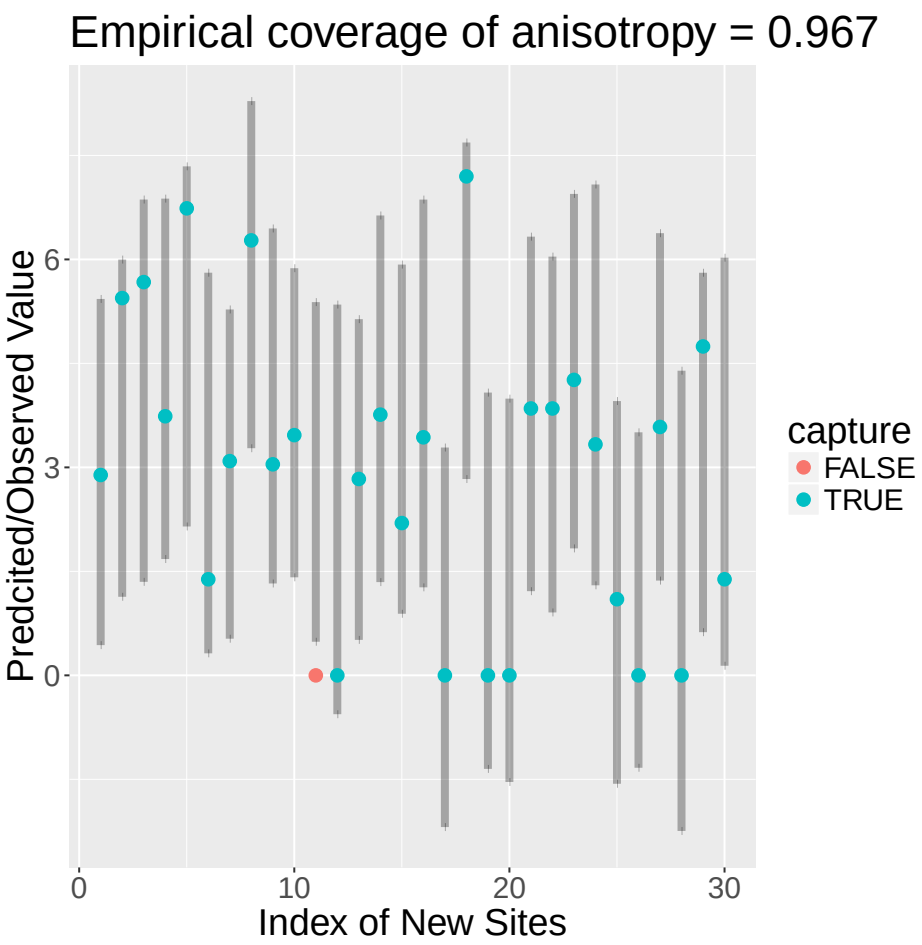

Figure 5.6: Model comparison: empirical coverage of isotropy and anisotropy

Figure 5.7 shows the posterior predictive distribution and PMSE for 4 randomly selected hold out sites under isotropy and anisotropy. The vertical line represents the observed value. Under anisotropy, the predictive distributions are more closely concentrated around the observed value, and the PMSE's are smaller.

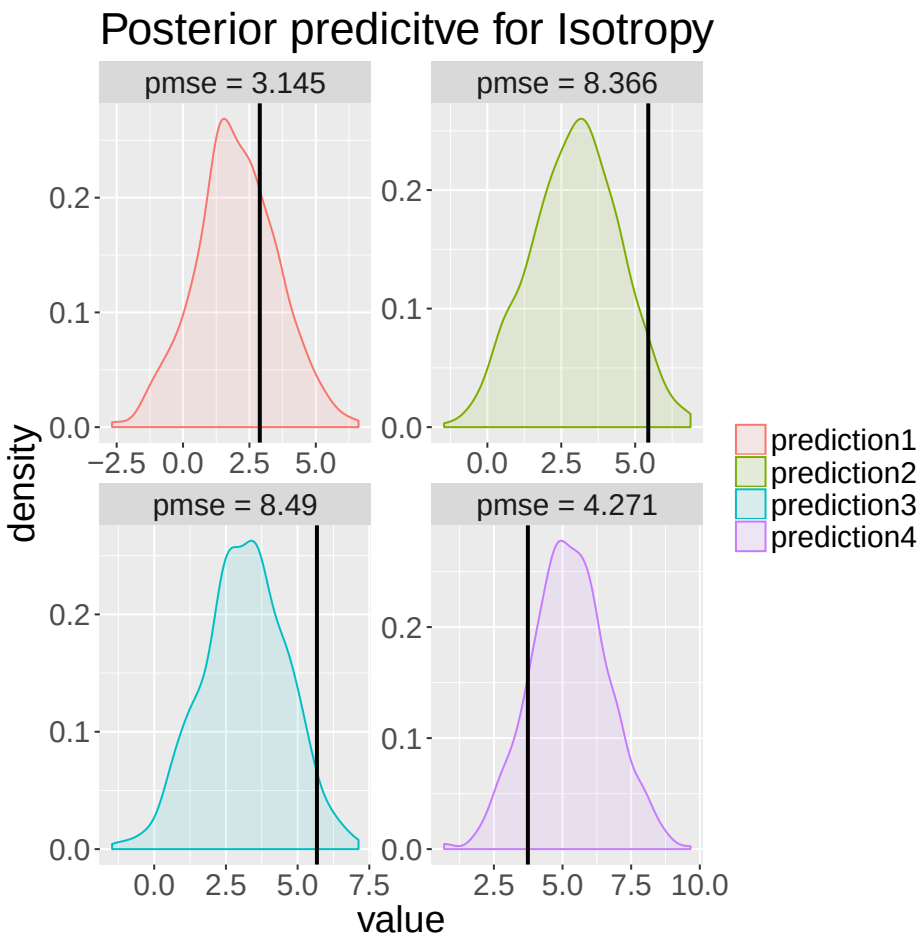

Figure 5.7: Model comparison: posterior predictive distribution and PMSE for 4 hold out sites under isotropy and anisotropy. Vertical line represents the observed value.

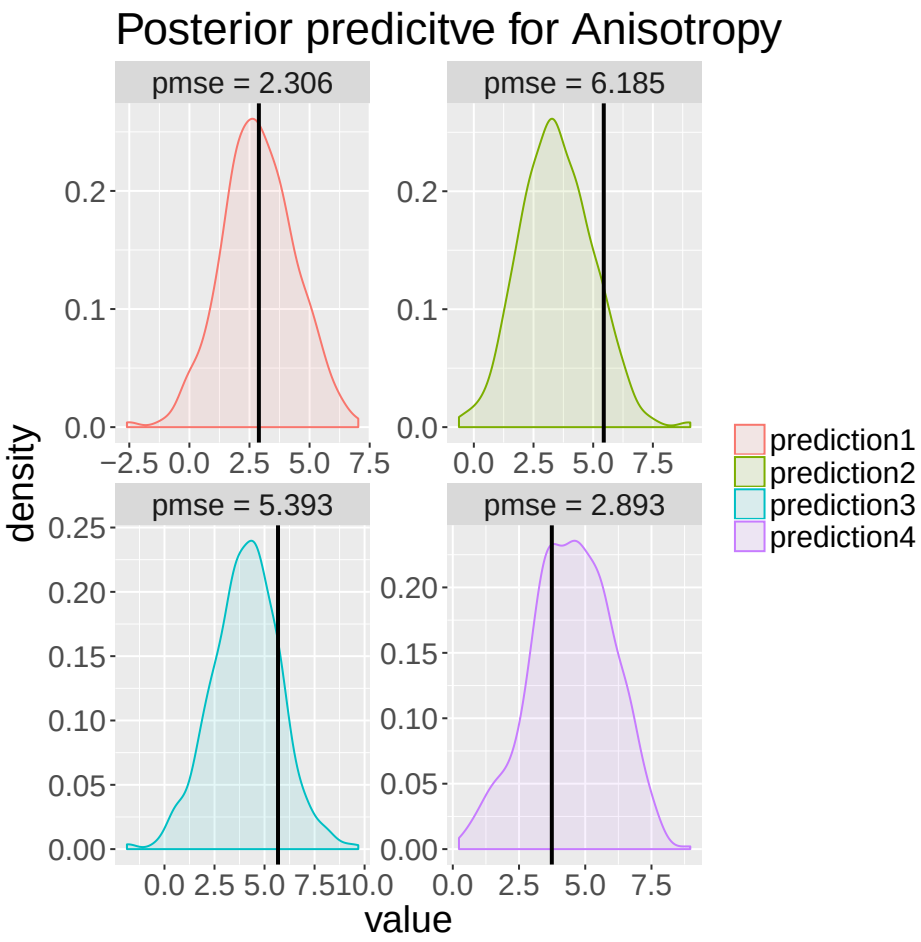

Figure 5.8: Model comparison: posterior predictive distribution and PMSE for 4 hold out sites under isotropy and anisotropy. Vertical line represents the observed value.

Chapter 6

Discussion and Future Work

This paper explored when and how much geometric anisotropic models for random effects in geostatistical settings improve predictive performance over isotropic models. We compared the predictive performance of isotropic and anisotropic models for data generated under anisotropy, using different values of the parameters in the data generating model, including different sample sizes, different choices of anisotropy ratio, and different scales and ratios of spatial variance and pure error. We found that anisotropy yields better predictive performance when the data significantly departs from isotropy (anisotropy ratio is much greater than one), and that the improvement is more prominent when the anisotropy ratio is higher and when spatial variance is higher compared to pure error, regardless of sample size. The anisotropic model yields much better predictive results on the real scallop catches data, which have been suggested to exhibit anisotropic behavior. We performed full Bayesian inference on all model parameters using a Metropolis-Hastings algorithm for model fitting.

Future work involves extending the geometric anisotropic model assessment to multivariate observations at locations and to space-time settings. We will also explore different ways of constructing stationarity by using different covariance structures. In particular we are interested in exploring product covariance function, i.e. the product of one-dimensional covariance function in the x-coordinate and a one-dimensional covariance function in the y-coordinate, which takes the following form:

$$C(d) = \begin{cases} \tau^2 + \sigma^2 & \text{if } d = 0 \\ \sigma^2 \text{exp}(-\phi_x d_x)\text{exp}(-\phi_y d_y) & \text{if } d > 0 \end{cases}$$

where d_x is the distance between x-coordinates and d_y is the distance between y-coordinates.

References

Allard, D., Senoussi, R., & Porcu, E. (2016). Anisotropy models for spatial data. *Mathematical Geosciences*, *48*(3), 305–328.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Taylor; Francis.

Budrikaite, L., & Ducinskas, K. (2005). Modelling of geometric anisotropic spatial variation. *Mathematical Modeling and Analysis*, 361–366.

Ecker, M. D., & Gelfand, A. E. (1999). Bayesian modeling and inference for geometrically anisotropic spatial data. *Mathematical Geology*, *31*(1), 67–83.

Eriksson, M., & Siska, P. P. (2000). Understanding anisotropy computations. *Mathematical Geology*, *32*(6), 683–700.

Finley, A. O., Banerjee, S., & Carlin, B. P. (2007). spBayes: An R package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software*, *19*(4), 1–24. Retrieved from `http://www.jstatsoft.org/v19/i04/`

Kazianka, H. (2013). Objective bayesian analysis of geometrically anisotropic spatial data. *Journal of Agricultural, Biological, and Environmental Statistics*, *18*(4), 514–537.

Porcu, E., Gregori, P., & Mateu, J. (2006). Nonseparable stationary anisotropic space–time covariance functions. *Stochastic Environmental Research and Risk Assessment*, *21*, 113–122.

Zimmerman, D. L. (1993). Another look at anisotropy in geostatistcs. *Mathematical Geology*, *25*(4), 453–470.