Explore Geometric Anisotropy for Point Referenced Spatial Data

Angie Shen

October 8, 2017

1 Introduction

2 Models

2.1 Isotropic point-level models

Suppose $\{Y(s) : s \in D\}$ is a stochastic process where D is a fixed subset of 2-dimensional Euclidean space. Even though Y(s) exists on an infinite dimensional function space, in reality we can only observe data at a finite number of locations. 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. Data generated by such spatial process $Y(s)$ where s varies continuously over D is categorized as *point-referenced data*.

Usually we model $Y(s)$ with *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 $n \geq 1$, any site 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 a finite number of locations, we can infer about the process at an uncountable number of locations by specifying the association between locations through *structured* dependence. Suppose we assume the random variables at two locations depend on the *distance* of the locations. One commonly used covariance specification is the exponential model, where the covariance between measurements at two locations is an exponential function of the distance between the two 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*. Suppose we have observations $\mathbf{Y} = \{Y(s_1), ..., Y(s_n)\}$. We assume a multivariate normal model where

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

where N_n denotes the n-dimensional normal distribution, μ is the global mean, and $(\Sigma(\theta))_{ii'}$ gives the covariance between $Y(s_i)$ and $Y(s_{i'})$. If we choose an isotropic exponential correlation function, then

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

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

2.2 Geometric anisotropy

In spatial anisotropy, correlation function does not simply depend on the distance between locations, but depends on separation vector between locations. The spatial correlation between locations depend on their direction in relation to each other in addition to distance. A commonly used set of anisotropic covariance functions are geometric covariance functions where

$$Cov(s - s') = \sigma^2 \rho(h^T B h) \tag{3}$$

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 range parameter. Solving for $\rho = 0.05$ gives us the range in each direction, and the contours of the range form an ellipse.

While there are a few ways of parametrizing range geometric anisotropy, we choose a more intuitive approach where the covariance function is defined as:

$$\Sigma(\theta)_{ii'} = \sigma^2 \text{exp}(-\phi(h_{ij}^T B h_{ij})^{1/2}) + \tau^2 I(i = i') \tag{4}$$

where

$$B = AA^T, A = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} 1/r & 0 \\ 0 & 1 \end{bmatrix} \tag{5}$$

where $h_{ij} = s_i - s_j$, 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 rotate and stretch the range of the covariance. When B is an identity matrix, the covariance function is isotropic. Figure 1 shows ellipses with different rotation angles and major-to-minor axis ratios.

For any correlation function $\rho(h^T B h)$ where here $B = \phi A^T A$ with A defined through α and r, 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 exponential correlation function, we have

$$\text{exp}(-c_\theta (h_\theta^T B h_\theta)^{1/2}) = 0.05 \tag{6}$$

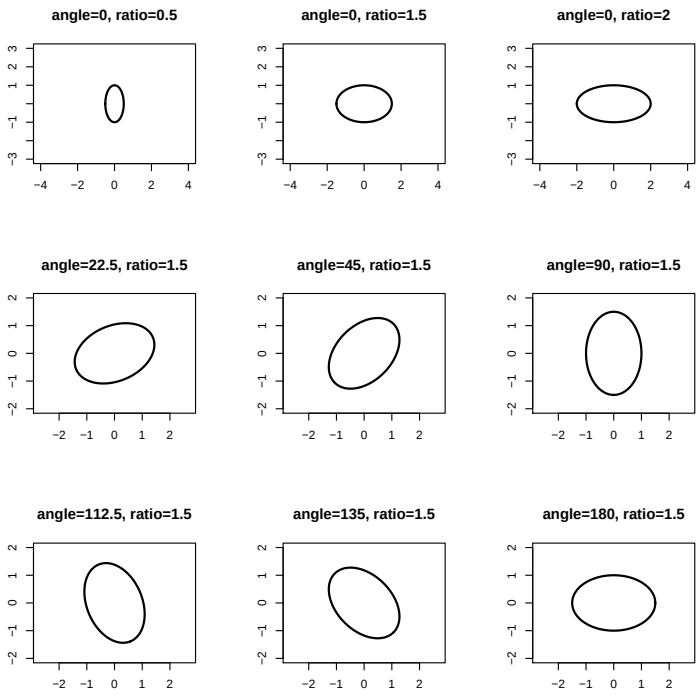

Figure 1: Ellipses with different angles and major-to-minor axis ratios

3 Methodology

3.1 Model Fitting

We use the Bayesian framework to perform inference on model parameters and prediction for new locations. We have

$$\mathbf{Y} \sim N_n(\mu\mathbf{1}, \Sigma(\theta)) \tag{7}$$

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), we complete the Bayesian formulation by assuming the prior takes the form

$$\pi(\mu, \theta) = \pi(\mu, \sigma^2, \tau^2, \phi, \alpha, r) = \pi(\mu)\pi(\sigma^2)\pi(\tau^2)\pi(\phi)\pi(\alpha)\pi(r) \tag{8}$$

assuming all the parameters are independent.

We use non-informative $N(0, 1000)$ prior for μ. We use non-informative 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 center and infinite prior variance. We put a uniform$(0, \pi)$ prior on orientation of the ellipse α and a uniform$(0.2D, 0.7D)$ on range $1/\phi$, where D is the maximum distance between locations. Range is usually around half the maximum distance between locations.

We use Metropolis-Hastings (MH) sampling algorithm to approximate the posterior distribution of all parameters $p(\theta|y)$. MH algorithm is useful when we cannot directly sample from the posterior distribution. Assume we have a collection of $\{\theta^{(1)}, ..., \theta^{(s)}\}$. To add 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 MH 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.

Table 1 lists the proposal and prior distributions we use. For anisotropy angle α we use normal proposals, divide them by π and take the remainder so that the proposals will be between 0 and π. For anisotropy ratio r, we use truncated normal distribution from 0 to infinity. For σ^2, τ^2 and ϕ we use log normal distribution. We also experimented with using log normal proposal for r, but the sampler does not behave as well as using truncated normal proposals. Because α and r are highly correlated, we update them jointly. Finally, we use Gibbs sampler to update μ as the normal prior on μ and the likelihood given all other parameters are conjugate.

	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

Table 1: Proposal and prior distributions for Metropolis algorithm

3.2 Kriging

An observation at location we want to predict s_0 follows the following distribution:

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

Prediction approach 1: sample spatial random effect w

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

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

Prediction approach 2: marginalize out spatial random effect w

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

4 Model Comparison

4.1 Predictive Mean Square Error

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

4.2 Empirical Coverage

One method of assessing model calibration is assessing how well credible intervals, derived from the posterior predictive distributions of the responses, capture the true/observed values.

4.3 Continuous Rank Probability Score (CRPS)

To look at how concentrated the predictive distribution of $y(s_0)$ is around the true value, we look at the metric CRPS, 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 $Y(s_0)$)(empirical CDF). We could sum the CRPS over multiple prediction locations for model comparison.

5 Simulation Study

We designed a set of simulations to study whether data can differentiate between anisotropic and isotropic models, especially in terms of predictive performance. We are also interested in studying how sensitive the data is to sample size, difference choices or anisotropy angle and ratio, and different magnitude and ratio of spatial and non-spatial variance.

We generate locations randomly distributed 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 0.3, around half of the maximum distance between the points. Decay ϕ is 3/0.3 = 10. We test different combinations of anisotropy angle and ratio, as well as spatial and non-spatial variances, and apply them to 2 different sample sizes.

SHOW ELLIPSE OF CORRELATION FUNCTION POSTEIOR PREDICTIVE FOR A HOLD OUT SHOW RELIZATION, CIRCLE FITTED, XX HOLDOUT

	$\alpha = 60°$	$\alpha = 150°$
$r = 1.5$		
$r = 4$		

	$\sigma^2/\tau^2 = 1$	$\sigma^2/\tau^2 = 5$
$\sigma^2 = 0.2$		
$\sigma^2 = 1$		

Table 2: n=100

	$\alpha = 60°$	$\alpha = 150°$
$r = 1.5$		
$r = 4$		

	$\sigma^2/\tau^2 = 1$	$\sigma^2/\tau^2 = 5$
$\sigma^2 = 0.2$		
$\sigma^2 = 1$		

Table 3: n=500

Put an inverse gamma prior on τ^2 and σ^2 and a uniform prior on ϕ. Then

$$\tau^2 \sim \text{IG}(1, 1)$$
$$\sigma^2 \sim \text{IG}(1, 1)$$
$$\phi \sim \text{Unif}(3/1, 3/0.1)$$

I implemented the model using JAGS with 2000 iterations, 1000 burn-in, 10% thinning. Gelman stats is 1.31. The results ($\sigma^2 = 0.62, \tau^2 = 0.14, \phi = 3.83$) were fairly similar to the truth. Results from spBayes package is $\sigma^2 = 0.65, \tau^2 = 0.14, \phi = 3.79$

I implemented the model using JAGS with 10,000 iterations, 3 chains, 10% thinning and 2,000 burn-in. We end up with 1,000 samples. The posterior mean and 95% confidence interval are:

Parameters	μ=0	ϕ=3.75	σ^2=1	τ^2=0.05	$\alpha = 1.047$	r=4
Isotropy	0.63 (0.04, 1.20)	4.40 (3.04, 7.64)	0.53 (0.30, 0.88)	0.15 (0.09, 0.23)	NA	NA
Anisotropy	0.54 (-0.27, 1.26)	4.84 (3.06, 9.40)	0.67 (0.34, 1.15)	0.15 (0.09, 0.22)	1.07 (0.85, 1.27)	5.15 (2.48, 7.77)

After fitting the model, we plot the range in each direction between 0 and 2π. We obtain posterior samples of B, and hence we get an entire posterior distribution for the range in a given direction. We calculate the posterior mean for the range in that direction by averaging the ranges over the samples. We also plot the 0.95 credible interval by taking the 0.025 and 0.975 quantile of the ranges over all samples.

I plot the range c_θ for θ in $(0, \pi)$.

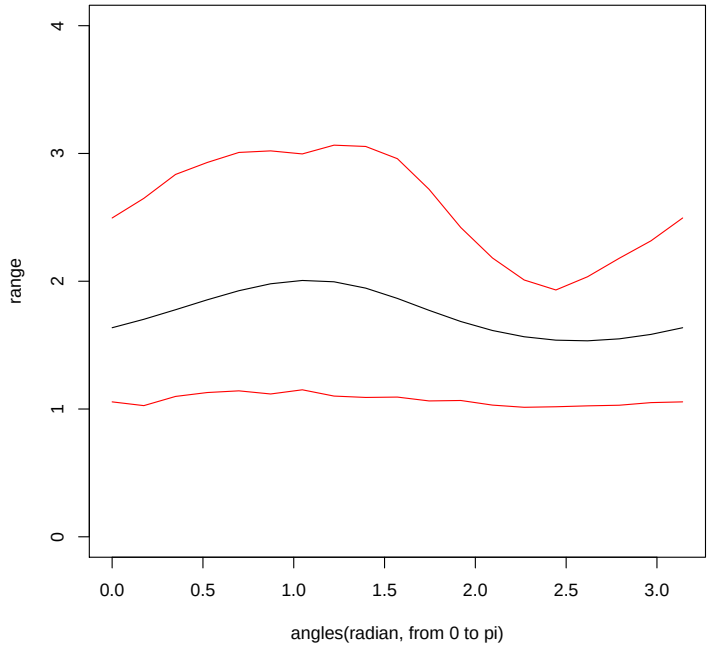

Figure 2: Range at different angles between 0 and 2π

5.1 Prediction Evaluation

I fit isotropic and anisotropic model to data generated under anisotropic model, and compare them using three metrics: Predictive Mean Square Error (PMSE), Empirical Coverage (EC), and Continuous Rank Probability Score (CRPS).

Model	EC	PMSE	CRPS
Isotropy	90%	0.275	0.034
Anisotropy (approach 1)	97.5%	0.212	0.024
Anisotropy (approach 2)	97.5%	0.216	0.024

Table 4: Model Comparison: n=100, angle=60, ratio=4

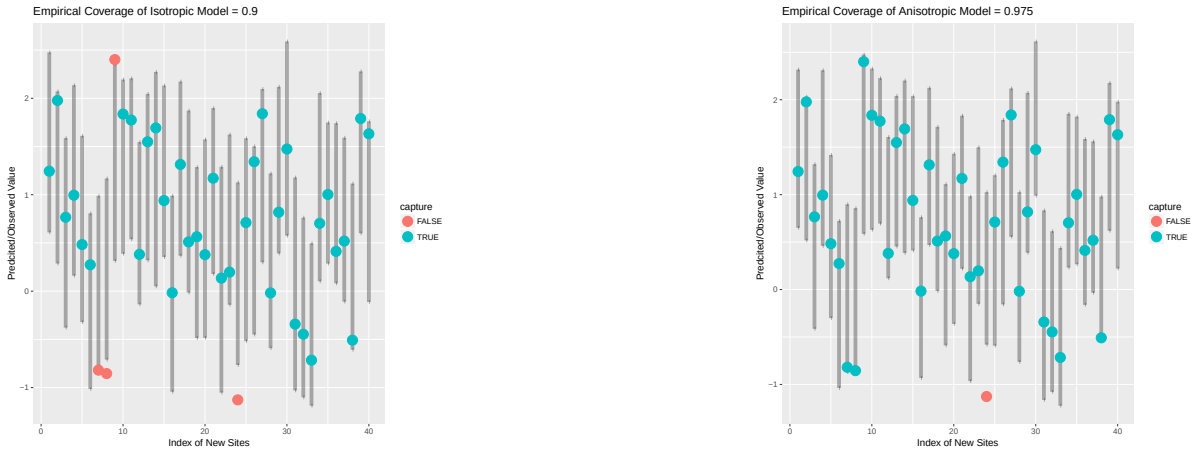

Figure 3: Empirical Coverage

6 Product Covariance

We represent an anisotropic model wth product covariance function, i.e. the product of 1-dimensional covariance function in the x-coordinate and a 1-dimensional covariance function in the y-coordinate. The covariance function is

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

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

I use the following simulation settings

$$\sigma^2 = 1$$
$$\phi_x = 4$$
$$\phi_y = 8$$
$$\tau^2 = 0.05$$

7 Real Data Example

8 Conclusion