A review and recreation of results in 'Geostatistical inference under preferential sampling'

John Paige

Department of Statistics, University of Washington Seattle, WA, 98195, USA

Abstract

This report reviews and attempts to recreate the results of Diggle et al. [2010], which incorporates preferentiality modeled as a log-Gaussian Cox process into classical geostatistical methods. We consider the work's literary context and possible avenues for future research.

1 Background and Motivation

As discussed in Diggle et al. [2010], geostatistics considers stochastic processes varying in space. In some applications the process, S, might be measured at a set of discrete locations, say $X = \{x_1, x_2, ..., x_n\}$. We often assume X is a subset of $\mathbb{R}^2$, as in Diggle et al. [2010]). In other settings, S might be measured as a set of averages or a quantile function over different regions. Here, we consider the following model for measured observations, $Y_i \in \mathbb{R}$, of the spatial process:

$$Y_i = \mu + S(x_i) + Z_i, \tag{1}$$

where $i \in \{1, ..., n\}$, μ is the mean of the measured process, and the Z_i are independent and identically distributed white noise with zero mean and variance τ^2. Here, the Z_i can be interpreted as independent fine-scale measurement errors. Note that in the above model we assume $S(x)$ is multivariate normal with zero mean.

Variance of the spatial process is usually defined in terms of a correlation function, ρ, depending on locations and a set of parameters, $\vec{\theta}$:

$$\text{Var}\left(S(x) - S(x')\right) = \sigma^2 - \sigma^2 \rho(x, x' \,|\vec{\theta}). \tag{2}$$

Here σ^2 is the variance of the field of interest. Correlations in S are modeled under the assumption of stationarity and isotropy in the case of Diggle et al. [2010]. Under

stationarity, the correlation becomes a function of only $x - x'$. Under isotropy, the correlation function does not depend on the angle of $x - x'$, so with both assumptions we have:

$$\text{Var}\left(S(x) - S(x')\right) = \sigma^2 - \sigma^2 \rho(|x - x'| \mid \vec{\theta}). \tag{3}$$

The positive definite function $\rho(\cdot|\vec{\theta})$ represents a correlation depending on distance conditional on the parameters $\vec{\theta}$. The variance in (3) is called a variogram. Diggle et al. [2010] uses the Matérn correlation function,

$$\rho(u \mid \phi, \kappa) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)}(u/\phi)^{\kappa}K_{\kappa}(u/\phi),$$

where $\kappa > 0$ is a smoothness parameter, $\phi > 0$ is the spatial scale parameter, and $K_{\kappa}(\cdot)$ is the modified Bessel function of the second kind of order κ. As noted in Diggle et al. [2010], the Matérn class of covariance functions is commonly used and flexible, containing exponential correlation as a special case when $\kappa = 0.5$. One difficulty when using a Matérn covariance model is that ϕ and κ are sometimes difficult to estimate jointly, especially κ. Hence, [Diggle et al., 2010] fixes κ at 0.5 in the inference.

There are several advantages to modeling the variogram in a spatial model. It provides a simple interpretation relating correlations and variance in S over space as a function of distance. In addition, it is easy to estimate assuming stationarity and under certain distributional assumptions of the sample locations. A point cloud method of estimating empirical variogram works by plotting the distance between observations y_i and y_j versus $(y_i - y_j)^2/2$, which is the method of moments estimator (and nonparametric MLE) for the variance based on two observations assuming the y_i have constant expectation [Diggle et al., 2010]. If smoothed nonparametric empirical variogram estimates are desired, we could partition the possible distances between observation locations into bins, and estimate the variance between differences of observation values when the corresponding distances are in any given bin. Examples of binned empirical variograms are given in Figure 5.

Classical geostatistical methods assume that the observation locations are independent of the measured spatial field [Diggle et al., 2010]. However, in the case of certain datasets, this might not be the case. Consider a tornado chaser trying to measure tornado wind speeds. It would be a terrible idea to try to infer average wind speeds in a region from the tornado chaser's wind speed data since the data they collect is much

more likely to be in locations and times where the wind speeds are abnormally high. This is what Diggle et al. [2010] refers to as *preferential sampling*. Recent examples of scientific studies using preferential sampling involve sampling the distribution of a species in order to better determine its habitat [Pati et al., 2011] and pollutant and environmental hazard monitoring [Lee et al., 2011], although preferential sampling occurs in more than just spatial statistics. Consider doctor visit times, where patients might be more likely to visit when sick, or optimizing any function. McCullagh [2008] notes the importance of taking into account preferential sampling in Generalized Linear Models (GLMs).

In order to account for preferential sampling in the context of spatial statistics, Isaaks and Srivastava [1988] and Srivastava and Parker [1989] proposed a non-ergodic variogram estimator as an alternative to classical estimators that they claimed was more robust to nonstationary and preferential data. However, Curriero et al. [2002] found that the non-ergodic estimators 'possess no clear advantage' over the traditional estimators, and in fact performed worse in the cases they studied. Schlather et al. [2004] note that if S is stationary, then $M_k(h) \equiv E[S(x)^k | x, x + h \in X]$ is constant if the sampling process is non-preferential, since the expectation does not depend on $x + h$. However the expectations might not be constant under preferential sampling. Schlather et al. [2004] then defines tests for preferentiality assuming that S is stationary based on $M_1(h)$ and $M_2(h)$ using simulations under models assuming non-preferentiality.

In addition to analyzing variations in observation *values* over space, it is common to study the patterns in the observation *locations*, as in Diggle et al. [2010]. Point processes are stochastic processes used to model random distributions of countably many points throughout a spatial domain (here we assume a bounded subset of $\mathbb{R}^2$) [Gelfand et al., 2010]. Common models for point processes include homogeneous or inhomogeneous Poisson processes, Cox processes, and Markov point processes [Gelfand et al., 2010]. A homogeneous Poisson process with rate (intensity) $\lambda > 0$ is a point process such that for bounded Borel sets $B, B' \subset \mathbb{R}^2$ the following conditions are satisfied:

1. the number of points in a bounded Borel set, B, is a random variable given by $N(B) \sim Pois\left(\lambda \mu(B)\right)$, where μ is the Lebesgue measure, and

3

2. $N(B) \perp\!\!\!\perp N(B')$ when $B \cap B' = \emptyset$

An inhomogeneous Poisson process is the same, except it has a rate $\lambda(x)$ that varies in space so that

$$N(B) \sim Pois\left(\int_B \lambda(x)\ dx\right)$$

Cox processes [Cox, 1955] (also known as doubly stochastic Poisson processes) are generalizations of inhomogeneous Poisson processes with the following properties:

1. Random rate $\Lambda = \left\{\Lambda(x) : x \in \mathbb{R}^2\right\}$ is a nonnegative stochastic process, and

2. In any fixed realization, $\Lambda(x) = \lambda(x) : x \in \mathbb{R}^2$, the point process is an inhomogeneous Poisson process with rate $\lambda(x) \geq 0$.

In the case of

$$\Lambda(x) = \exp\left\{\alpha + \beta S(x)\right\}, \tag{4}$$

where $S(x)$ is the stochastic Gaussian process defined in (1), the resulting point process is known as a log-Gaussian Cox process (LGCP) [Diggle et al., 2010].

Just as the variogram can measure the level of correlation in observation values through space, the K-function can be helpful when analyzing the level of correlation in observation locations. The K-function is defined as $K(s) = \lambda^{-1}E[N_0(s)]$ for homogeneous point processes, where $E[N_0(s)]$ denotes the expected number of points with distance s from another point. Under complete spatial randomness (CSR) (*i.e.* if it is consistent with a homogeneous Poisson process model) the K function has the form $K(s) = \pi s^2$, but when the empirical K function is above or below πs^2, the data is indicative of clustering or repulsion respectively. The K function can be used in Monte Carlo goodness of fit tests for point process models [Gelfand et al., 2010, Section 18.3]. It can also be used to test whether the data follows complete spatial randomness (CSR) or whether there is clustering or repulsion among the observation locations. The K function for LGCPs has the form:

$$K(s) = \pi s^2 + 2\pi \int_0^s \gamma(u)u\ du \tag{5}$$

where $\gamma(u) = \exp\left\{\beta^2\sigma^2\rho(u;\phi,\kappa)\right\} - 1$ is proportional to the covariance of Λ, $\mathrm{Cov}\left(\Lambda(x), \Lambda(x+u)\right)$ under the parameterization of (4). Derivation of (5) is given in Appendix A in Sec. 6.4.

The main contribution of Diggle et al. [2010] is the authors' use of a LGCP model for sample locations in conjunction with the spatial process model in Eq 1 for data being measured. They give a relatively simple parameter fitting procedure by maximizing a tractable Monte Carlo likelihood and give examples of the dangers that could occur under a 'naive' geostatistical analysis that assumed non-preferentiality. Further, they give new tests for preferentiality and compute the results of a Monte Carlo goodness of fit based on the K-function under their proposed model.

2   Methods

Diggle et al. [2010] studies lead conentration data in Gallicia, a region in northern Spain. The data, with sample locations shown in Fig. 1, is sampled in October 1997 and July 2000. In 1997, the sample locations are concentrated in the north, whereas in 2000 the sample locations are on a relatively regular lattice. There are 63 samples in 1997 and 132 in 2000 with an apparent mean-variance relationship that suggested the use of a log transformation of the data. There appear to be two outliers in the samples from 2000, which the authors of Diggle et al. [2010] replace with the mean of the rest of the log-transformed observations from that year. All data used in this report is available at www.lancs.ac.uk/staff/diggle/.

	Untransformed		Log-transformed	
	1997	2000	1997	2000
Number of locations	63	132	63	132
Mean	4.72	2.06	1.44	0.66
Standard deviation	2.21	0.91	0.48	0.43
Minimum	1.67	0.80	0.52	-0.22
Maximum	9.51	6.00	2.25	2.16

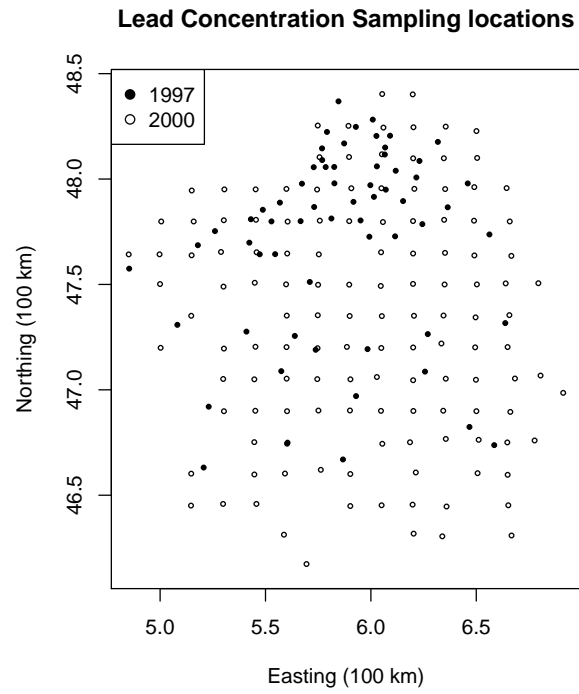Table 1: Lead levels in $\mu(g \text{ dry weight})^{-1}$ for the given response transformations and years.

Lead Concentration Sampling locations

Figure 1: Lead concentration data sample locations for 1997 and 2000.
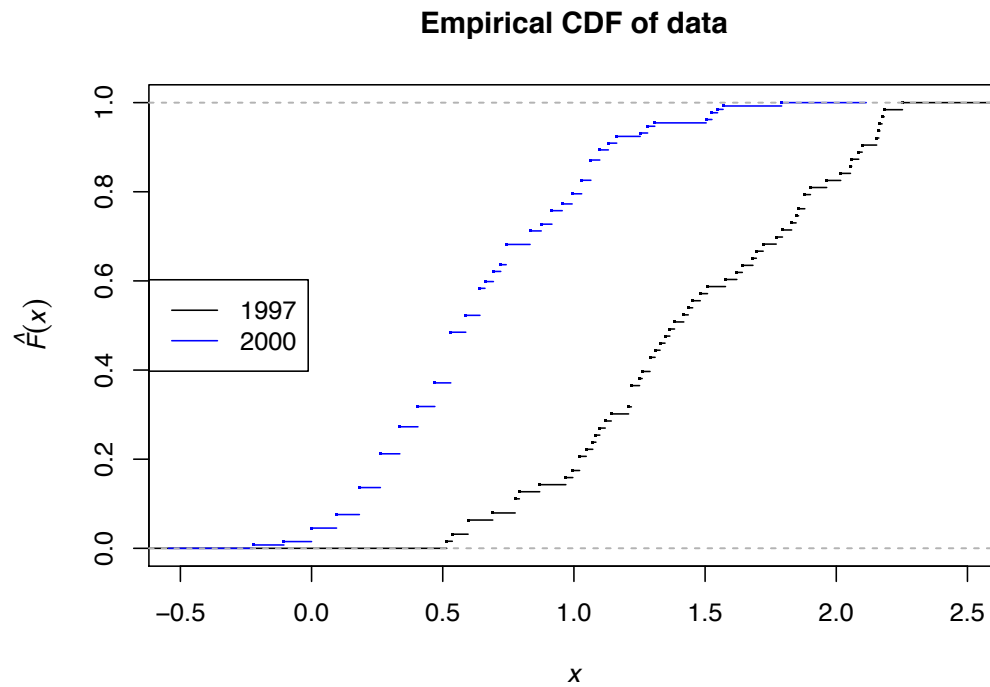
Empirical CDF of data

Figure 2: Lead concentration empirical distribution functions for 1997 and 2000.

2.1 Unit square simulations

In addition to studying lead concentration data, Diggle et al. [2010] simulates data on the unit square under three different sampling schemes to determine the effects of preferential sampling on classical methods in geostatistics. In the first sampling scheme, points are sampled from the unit square uniformly, or equivalently, under a homogeneous Poisson process. In the second, samples are taken under a clustered scheme, where sample locations are chosen via a preferential model with respect to a response that is independent of that being measured but with identical statistical properties. In the third and final sampling scheme, the locations are chosen preferentially. In the case of the clustered and preferential sampling methods, we chose a random LGCP intensity, $\Lambda(x) = \exp\{4 + 2S(x)\}$, where x is a location on the unit square, $ES(x) = 0$, and $\text{Var}(S(x)) = 1.5$. The values of the response are given by $4 + S(x)$, and the correlation function of $S(x)$ is Matérn with range $\phi = .3$ and smoothness $\kappa = 1$. Examples of the sampling schemes are shown in Figure 3.

I compute the binned empirical variogram estimator discussed in Sec. 1 using the `fields` package. I generate 500 simulations under each sampling scheme and use them to estimate the variogram bias and standard deviations under each scheme. I ensure that exactly 100 observations are sampled in each simulation using independent and with equal probability random thinning for the cluster and preferentially sampled data, since this process holds relative intensities constant across the spatial domain equal as shown in Diggle et al. [2013].

In order to gauge the affect of each sampling methodology on the classical predictor, I compute the conditional mean $E[S(x)|\vec{Y}]$ at $x = (.49, .49)$ for each simulation, comparing bias and standard deviations under each sampling scheme. When calculating the predictor I use two models. In the first, I use maximum likelihood to estimate parameters under the classical geostatistical model ignoring preferentiality. In the second, I set the simulation model parameters to be those of the preferential model MLEs for the 1997 data under the joint model obtained by Diggle et al. [2010]: $\mu = 1.515$, $\sigma^2 = 0.138$, $\phi = .313$, $\kappa = 0.5$, $\beta = -2.198$, and $\tau^2 = 0.059$. I then set the plug-in parameters used when generating the draws from the non-preferential predictive distribution to be the parameters used in the simulations.

2.2   Model inference

In order to fit the parameters of the classical and preferential models, I use maximum likelihood. The joint likelihood of S, X, and Y can be written

$$f_{S,X,Y}(s,x,y) = f_S(s)f_{X|S}(x|s)f_{Y|S,X}(y|s,x) = f_S(s)f_{X|S}(x|s)f_{Y|S(X)}(y|s(x)) \quad (6)$$

where $S(X) = \{S(x_1), ..., S(x_n)\}$. In the case of the classical model, where $S \perp\!\!\!\perp X$, then the joint distribution of S, X, and Y is:

$$f_{S,X,Y}(s,x,y) = f_S(s)f_X(x)f_{Y|S}(y|s)$$

Hence, under no preferentiality, we can marginalize X to get

$$f_{S,Y}(s,y) = f_S(s)f_{Y|S}(y|s)$$

In fact, classically the likelihood that is optimized is $f_{S|Y}(s|Y)$, since the observations are fixed so $f_Y(y) = \mathbb{1}_{(Y=y)}$. Under preferentiality, however, we cannot simplify (6) in the same way since we cannot ignore the dependence of S on X in the $S(X)$ term. However, Diggle et al. [2010] shows that the joint likelihood is still tractable. Let $S = \{S_0, S_1\}$ be that values of S over a finely spaced regular lattice covering the spatial domain, where $S_0 = S(X)$ is the values of S at the observation locations whereas S_1 is the values of S at other locations on the lattice. The authors note that, where θ is the vector of model parameters, the likelihood can be written

$$L(\theta) = E_{S|Y}\left[f_{X|S}(x|s)\frac{f_{Y|S_0}(y|s_0)}{f_{S_0|Y}(s_0|y)}f_{S_0}(s_0)\right]$$

This is derived in the Appendix in Sec. 6.1 and can be approximated with the Monte Carlo estimate

$$L(\theta) \approx \frac{1}{m}\sum_{j=1}^{m} f_{X|S}(x|S_j)\frac{f_{Y|S_0}(y|S_{0j})}{f_{S_0|Y}(S_{0j}|y)}f_{S_0}(S_{0j}) \quad (7)$$

where $S_j = \{S_{0j}, S_{1j}\}$ is the jth simulation of S conditional on Y, and S_{0j} are the values of S_j at the observation locations. In order to simulate S_j, Diggle et al. [2010] suggests the following. First, let $X^* = \{x_1^*, ..., x_N^*\}$ be the N locations of the points on the regular lattice over the domain, and let C be the $n \times N$ matrix where the ith row of C has $N - 1$ 0's and a single 1 identifying the index of x_i in X^* (hence, $S_0 = CS$). Given the parameters, θ, we have $S \sim \text{MVN}(0, \Sigma(\theta))$ and $Y \sim \text{MVN}(\mu, \Sigma_0(\theta))$ where

8

$\Sigma_0(\theta) = C\Sigma(\theta)C' + \tau^2 I$. From here, we will drop θ from the notation, noting that the covariance matrices Σ and Σ_0 still do depend on the parameters, θ. Then if we simulate S marginally and $Z \overset{iid}{\sim} \mathrm{MVN}(0, \tau^2 I)$, Diggle et al. [2010] simulates S_j with the formula

$$S_j \equiv S + \Sigma C' \Sigma_0^{-1}(y - \mu + Z - CS) \tag{8}$$

This is based on Rue and Held [2005, Chapter 2], and also on the posterior sampling results in Eidsvik et al. [2009]. I verify the formula along with deriving the terms in the Monte Carlo likelihood in the Appendix in Sec. 6.2. Note that this formula does not require the inversion of the $N \times N$ matrix Σ. In fact, using the efficient Cholesky decomposition and forward and backsolve algorithms we can even avoid direct inversion of Σ_0. We can also efficiently sample the N-dimension vector S using the circulant embedding algorithm, which requires only order $\mathcal{O}(N^2)$ floating point operations [Wood and Chan, 1994]. In this analysis, I draw from the marignal distribution of S using the `RandomFields` package in `R`.

Once the shared latent model likelihood is factored into the form given in (7), evaluation of the likelihood becomes tractable with the simulation method given above. I evaluate the random intensity on the grid, X^*, as $\Lambda(x) = \exp\{\alpha + \beta S_j\}$ where α is chosen as the maximum likelihood estimator conditional on the other parameters $\theta = (\mu_{97}, \mu_{00}, \sigma, \tau, \phi, \beta)$. Here μ_{97} is the mean for the 1997 log-transformed response and μ_{00} is the mean log response for the 2000 data. Using our lattice-based approximation of $\Lambda(x)$, I apply the inhomogeneous Poisson process likelihood:

$$L(X|\Lambda) = \prod_{i=1}^n \Lambda(x_i) \left(\int \Lambda(x) \, dx\right)^{-n}$$

where the integral is taken over the entire domain and is calculated numerically after computing $\Lambda(x)$ over X^*. The other portions of the likelihood can be calculated using Gaussian densities derived in Sec. 6.2.

I use the Nelder-Mead simplex algorithm [Nelder and Mead, 1965] for likelihood maximization with 10,000 Monte Carlo iterations in our preferential likelihood approximation. I fit parameters under a standard geostatistical model for 1997 and 2000 separately before fitting a joint model. I do the same under the preferential model, noting that the year 2000 data is not preferential since it is composed of regularly

spaced observations on a grid. The optimization is performed using the `optim` R function, and I calculate standard errors and the correlation matrix for the MLEs under the joint preferential model by computing the Hessian at the optimum.

As in Diggle et al. [2010], data from 1997 is assumed to be independent from data measured in 2000. This is somewhat problematic, although it is unclear that the additional complexity of modeling that would be required for modeling intertemporal dependence would substantially improve the results, and it would complicate the likelihood evaluation. Since there are significant differences between concentrations in 1997 and 2000, the assumption is slightly less problematic.

2.3   Goodness of fit

As mentioned in Sec. 1, K-functions are useful tools for goodness of fit tests. I perform a goodness of fit test on the preferential model by comparing the empirical K-function of the 1997 data with th theoretical K-function of an LGCP given by the parameters fit under the joint preferential model for the 1997 and 2000 data. Using the `envelope` function in the `spatstat` package, I generate 99 simulations of the fit LGCP. Using the test statistic

$$T = \int_0^{0.25} \frac{\left(\hat{K}(s) - K(s)\right)^2}{v(s)} \, ds \tag{9}$$

where $v(s)$ is the variance of $\hat{K}(s)$ estimated from the 99 simulations, I compare the rank of the empirical test statistic to the test statistic calculated from each ofthe 99 simulations. The p-value for the test is the rank (from 1 to 100) out of the 100 calculated test statistics divided by 100 [Gelfand et al., 2010].

2.4   Predictions

After the parameters are fit under the classical and preferential models I generate 10,000 simulations of $\exp\{S(x)\}$, the untransformed lead concentrations, conditional on Y for the 1997 data under each model. Normally, the conditional mean would be used as the predictor, but skewness in the distribution of the untransformed lead concentrations makes percentile comparisons between the models preferred. I therefore compare the 5%, 50%, and 95% predictive distributions. In addition, I compare estimates of the

proportion of Gallicia with lead concentration exceeding 3, 5, and 7 μg (g dry weight)$^{-1}$ between the two models.
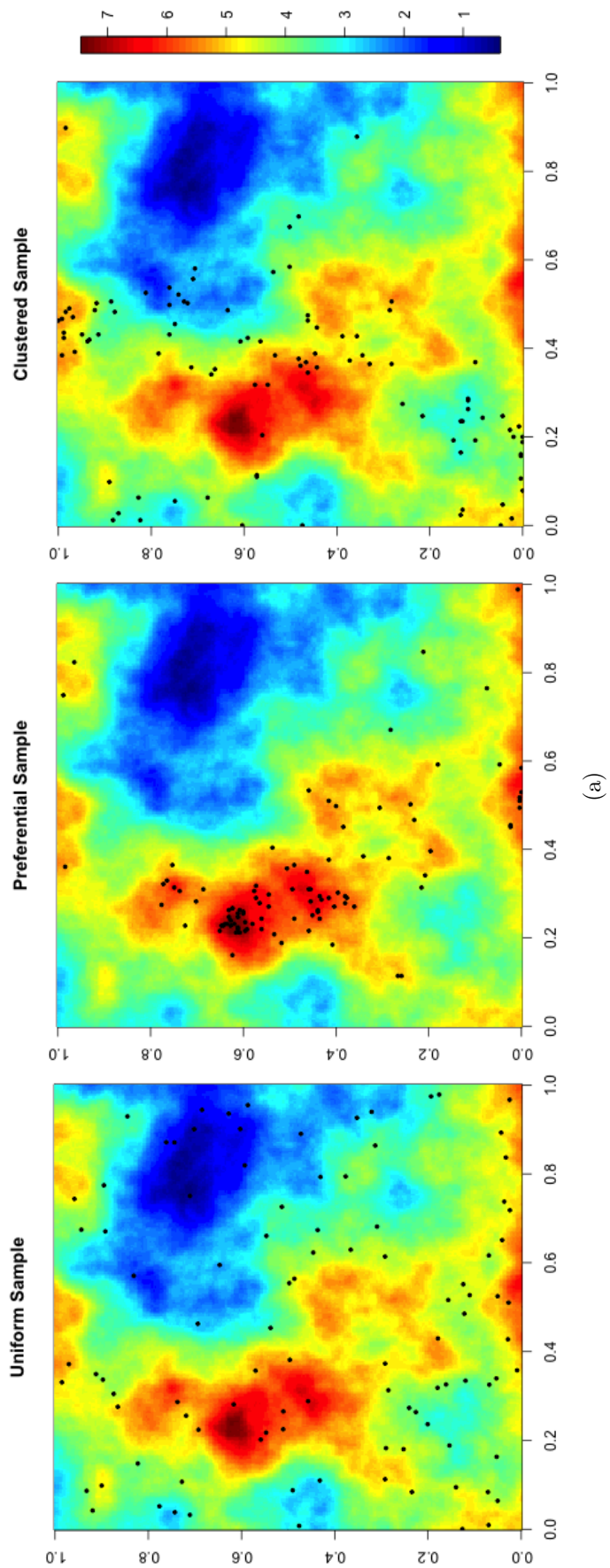
Figure 3: (a) Uniform, (b) preferential, and (c) clustered sampling schemes (the black dots) on the unit square plotted with the associated stochastic process $\mu + S(x)$. Here $\mu = 4$, $\mathrm{Var}(S) = 1.5$, and S is mean zero with Matérn covariance structure with range $\phi = 0.3$ and smoothness $\kappa = 1$.

3 Results

3.1 Unit square simulations

Figure 4 shows the bias of the classical empirical variogram estimates based on the 500 simulations on the unit quare. While the uniform and clustered variograms do not appear significantly biased, the preferential sampling scheme seems to cause substantial bias in the variogrm estimates. The magnitude of the bias increases as a function of distance. In addition, the standard deviation of the preferential estimate of the variogram is smaller than the non-preferential clustered sampling scheme. Hence, preferentiality is not accounted for, then then relatively small standard deviation could lead to a false sense of security in the variogram estimates.

Model	Parameter	Confidence intervals		
		Completely random	Preferential	Clustered
1	Bias	(-0.029, 0.038)	(0.956, 1.123)	(-0.074, 0.064)
1	RMSE	(0.354, 0.410)	(1.318, 1.501)	(0.717, 0.851)
2	Bias	(-0.040, 0.030)	(-0.265, -0.195)	(-0.040, 0.032)
2	RMSE	(0.375, 0.425)	(0.434, 0.491)	(0.382, 0.432)

Table 2: 95% Confidence intervals for the given parameters under the given models and sampling schemes.

3.2 Model inference

Under a non-preferential model initialized at reasonable but naive values, I estimate the model parameters under seperate models for the two years and also jointly, assuming a shared correlation range ϕ, as well as shared variance parameters σ^2 and τ^2. The fit MLEs are given in Table **??**. While Diggle et al. [2010] do not give parameter estimates to compare these to, the MLEs fit in this analysis from the joint non-preferential model are very similar to the estimates from Diggle et al. [2010] using the joint preferential model, which are $\hat{\mu}_{97} = 1.51$, $\hat{\mu}_{00} = 0.76$, $\hat{\sigma}^2 = 0.14$, $\hat{\phi} = 0.31$, and $\hat{\tau}^2 = 0.06$ (and $\hat{\beta} = -2.20$). The joint model MLE estimates and standard errors are shown in Table 3
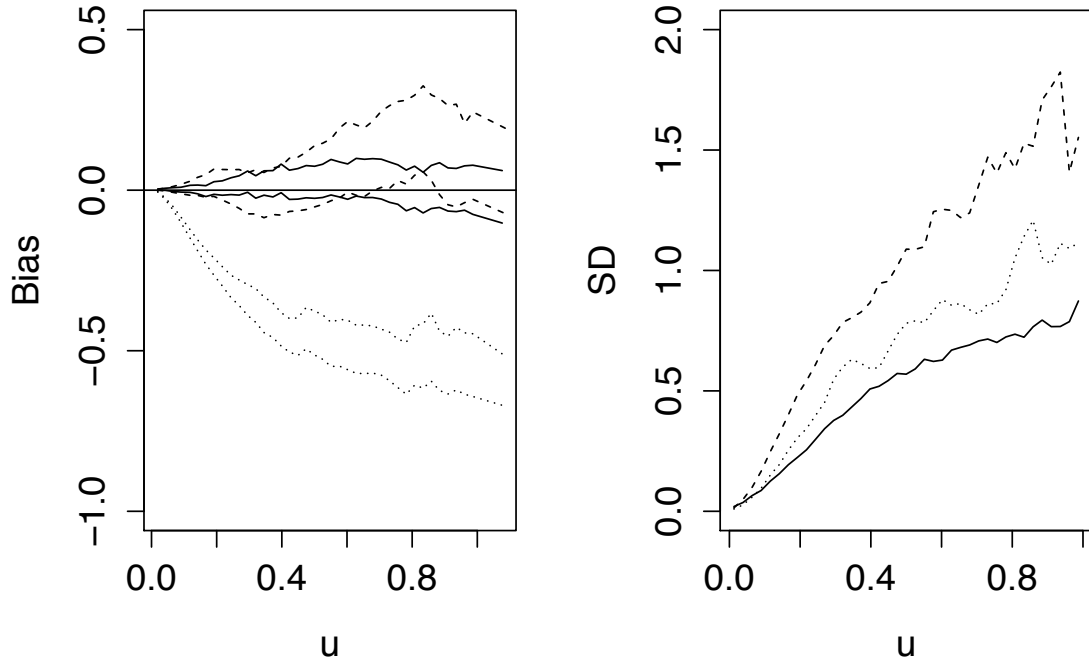
13

Figure 4: Based on 500 simulations with 100 observations each on the unit square, plotted are (a) estimated empirical variogram bias ±2 standard errors as functions of distance, u, for the uniform (solid), clustered (dashed) and preferential (dotted) sampling schemes and (b) standard deviations of the empirical variogram estimates.

along with a correlation matrix computed using the hessian fit using the `optim` function in R. Since the parameters are optimized on a log scale, the log scale standard errors are transformed to the linear scale using the delta method.

As in Diggle et al. [2010], I find that the fit parameters for the two years are somewhat different. However, unlike in Diggle et al. [2010], when using a likelihood ratio test to determine if differences in the parameters between the years warrant separate models, I get a p value of 0.007 under 3 degrees of freedom. This suggests that a joint model is problematic. The difference between these results and Diggle et al. [2010] is likely due to outliers in the 2000 log-lead observations, and is further discussed in the Appendix in Sec. 6.5. Since each separate year better identifies complementary sets of parameters, a joint model may better estimate some paramters in spite of the results from the likelihood ratio test to fit a generalizeable model. Interestingly, while preferentiality seems to bias variogram estimates downard to smaller variances, the 1997 data has larger variance than the 2000 data as evidenced by Figure 5. This further suggests that a joint model might not fit the data well. It may however be necessary in order for the preferential model to be well identified.

Parameter	Estimate	Standard error	Correlation matrix				
μ_{97}	1.55	0.018	1.00	0.00	0.02	-0.00	-0.02
μ_{00}	0.73	0.014		1.00	0.10	0.20	0.07
σ	0.37	0.006			1.00	0.39	-0.37
ϕ	0.31	0.039				1.00	0.47
τ	0.23	0.005					1.00

Table 3: Parameter estimates and uncertainties calculated using the fit hessian and the delta method for the non-preferential joint model. Note that the σ, ϕ, and τ parameters were estimated on a log scale, and while the estimate and standard error columns account for this, the correlation matrix is has entries with respect to the parameters' log-transformed values.
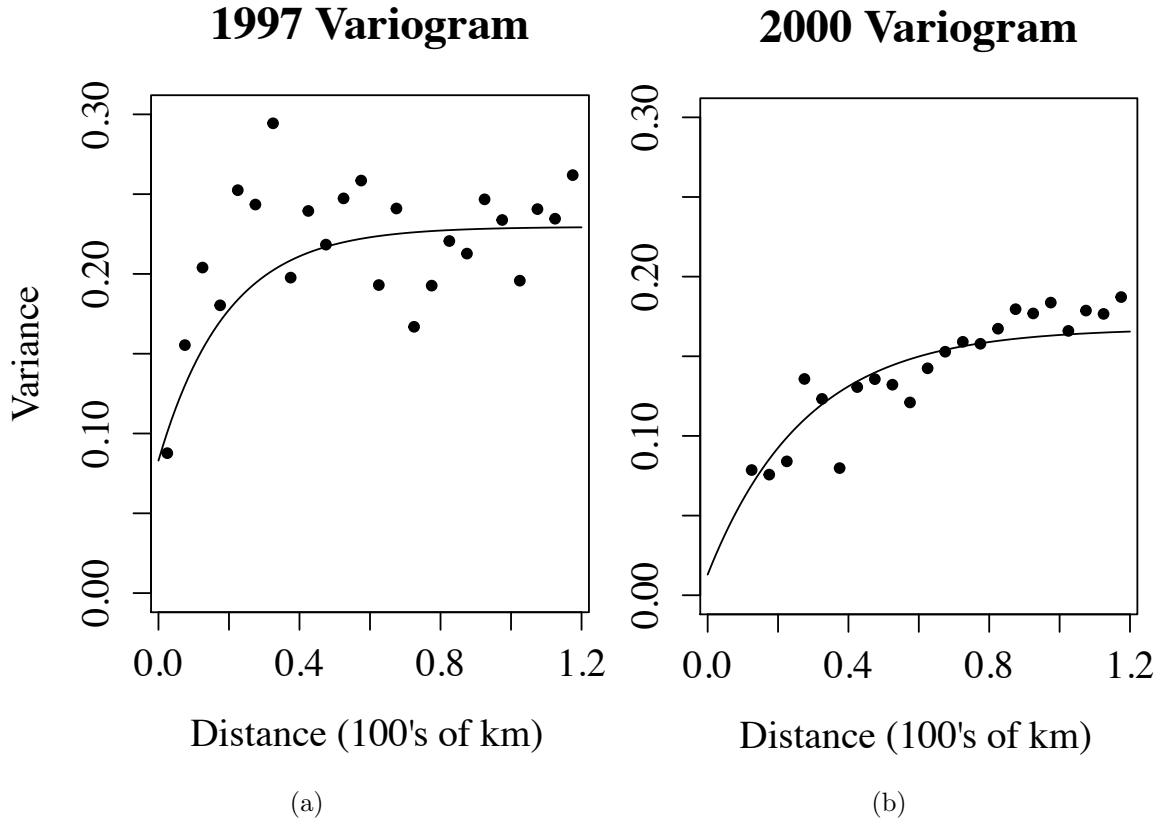
Figure 5: Empirical and theoretical variograms fit using maximum likelihood for the (a) 1997 log-lead concentration data and (b) 2000 log-lead concentration data with outliers replaced by the mean of the rest of the 2000 data.
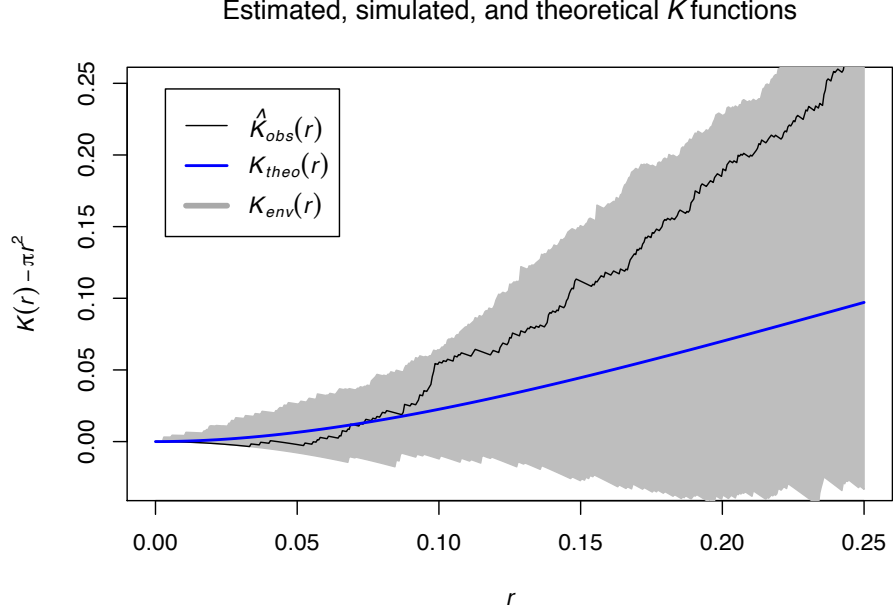
Figure 6: Empirical K-function (black) plotted against the fitted theoretical K-function (blue). The gray envelope is generated using 99 simulations from the log-Gaussian Cox Process model.

3.3 Goodness of fit

The envelope from the empirical and theoretical K-functions from 99 simulations from the theoretical LGCP model under the MLE plug-in parameters is shown along with the empirical K-function from the 1997 data in Figure 6. The results are comparable to Figure 6 from Diggle et al. [2010], although instead of the empirical K function from the 1997 data being just outside of the simulation envelope, mine is just within. Using the Monte Carlo test statistic defined in (9), I get a p-value of 0.07. This implies that up to distances of 25 kiliometers, the LGCP model is decent at capturing clustering behavior in the observation locations due to preferentiality. While the preferential model accurately shows that the K-function is above what would be expected under homogeneous sampling (corresponding to the horizontal line at 0 in Figure 9), there may be additional sources of clustering not accounted for by the LGCP model.

3.4   Predictions

Predictions for the 1997 lead-concentration field are given using the 10,000 draws from the predictive distribution using the plug-in joint preferential and non-preferential MLEs. The 5th, 50th, and 95th quantiles are shown in Figure 8, and the predicted areal proportions exceeding 3, 5, and 7 $\mu g/(g$ dry weight) lead concentrations are given in Figure 7. I generate one set of predictions using the parameters I fit from the standard, non-preferential joint model, while the other set is uses the preferential model MLEs from Diggle et al. [2010]. As noted in Sec. 3.2, these sets of parameters are very similar. It should therefore be unsurprising that the results look nearly identical.

It is important to note that, as in Diggle et al. [2010], I draw predictions from the non-preferential predictive distribution condtional on the plug-in parameters by simulating the S_j's defined in (8). This is problematic, since this predictive distribution is different from the true preferential predictive distribution, which would incorporate the information from the LGCP model. Although the true preferential predictive distribution $(f_{\vec{S}|\vec{X},\vec{Y}}(\vec{s}|\vec{x},\vec{y}))$ is not used in Diggle et al. [2010], I have attempted to derive it in the Appendix in Sec. 6.3. While the results seem promising, they are currently incomplete.
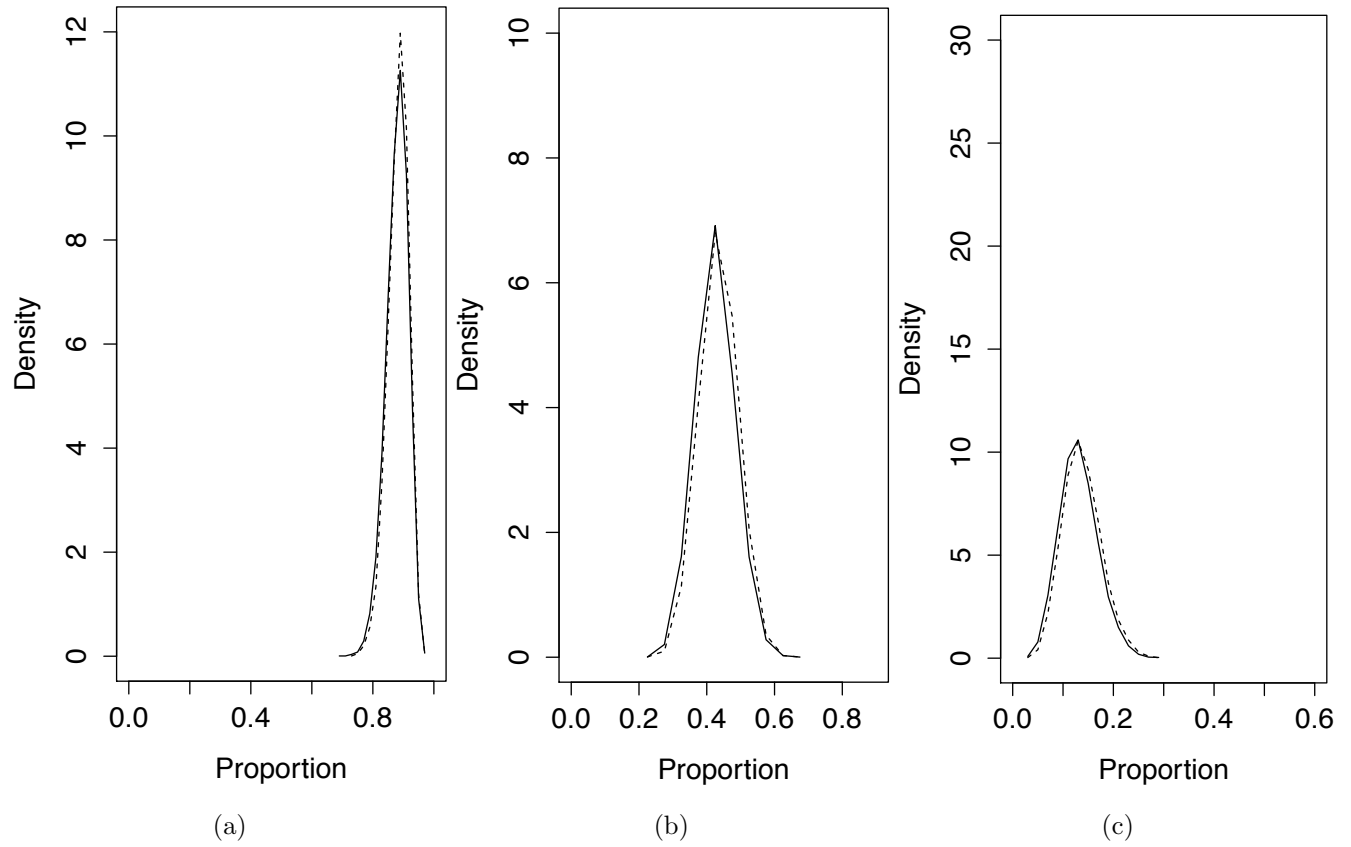
Figure 7: Areal proportion of predicted 1997 lead concentration exceeding (a) 3, (b) 5, and (c) 7 $\mu g/(g$ dry weight). Results are given for both the non-preferential (dashed lines) and preferential (solid lines) prediction.
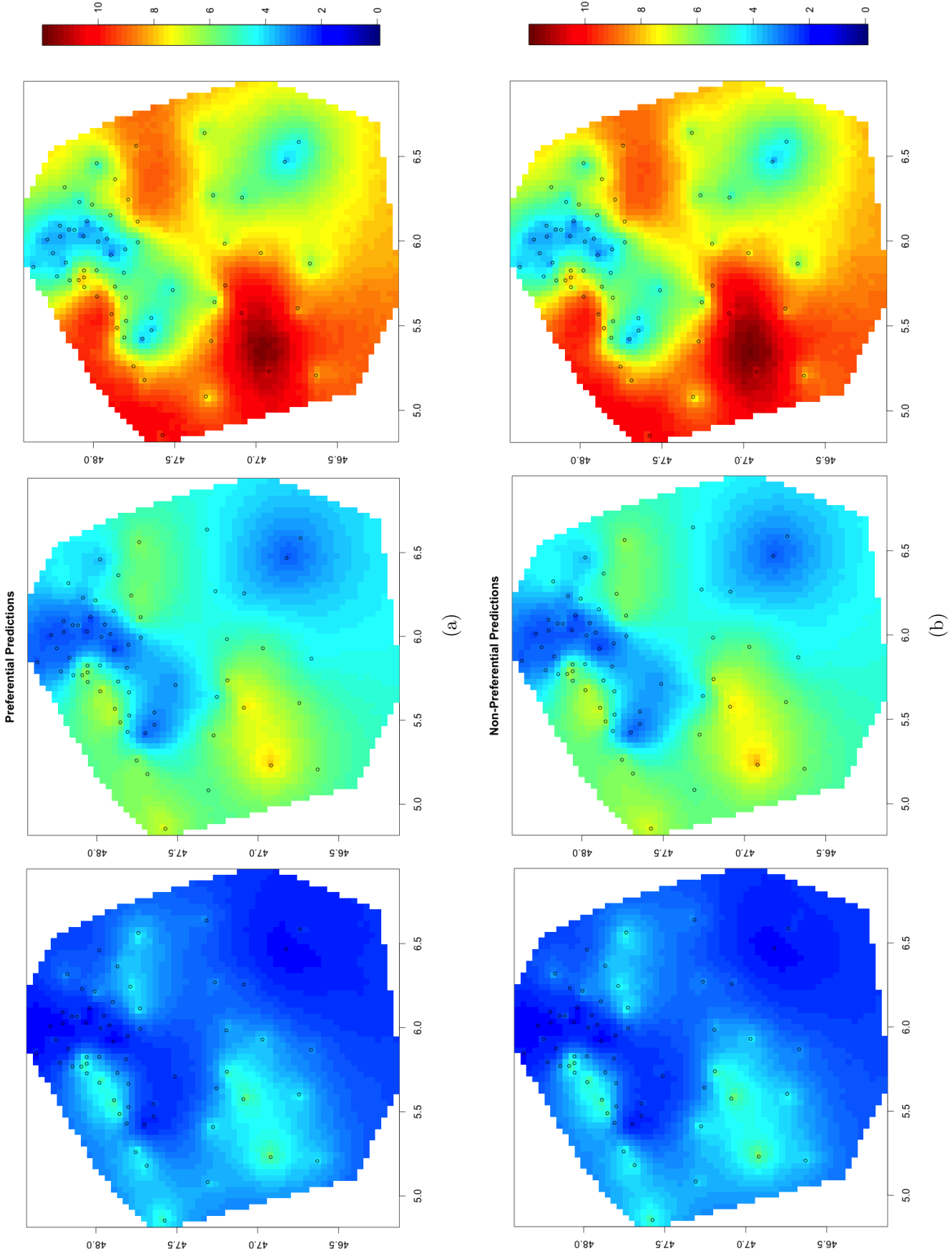
Figure 8: 5th (left), 50th (center) and 95th (right) quantiles from the predictive distribution using the MLEs from the (a) preferential, and (b) non-preferential predictive distributions generated using the plug-in joint model MLEs. The 1997 data locations are plotted as black circles.

4   Discussion

As shown in Sec. 3.1, preferential sampling can have a substantial impact on classical techniques for covariance parameter and variogram estimation as well as on predictions. However, the variance of the field of interest seems to play an important role in the affect of the preferentiality. For instance, in Table 2 the bias is less under model 2 in spite of using a preferentiality parameter, β, that is larger in magnitude than in model 1 (here $\beta = -2.20$ rather than 2). As noted in Diggle et al. [2010], the difference in the bias is likely due to the difference in variance, since $\sigma^2 = 0.14$ in model 2 and 1.5 in model 1. Larger variations in the true process means there are more chances for regions with extreme values that promote clustering or repulsion.

While the LGCP model using the preferential parameter from Diggle et al. [2010] seems to fit the data well enough using the test statistic defined in (9), it is unclear what the correct cutoff distance for the statistic should be. Diggle et al. [2010] uses 25 kilometers as the cutoff, but the spatial domain of interest is roughly 200 kilometers in diamter, and the fit range parameter is $\phi = .31$, which corresponds to a range of 31 kilometers. Under the fit covariance model, data will then be at least 5% correlated until about 93 kilometers away. Perhaps a larger maximum distance in (9) would give more information regarding how well the LGCP model truly fits the data.

The predictions correspond well to the joint preferential model predictions given in Diggle et al. [2010]. It is important to note, however, that the data locations are sparse in all of the spatial domain except for the north, and the data is especially sparse in the south and east regions. Assumptions of stationarity and constant expectation may be especially problematic when predicting far from the data, so predictions farther south than the 45.5 kilometer northing portion of the predictive maps for instance are likely not as certain as claimed.

5   Conclusions

Although classical geostatistical methods often ignored preferential sampling, this analysis based on Diggle et al. [2010] shows that ignoring the preferential nature of the sampling procedure can lead to biased results. Simulation results indicate that pref-

erentiality leads to negatively biased empirical variogram estimates with more bias as the distances of the variogram bins increase. In addition, the simulation study suggests that generated predictions are biased after the parameters are fit using maximum likelihood. When preferentiality is positive ($\beta > 0$ in an LGCP model) the predictions tend to be too high, while when preferentiality is negative ($\beta < 0$) the predictions tend to be too low.

Additional testing would help to show the usefulness of the preferential model proposed by Diggle et al. [2010]. While the simulation study shows that non-preferential models perform poorly in a preferential setting, there is no testing of the preferential model. The same test as performed in the simulation study (the results of which are shown in Figure 4 and Table 2) could test how well the the preferential model is able to estimate covariance parameters and make predictions on the unit square under a known model.

The main contribution of Diggle et al. [2010] is the combination of the LGCP model for preferentiality with the classical spatial statistical model for geostatistical data in (1). The Monte Carlo likelihood technique makes the likelihood tractable, and the flexibility of LGCP models could make the preferential model proposed by Diggle et al. [2010] applicable in many settings. However, the lack of use of a true predictive distribution under the model and instead using the plug-in MLEs under the non-preferential predictive distribution is problematic.

More recently, MCMC methods for inference and prediction with geospatial data under a LGCP model for preferentiality have been applied in the `lgcp` R package [Taylor et al., 2011]. Also, Diggle et al. [2013] discusses moment-based, likelihood, and Markov chain Monte Carlo (MCMC) techniques for performing inference, and the `INLA` package in R can be used for the same purpose. Future directions may lie in improving MCMC methods and extending this model to include covariates, perhaps as in Eq. (12) of Diggle et al. [2010].

References

JP Chilès and P Delfiner. Geostatistics: Modeling spatial uncertainty. *New York: Wiley*, 1999.

David R Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 129–164, 1955.

Noel Cressie. Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, 17(5):563–586, 1985.

Noel Cressie. *Statistics for spatial data*. New York: Wiley, 1991.

Frank C Curriero, Michael E Hohn, Andrew M Liebhold, and Subhash R Lele. A statistical evaluation of non-ergodic variogram estimators. *Environmental and Ecological Statistics*, 9(1):89–110, 2002.

Peter Diggle, Raquel Menezes, and Ting-li Su. Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(2):191–232, 2010.

Peter J Diggle. *Statistical analysis of spatial point patterns*. Edward Arnold, 2003.

Peter J Diggle, Paula Moraga, Barry Rowlingson, Benjamin M Taylor, et al. Spatial and spatio-temporal log-gaussian cox processes: Extending the geostatistical paradigm. *Statistical Science*, 28(4):542–563, 2013.

Jo Eidsvik, Sara Martino, and Håvard Rue. Approximate Bayesian inference in spatial generalized linear mixed models. *Scandinavian journal of statistics*, 36(1):1–22, 2009.

Alan E Gelfand, Peter Diggle, Peter Guttorp, and Montserrat Fuentes. *Handbook of spatial statistics*. CRC press, 2010.

EH Isaaks and R Mohan Srivastava. Spatial continuity measures for probabilistic and deterministic geostatistics. *Mathematical geology*, 20(4):313–341, 1988.

Duncan Lee, Claire Ferguson, and E Marian Scott. Constructing representative air quality indicators with measures of uncertainty. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(1):109–126, 2011.

Peter McCullagh. Sampling bias and logistic models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):643–677, 2008.

John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

Debdeep Pati, Brian J Reich, and David B Dunson. Bayesian geostatistical modelling with informative sampling locations. *Biometrika*, 98(1):35–48, 2011.

Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC Press, 2005.

Martin Schlather, Paulo J Ribeiro, and Peter J Diggle. Detecting dependence between marks and locations of marked point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):79–93, 2004.

R Mohan Srivastava and Harry M Parker. Robust measures of spatial continuity. In *Geostatistics*, pages 295–308. Springer, 1989.

Benjamin M Taylor, Tilman M Davies, Barry S Rowlingson, and Peter J Diggle. lgcp an r package for inference with spatio-temporal log-Gaussian Cox processes. *arXiv preprint arXiv:1110.6054*, 2011.

Andrew TA Wood and Grace Chan. Simulation of stationary Gaussian processes in [0, 1] d. *Journal of computational and graphical statistics*, 3(4):409–432, 1994.

6 Appendix

6.1 The preferential likelihood as a Monte Carlo problem

$$
\begin{aligned}
L(\theta) = f_{X,Y}(x,y) &= \int f_S(s) f_{X|S}(x|s) f_{Y|S,X}(y|s,x) \ ds \\
&= \int f_{X|S}(x|s) f_{Y|S,X}(y|s,x) \frac{f_{S|Y}(s|y)}{f_{S|Y}(s|y)} f_S(s) \ ds \\
&= \int f_{X|S}(x|s) f_{Y|S_0}(y|s_0) \frac{f_{S|Y}(s|y)}{f_{S_0|Y}(s_0|y) f_{S_1|S_0,Y}(s_1|s_0,y)} f_{S_0}(s_0) f_{S_1|S_0}(s_1|s_0) \ ds \\
&= \int f_{X|S}(x|s) \frac{f_{Y|S_0}(y|s_0)}{f_{S_0|Y}(s_0|y)} f_{S|Y}(s|y) f_{S_0}(s_0) \ ds \\
&= E_{S|Y} \left[f_{X|S}(x|s) \frac{f_{Y|S_0}(y|s_0)}{f_{S_0|Y}(s_0|y)} f_{S_0}(s_0) \right]
\end{aligned}
$$

6.2 Terms of the preferential likelihood

$[X|S]$, the inhomogeneous Poisson process likelihood, is given by:

$$L(\vec{\theta}) = \left(\prod_{i=1}^{n} \Lambda(x_i)\right) \left(\int \Lambda(x) \; dx\right)^{-n}$$

$$= \left(\prod_{i=1}^{n} \exp\{\alpha + \beta S(x_i)\}\right) \left(\int \exp\{\alpha + \beta S(x)\} \; dx\right)^{-n}$$

Since Diggle et al. [2010] does not specify how to choose α, I derive the profile MLE using the MLE for the canonical parameter of the Poisson distribution. First, note that if $X \sim Pois(a)$, then $\hat{a}_{MLE} = X_1$ if we only have one observation. Since the number of observations, say n, in the 1997 data is distributed $Pois\left(\int \Lambda(x) \; dx\right) \overset{d}{=} Pois(A \cdot E\Lambda)$ where A is the area of the spatial domain, our MLE for $E\Lambda$ is $\frac{n}{A}$. But $\Lambda(x)$ is distributed lognormal, so

$$E\Lambda = E\exp\{\alpha + \beta S\}$$

$$= \exp\{\alpha + \beta^2 \sigma^2/2\}$$

We can therefore easily solve for a profile estimate of α:

$$\frac{n}{A} = \exp\{\hat{\alpha} + \beta^2 \sigma^2/2\}$$

$$\log(\frac{n}{A}) = \hat{\alpha} + \beta^2 \sigma^2/2$$

$$\hat{\alpha} = \log(\frac{n}{A}) - \beta^2 \sigma^2/2$$

Since S is simulated on a fine lattice, we can compute Λ over the lattice as well as its mean.

The distribution of $S|Y$ can be derived using the fact that they are jointly normal. To see this, note that $Y_i = \mu + S(x_i) + Z_i$, where $Z_i \perp\!\!\!\perp S$. The independence of Z_i and S and their respective distributions being Gaussian implies S and Y are jointly normal. Given this fact, we can use the standard conditional distribution for jointly multivariate normal random vectors [Rue and Held, 2005, Chapter 2]. If random vectors $\vec{A}$ and $\vec{B}$ are jointly normal such that

$$\begin{pmatrix} \vec{A} \\ \vec{B} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \vec{\mu}_A \\ \vec{\mu}_B \end{pmatrix}, \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{pmatrix}\right)$$

then

$$\vec{\mu}_{A|B} = \mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(\vec{B} - \vec{\mu}_B)$$

$$\Sigma_{A|B} = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}$$

Since

$$\begin{pmatrix} \vec{S} \\ \vec{Y} \end{pmatrix} = \begin{pmatrix} \vec{S} \\ \vec{S}_0 + \vec{Z} + \vec{\mu} \end{pmatrix}$$

$$\sim \text{MVN}\left(\begin{pmatrix} \vec{0} \\ \vec{\mu} \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma C' \\ C\Sigma & \Sigma_0 \end{pmatrix} \right)$$

we apply the conditional formulas to get $\vec{S}|\vec{Y}$ is multivariate normal with expectation and covariance given by:

$$\vec{\mu}_{\vec{S}|\vec{Y}} = \vec{\mu}_S + \Sigma C'\Sigma_0^{-1}(\vec{Y} - \vec{\mu})$$

$$= \Sigma C'\Sigma_0^{-1}(\vec{Y} - \vec{\mu})$$

$$\Sigma_{\vec{S}|\vec{Y}} = \Sigma - \Sigma C'\Sigma_0^{-1}C\Sigma$$

Put $\widetilde{\Sigma}_0 = \Sigma_0 - \tau^2 I = C\Sigma C'$. Since $\vec{S}_0 = C\vec{S}$, we then get:

$$\vec{\mu}_{\vec{S}_0|\vec{Y}} = C\vec{\mu}_{\vec{S}|\vec{Y}}$$

$$= C\Sigma C'\Sigma_0^{-1}(\vec{Y} - \vec{\mu})$$

$$= \widetilde{\Sigma}_0\Sigma_0^{-1}(\vec{Y} - \vec{\mu})$$

Similarly,

$$\Sigma_{\vec{S}_0|\vec{Y}} = C\Sigma C' - C\Sigma C'\Sigma_0^{-1}C\Sigma C'$$

$$= \widetilde{\Sigma}_0 - \widetilde{\Sigma}_0\Sigma_0^{-1}\widetilde{\Sigma}_0$$

$$= \widetilde{\Sigma}_0 - (\Sigma_0 - \tau^2 I)\Sigma_0^{-1}(\Sigma_0 - \tau^2 I)$$

$$= \widetilde{\Sigma}_0 - (\Sigma_0 - 2\tau^2 I + \tau^4\Sigma_0^{-1})$$

$$= \tau^2 I - \tau^4\Sigma_0^{-1}$$

$$= \tau^2(I - \tau^2\Sigma_0^{-1})$$

Hence,

$$\vec{S}|\vec{Y} \sim \text{MVN}\left(\Sigma C'\Sigma_0^{-1}C\Sigma,\ \Sigma - \Sigma C'\Sigma_0^{-1}C\Sigma\right)$$

$$\vec{S}_0|\vec{Y} \sim \text{MVN}\left(\widetilde{\Sigma}_0\Sigma_0^{-1}C\Sigma,\ \tau^2(I - \tau^2\Sigma_0^{-1})\right)$$

Note that this assumes a preferential model, since the observation locations $\vec{X}$ give no additional information from the log-Gaussian Cox Process model.

The S_j's given by (10) in Diggle et al. [2010] have the same distribution as $\vec{S}|\vec{Y}$. It is clearly multivariate normal, since the Z vector of simulated measurement noise is independent of $\vec{Y}$ and $\vec{S}$. I will denote the simulated Z vector from (10) from Diggle et al. [2010] as $\tilde{Z}$ from now on to emphasize the fact that it is independent from the Z_i's in (1). Then, conditional on the data $\vec{Y}$,

$$
\begin{aligned}
E\left[S_j\right] &= E\left[S + \Sigma C'\Sigma_0^{-1}(\vec{Y} - \vec{\mu} + \tilde{Z} - C\vec{S})\right] \\
&= \vec{0} + \Sigma C'\Sigma_0^{-1}(\vec{Y} - \vec{\mu} + \vec{0} - C\vec{0}) \\
&= \Sigma C'\Sigma_0^{-1}(\vec{Y} - \vec{\mu}) \\
&= \vec{\mu}_{\vec{S}|\vec{Y}}
\end{aligned}
$$

Similarly,

$$\text{Var}(S_j) = \text{Var}\left(S + \Sigma C' \Sigma_0^{-1}(\vec{Y} - \vec{\mu} + \tilde{Z} - C\vec{S})\right)$$

$$= \text{Var}(S) + \text{Var}\left(\Sigma C' \Sigma_0^{-1}(\vec{Y} - \vec{\mu} + \tilde{Z} - C\vec{S})\right)$$

$$+ \text{Cov}\left(S, \Sigma C' \Sigma_0^{-1}(\vec{Y} - \vec{\mu} + \tilde{Z} - C\vec{S})\right)$$

$$+ \text{Cov}\left(\Sigma C' \Sigma_0^{-1}(\vec{Y} - \vec{\mu} + \tilde{Z} - C\vec{S}), S\right)$$

$$\text{Var}(S) = \Sigma$$

$$\text{Var}\left(\Sigma C' \Sigma_0^{-1}(\vec{Y} - \vec{\mu} + \tilde{Z} - C\vec{S})\right) = \Sigma C' \Sigma_0^{-1} \text{Var}\left(\vec{Y} - \vec{\mu} + \tilde{Z} - C\vec{S}\right) \Sigma_0^{-1} C\Sigma$$

$$= \Sigma C' \Sigma_0^{-1}\left(\text{Var}\left(\tilde{Z}\right) + \text{Var}\left(\vec{S_0}\right)\right) \Sigma_0^{-1} C\Sigma$$

$$= \Sigma C' \Sigma_0^{-1} \Sigma_0 \Sigma_0^{-1} C\Sigma$$

$$= \Sigma C' \Sigma_0^{-1} C\Sigma$$

$$\text{Cov}\left(S, \Sigma C' \Sigma_0^{-1}(\vec{Y} - \vec{\mu} + \tilde{Z} - C\vec{S})\right) = \text{Cov}\left(S, \vec{Y} - \vec{\mu} + \tilde{Z} - C\vec{S}\right) \Sigma_0^{-1} C\Sigma$$

$$= \text{Cov}\left(S, -C\vec{S}\right) \Sigma_0^{-1} C\Sigma$$

$$= -E\left[SS'\right] C' \Sigma_0^{-1} C\Sigma$$

$$= -\Sigma C' \Sigma_0^{-1} C\Sigma$$

$$\text{Cov}\left(\Sigma C' \Sigma_0^{-1}(\vec{Y} - \vec{\mu} + \tilde{Z} - C\vec{S}), S\right) = -\Sigma C' \Sigma_0^{-1} C\Sigma$$

$$\text{Var}(S_j) = \Sigma - \Sigma C' \Sigma_0^{-1} C\Sigma$$

We therefore find that the S_j's have the required conditional distribution.

6.3   The correct predictive distribution

The S_j's given in (10) of Diggle et al. [2010] have the same distribution as would be expected under a clasical, non-preferential model as shown earlier in Appendix A in Sec. 6.2. However, it is unclear what to true distribution under the preferential model is. We can write

$$f_{\vec{S}|\vec{X},\vec{Y}}(\vec{s}|\vec{x},\vec{y}) \propto f_{\vec{X},\vec{Y}|\vec{S}}(\vec{x},\vec{y}|\vec{s}) f_{\vec{S}}(\vec{s})$$

$$= f_{\vec{Y}|\vec{X},\vec{S}}(\vec{y}|\vec{x},\vec{s}) f_{\vec{X}|\vec{S}}(\vec{x}|\vec{s}) f_{\vec{S}}(\vec{s})$$

But $f_{\vec{Y}|\vec{X},\vec{S}}(\vec{y}|\vec{x},\vec{s})f_{\vec{S}}(\vec{s})$ is the joint likelihood of $\vec{S}$ and $\vec{Y}$ under a non-preferential model. As given in Sec. 6.2,

$$\begin{pmatrix} \vec{S} \\ \vec{Y} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \vec{0} \\ \vec{\mu} \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma C' \\ C\Sigma & \Sigma_0 \end{pmatrix} \right)$$

with joint likelihood

$$f_{\vec{S},\vec{Y}}(\vec{s},\vec{y}) \propto \exp\left\{ -\frac{1}{2}\begin{pmatrix} \vec{S} \\ \vec{Y} - \vec{\mu} \end{pmatrix}' \begin{pmatrix} \Sigma & \Sigma C' \\ C\Sigma & \Sigma_0 \end{pmatrix}^{-1} \begin{pmatrix} \vec{S} \\ \vec{Y} - \vec{\mu} \end{pmatrix} \right\}$$

Now, letting A be the area of the spatial domain,

$$f_{\vec{X}|\vec{S}}(\vec{x}|\vec{s}) = \left(\prod_{i=1}^{n} \exp\left\{ \alpha + \beta S(x_i) \right\} \right) \left(\int \exp\left\{ \alpha + \beta S(x) \right\} \, dx \right)^{-n}$$

$$= \exp\left\{ n\alpha + \beta \sum_{i=1}^{n} S(x_i) \right\} \left(AE\left[\exp\left\{ \alpha + \beta S(x) \right\} \right] \right)^{-n}$$

$$= \exp\left\{ n\alpha + \beta \sum_{i=1}^{n} S(x_i) \right\} \left(A\exp\left\{ \alpha + \beta^2 \sigma^2/2 \right\} \right)^{-n}$$

$$\propto \exp\left\{ \beta \sum_{i=1}^{n} S(x_i) \right\}$$

$$= \exp\left\{ \vec{\beta}' C' IC\vec{S} \right\}$$

We therefore find the true preferential predictive distribution is

$$f_{\vec{S}|\vec{X},\vec{Y}}(\vec{s}|\vec{x},\vec{y}) \propto \exp\left\{ \vec{\beta}' C' IC\vec{S} - \frac{1}{2}\begin{pmatrix} \vec{S} \\ \vec{Y} - \vec{\mu} \end{pmatrix}' \begin{pmatrix} \Sigma & \Sigma C' \\ C\Sigma & \Sigma_0 \end{pmatrix}^{-1} \begin{pmatrix} \vec{S} \\ \vec{Y} - \vec{\mu} \end{pmatrix} \right\}$$

From here, however, I am not sure how to proceed. Since the term added to the exponent of the Gaussian likelihood is linear in $\vec{S}$, it looks like the terms in the exponent can be combined, but it is unclear how to do so.

6.4 K-function of the log-Gaussian Cox process

Given a point process, $N(\cdot)$, where $N(B)$ is a random variable giving the number of points in set B, Diggle [2003, Chapter 4.2] defines both the *first-order* intensity function,

$$\lambda(x) = \lim_{|dx| \to 0} \frac{E(N(dx))|}{|dx|}$$

and the *second-order* intensity,

$$\lambda_2(x, y) = \lim_{|dx|, |dy| \to 0} \frac{E(N(dx)N(dy))|}{|dx||dy|}$$

where dx and dy are open sets in our space containing x and y and the above norms in the denominator denote the Lebesgue measure. We further define the *conditional intensity*,

$$\Lambda_c(x) = \frac{\lambda_2(x, y)}{\lambda(y)},$$

which represents the intensity at point x conditional on there begin a point at location y. For a Cox process, we have by Diggle [2003, Chapter 5.5] that

$$\lambda(x) = E\left[\Lambda(x)\right]$$

$$\lambda_2(x, y) = E\left[\Lambda(x)\Lambda(y)\right]$$

$$= \text{Cov}\left(\Lambda(x), \Lambda(y)\right) + E\left[\Lambda(x)\right]^2$$

Under stationarity, we let $\lambda \equiv E\left[\Lambda(x)\right]$, and put $\gamma(t) \equiv \text{Cov}\left(\Lambda(x), \Lambda(x + t)\right)$. As given in Diggle et al. [2013], we then get $\gamma(t) = \lambda^2 \left(\exp\left\{\beta^2\sigma^2\rho(u; \phi, \kappa)\right\} - 1\right)$. Note that $\gamma(t)$ as used here is not the same $\gamma(t)$ in (11) of Diggle et al. [2010].

The K-function of a stationary point process is defined as

$$K(t) = \lambda^{-1}E\left[N_0(t)\right]$$

when t is taken as a distance, and $N_0(t)$ is the number of points in the point process that are in a disc of radius t centered at any point in the point process. Note that we do not count the point in the center of the disc. Hence, following Diggle [2003, Chapter

4.2],

$$K(t) = \lambda^{-1} E\left[N_0(t)\right]$$
$$= \lambda^{-1} \int_0^{2\pi} \int_0^t \lambda_c(t)t \ dt \ d\theta$$
$$= 2\pi\lambda^{-1} \int_0^t \frac{\lambda_2(t)}{\lambda}t \ dt$$
$$= 2\pi\lambda^{-2} \int_0^t (\gamma(t) + \lambda^2)t \ dt$$
$$= 2\pi\lambda^{-2} \int_0^t \lambda^2 t \ dt + 2\pi\lambda^{-2} \int_0^t \gamma(t)t \ dt$$
$$= \pi t^2 + 2\pi\lambda^{-2} \int_0^t \gamma(t)t \ dt$$
$$= \pi t^2 + 2\pi\lambda^{-2} \int_0^t \lambda^2 (\exp\left\{\beta^2\sigma^2\rho(u;\phi,\kappa)\right\} - 1)t \ dt$$
$$= \pi t^2 + 2\pi \int_0^t (\exp\left\{\beta^2\sigma^2\rho(u;\phi,\kappa)\right\} - 1)t \ dt$$

This is (11) from Diggle et al. [2010].

6.5   Variogram fitting

As noted in Diggle et al. [2010], there are two observations from the 2000 data that are outliers. Using the authors' proposed method, I replace the outliers with mean of the other observations. This leads to statistical properties of the dataset matching those given in Table 2 from Diggle et al. [2010]. When fitting the variograms using this data, I get the results shown in Figure 5, and the empirical 2000 variogram has smaller variance than that given in Diggle et al. [2010]. I replot the empirical variogram using the raw data with the fit from the corrected data in Figure 9. Although the empirical variogram appears the match that of Diggle et al. [2010], it is using the uncorrected data that includes the outliers. If the parameters for the 2000 data were fit using the raw data, it could explain some of the the difference in the likelihood ratio results testing the applicability of a joint model versus two separate models for 1997 and 2000.
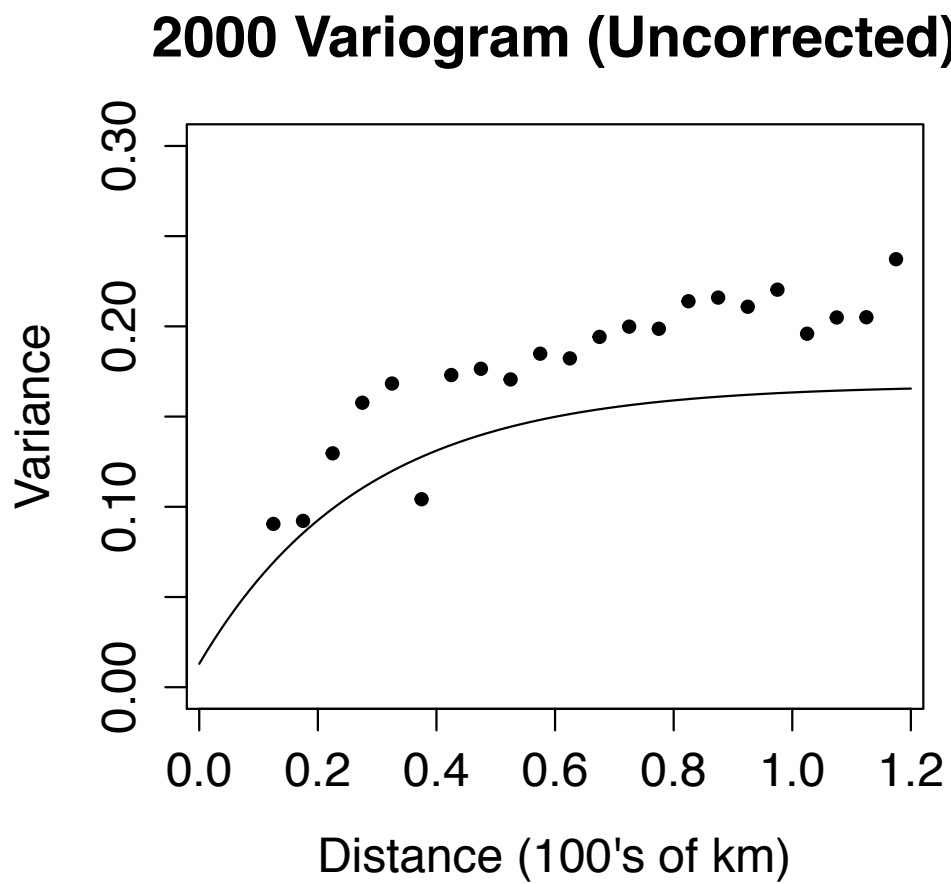
Figure 9: Empirical variogram (dots) for uncorrected year 2000 log-lead data plotted against the fitted theoretical Matérn variogram for the corrected log-lead data (solid line).