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

John Paige

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

1 Background and Motivation

(VM's comment: **I would try to keep introduction as free from formulae as possible. Try to explain math in words.**) (JP's comment: **I removed some unnecessary math notation, although there's still a decent amount left. I tried to emulate the paper to an extent starting off with the model, but I've tried to describe the meaning of the equations in words for the most part here.**) 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 $X = \{x_1, x_2, ..., x_n\}$ (often a subset of $\mathbb{R}^2$, as in Diggle et al. [2010]). In others, 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. Assuming a constant mean is not restrictive, since we can model μ as a function in space using a generalized linear regression framework, where errors are correlated due to spatial correlations in S.

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:

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

The positive definite function $\rho(\cdot|\vec{\theta})$ represents a correlation depending on distance conditional on the parameters $\vec{\theta}$. The variance in Eq. (3) is called a variogram function. 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 very flexible, containing exponential correlation as a special case when $\kappa = 0.5$, and is commonly used in geostatistics. One difficulty when using a Matérn covariance model is that ϕ and κ are sometimes difficult to estimate jointly, especially κ [Diggle et al., 2010].

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 for the variance based on two observations assuming the y_i have zero expectation [Diggle et al., 2010]. If the y_i are not mean zero, we can simply subtract off μ from Eq. (1). Additionally, we could partition the possible distances between observation locations into sufficiently fine bins, and estimate the variance between differences of observation values when the corresponding distances are in any given bin. Of course, if the bins are too fine, then our variogram estimates within each distance bin would be poor. More advanced methods for estimating variograms are given in Chilès and Delfiner [1999, section 2.2], Cressie [1985], and Cressie [1991, section 2.4].

In classical geostatistical analysis, it is assumed 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. This is because 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 polluatant 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). (VM's comment: **I would beef up literature review with connections to other fields, where something like preferential sampling comes up. For example, informative doctor visit times in medical statistics. The discussion of the paper may provide more connections.**) (JP's comment: **Just added some more references to uses of preferential sampling.**)

In order to account for preferential sampling in a spatial statistical context, 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 non-stationary 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 through-out 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 $N(B) \sim Pois(\lambda\mu(B))$, where μ is the Lebesgue measure, and

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 general-izations of inhomogeneous Poisson processes with the following properties:

1. Random rate $\Lambda = \{\Lambda(x) : x \in \mathbb{R}^2\}$ 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

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

where $S(x)$ is the stochastic Gaussian process defined in Eq. (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 pro-cesses, 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 the covariance of Λ, $\text{Cov}\left(\Lambda(x),\Lambda(x+u)\right)$ under the parameterization of Eq. (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. Note that all data used in this report is available at www.lancs.ac.uk/staff/diggle/.
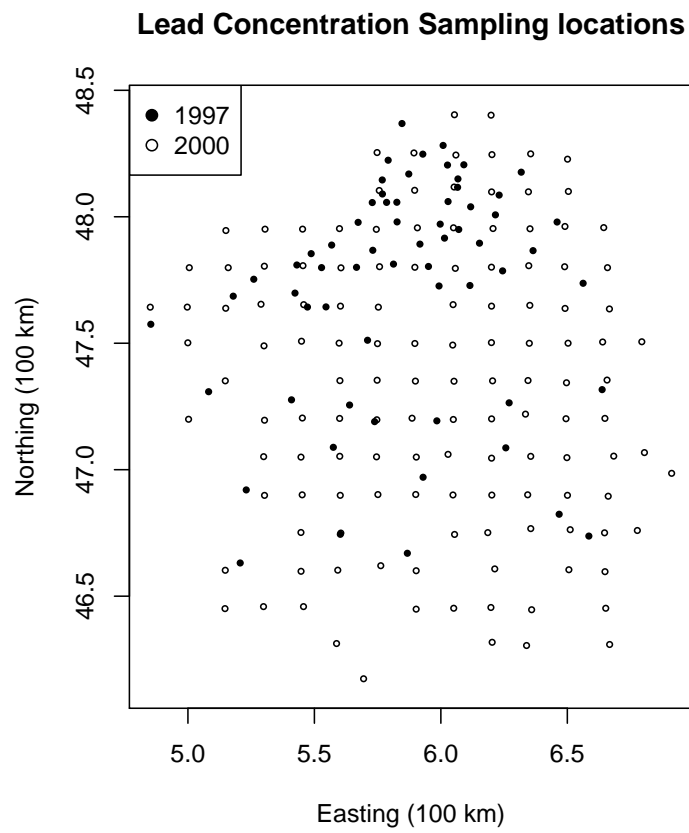
Figure 1: Lead concentration data sample locations for 1997 and 2000.

	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.

2.1 Unit square simulations

In addition to using lead concentration data, Diggle et al. [2010] simulates data on the unit square under three different sampling schemes in order to test for the effects of preferential sampling. 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 intensity from a LGCP, $\Lambda(x) = \exp\{4 + 2S(x)\}$, 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 $\kappa = 1$.

We estimate variograms using the point-cloud based method of moment estimator discussed in Sec. 1 with functions from the `fields` package. We generate 500 simulations under each sampling scheme and use them to estimate the variogram bias and standard deviations under each scheme. Note that we 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 Kriging pre-
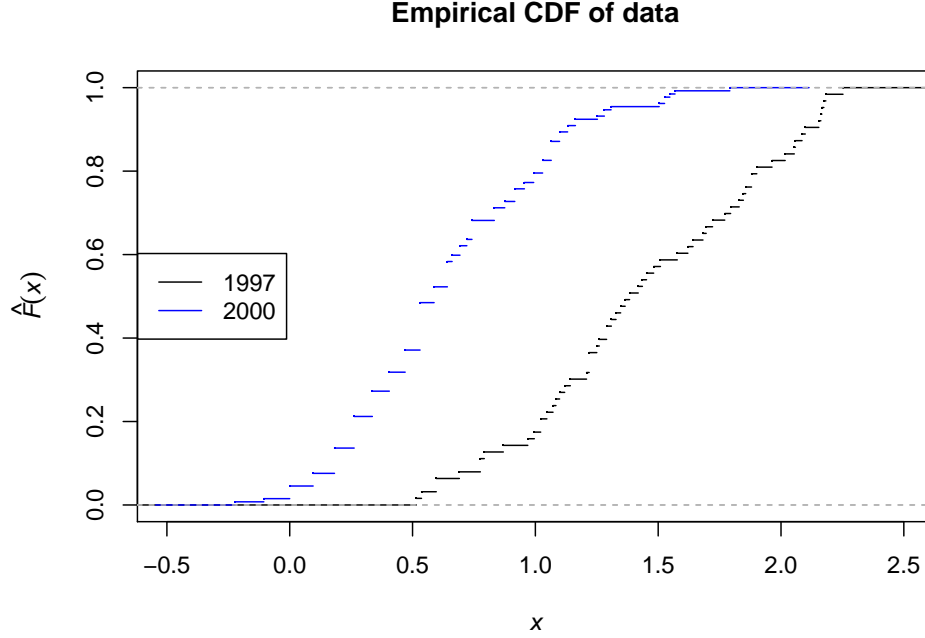
Empirical CDF of data

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

dictor, we 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 Kriging predictor we use two models. In the first, we use maximum likelihood to estimate parameters under the classical geostatistical model ignoring preferentiality. In the second, we set the simulation model parameters to be those of the maximum likelihood estimates of the preferential model parameters 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 $\tau^2 = 0.059$. We then set the plug-in parameters used when generating the ordinary kriging predictor to be the parameters used in the simulations.

2.2   Model inference

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

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

8

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 from a Frequentist perspective $f_Y(y) = \mathbb{1}_{(Y=y)}$. Under preferentiality, however, we cannot simplify Eq. (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,

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

This can be approximated with the Monte Carlo estimate

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

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, put $X^* = \{x_1^*, ..., x_N^*\}$ 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, $X = CX^*$). Given the parameters, θ,

we have $S \sim \text{MVN}(0, \Sigma(\theta))$ and $Y \sim \text{MVN}(\mu, \Sigma_0(\theta))$ where $\Sigma_0(\theta) = C\Sigma(\theta)C' + \tau^2 I$. 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 $Z \stackrel{iid}{\sim} \text{MVN}(0, \tau^2 I)$, 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]. It can easily be verified using the 'sweep' operator from Rue and Held [2005, Chapter 2]. Since the above Z vector is independent of S, they are jointly normal. Their conditional is therefore normal with mean

$$\mu_{S|Y} = ES + \Sigma C' \Sigma_0^{-1}(y - \mu - ES_0) = \Sigma C' \Sigma_0^{-1}(y - \mu)$$

and variance

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

It is easy to verify that Eq. (8) is normal with the correct mean and variance. Note that this formula does not require the inversion of Σ, which is the much larger size of $N \times N$. In fact, using the efficient Cholesky decomposition and forward and backsolve algorithms we can even avoid inversion of Σ_0. Although this algorithm does require sampling from the marginal distribution of the N-dimension vector S, this can be performed efficiently using the circulant embedding algorithm, which requires only order $\mathcal{O}(N^2)$ floating point operations [Wood and Chan, 1994]. We generate samples 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 Eq. (7), evaluation of the likelihood becomes tractable with the simulation method given above. We evaluate the random intensity on our 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. Conditional on these, $\hat{\alpha} = \log(n/\text{Area}) - \hat{\beta}^2 \hat{\sigma}^2/2$, where n is the number of 1997 samples and Area is the area of the domain. This is based on the expectation of a Lognormal distribution (due to the exponentiation of S, which is Gaussian), and the MLE

for Poisson distribution parameter λ. This estimator for α is both the method of moments estimator and the MLE due to MLE invariance conditional on θ. Using our lattice-based approximation of $\Lambda(x)$, we apply the inhomogeneous Poisson process likelihood given in Diggle et al. [2013]:

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

where the integral is taken over the entire domain and is calculated numerically after computing $\Lambda(x)$ over X^*.

The $f_{Y|S_0}$ and $f_{S_0|Y}$ portions of the likelihood can be calculated using multivariate normal densities. Conditional on S_0, the Y_i are distributed $Y_i \overset{iid}{\sim} \mathcal{N}(\mu + S(x_i), \tau^2)$. On the other hand, $S|Y$ follows a correlated multivariate normal structure given by

$$S|Y \sim \mathrm{MVN}\big(\Sigma C' \Sigma_0^{-1}(Y - \mu), \Sigma - \Sigma C' \Sigma_0^{-1} C \Sigma\big)$$

So the distribution of $S_0|Y$ is

$$S_0|Y \sim \mathrm{MVN}\big(C\Sigma C' \Sigma_0^{-1}(Y - \mu), C\Sigma C' - C\Sigma C' \Sigma_0^{-1} C\Sigma C'\big)$$
$$\overset{d}{=} \mathrm{MVN}\big(C\Sigma C' \Sigma_0^{-1}(Y - \mu), \tau^2(I - \tau^2 \Sigma_0^{-1})\big)$$

Note that the last equality follows after some simplification from the fact that $C\Sigma C' = \Sigma_0 - \tau^2 I$. Similarly, the $f_{S_0}(S_{0j})$ term in Eq. (7) can be calculated using the fact that, marginally, $S_0 \sim \mathrm{MVN}(0, C\Sigma C')$.

We use the Nelder-Mead simplex algorithm [Nelder and Mead, 1965] for likelihood maximization with 10,000 Monte Carlo iterations in our preferential likelihood approximation (**Note: will have to update this when I find out how many iterations my code can achieve in a reasonable amount of time. Right now 10,000 iterations would take on the order of days due mainly to inefficient calling of the circulant embedding function in R, it seems...**). We fit parameters under a standard geostatistical model for 1997 and 2000 separately before fitting a joint model. We 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` function, and we 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 seems 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 the authors found 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. We 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, we generate 99 simulations of the fit LGCP. Using the test statistic

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

where $v(s)$ is the variance of $\hat{K}(s)$ estimated from the 99 simulations, we 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 we 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. We therefore compare the 5%, 50%, and 95% predictive distributions. In addition, we compare kernel density estimates of the proportion of Gallicia with lead concentration exceeding 3, 5, and 7 μg (g dry weight)$^{-1}$ between the two models.

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, 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.

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.