Bayesian Analysis (2010) **5**, Number 1, pp. 97–122

Analyzing Spatial Point Patterns Subject to Measurement Error

Avishek Chakraborty* and Alan E. Gelfand†

Abstract. We address the issue of inference for a noisy point pattern. The unobserved true point process is modelled as a nonhomogeneous Poisson process. For modeling the underlying intensity surface we use a scaled Gaussian mixture distribution. The noise that creeps in during the measurement procedure causes random displacement of the true locations. We consider two settings. With a bounded region of interest, (i) this displacement may cause a true location within the boundary to be associated with an 'observed' location outside of the region and thus missed and (ii) we have the possibility in (i) but also vice versa; the displacement may bring in an observed location whose true location lies outside the region. Under (i), we can only *lose* points and, depending on the variability in the measurement error as well as the number of true locations close to boundary, this can cause a significant number of locations to be lost from our recorded set of data. Estimation of the intensity surface from the observed data can be misleading especially near the boundary of our domain of interest. Under (ii), the modeling problem is more difficult; points can be both lost and gained and it is challenging to characterize how we may *gain* points with no data on the underlying intensity outside the domain of interest. In both cases, we work within a hierarchical Bayes framework, modeling the latent point pattern using a Cox process and, given the process realization, introducing a suitable measurement error model. Hence, the specification includes the true number of points as an unknown. We discuss choice of measurement error model as well as identifiability problems which arise. Models are fitted using an markov chain Monte Carlo implementation. After validating our method against several synthetic datasets we illustrate its application for two ecological datasets.

Keywords: Gaussian mixture model, measurement error model, intensity surface, Markov chain Monte Carlo, Neymann-Scott process, nonhomogeneous Poisson process

1 Introduction

Spatial point processes (Diggle 2003) are employed to model a set of random locations within a bounded region where some particular event or set of events has been observed to take place. These locations are represented through coordinates in, say, $D \subset \mathbb{R}^d$ for some $d > 0$; $d = 2$ provides the usual spatial setting. Illustrative examples arise in ecology where points might denote species locations, in disease mapping where points denote locations of disease cases, and in the development of cities where points denote locations of building construction. See, e.g., Baddeley et al. (2005) and Gatrell

*Department of Statistical Science, Duke University, Durham, NC, mailto:ac103@stat.duke.edu
†Department of Statistical Science, Duke University, Durham, NC, mailto:alan@stat.duke.edu

 DOI:10.1214/10-BA504

et al. (1996) for further examples. We may be interested in learning how incidence and prevalence for a particular species varies over the region. We may be interested in the extent of disease spread in different parts of a region. The "null" behaviour assumed for point patterns is complete spatial randomness, characterized by a homogeneous Poisson process. Alternatives are often specified through a nonhomogeneous Poisson process characterized by an intensity function which is the approach we adopt here.

More specifically, an "intensity" surface is a natural concept for point patterns to reflect expectation of more points in some portion of the region, fewer in others. Models based on nonconstant intensity surfaces are called nonhomogeneous Poisson process (NHPP). When these intensity surfaces are realizations of a stochastic process, the models are usually referred to as Cox processes. See, e.g., Cressie (1993), Møller and Waagepetersen (2002). General model specification for intensities is usually supplied through parametric representations or as a process realization, e.g., from a Gaussian process. Intensities can be designed to capture specific mechanistic behaviour. For instance, the case where points related to each other tend to stay close and produce aggregated patterns (like offspring of the same parents) is generally addressed through cluster processes, which typically use an initial point process model convolved with a growth function to produce the aggregating behaviour. Discussion and interesting applications can be found in, e.g., Neyman and Scott (1958) and Faÿ et al. (2006). See section 4.1 for details.

Customarily, the point pattern literature starts with a set of locations as raw data and fits a point process model to this set Our contribution is to consider the setting where the observed locations are measured with error and we seek to assess the resultant effect on the object of our interest, the intensity function. Intuitively, adding noise will "blur" the intensity surface, making detection of its features more difficult. This problem is increasingly relevant as more and more automated map construction systems are coming into play, generating large amounts of spatial data. The degree of accuracy attached to such procedures influences the quality of databases generated using them. It is quite likely that, in recording locations, measurement error is introduced. Noise that creeps into the data comes from the degree of accuracy of the measuring instrument as well as factors influencing detection of event occurrences within the region such as thinning and censoring.

Measurement error is well-studied by now. Fuller (1987) explored measurement error in normal regression models in detail. Relevant work in GLMs can also be found in Stefanski and Carroll (1987). Most of the literature in this area focuses on epidemiological studies where the exposure information is assumed to be recorded with error. The Bayesian perspective on this problem takes the approach of relating data, parameters, and unmeasured variables through a graphical model structure and implementing Markov chain Monte Carlo (MCMC) to fit that model specification. Such models were investigated by, e.g., Richardson and Gilks (1993) and Mallick and Gelfand (1995). In our setting, if for instance, data on locations of trees are collected say, by imaging the forest from above, then the locations of trees in the image can differ from the true loca-

tions. This might be due to the quality of the imaging procedure, the height at which the device was placed, or other factors such as air transparency, clouding, or congestion within the forest. Retrieving the original set of locations may be of interest in this situation. But, as noted above, inference about the intensity surface will be degraded and we seek to quantify the increased uncertainty. A novel version of our problem arises in data confidentiality situations, where location acts as an identifier for a record in a database. If we seek to publicly release such databases then location must be perturbed by adding some random noise. Understanding the implications of perturbation enables specification of mechanisms that will retain essentially the same spatial structure under perturbation.

Modeling point patterns using intensities requires restriction to a bounded subset of the plane. As a result, such measurement noise can push locations in and out of the study domain. Thus we are not only observing a noisy version of the original realization, but it is also possible that we are missing some of the true events and also observing some which are not truly in the study region. For patterns having high event aggregation near the boundary of the region, this problem can be quite significant. In Sections 3, 4 and 5 we show different illustrative examples. Expressed in different words, in our setting, measurement error results in a form of censoring to yield the actual dataset Modeling of censored data is common in the survival analysis literature where one observes the event exactly only if it occurs in a certain time interval. Relevant literature can be found in Breslow (1974), Cox and Oakes (1984), Sinha and Dey (1997), Kalbfleisch (1978). Below, similarities and differences are pursued further.

There is a small previous literature on degraded point patterns. That is, points within the study region may be contaminated with points not belonging to the true pattern, points may be lost, and points may be subject to displacement. A general description is that the observed pattern is a random transformation of the true pattern. For instance, Diggle (1993) viewed the transformation as a conditionally independent random deformation of the true pattern and examined its effect on the familiar K function which is used to capture the expected number of points within a given distance of a given point (Ripley 1977). Intensity estimation from a dataset that is incompletely geocoded can be found in Zimmerman (2008). We look at the problem as a two-stage specification - model the true pattern and given the true pattern, model the random transformation. Work in this spirit appears in Lund and Rudemo (2000) and Lund et al. (1999). There, the distinction is made between inferring about the properties of the true point pattern and reconstructing the true point pattern. In particular, Lund and Rudemo (2000) formulate the problem as one of maximum likelihood estimation. They assume that both the true and observed patterns are available. Under a conditional likelihood that incorporates thinning, displacement, censoring, and superposition, and is induced under a cluster process for the true pattern, they obtain maximum likelihood estimates (MLE's) of the noise model parameters (using likelihood approximations). See, also, Baddeley and Van Lieshout (1993, sec. 5) in this regard. On the other hand, Lund et al. (1999), using a similar degradation model, assume that the true pattern is unknown but

that the model parameters are known (say, through training data) and infer about the true pattern. A Gibbs point process with a pairwise interaction function provides the prior for the true pattern.

Within the fully hierarchical framework we adopt, there is no need to separate the point process parameter estimation and pattern reconstruction problems. Both can be addressed through suitable posterior inference and we show how this can be done for fairly general Cox processes. We can infer about uncertainty (which is not available in the MLE approach) and the only prior knowledge we assume relates to the extent of measurement error. We consider two scenarios. First, we assume that events can only occur inside D so a shift of a location due to noise can only throw a point from D to D^c. But, since no event is allowed to take place outside D, each of the noisy locations observed corresponds to some true location in D. What we are assuming is that there will not be any intrusion of points from D^c to D. This requires that D is reasonably well 'isolated' from other possible areas where the event is expected to be observed. Events in those areas have no impact on our experiment within D. We term this setting an "island" model; Section 3 is devoted to formal development. We employ an intensity surface which is a scaled mixture model where the scale parameter captures the expected number of points in D. Then, we remove this restriction by assuming that our region of interest is actually a subset of a bigger region of possible event findings (e.g., mapping tree locations in a specific part of a forest). Now events outside can also enter D because of noise; we refer to this case as a "subregion" model and address it in section 4. We note that the island model is less useful in practice than the subregion model but is inferentially much easier. The subregion model presents a more difficult problem since it is unclear what the "full" region should be as well as how to characterize the intensity outside of the subregion. Here, we employ an intensity surface that is driven by a Poisson cluster process model, in fact a Neymann-Scott process. In Lund and Rudemo (2000) and in Lund et al. (1999) this issue is avoided through the use of an artificial superposition intensity (in fact, a known constant intensity) which simply adds random locations in D.

We consider measurement error in additive form. It is evident that measurement error will tend to result in a more scattered point pattern for the observed data than for the true pattern and this will become even more so with increasing uncertainty in the measurement error process. Again, the impact of noise on event locations is that it can take points within the region to outside and vice versa. Thus, the observed number of points within the study region, particularly near its boundary can be quite different from that in the true pattern and the extent of difference will depend on the magnitude of the variability of the measurement error. For a bounded study domain D and a true location x, we assume the recorded location $y = x+\epsilon$, where ϵ is the measurement error. There may be other noise mechanisms like false detections and/or random missingness but our discussion is limited to such displacement mechanisms. Also, as noted in Lund et al. (1999), we can introduce a systematic displacement u, writing $y = u + x + \epsilon$, if appropriate.) Hence, we can have (i) $x \in D$, $y \in D$, (ii) $x \in D$, $y \in D^c$ or (iii) $x \in D^c$,

$y \in D$.

We propose the usual measurement error model (MEM) specification, i.e., an error model of the form $f(y|x, \theta)$. The Berkson error model works with the form $f(x|y, \theta)$. The latter is computationally attractive to work with because, by conditioning on observed y's, we have fixed locations to insert into a spatial correlation function, hence into the associated covariance matrix. With the MEM we would have unknown locations in this matrix. Since the inverse of this matrix appears in the Gaussian likelihood, computation becomes very challenging. For this reason, Barber et al. (2006) have used the Berkson model in constructing maps for feature locations. But, for us, x is viewed as latent, a member of a point pattern whose intensity we seek to infer about. So, the MEM is required for our hierarchical specification and our main concern turns to how to specify a classical measurement error model which enables feasible computation and produces sensible results. Barber et al. (2006) discuss the relative features of these two approaches. Another discussion on comparison of these two specifications, in the context of assessing radon exposure, can be found in Heid et al. (2004).

The dataset is viewed as arising from a point process model along with a measurement error model. So, the observational data by itself will be unable to separate the uncertainty in these two components. We need an informative prior for the uncertainty in at least one of them. This issue is of separating modeling error from measurement error is common to measurement error modeling in general. The degree of accuracy/uncertainty of a measuring instrument and its variation in performance across different conditions can sometimes be obtained from the manufacturer. In cases such as animal counts in conservation related studies, we may have some prior knowledge about the average range of movement for an animal around its habitat. More generally, Richardson (1995) talks about different sources for obtaining additional information about the measurement error process. One way is repeated measurement, not an option in our setting. Another approach - the inclusion of a validation group or a training dataset - can be of help here. Use of training datasets is familiar in the statistical literature, e.g. Helmers and Bunke (2003). We may run a controlled experiment where we have both true and observed locations using the same measurement procedure. With an appropriate (perhaps spatial) model we can learn about measurement error variability.

We assume conditionally independent homogeneous displacements in the MEM scenario as employed in Diggle (1993), i.e., Ω, the covariance matrix for ϵ, is constant across x. Of course, marginally, the y's are spatially dependent. In some contexts we might imagine that the error variability has spatial structure. That is, points closer to each other are exposed to similar levels of factors that affect location accuracy, so the extent of shift from the true values are also expected to be similar. Also, the variability of noise induced at location x can be influenced by factors present there. Thus, the homogeneous covariance Ω would be replaced with a $\Omega(\mathbf{z}(x))$ for some covariate vector, $\mathbf{z}(x)$. For example in the case of imaging, the elevation and slope at a particular location may affect how much error we are likely to make in capturing that location. However, since the x's

are unknown, such specification produces a complicated posterior full conditional for x and, in our experience, overall model fitting is unstable.

Finally, in most cases studied in the MEM literature, we have a set of observations {Y} generated by mixing some noise distribution with an actual set of underlying observations {X} and the job is to learn about the distribution of X from the data Y. We have identified (X, Y) pairs. In our setting, the number of X's differs from the number of Y's and we have no pairing. This issue is formalized in Lund and Rudemo (2000) and in Lund et al. (1999) through the introduction of a matching s which matches displaced y's with associated x's. s is unknown but marginalization requires summing over a very large number of possible s's. Instead, Lund et al. (1999) retain s as a latent variable, employing a demanding reversible jump MCMC to sample it. In both our island model and our subregion model, we are able to circumvent this issue, as we clarify below.

The format of the paper is as follows. Section 2 develops nonhomogeneous Poisson process model for unobserved points. Section 3 deals with development of the Bayesian modeling of noisy locations, prior specification, posterior computation and inferences in the case of island types of datasets. Two illustrative simulated examples are given. Section 4 generalizes the earlier methodology to the case of a subregion model, again with an example. Section 5 presents application of the methodology to a couple of ecological datasets. Section 6 highlights some of the possible directions of extending our work in terms of modeling and computation.

2 Intensity function modeling

To model the set of true locations inside a bounded domain $D \in \mathbb{R}^2$, we use a spatial nonhomogeneous Poisson process, NHPP (Van Lieshout 2000) with intensity $\lambda : D \to \mathbb{R}^+ \cup \{0\}$, i.e. we are making two basic assumptions about distribution of points in D.

(i) Given any Borel set $B \subseteq D$, number of locations inside B, $\mathcal{N}(B) \sim Poi(\int_B \lambda(s)ds)$

(ii) If $B_1, B_2, ..., B_k$ are disjoint Borel subsets of D for any $k \in \mathbb{N}$, then $\mathcal{N}(B_1), \mathcal{N}(B_2), ..., \mathcal{N}(B_k)$ are independent.

We need $\lambda(\cdot)$ to be a Borel measure, and $\int_D \lambda(x)\, dx < \infty$. In the subsequent analysis, we will work with *simple* point processes, i.e., processes that don't allow multiple replications of the same location.

To specify the likelihood associated with a realization $\{s_1, s_2, ..., s_n; n \in \mathbb{N} \cup \{0\}, s_i \in D\}$, we use the fact that, conditional on number of events in D, locations inside D are

independent draws from $\lambda(\cdot)$, normalized to a density over D. Thus we have,

$$L(\lambda(s), s \in D|n; s_1, s_2, ..., s_n) \quad \propto \quad e^{-\int_D \lambda(s)\, ds} \frac{\prod_{i=1}^{n} \lambda(s_i)}{n!} \tag{1}$$

We now focus on modeling the intensity function $\lambda(\cdot)$ on D. From the conditional independence property stated above, the NHPP can be thought as a two stage process, first determining the count and then conditional on the count, generating that number of points from a density over D. Consistent with that, we employ a specification that proposes a density surface over D and elevates the surface to the level of the expected actual count. That is, we model $\lambda(s) = \lambda f(s)$, $s \in D$ as proposed in Kottas and Sansó (2007). This separable λf formulation is easy to interpret; separate parameters take care of the elevation and the orientation of the surface, respectively. Here λ controls the expected number while f is a density which integrates to 1 over D and determines how they should be located. In different words, $\lambda(\cdot)$ provides absolute intensity while $f(\cdot)$ provides relative intensity.

Flexibility in the choice of f allows for a wide range of specifications. Mixture models provide a path. We can provide mixture distributions with a fixed number of components or an unknown, say random number of components. In the same spirit, one can consider a nonparametric choice using Dirichlet process, as in Kottas and Sansó (2007) or Ji et al. (2009). In the sequel, we choose f as Gaussian mixture distribution restricted to D with fixed number of components, anticipating model comparison across various choices for the number of components.

3 The Island Model

Under the island model we assume our study region D contains the support of the true point process i.e. $\mathrm{P}(x \in D^c) = 0$ for any event location x. So now we we can only have (i) $x \in D$, $y \in D$, (ii)$x \in D$, $y \in D^c$.

3.1 Model Specification

Again, in a bounded region $D \subset \mathbb{R}^2$, we assume n observed event locations $(y_1, y_2, ..., y_n)$, which are a noisy version of a set of m actual locations $(x_1, x_2, ..., x_m)$ representing the complete realization of the point pattern in D. So, m is unknown but $m \geq n$ and, when we recorded our observation, $(m-n)$ of these locations fell outside of D. For $x \in D$ we adopt the Gaussian noise distribution suggested in Section 1. Also, conditional on the true location x, y is independent of every other location. Relabeling the x's so that, for $i = 1, 2, ..., n$, x_i is the true location corresponding to y_i with the last $(m-n)$ x's corresponding to y locations outside D, we obtain the following model:

$$\begin{aligned} y_i &\overset{ind}{\sim} \phi_2(\cdot; x_i, \Omega),\ i = 1, 2, ..., n \\ \pi(x_1, x_2, ..., x_m) &= NHPP(\lambda(.)) \\ \lambda(x) &= \lambda f_D(x) = \lambda \sum_{k=1}^{K} q_k \phi_{2,D}(x|\mu_k, \Sigma_k) \end{aligned} \tag{2}$$

where $\phi_{2,D}$ denotes the restriction of the bivariate normal density to D and q_k are the mixing weights. We refer to (2) as the Island measurement error model.

In all subsequent analysis, we work with fairly vague priors: diffused bivariate normals for μ's, inverse Wishart with small degrees of freedom for Σ's, Dirichlet with uniform cell weights for $\mathbf{p}$, and a flat prior on $\mathbb{R}^+$ for λ. However since point patterns such as species distribution do not change rapidly over time, one might propose to use past records to construct the prior. But, as we show in next few examples, if the past data was inclusive of measurement error, using a moderate or strong prior centered around that information can lead to posterior inference quite different from the truth. However, under the λf intensity formulation and a mixture such as f, past data or external information may be useful with regard to the number of mixture components.

3.2 Computational Details

Since our primary inference objective is to estimate the intensity surface, posteriors for $\{q_{1:k}\}$, $\{\mu_{1:k}\}$, $\{\Sigma_{1:k}\}$ and λ are sought. Note that the intensity surface involves a Gaussian probability density function (pdf) which has to be truncated within D. We start with the likelihood computation. It has two parts, one from the observed locations $(y_1, y_2, ..., y_n)$ (say L_1), the other from the unobserved y's known to be in D^c (say L_2). Upon associating the x_i's with y_i's, the likelihood takes the form $L = L_1 L_2$ (as in Lund and Rudemo 2000; Lund et al. 1999) where $L_1 = \prod_{i=1}^{n} \phi_2(y_i|x_i, \Omega)$ and $L_2 = \prod_{i=n+1}^{m} \bar{\Phi}_2(D; x_i, \Omega)$, with $\bar{\Phi}_2(A; a, B) = P(x \notin A | x \sim N(a, B))$

In writing the NHPP prior we assume that the first n of the x's are identified with the observed y's. In fact, there are $\frac{m!}{(m-n)!}$ possible matchings which have been collapsed into a single case. So, the prior density is, in fact,

$$\pi(x_{1:n}, x_{n+1:m}) = \frac{m!}{(m-n)!} \lambda^m \prod_{i=1}^{m} f_D(x_i) \frac{e^{-\lambda}}{m!} \tag{3}$$

In the sequel we assume Ω for the measurement error process is known, obtained in some fashion following the discussion in Section 1, and is suppressed in our notation.

Hence, the full posterior for the model parameters becomes

$$\begin{aligned}
&\pi(m, x_{1:m}, \lambda, \mu_{1:K}, \Sigma_{1:K}, q_{1:k} \mid y_{1:n}) \\
&\propto \binom{m}{n} e^{-\lambda} \frac{\lambda^m}{m!} \prod_{i=1}^{m} f_D(x_i|\mu_{1:K}, \Sigma_{1:K}, q_{1:K}) \prod_{i=1}^{n} \phi_2(y_i \mid x_i) \\
&\quad \times \prod_{i=n+1}^{m} \bar{\Phi}_2(D \mid x_i) \pi(\lambda, \mu_{1:K}, \Sigma_{1:k}, q_{1:K}) \qquad (4)
\end{aligned}$$

We implement a Markov chain Monte Carlo (MCMC) algorithm to fit the model and sample from the posterior full conditionals in the following sequence: (i) $\lambda, \mu_{1:k}, \Sigma_{1:k}, q_{1:k}$ $|x_{1:m}, m$ and (ii) $x_{1:m}, m|\lambda, \mu_{1:k}, \Sigma_{1:k}, q_{1:k}$ via $m|\lambda, \mu_{1:K}, \Sigma_{1:K}, q_{1:K}$ followed by $x_{1:m}|m, \lambda$, $\mu_{1:K}, \Sigma_{1:K}, q_{1:K}$.

With f_D as in (2), the full conditional for $\mu_{1:K}, \Sigma_{1:K}, q_{1:K}$ becomes

$$\pi(\mu_{1:K}, \Sigma_{1:K}, q_{1:K}|m, x_{1:m}) \quad \propto \quad \frac{\prod_{i=1}^{m} \sum_{j=1}^{K} q_j \phi_2(x_i|\mu_j, \Sigma_j)}{(\sum_{j=1}^{K} q_j \Phi_2(D|\mu_j, \Sigma_j))^m} \pi(\mu_{1:K}, \Sigma_{1:K}, q_{1:K}) \quad (5)$$

In the absence of truncation, this is routinely sampled using augmentation with latent mixture component indicator variables (Diebolt and Robert 1994). For a truncated Gaussian distribution, the posterior is non-standard; we use a nontruncated version of the distribution as proposal for the Metropolis step. One of the issues in the data augmentation method is the exchangeability of components or *label switching*. To identify each (μ_k, Σ_k, q_k), one needs to put an order constraint on the set of parameters that can efficiently distinguish each component. For example, putting an order restriction on the component weights can work well only if no pair of components have weights close to each other. One can choose scalar functions of component parameters say, $||\mu_i||$ or $\mu_i^T \Sigma_i^{-1} \mu_i$, which are most likely to be distinct for different components except in very pathological examples. For the applications done in this article, where component weights were similar, we arranged by first component of the mean. See Stephens (2000) for a review of this problem as well as techniques for handling it.

The full conditional distribution for m can be simplified by integrating out the x's yielding

$$\begin{aligned}
&\pi(m|\lambda, \mu_{1:K}, \Sigma_{1:K}, q_{1:K}, y_{1:n}) \\
\propto \quad &\prod_{i=n+1}^{m} \int_D \bar{\Phi}_2(D; x_i, \Sigma) f_D(x_i) dx_i \binom{m}{n} \frac{\lambda^m}{m!} \\
\propto \quad &\left(\int_D \bar{\Phi}_2(D; x, \Sigma) f_D(x) dx \right)^{m-n} \frac{\lambda^{m-n}}{(m-n)!} \qquad (6)
\end{aligned}$$

i.e., m is distributed as $n+v$ where $v \sim Poi(\lambda \int_D \bar{\Phi}_2(D; x, \Sigma) f_D(x) dx)$. This matches our intuition since λ is the expected number of observations in D and $\int_D \bar{\Phi}_2(D; x, \Sigma) f_D(x) dx$ is the probability that a random location from this point process will be thrown outside of D because of noise.

For $x_1, ..., x_n$, the full conditionals turn out to be Gaussian. For $x_{n+1}, x_{n+2}, ..., x_m$, the full conditional consists of contributions from f and $\bar{\Phi}_2$. We generate samples from f and employ the accept-reject method with $\bar{\Phi}_2$.

Next, we attempt to provide insight to clarify that, with informative knowledge of the measurement error uncertainty, we can expect good behaviour in the estimation of the point process intensity. We investigate analytically, whether we can expect to retrieve the true number of observations m or corresponding parameter λ from the Gibbs sampler. Suppose we use a starting value of $m_0 = n$ in our MCMC. At each iteration we are simulating m from a shifted Poisson distribution with mean parameter $\lambda(1-p)$ and shift n, where $p = \int_D \Phi_2(D; x, \Sigma) f_D(x) dx$. Notice that p does not depend on either m or λ; it depends only on the parameters of f and Σ. Under a noninformative prior for λ, at each stage, λ is updated from a Gamma$(m+1, 1)$ distribution. Then, iteratively,

$$E(m_t | y, m_{t-1}, p_{t-1}) \quad = \quad n + (m_{t-1} + 1)(1 - p_{t-1})$$

So $E(m_t - m_{t-1} | y, m_{t-1}, p_{t-1}) = n + (1 - p_{t-1}) - m_{t-1} p_{t-1}$. Thus $E(m_t - m_{t-1} | y, m_{t-1}, p_{t-1}) \geq 0$ if and only if $n + (1 - p_{t-1}) \geq m_{t-1} p_{t-1}$ which we can rewrite as $m_{t-1} \leq \frac{n}{p_{t-1}} + \frac{(1-p_{t-1})}{p_{t-1}}$. Also, $E(n|\lambda, p) = \lambda p$, so $\mathrm{E}(\frac{n}{p}|\lambda, p) = \lambda$. Thus, if the p_t sequence is well-behaved then we can expect $\frac{n}{p_{t-1}}$ to be close to λ and thus, to learn about m. This suggests that m_t is expected to increase from its starting value but eventually we expect m_t to behave well since $E(m_t|\lambda, p)$ will converge to $\lambda + \frac{1-p}{p}$.

3.3 Examples

We consider two simulation examples to illustrate our methodology. In the first case, we generate data from an intensity surface with f as a bivariate normal distribution (not a mixture) with parameter values given in Table 1. Our observation window is a unit square in $\mathbb{R}^2$, [0,1]×[0,1] so f is normalized to this square and the expected number of points is 200. We simulate the x's and then add a Gaussian zero-mean noise with dispersion $\left(\begin{smallmatrix} 0.036 & 0.002 \\ 0.002 & 0.021 \end{smallmatrix}\right)$ to obtain the y's. Initially there were 197 points in the window, but, with the addition of noise, only 180 were left, so we have lost about 8% of the points. (The expected fraction of points lost can be calculated using the intensity and error parameters. It turns out to be 8.34%.) Figure 1 shows the original and perturbed point patterns. One can clearly see the increased spread in the perturbed pattern.

Again, the inference goal is to learn about m, λ, and f, and to display the estimated

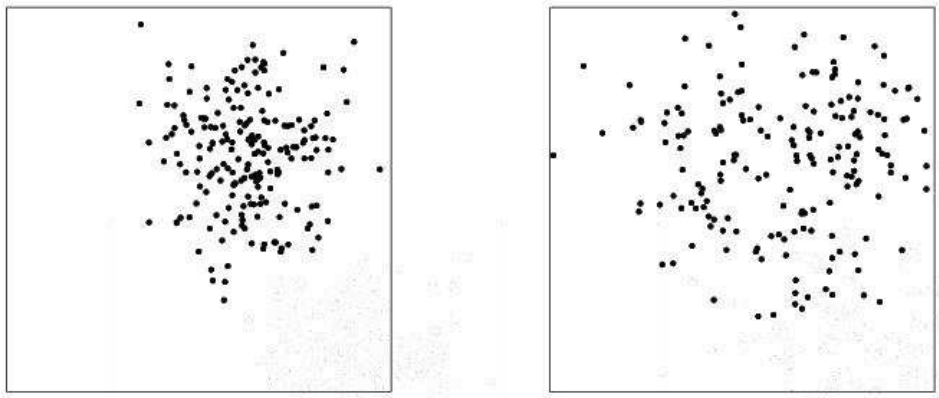

Figure 1: Original (*left*) and Perturbed (*right*) point patterns

intensity with associated uncertainty. We also offer a comparison, fitting the island model and fitting a NHPP model assuming there is no measurement error. Table 1 and Figure 2 show the comparison between the two models (parameters that noticeably differ are in **bold**). As expected, the Island model substantially improves over the NHPP model; also, it is able to retrieve the model parameters which generated the actual point pattern. Plots of the actual intensity along with the estimated intensity using the Island model and the NHPP model reveal the benefits of the Island model.

Parameters	$\mu(1)$	$\mu(2)$	$\Sigma(1,1)$	$\Sigma(1,2)$	$\Sigma(2,2)$	λ
True Values	0.64	0.61	0.016	0.0007	0.020	200
Island Model Estimates	0.6366 (0.5969, 0.6730)	0.6085 (0.5798, 0.6393)	0.0165 (0.0087, 0.0263)	0.0002 (−0.0069, 0.0062)	0.0148 (0.0060, 0.0228)	199.3371 (168.2863, 233.7119)
Noiseless NHPP Estimates	**0.6073** (0.5786, 0.6368)	0.5985 (0.5735, 0.6237)	**0.0422** (0.0343, 0.0524)	0.0005 (−0.0047, 0.0063)	**0.0313** (0.0257, 0.0384)	**181.0947** (155.4500, 208.2373)

Table 1: Comparison of models with and without measurement error in case of bivariate Gaussian intensity. Point estimates are given with 95% equal tail interval estimates in parentheses.

In particular, the posterior center for $\mu(1)$ is farther from the true value than that for $\mu(2)$. This illustrates the fact that the original sample lost more points in the horizontal direction than in the vertical one. For $\Sigma(1,1)$ and $\Sigma(2,2)$ in the NHPP analysis the 95% posterior intervals are far from the true value. The increased uncertainty in the parameter inference under the Island model is expected due to the added uncertainty associated with the observed locations.

Next, we take f to be a 2-component normal mixture distribution (see Table 2) within

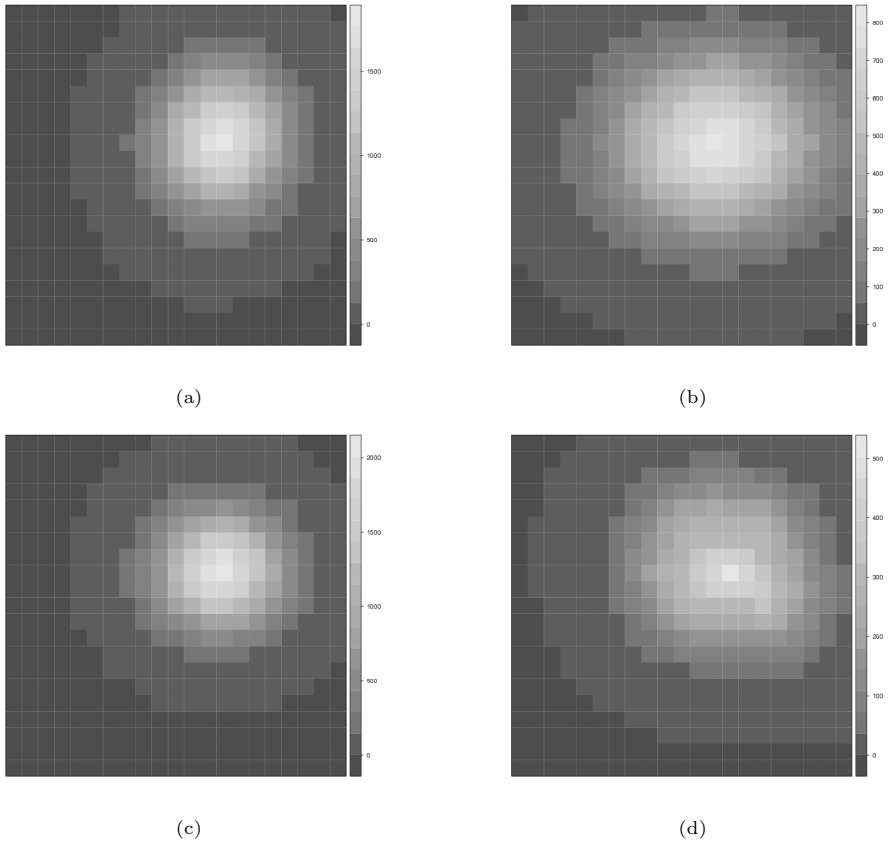

Figure 2: Model Analysis : (a) actual intensity surface, (b) its estimate based on noiseless NHPP, (c) posterior intensity estimate from Island model, (d) uncertainty of estimated intensity

the unit square and contaminate it with Gaussian noise having dispersion matrix as $\left(\begin{smallmatrix} 0.023 & 0.002 \\ 0.002 & 0.019 \end{smallmatrix}\right)$ similar to the previous example. Now, there were 199 points initially in the window, but after noise addition only 177 are left, roughly 11% loss of the points. (The expected fraction of points lost can again be calculated using the intensity and error parameters. It turns out to be 10.48%.) From Figure 3, apart from the increased spread in the noisy pattern, the bimodality of the intensity essentially disappears. Similar to Example 1, we fit the island model as well as the noiseless NHPP. Included in Table 2 is the comparison between the models while Figure 4 provides comparison of the estimated intensities. Again, we see the benefit of the measurement error model. As expected, estimation of the Σ's along with q and λ was severely affected by the noise. The effect on the μ's is noteworthy. Fitting a mixture model directly to that data likely caused the μ's to shift a bit to adjust for the overlap. Also in panels (c)

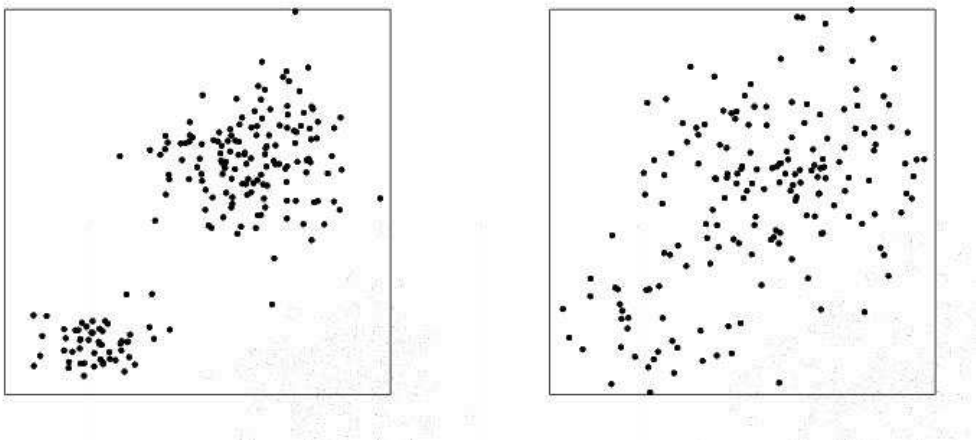

Figure 3: Original (*left*) and Perturbed (*right*) point patterns

and (d), the spatial pattern for the posterior uncertainty follows that of the posterior mean. This is intuitively sensible since, for any areal unit the count of points follows a Poisson distribution where the mean equals the variance. In Table 2 we can see the 95% credible interval for $\mu_1(1)$ produced by the noiseless NHPP excludes the true value (again parameters that noticeably differ are in **bold**).

Parameters	$\mu_1(1)$	$\mu_1(2)$	$\mu_2(1)$	$\mu_2(2)$	q	λ
True Values	0.64	0.61	0.25	0.14	0.71	200
Island Model Estimates	0.6291 (0.5855, 0.6689)	0.5965 (0.5656, 0.6291)	0.2454 (0.1713, 0.3243)	0.1575 (0.0889, 0.2292)	0.7238 (0.6138, 0.8140)	200.7096 (168.4630, 238.9672)
Noiseless NHPP Estimates	**0.6053** (0.5697, 0.6389)	0.5821 (0.5524, 0.6123)	0.2546 (0.2002, 0.3069)	0.1694 (0.1282, 0.2125)	**0.8150** (0.7461, 0.8771)	**177.7389** (152.5286, 204.9620)
Parameters	$\Sigma_1(1,1)$	$\Sigma_1(1,2)$	$\Sigma_1(2,2)$	$\Sigma_2(1,1)$	$\Sigma_2(1,2)$	$\Sigma_2(2,2)$
True Values	0.016	0.0007	0.018	0.007	0.0005	0.002
Island Model Estimates	0.0153 (0.0068, 0.0275)	0.0003 (−0.0060, 0.0081)	0.0116 (0.0040, 0.0206)	0.0105 (0.0038, 0.0176)	0.0004 (−0.0050, 0.0072)	0.0037 (0.0015, 0.0060)
Noiseless NHPP Estimates	**0.0339** (0.0263, 0.0432)	0.0037 (−0.0019, 0.0098)	**0.0271** (0.0207, 0.0352)	**0.0175** (0.0091, 0.0306)	−0.0033 (−0.0093, 0.0017)	**0.0096** (0.0051, 0.0172)

Table 2: Comparison of models with and without measurement error in case of bivariate Gaussian mixture intensity. Again, point estimates with 95% interval estimates in parentheses.

4 The Subregion Model

Here, we allow the possibility that all three errors are potentially present in our data, i.e., shift within D and displacement to and from D.

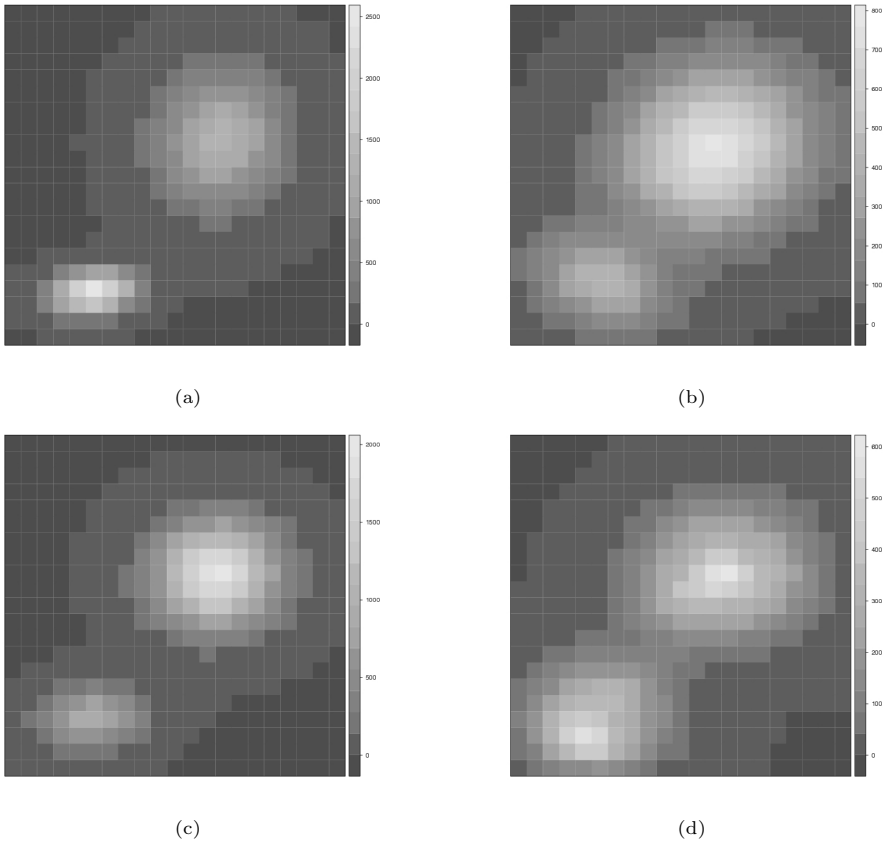

Figure 4: Model Analysis : (a) actual intensity surface, (b) its estimate based on noiseless NHPP, (c) posterior intensity estimate from Island model, (d) uncertainty of estimated intensity

4.1 Model Specification

Our formulation of this problem envisions a larger region $\bar{D} \supseteq D$, such that if a point falls outside $\bar{D}$, then the probability is negligible that the noise can bring it inside D. With our assumed knowledge about the noise dispersion, this can be done, for example by enclosing D in an ellipse of the form $\bar{D} = \{y \in \mathbb{R}^2 : (y-x)^T \Sigma(x)^{-1}(y-x) \leq c\}$ or alternatively employing a big enough rectangle so that in either case, the chance that the noise can take a point inside D outside of that rectangle is sufficiently small.

Given $\bar{D}$, we need to model the intensity surface on it. Our objective is to estimate the

point process intensity only within D and we are introducing $\bar{D}$ only as an artifice; with no observations in $\bar{D} - D$ our prior assumption will drive the intensity over it. Still, the intensity surface needs to be defined on the whole of $\bar{D}$, not just D. Below, we propose a plausible specification, recognizing that the effects of this specification are confounded with the effects of the noise process.

Conditional on n and $f_{\bar{D}}$, the expected number of actual points is

$$n\frac{\int_D f_{\bar{D}}(x)dx}{\int_{\bar{D}} \bar{\Phi}_2(D;x,\Omega)f_{\bar{D}}(x)dx} = n\frac{\int_D f_{\bar{D}}(x)dx}{\int_D \bar{\Phi}_2(D;x,\Omega)f_{\bar{D}}(x)dx + \int_{\bar{D}-D} \bar{\Phi}_2(D;x,\Omega)f_{\bar{D}}(x)dx} \quad (7)$$

If the second integral in the denominator on the right side of (7) is not small, then it can have a consequential effect on our inference. To deal with this problem, we adopt a version of a Neyman-Scott cluster process over all of $\bar{D}$ with the restriction that the cluster centers are in D. Evidently, this allows true locations to be in $\bar{D} - D$. In fact, as we now argue, apart from this restriction, this process model is essentially that of our island model.

Recall that a Neyman-Scott cluster process model (Neyman and Scott 1958; Stoyan 1992) over D is built in 3 stages

(**1**) Generate $K \sim Poi(\lambda)$ and $(\mu_1, \mu_2, ...\mu_k)$ i.i.d $\sim Unif(D)$
(**2**) Conditional on K, generate $N_1, N_2, ..., N_K$ i.i.d $\sim g$
(**3**) Conditional on $N_1, N_2, ..., N_K$, draw $x_1, x_2, ..., x_{N_i}$ i.i.d $\sim h(x; \mu_i, \Sigma^{(0)})$ for $1 \leq i \leq K$

A closer look at these steps reveals that, conditional on step (1), we can rewrite steps (2) & (3) as,
(**2**, **3**)$'$ Conditional on $K, (\mu_1, \mu_2, ...\mu_K)$, generate $N \sim g_K$ and generate $x_1, x_2, ..., x_N$ i.i.d $\sim \sum_{i=1}^{K} \frac{1}{K} h(x; \mu_i, \Sigma^{(0)})$,

(where g_K is the distribution of sum of K i.i.d. variates from g and is easy to obtain when g is a member of the exponential family). This remark connects us to the earlier λf formulation for the intensity surface, with λ being the parameter for g_K and f being the Gaussian mixture, now with a common dispersion structure and equal weights across components. Presuming there are clusters well within the region which have not suffered loss of points because of the noise, we can expect to learn about the common Σ and, hence, learn about clusters close to the boundary which are more affected with regard to loss of points by the noise. Interpreting a Neyman-Scott process through mixtures, it emerges that the only difference between our modeling here and that of the previous section is that, there, K was fixed rather than random. In fact, to simplify model fitting, here, we take K as fixed below, suggesting the same guidance regarding choice as we

did above.

4.2 Computational Details

As before, assume we observe a set of n points $y_1, y_2, ..., y_n$ within D. Actually, there were $m_{\bar{D}}$ points $(x_1, x_2, ..., x_{m_{\bar{D}}})$ within $\bar{D}$, out of which m_D fell in D. By construction of $\bar{D}$, noise displacement can not take a point in D to $\bar{D}^c$ and vice versa. Unlike before, n can be greater or less than m_D, depending on the extent of shift of locations from D to $\bar{D} - D$ and the vice versa. In particular, conceptually, we can imagine b true locations in D that have been displaced to $\bar{D} - D$ and c locations in $\bar{D} - D$ that have been displaced to D with implicit constraints that $0 \leq b \leq min(m_D, m_{\bar{D}} - n)$ and $0 \leq c \leq min(m_{\bar{D}} - m_D, n)$. The net change to D is $c-b$, to $\bar{D}-D$ is $b-c$. Equivalently, $n = m_D + c - b$ and $m_{\bar{D}} - n = m_{\bar{D}} - (m_D + c - b)$. Note that we do not need to identify $c-b$ or m_D. We simply need to connect n of the true x's with observed y's and $m_{\bar{D}} - n$ of the x's with unobserved y's. So, as with the Island model, we can again write the likelihood as product of these two pieces of information,

$$\begin{aligned} L &= L_1 L_2 \\ L_1 &= \prod_{i=1}^{n} \phi_2(y_i|x_i, \Omega, \beta) \\ L_2 &= \prod_{i=n+1}^{m_{\bar{D}}} \bar{\Phi}_2(D; x_i, \Omega, \beta) \end{aligned} \tag{8}$$

Notice that the x corresponding to an observed y may come from any part of $\bar{D}$, so it is again clear that we need to assign a prior intensity surface on the entire $\bar{D}$. As with the Island model, label the x's, so that the first n of them correspond to y's in the same order. Assigning a prior $\lambda_{\bar{D}}(s) = \lambda f_{\bar{D}}(s)$ on all of $\bar{D}$, will render the full conditional distributions,

$$\begin{aligned} \pi(x_i|...) &\propto \phi_2(y_i|x_i, \Omega, \beta) f_{\bar{D}}(x_i),\ i = 1, 2, ..., n \\ \pi(x_i|...) &\propto \bar{\Phi}_2(D; x_i, \Omega, \beta) f_{\bar{D}}(x_i), i = n+1, n+2, ..., m_{\bar{D}} \end{aligned} \tag{9}$$

Thus, for an observed point, the position of x is governed by both the prior intensity and a centering around the observed position. Any y falling well within D generates a posterior for x that has little mass outside of D so x is most likely to be simulated inside D. On the other hand, for a y close to the boundary with regard to measurement error, there will be a considerable chance for that x to be simulated from $\bar{D} - D$, and thus to be identified as an intruding point. Conversely, for x corresponding to a missing point y, $\Phi_2(D; x, \Omega, \beta)$ puts higher weight on the fact that x was either close to boundary of D or was a point outside of D. The relative chances depend on $f_{\bar{D}}$.

Estimation of the μ's and Σ's can again be done using data augmentation. With equal component weights, ordering the first element of the component means identifies the components. Noteworthy is the posterior full conditional for $m_{\bar{D}}$,

$$\begin{aligned}
&\pi(m_{\bar{D}}|\lambda, \mu_{1:k}, \Sigma^{(0)}, y_{1:n}) \\
&\propto \prod_{i=n+1}^{m_{\bar{D}}} \int_{\bar{D}} \Phi_2(\bar{D} - D; x_i, \Sigma) f_{\bar{D}}(x_i) dx_i \, \frac{m_{\bar{D}}!}{(m_{\bar{D}} - n)!} \frac{\lambda^{m_{\bar{D}}}}{m_{\bar{D}}!} \\
&\propto \left(\int_{\bar{D}} \Phi_2(\bar{D} - D; x, \Sigma) f_{\bar{D}}(x) dx \right)^{m_{\bar{D}} - n} \frac{\lambda^{m_{\bar{D}} - n}}{(m_{\bar{D}} - n)!} \qquad (10)
\end{aligned}$$

The only change from equation (6) is that when we integrate out the x's, it has to be over $\bar{D}$. Thus, $(m_{\bar{D}} - n)$ is to be sampled from $Poi(\lambda \int_{\bar{D}} \Phi_2(\bar{D} - D; x, \Sigma) f_{\bar{D}}(x) dx)$, this being the expected number of observations from $\bar{D}$ that are displaced outside D by the noise. The quantity of principal interest to us is m_D or its expectation $\lambda \int_D f_{\bar{D}}(x) dx$, the expected number of observations within D. The posterior distribution of the former can be estimated by obtaining m_D at each step of the MCMC run from the simulated set of x's. The latter can be obtained either by averaging these posterior samples or by inserting posterior estimates of the intensity parameters in the expression for the expectation.

4.3 An Example

As an example, suppose the actual domain of the locations is $\bar{D} = (0, 2.5) \times (0, 2.5)$. However, our interest is to explain the event occurrences within the lower left subregion $D = (0, 2) \times (0, 2)$. Following the above, we take f to be a 3-component normal mixture (Table 3) with uniform weights and common dispersion matrix across all components. Then we add to it Gaussian noise having covariance matrix $\left(\begin{smallmatrix} 0.016 & 0 \\ 0 & 0.007 \end{smallmatrix} \right)$ In our generated sample (Figure 5) there were 991 points in the bigger window out of which 867 fell into the domain of observation, around 87.49%. (The expected fraction of points in D can again be calculated using the intensity and error parameters. It turns out to be 87.94%.) After noise addition, points moved in and out of D yielding 845 observed points.

As before we fitted the subregion model as well as a NHPP without noise to the observed dataset and we compare their performance in Table 3 and Figure 6. (again, parameters that noticeable differ are in **bold**). As in Figure 2 and 4, the noise-free analysis shown in Figure 4.2 produced an intensity surface with larger spread around the modes. Due to loss of points in the x- direction, estimation of $\mu_2(1)$ was most affected. In the case of the noiseless NHPP, its 95% posterior credible interval substantially misses the true value. Also λ_D, the expected frequency of points within D was affected by the noise addition. The NHPP analysis estimates it to be close to the observed number of records. The subregion model accounts for the loss and gain of points and yields an estimate which, though a bit different from the actual value, is much closer to it compared to

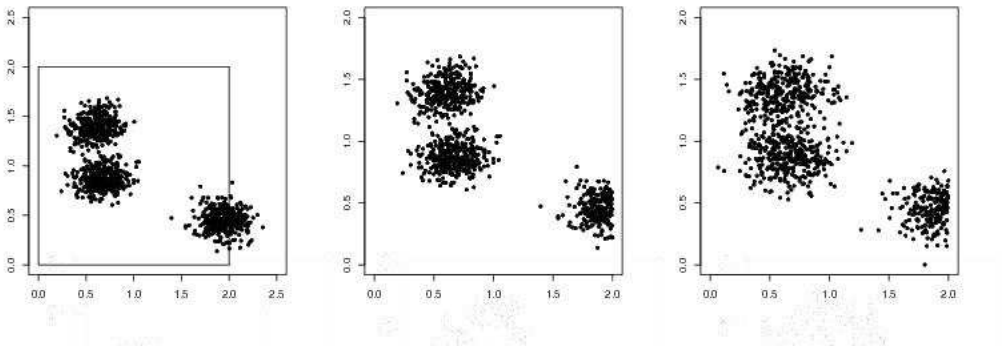

Figure 5: Original locations in whole domain (*left*) and subregion (*center*), Perturbed locations inside subregion (*right*)

the NHPP estimate. Ongoing experimentation with the subregion model will enable us to better assess its performance.

Parameters	$\mu(1)(1)$	$\mu_1(2)$	$\mu_2^{(}1)$	$\mu_2(2)$	λ_D
True Values	0.60	1.40	1.95	0.45	870.60
Subregion Model Estimates	0.6156 (0.5948, 0.6362)	1.3933 (1.3778, 1.4099)	1.9371 (1.9014, 1.9690)	0.4494 (0.4304, 0.4697)	857.9236 (831.4508, 884.0664)
Noiseless NHPP Estimates	0.6158 (0.5959, 0.6355)	1.3927 (1.3766, 1.4094)	**1.8285** (1.8046, 1.8534)	0.4459 (0.4279, 0.4631)	**845.5023** (789.1823, 905.1914)
Parameters	$\mu_3(1)$	$\mu_3(2)$	$\Sigma(1,1)$	$\Sigma(1,2)$	$\Sigma(2,2)$
True Values	0.68	0.86	0.020	0.0007	0.011
Subregion Model Estimates	0.6527 (0.6230, 0.6824)	0.8717 (0.8557, 0.8875)	0.0202 (0.0168, 0.0238)	0.0012 (−0.0007, 0.0034)	0.0102 (0.0086, 0.0121)
Noiseless NHPP Estimates	0.6531 (0.6238, 0.6837)	0.8712 (0.8550, 0.8868)	**0.0323** (0.0292, 0.0355)	0.0013 (−0.0006, 0.0032)	**0.0170** (0.0153, 0.0189)

Table 3: Comparison of models with and without measurement error in case of 3 component subregion model

5 An Ecological Data Application

A long-standing issue in the ecological literature is to learn about the distribution of various species of interest within a particular geographic region. See, e.g., Rosenzweig (1995) and Gaston (2003). In many settings the raw data consist of recorded locations over the region where the species was observed. One area of massive ecological field data collection is the Cape Floristic Region (CFR) in South Africa. Spread over an area of $\sim 9 \times 10^4$ sq.km., this region hosts thousands of plant varieties and is a global hotspot for biodiversity research. Data are collected on presence locations for several of these plant species with the goal of using environmental and soil type factors to explain the

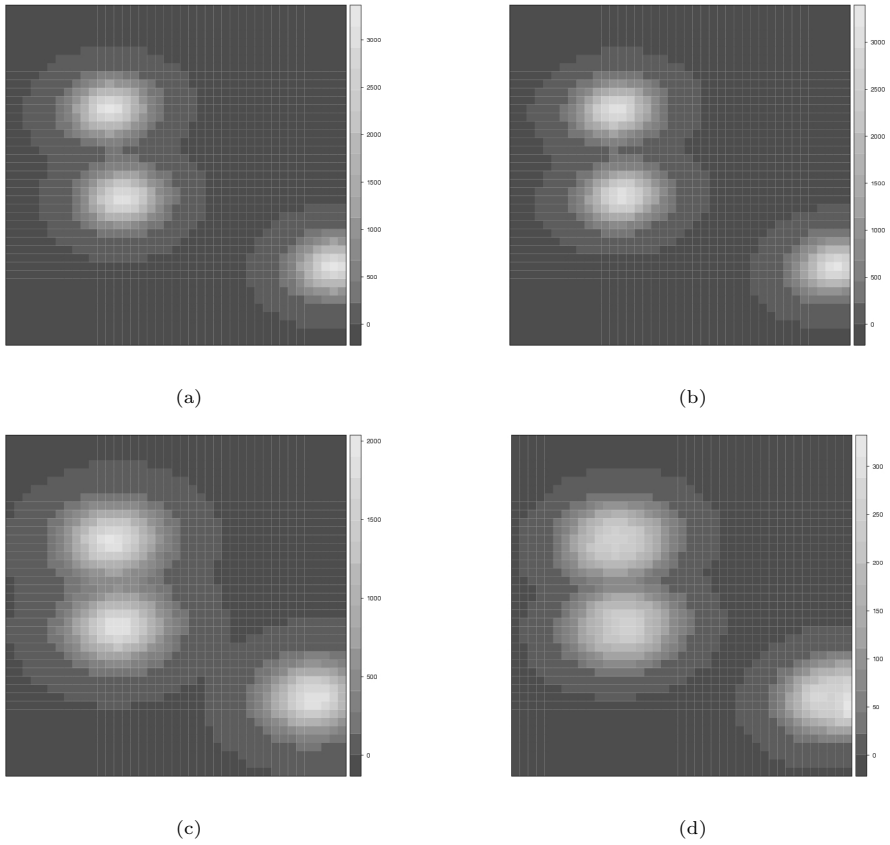

Figure 6: Model Analysis : (a) actual intensity surface, (b) its estimate based on noiseless NHPP, (c) posterior intensity estimate from Subregion model, (d) uncertainty of estimated intensity

observed pattern of locations for each species. The locations of presence are recorded using a global positioning system (GPS) and/or topographical maps, both of which are subject to roughly known degrees of measurement error (the latter even more than the former). Often the point level data are gridded to cell counts in order to model them with covariate information; the cells are determined by the areal resolution at which this information is available. One such modeling example can be found in Gelfand et al. (2005). Hence inaccuracy in recording exact locations can result in a grid level presence summary different from the truth.

Here we use the CFR data with two species, *Mimetes hirtus* (*MIHIRT*) and *Protea cryophila* (*PRCYRO*) in order to explore sensitivity of their abundance patterns with respect to different degrees of assumed noise variability. For *MIHIRT*, we looked

at its observed presence pattern within a rectangular subregion $\mathcal{D}_1 = [18.35, 19.05] \times [-34.36, -34.21]$ and for *PRCYRO*, $\mathcal{D}_2 = [19.145, 19.195] \times [-32.512, -32.357]$. In both cases these regions are defined in terms of degrees of latitude and longitude. The resulting patterns consist of 131 and 51 locations respectively as displayed in Figure 7. We work with the point patterns directly, avoiding gridding.

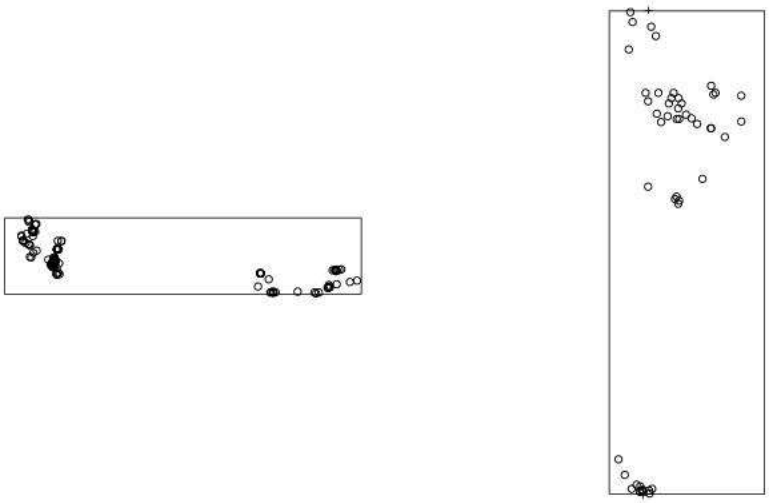

Figure 7: Observed locations for (*left*) MIHIRT within $\mathcal{D}_1$ and (*right*) PRCYRO within $\mathcal{D}_2$

These two species are not prevalent over the CFR; they have small and disjoint ranges and the illustrative subregions offer proposed envelopes for each species. We further assume that these envelops provide *hard* boundaries for the respective species distributions and that we have only sampled within these envelopes. Hence, we adopt the island model to handle measurement error in the presence pattern within $\mathcal{D}_i, i = 1, 2$. In what follows we analyze how the difference between the estimated true abundance and the observed pattern changes with change in noise variability. Along with noise free analysis, measurement error was tried with a scale matrix $\sigma_e^2 I_2$ at three different levels of $\sigma_e = 0.005, 0.010, 0.020$. At the scale of these regions, these correspond to error ranges ($3\sigma_e$) of roughly $27, 54$ and 108 meters in any direction from the true location. These levels are plausible for a field experiment as the first one corresponds to standard

GPS accuracy and the latter two are appropriate for topographical methods where shift may even be well above 100 meters. Each dataset was fitted with a two component bivariate normal mixture as suggested from the observed locations. We provide the posterior estimates of the intensity surfaces under different σ_e in Figures 8 and 9 with parameter summaries for all the models in Table 4.

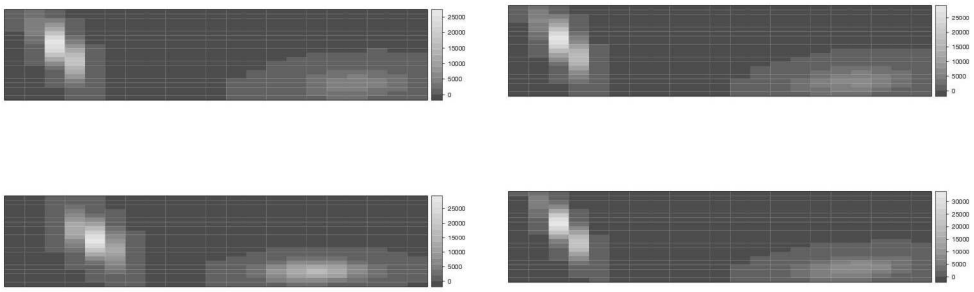

Figure 8: (*clockwise from left*) Change in estimated *MIHIRT* intensity surfaces with increasing $\sigma_e = (0, 0.005, 0.010, 0.020)$

From the figures, higher noise variability produces a tighter estimate of the true intensity surface. This is sensible, because precision for the true surface increases when we back out more noise. Table 4 shows that, with increasing σ_e, interval widths for the intensity parameters are also increasing, which is reasonable to expect. The changes in parameter estimates for varying σ_e are on the order of 10^{-3} or less. At the resolution of minute-by-minute grid cells ($\sim$ 1.5 km $\times$ 1.8 km over the study region) usually used to model this kind of data (Gelfand et al. 2005), this can imply errors of 50 to 100 meters. Table 4 reveals differences in sensitivity of species intensity across the different levels of σ_e. The sensitivity varies across mixture component too. The covariance structure is much more sensitive to change in noise than the location parameters. Mixture weights also change with σ_e. For *MIHIRT*, with increasing σ_e, the first component loses weight. This implies that points are being gained by the other component, revealing its higher sensitivity to error. For *PRCYRO*, moderate measurement error suggests a lower weight for component 1 but at the largest σ_e it shows the opposite trend. Finally, by looking at λ, we see how the expected number of lost points changes with uncertainty. The effect is relatively more pronounced for the less abundant species, *PRCYRO*.

6 Discussion

In this article we have discussed different issues in modeling a noisy point pattern. Additional variability enters into the realization and thus, into the model. This increased variability is reflected in wider posterior credible sets for intensity parameters. However,

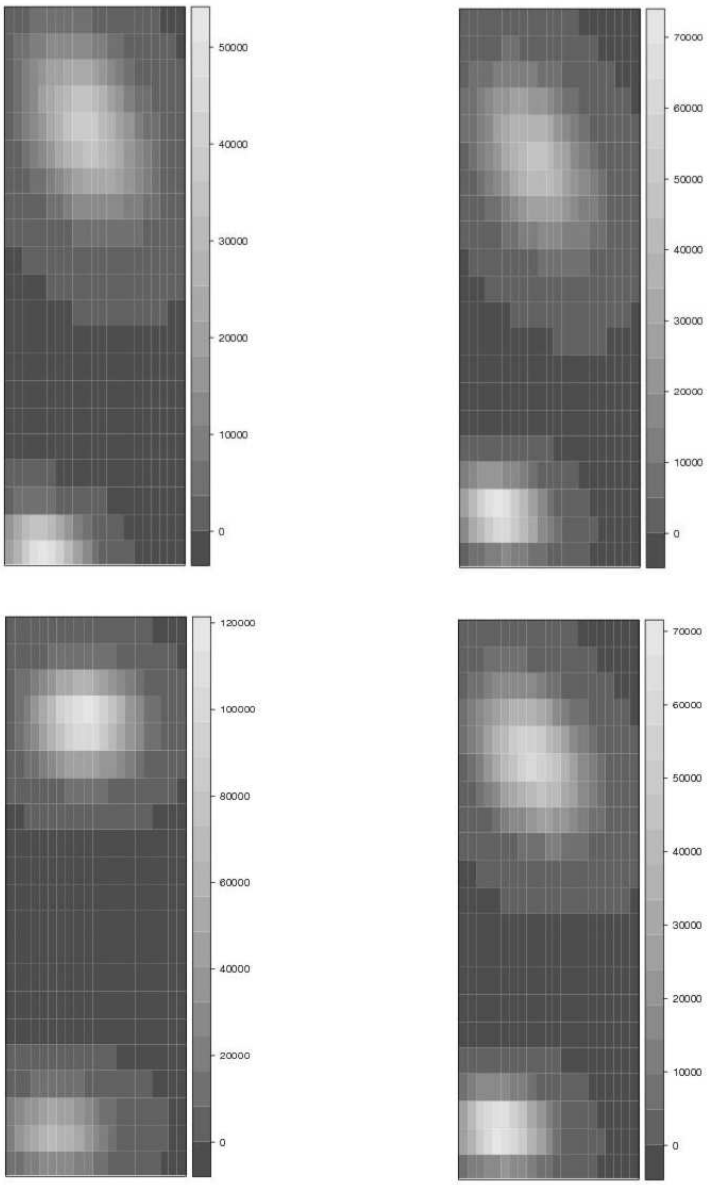

Figure 9: (*clockwise from left*) Change in estimated *PRCYRO* intensity surfaces with increasing $\sigma_e = (0, 0.005, 0.010, 0.020)$

as seen in the synthetic examples earlier, the gain for this added uncertainty is that the posterior centers are closer to the truth than their noiseless counterparts. Again, the λf decomposition of the intensity surface is very general, so one may study different

Parameters	Species							
	MIHIRT				PRCYRO			
σ_e	0	0.005	0.010	0.020	0	0.005	0.010	0.020
$\mu_1(1)$	18.4302 (0.0101)	18.4296 (0.0106)	18.4320 (0.0115)	18.4890 (0.5175)	19.1672 (0.0066)	19.1656 (0.0171)	19.1651 (0.0167)	19.1666 (0.0164)
$\mu_1(2)$	−34.2744 (0.0123)	−34.2734 (0.0135)	−34.2732 (0.0142)	−34.2814 (0.0858)	−32.3886 (0.0111)	−32.3969 (0.1277)	−32.3936 (0.1263)	−32.3831 (0.0158)
$\mu_2(1)$	18.9355 (0.0392)	18.9351 (0.0440)	18.9308 (0.0465)	18.8725 (0.5278)	19.1547 (0.0076)	19.1553 (0.0171)	19.1558 (0.0158)	19.1582 (0.0194)
$\mu_2(2)$	−34.3356 (0.0113)	−34.3384 (0.0121)	−34.3402 (0.0142)	−34.3372 (0.0829)	−32.5101 (0.0063)	−32.4999 (0.1254)	−32.5009 (0.1241)	−32.5027 (0.0130)
$\Sigma_1(1,1)$	0.00061 (0.00035)	0.00062 (0.00041)	0.00058 (0.00044)	0.00080 (0.00389)	0.00010 (0.00009)	0.00010 (0.00014)	0.00009 (0.00013)	0.00009 (0.00012)
$\Sigma_1(1,2)$	−0.00052 (0.00037)	−0.00056 (0.00044)	−0.00058 (0.00050)	−0.00041 (0.00096)	−0.00005 (0.00011)	−0.00005 (0.00014)	−0.00004 (0.00016)	−0.000003 (0.00011)
$\Sigma_1(2,2)$	0.00088 (0.00051)	0.00093 (0.00062)	0.00089 (0.00073)	0.00062 (0.00109)	0.00027 (0.00026)	0.00024 (0.00038)	0.00018 (0.00031)	0.00008 (0.00014)
$\Sigma_2(1,1)$	0.00436 (0.00382)	0.00443 (0.00397)	0.00434 (0.00402)	0.00333 (0.00540)	0.00005 (0.00008)	0.00006 (0.00009)	0.00006 (0.00011)	0.00009 (0.00016)
$\Sigma_2(1,2)$	0.00024 (0.00081)	0.00028 (0.00079)	0.00026 (0.00075)	0.00006 (0.00112)	−0.00000 (0.00005)	−0.00000 (0.00009)	−0.00000 (0.00007)	0.00000 (0.00009)
$\Sigma_2(2,2)$	0.00034 (0.00030)	0.00031 (0.00028)	0.00021 (0.00025)	0.00015 (0.00092)	0.00004 (0.00006)	0.00005 (0.00026)	0.00005 (0.00017)	0.00006 (0.00010)
q	0.6996 (0.1538)	0.6687 (0.1724)	0.6428 (0.2098)	0.5904 (0.2126)	0.7176 (0.2457)	0.6231 (0.2614)	0.6260 (0.2531)	0.6885 (0.2776)
λ	131.7379 (45.0589)	141.0974 (52.8167)	150.5222 (68.6860)	169.4646 (93.7239)	51.8881 (29.8167)	65.8787 (43.4248)	72.3167 (44.7333)	94.0949 (52.3775)

Table 4: Estimated intensity parameters and associated 95% interval width under different scales of error for species data

sorts of specifications for f, as noted in Section 2.

Instead of a single point process realization over a region, we can also think of spatio-temporal point patterns. In that setting, how the measurement error variability and the intensity surface evolve over time are subjects of interest. Another application is to marked point patterns. If the different patterns are affected by the same noise distribution, a joint model for the underlying true locations using interaction/dependence can be investigated. Lastly, as we mentioned above, not knowing the actual locations currently restricts our scope of modeling for the intensity surface, e.g., modeling as a process realization leads to infeasible computation. Finding feasible model fitting methods for this scenario will enable us to investigate wider classes of intensities. We also note that the examples presented here assumed rectangular regions to simplify computing of multi-dimensional integrals on our study domain. For more general regions, numerical integrations will be required. Since, within geographic information system (GIS) software, regions are typically described through polygonal curves, this should facilitate such integration.

Finally, a problem which is similar to ours is that of developing random set models for describing evolution of cells. There, we also have unobserved disc centers modelled as a realization of point process. Instead of modeling $\pi(y|x)$ and $\pi(x)$, where y and x stand for a noisy and a true location respectively, we have the same model structure but with y being an observed point in the disc with center x. Details of such models can be found in Baddeley and Møller (1989).

References

Baddeley, A., Gregori, P., Mateu, J., Stoica, R., and Stoyan, D. (eds.) (2005). *Case Studies in Spatial Point Process Modeling*. Lecture Notes in Statistics, Springer-Verlag, New York, 1st edition. 97

Baddeley, A. and Møller, J. (1989). "Nearest-neighbour Markov point processes and random sets." *International Statistical Review*, 57: 89–121. 120

Baddeley, A. and Van Lieshout, M. N. M. (1993). "Stochastic geometry models in high-level vision." *Journal of Applied Statistics*, 20: 231–256. 99

Barber, J., Gelfand, A. E., and Silander, J. A. (2006). "Modeling map positional error to infer true feature location." *Canadian Journal of Statistics*, 34(4): 659–676. 101

Breslow, N. (1974). "Covariance analysis of censored survival data." *Biometrics*, 30: 89–99. 99

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London, 1st edition. 99

Cressie, N. A. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Inc, New York, 2nd edition. 98

Diebolt, J. and Robert, C. P. (1994). "Estimation of finite mixtures through Bayesian sampling." *Journal of the Royal Statistical Society, Series B*, 56(2): 363–375. 105

Diggle, P. J. (1993). "Point process modeling in environmental epidemiology." In Barnett, V. and Turkman, K. F. (eds.), *Statistics for the Environment*, 89–110. John Wiley and Sons Ltd., Chichester. 99, 101

— (2003). *Statistical Analysis of Spatial Point Patterns*. Arnold, London, 2nd edition. 97

Faÿ, G., González-Arévalo, B., Mikosch, T., and Samorodnitsky, G. (2006). "Modeling teletraffic arrivals by a Poisson cluster process." *Queueing Systems: Theory and Applications*, 54: 121–140. 98

Fuller, W. A. (1987). *Measurement Error Models*. John Wiley & Sons, Inc., New York, 1st edition. 98

Gaston, K. J. (2003). *The Structure and Dynamics of Geographic Ranges*. Oxford University Press, Oxford, 1st edition. 114

Gatrell, A. C., Bailey, T. C., Diggle, P. J., and Rowlingson, B. S. (1996). "Spatial point pattern analysis and its application in geographical epidemiology." *Transactions of the Institute of British Geographers*, 21: 256–274. 97

Gelfand, A. E., Schmidt, A. M., Wu, S., Silander, J. A., Latimer, A., and Rebelo, A. G. (2005). "Explaining species diversity through species level hierarchical modeling." *Journal of the Royal Statistical Society, Series C*, 54: 1–20. 115, 117

Heid, I. M., Küchenhoff, H., Miles, J., Kreienbrock, L., and Wichmann, H. (2004). "Two dimensions of measurement error: classical and Berkson error in residential radon exposure assessment." *Journal of Exposure Analysis and Environmental Epidemiology*, 14(5): 365–377. 101

Helmers, M. and Bunke, H. (2003). *Generation and use of synthetic training data in cursive handwriting recognition*, volume 2652. Springer, Lecture Notes in Computer Science, Berlin/Heidelberg. 101

Ji, C., Merl, D., Kepler, T. B., and West, M. (2009). "Spatial mixture modelling for unobserved point processes: Application to immunofluorescence histology." *Bayesian Analysis*, 4: 297–316. 103

Kalbfleisch, J. D. (1978). "Nonparametric Bayesian analysis of survival time data." *Journal of the Royal Statistical Society, Series B*, 40: 214–221. 99

Kottas, A. and Sansó, B. (2007). "Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis." *Journal of Statistical Planning and Inference*, 137(10): 3151–3163. 103

Lund, J., Penttinen, A., and Rudemo, M. (1999). "Bayesian analysis of spatial point patterns from noisy observations." Technical report, Department of Mathematics and Physics, The Royal Veterinary and Agricultural University, Copenhagen. 99, 100, 102, 104

Lund, J. and Rudemo, M. (2000). "Models for point processes observed with noise." *Biometrika*, 87(2): 235–249. 99, 100, 102, 104

Mallick, B. and Gelfand, A. E. (1995). "Bayesian analysis of semiparametric proportional hazards models." *Biometrics*, 51: 843–852. 98

Møller, J. and Waagepetersen, R. P. (2002). "Statistical inference for Cox processes." In Lawson, A. B. and Denison, D. (eds.), *Spatial Cluster Modeling*, 37–60. Chapman and Hall/CRC. 98

Neyman, J. and Scott, E. L. (1958). "Statistical approaches to problems of cosmology (with discussion)." *Journal of Royal Statistical Society, Series B*, 20: 1–43. 98, 111

Richardson, S. (1995). "Measurement Error." In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.), *Markov chain Monte Carlo in Practice*, 401–418. Chapman and Hall, Boca Raton. 101

Richardson, S. and Gilks, W. R. (1993). "A Bayesian approach to measurement error problems in epidemiology using conditional independence models." *American Journal of Epidemiology*, 138(6): 430–442. 98

Ripley, B. D. (1977). "Modeling spatial patterns. (with discussion)." *Journal of Royal Statistical Society, Series B*, 39(2): 172–212. 99

Rosenzweig, M. L. (1995). *Species Diversity in Space and Time*. Cambridge University Press, Cambridge, 1st edition. 114

Sinha, D. K. and Dey, D. K. (1997). "Semiparametric Bayesian analysis of survival data." *Journal of the American Statistical Association*, 92: 1195–1212. 99

Stefanski, L. A. and Carroll, R. J. (1987). "Conditional scores and optimal scores for generalized linear measurement-error models." *Biometrika*, 74: 703–716. 98

Stephens, M. (2000). "Dealing with label switching in mixture models." *Journal of the Royal Statistical Society, Series B*, 62(4): 795–809. 105

Stoyan, D. (1992). "Statistical estimation of model parameters of planar neyman-scott cluster processes." *Metrika*, 39: 67–74. 111

Van Lieshout, M. N. M. (2000). *Markov point processes and their applications*. Imperial College Press, London, 1st edition. 102

Zimmerman, D. L. (2008). "Estimating the intensity of a spatial point process from locations coarsened by incomplete geocoding." *Biometrics*, 64: 262–270. 99

Acknowledgments

The authors wish to thank Andrew Latimer and Anthony Rebelo for providing the datasets as well as helpful feedback about the application in section 5. This work was supported in part by NSF DEB 0516198.