Estimating Abundance from Counts in Large Data Sets of Irregularly-Spaced Plots using Spatial Basis Functions

October 12, 2014

Abstract

Monitoring plant and animal populations is an important goal for both academic research and management of natural resources. Successful management of populations often depends on obtaining estimates of their mean or total over a region. The basic problem considered in this paper is the estimation of a total from a sample of plots containing count data, but the plot placements are spatially irregular and non randomized. Our application had counts from thousands of irregularly-spaced aerial photo images. We used change-of-support methods to model counts in images as a realization of an inhomogeneous Poisson process that used spatial basis functions to model the spatial intensity surface. The method was very fast and took only a few seconds for thousands of images. The fitted intensity surface was integrated to provide an estimate from all unsampled areas, which is added to the observed counts. The proposed method also provides a finite area correction factor to variance estimation. The intensity surface from an inhomogeneous Poisson process tends to be too smooth for locally clustered points, typical of animal distributions, so we introduce several new overdispersion estimators due to poor performance of the classic one. We used simulated data to examine estimation bias and to investigate several variance estimators with overdispersion. A real example is given of harbor seal counts from aerial surveys in an Alaskan glacial fjord.

KEY WORDS: sampling, change-of-support, spatial point processes, intensity function, random effects, Poisson process, overdispersion

1 Introduction

Monitoring plant and animal populations is an important goal for both academic research and management of natural resources. Successful management of populations often depends on estimates of their mean or total over a region. Historically, this has been the purview of sampling theory using simple random sampling, stratified random sampling, etc., which are design-based methods. For design-based methods, sample units are chosen at random, measurements are made or observed from the sample units, and inference is derived from the inclusion probability for sample units (i.e., Horwitz-Thompson estimation). For overviews, see Cochran (1977) or Thompson (1992). An alternative approach developed in the early 1960's called geostatistics includes methods such as block kriging (Gandin, 1959, 1960; Matheron, 1963), which also estimates a regional total. These methods rely on an assumption about a stochastic process that generated the realized observations, and are hence "model-based." Model-based inference relies on estimating parameters for the assumed model, and then forming probability statements (confidence intervals, prediction intervals, etc.) from the fitted model. In this paper we pursue the model-based approach because samples cannot always be drawn randomly. In particular, we consider counts from aerial photographs, which are difficult and inefficient to randomize, with the basic problem being the estimation of the total count for a region.

The goals and context of this paper are shown in Figure 1. The situation for block kriging is shown in Figure 1A. Let the spatial region of interest be R. Any particular location in R is given by x- and y-coordinates contained in the vector $\mathbf{s} = [s_x, s_y]'$, and the random variable at the ith location is denoted $Y(\mathbf{s}_i)$. We assume a spatial random field $\{Y(\mathbf{s}) : \mathbf{s} \in R\}$ (Cressie, 1993, pg. 30). The set $\{Y(\mathbf{s})\}$ is continuous in space and hence infinite. We use the notation that for vector $\mathbf{x} = [x_1,\ x_2]'$, $\|\mathbf{x}\| \equiv \sqrt{x_1^2 + x_2^2}$, so $\|\mathbf{s}_i - \mathbf{s}_j\|$ is Euclidean distance between $\mathbf{s}_i$ and $\mathbf{s}_j$. If the correlation between $Y(\mathbf{s}_i)$ and $Y(\mathbf{s}_j)$ goes to one as $\|\mathbf{s}_i - \mathbf{s}_j\| \to 0$, then $\{Y(\mathbf{s}_i)\}$ will form a smooth (differentiable) but random surface. Suppose that n observed values from the random surface are contained in the vector $\mathbf{y} = [y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)]'$ (the solid circles in Figure 1A). Block kriging uses a linear combination $\boldsymbol{\lambda}'\mathbf{y}$ to predict the average or total in block A; $Y(A) \equiv \int_A Y(\mathbf{u})d\mathbf{u}$, where this integral is assumed to exist (Yaglom, 1962, pg. 23; Cressie, 1993, pg.106). The salient feature of block kriging is that a model of the autocorrelation of the spatial random field can be estimated from the point level data of the observations, and it is relatively easy to aggregate (through the integral) for an estimator of block A. However, extensions to count

data have been difficult because data are modeled on a transformed space but the integral is desired on the original space (e.g., see Cressie, 1993, p. 286). For example, Christensen and Waagepetersen (2002); Wikle (2002); Monestiez, Dubroca, Bonnin, Durbec, and Guinet (2006) develop maps from count data but do not attempt abundance estimates.

Counts are often obtained from plots, B_i in Figure 1B, that have substantial area, are in a regular grid, and exhaustively fill both R and A. Here, classical random sampling using design-based inferences are often employed, where a random sample of the observed count on the ith block, $y(B_i)$, are used to estimate a total. To correctly estimate variance, a finite population correction factor is employed, $\hat{\text{var}}(\hat{Y}(A)) = (\hat{\sigma}^2/n)(1-f)$ where $\hat{\sigma}^2$ is the sample variance, $f = n/N$ is the fraction of sampled units (n) to total sample units (N) within A, and $1-f$ is called the finite population correction factor. A model-based, finite-population version of block kriging for the situation shown in Figure 1B was developed by Ver Hoef (2000, 2008). In the strict sense, distance between samples is not well-defined because of the non-point nature of each sample B_i. However, models of autocorrelation are often built in this case by using the centroid of each plot. The main problem considered by Ver Hoef (2000, 2008) was that the traditional formulation of block kriging as shown in Figure 1A assumed an infinite population. Hence, if one were to estimate an autocorrelation model for Figure 1B, and then apply standard block kriging formulas (as has been done in the literature), there is no finite correction factor. For example, if we sampled all of the plots, then the prediction variance should be zero, but the traditional formulation of block kriging would have nonzero prediction variance. Hence, a finite population version was developed by Ver Hoef (2000, 2008), and it performed well and had proper confidence intervals in a variety of situations (Ver Hoef, 2002).

Now consider the situation in Figure 1C. Here, we would like to use counts from samples with substantial area, $y(B_i)$, to predict at $Y(A)$, but the sample units are not arranged in a regular grid that fills either R or A. Classical random sampling would usually be employed in this situation, with an estimator of the total being $|A|\bar{Y}(B)$, where $\bar{Y}(B)$ is the total count divided by the sampled area, and the variance estimated by $\text{var}(\hat{Y}(A)) = (\sigma^2/n)(1-f)$, where here $f = |a|/|A|$, with $|a|$ being the total area sampled. However, what if the samples $Y(B_i)$ are not randomly placed, and may in fact be in some regular pattern that does not form a regular grid that does not exhaustively fill both D and A? The basic problem considered in this paper is the prediction of a total in region A from a sample of $\{Y(B_i)\}$ as in Figure 1C, where sampling at random is not possible. This is the case for counts from photographs taken from aircraft. The National Marine Mammal Laboratory

of the NOAA-NMFS Alaska Fisheries Science Center developed aerial survey methods to estimate and monitor harbor seal populations in glacial fjords in Alaska. We give more details next as a motivating example.

1.1 Motivating example

To make the problem concrete, we consider an example that prompted the model development. Aerial surveys were flown over the ice haul-out area of harbor seals in Icy Bay, Alaska. Ice emanating from tidewater glaciers provides a dynamic expanse of floating ice on which the seals whelp and nurse pups and rest during the molting season. The aerial platform, a twin-engine Aero Commander Shrike, was flown at 1000 ft and ca. 100 knots on transects with variable spacing that were oriented in two main directions to sample the two main arms of the bay (Figure 2), covering about 79 sq km. A vertically-mounted camera (Nikon D1X with a 60 mm lens) captured an image approximately every 2 seconds through a portal, each covering about 80 $\times$ 120 m at the surface of the water. This firing rate, and the spacing of the transects, allowed for a gap between images of about 30 m end-to-end, and transects varied in spacing from side-to-side, but largely ensured that images were separated from each other; i.e., seals were sampled only once. The camera was usually turned off when flying over large areas of open water where hauled out seals would necessarily be absent. This survey was conducted on 20 May, 2004, in the afternoon (1300 to 1430 hr) when seals typically haul out in peak numbers. This is just one data set collected among dozens annually as part of a monitoring program for harbor seals.

Images were georeferenced and embedded as a raster layer in an ArcGIS (ESRI, 2009) project allowing individual seals to be spatially marked in a point layer by visually inspecting each image ($n = 2080$ images). Footprints showing the extent of each image were generated as polygons (Figure 2) in a separate layer and seal points were summed within and assigned to each centroid and exported for statistical analysis. The spatial extent of each image was assumed to be constant despite small random variation in altitude (max: $\pm$ 30 m) during the survey. The total spatial extent of each days survey effort, over which the intensity surface would be calculated, was delineated by creating a polygon that corresponded to: 1) the coastline of the bay (shorelines from Alaska Department of Natural Resources line shapefiles), 2) an estimate of the location of the face of the glaciers (by connecting points that marked the glacial terminus in each of the northernmost images from every transect, and 3) the extent of the ice field defined as the edge of the images where ice cover (by area)

dropped to $< 5\%$ by visual estimation. Areas of open water ($< 5\%$ cover) were delineated by donut-holes in the overall polygon where the spatial boundaries was defined by the outermost images in which ice cover increased to $\geqslant 5\%$. In other words, to minimize problems with selection bias, any area that could not contain a seal was eliminated by creating the proper boundary.

The 2080 images covered 25.3% of the study area. Of the 2080 images, 180 of them had nonzero counts, so about 91% were zero. A total of 1002 seals were observed in the 180 plots. A maximum of 44 seals were counted in a single photograph. The data are summarized spatially in Figure 2.

1.2 Previous work

Most of the previous work in this area has used Bayesian models. The literature has concentrated on producing smoothed maps of relative abundance, although going from those smooth maps to an abundance estimate would not seem difficult. In particular, Wikle (2002) developed Poisson-lognormal models for a continuous surface, but the counts were at a scale that could be considered points, whereas we have counts in plots with substantial areas. Thogmartin et al. (2004) used a tessellation to create spatial conditional autoregressive (CAR) models (Besag, 1974) for neighbors as a spatial random effects model, with the CAR random effects constant within the tessellation, but the model retained point level data for covariates. Royle et al. (2007) model on a subsample of a systematic grid, and include detection models, but ultimately use a continuous Poisson-lognormal model for counts. Barber and Gelfand (2007) also use Poisson-lognormal models with known covariates to model the intensity surface. Note that none of these methods include a finite population correction factor, and while they are all very attractive, we do not adopt any of them for reasons that we describe next.

1.3 Goals and Organization

Based on this introduction, we desire a total abundance estimator from data like the motivating example that will satisfy several practical conditions. 1) It must be fast to compute, robust, and require few modeling decisions, similar to classical survey methods. Annually, we compute dozens of estimates for data like the example and, depending on the size of the fjord, each may have thousands of photographs. 2) The estimator must use only counts within plots; actual spatial locations of animals are unknown. 3) We are interested in the

actual number of seals, not the mean of some assumed process that generated the data. Thus, the estimator must make use of the actual number of seals, and predict to those areas that are unsurveyed. 4) The variance estimator should have a population correction factor that shrinks to zero as the proportion of the study area that gets sampled goes to one. In our real example, we surveyed approximately 26% of the study area; in classical sampling, that directly reduces that variance by 26%. Some fjords that we have surveyed are up to 50% sampled. 5) The estimator should be approximately unbiased (demonstrated through simulations), and we want valid confidence intervals that cover the true number of seals the correct proportion of times; that is, we use this method dozens of times per year and desire confidence intervals in the frequentist sense. 6) It appears that there is nonstationary variance throughout the area, with large areas of zero counts (no seals). A variance estimator that accommodates this will be required. The goals of this manuscript are to develop an estimator to satisfy these criteria.

The rest of this paper is organized as follows. The estimator is developed from models for spatial point processes and generalized linear mixed models, so we begin with a brief review and then develop the estimator in Section 2. We provide some simulations to validate the estimator and compare variance estimators in Section 3. In Section 4 we use the estimator on the motivating example of aerial photographs taken of harbor seals in a glacial fjord in Alaska. We provide some concluding remarks in Section 5.

2 Model development

From the Introduction and motivating example, note that data are observed at an aggregated support level, but we need a model at the point support level. The reason should be clear; because of the possibility of unbalanced spatial sampling (Figure 1C), and a real example of it (Figure 2), we need to predict and then integrate an abundance density surface continuously throughout the unsampled area. To achieve this, we develop a model-based estimator motivated by an inhomogeneous point process (IPP) that has been integrated to yield a Poisson regression model. Part of attraction of this framework is that it allows inference on point level support from data on areal support, and then we use the point level support model to make our abundance estimate. We begin with a brief review of the IPP, describe how abundance is related to the intensity surface, and then draw the connection to Poisson regression.

2.1 Inhomogeneous point process model

Assume that locations $\mathbf{s} = [s_x, s_y]'$ of all individuals in A, say $\mathcal{S}^+ = (\mathbf{s}_1, \ldots, \mathbf{s}_N)$, are the result of an inhomogeneous Poisson process (IPP) with intensity function $\lambda(\mathbf{s}|\boldsymbol{\theta})$ that varies with $\mathbf{s}$, where $\boldsymbol{\theta}$ is a vector of parameters controlling the intensity function. The intensity function is defined as

$$\lambda(\mathbf{s}) = \lim_{|d\mathbf{s}|\to 0} \frac{E\left(T(d\mathbf{s})\right)}{|d\mathbf{s}|},$$

where $E(\cdot)$ is expectation, $T(R)$ is the total number of points in planar region R, and $|R|$ is the area of R. In general, when analyzing IPP data, all of the individuals would be located within A, and then inference about $\boldsymbol{\theta}$ could be made by maximizing the point process log-likelihood (e.g., Cressie, 1993, p. 655). However, this is difficult in our case because A cannot be surveyed in its entirety and individual locations are unknown. For example, a simulated point pattern is shown in Figure 3, and the plots form a disjointed window on the point pattern that is masked in the areas between the plots, and although we show the point locations within plots, we assume we only have a count for each plot. Ultimately, we will need to estimate the intensity function, but for now we proceed assuming that we have an estimate of the intensity function in the area between plots, as shown in Figure 3.

2.2 Estimating abundance

The primary quantity of interest is the abundance in a particular block $A \subseteq R$. Because the distribution of individuals is random under the model-based paradigm, there are two types of abundance to consider. First is the *expected* abundance in A, $\mu(A) = \int_A \lambda(\mathbf{u}|\boldsymbol{\theta})d\mathbf{u}$, and then there is the *realized* abundance $T(A)$ for a given realization of $\mathcal{S}^+$ from $\lambda(\mathbf{s}|\boldsymbol{\theta})$. Assuming an inhomogeneous point process, $T(A) \sim \text{Poi}(\mu(A))$, which is Poisson distribution with mean $\mu(A)$. An estimate of the *expected* abundance is $\hat{\mu}(A) = \int_A \hat{\lambda}(\mathbf{u}|\boldsymbol{\theta})d\mathbf{u}$, which is often based on plug-in methods from estimates $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$; i.e., $\hat{\mu}(A) = \int_A \lambda(\mathbf{u}|\hat{\boldsymbol{\theta}})d\mathbf{u}$.

For an estimate of the realized abundance, consider that the total abundance can be partitioned into observed and unobserved. Assume there are n sample units $B_i \in A$, and let $\mathcal{B} = \cup_{i=1}^n B_i$. Note that some B_i could be outside A but within R. It is also possible that some B_i straddle the boundary of A, though we will not consider that problem here. The region within A that was not sampled is $\mathcal{U} \equiv \overline{\mathcal{B}} \cap A$, where $\overline{\mathcal{B}}$ is the complement of $\mathcal{B}$. Then $T(\mathcal{B})$ is the number of observed points and $T(\mathcal{U})$ the number of unobserved points and $T(A) = T(\mathcal{B}) + T(\mathcal{U})$. The total $T(A)$ involves predicting $T(\mathcal{U}) \sim \text{Poi}(\mu(\mathcal{U}))$, where

$\mu(\mathcal{U}) = \int_{\mathcal{U}} \lambda(\mathbf{u}|\boldsymbol{\theta})d\mathbf{u}$. By substituting $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$, one can use the estimator $\hat{\mu}(\mathcal{U}) = \int_{\mathcal{U}} \lambda(\mathbf{u}|\hat{\boldsymbol{\theta}})d\mathbf{u}$, and, without any observations from $\mathcal{U}$, we use the mean as a predictor $\widehat{T}(\mathcal{U}) = \hat{\mu}(\mathcal{U})$. Hence an estimator of the total is

$$\widehat{T}(A) = T(\mathcal{B}) + \widehat{T}(\mathcal{U}), \tag{1}$$

for making inference to the realized abundance $T(A)$. Note that, as $T(\mathcal{B}) \to T(A)$, then $\widehat{T}(A) \to T(A)$, so an estimator of this form satisfies condition 3 in Section 1.3. Making such inferences involves first estimating the intensity function $\lambda(\mathbf{s}|\boldsymbol{\theta})$, and also incorporating the uncertainty of estimating $\lambda(\mathbf{s}|\boldsymbol{\theta})$. So our immediate goal is to infer $\lambda(\mathbf{s}|\boldsymbol{\theta})$ from data on areal support, which we describe in the next sections.

2.3 From IPP to Poisson Regression

Suppose that we have a smooth spatial surface $\lambda(\mathbf{s}|\boldsymbol{\theta})$ that varies with spatial location $\mathbf{s}$ and is controlled by parameters $\boldsymbol{\theta}$; this is the intensity surface. This surface may be integrated over some compact region, such as the plot B_i, and this forms the mean of an IPP. Let $Y(B_i)$ be a random variable for a count in B_i, then $Y(B_i) \sim \text{Poi}(\mu(B_i))$, where

$$\mu(B_i) = \int_{B_i} \lambda(\mathbf{u}|\boldsymbol{\theta})d\mathbf{u}. \tag{2}$$

Now, let $\mathbf{s}_i$ be the centroid of plot B_i. If the area of B_i is small compared to the survey area A, and if $\lambda(\mathbf{u}|\boldsymbol{\theta})$ is smooth (i.e., changing slowly within B_i), then Berman and Turner (1992) show that a reasonable approximation for (2) is,

$$\mu(B_i) = |B_i|\lambda(\mathbf{s}_i|\boldsymbol{\theta}), \tag{3}$$

where $|B_i|$ is the area of B_i. This is an important assumption and is part of the general problem of change-of-support; see Gotway and Young (2002), Banerjee, Carlin, and Gelfand (2004, Chapter 6) and Wikle and Berliner (2005). For example, Brillinger (1990, 1994) shows an early attempt at creating a continuous surface from count data in census tracts.

The mean of $Y(B_i)$ can then be modeled with a log link function, forming a GLM with offset $\log(|B_i|)$,

$$\log(\mu(B_i)) = \log(|B_i|) + \log(\lambda(\mathbf{s}_i|\boldsymbol{\theta})).$$

Now we use spatial radial-basis functions to model $\lambda(\mathbf{s}_i|\boldsymbol{\theta})$. Let $\mathbf{s}_i$ be the centroid of plot B_i.

Then

$$\log(\lambda(\mathbf{s}_i|\boldsymbol{\theta})) = \beta_0^* + \mathbf{z}(\mathbf{s}_i)'\boldsymbol{\gamma}, \tag{4}$$

where $\mathbf{z}(\mathbf{s}_i)$ is a vector of covariates at location $\mathbf{s}_i$ and $\boldsymbol{\gamma}$ is a parameter vector of fixed effects. The spatial basis functions will form the values of $\mathbf{z}(\mathbf{s}_i)$.

There has been increasing interest lately in spatial models that use radial basis functions. Suppose there is a set of fixed points in the study area, $\{\boldsymbol{\kappa}_j; j = 1, \ldots, K\}$, called "knots." Let $\mathbf{z}(\mathbf{s}_i)'$ be a row vector where the jth item contains a radial basis function value $C(\|\mathbf{s}_i - \boldsymbol{\kappa}_j)\|; \rho)$. For example, we will use $C(h; \rho) = \exp(-h^2/\rho); \rho > 0$, which is a Gaussian basis function. A flexible surface is created by taking a linear combination of the radial basis functions. The surface value at location $\mathbf{s}_i$ depends on parameters $\boldsymbol{\gamma}$ and ρ as $\mathbf{z}_\rho(\mathbf{s}_i)'\boldsymbol{\gamma}$, and we attach the subscript to show that values in $\mathbf{z}$ depend on ρ. Using radial basis functions can be viewed as a semiparametric approach to spatial modeling (Ruppert, Wand, and Carroll, 2003), and they have been used for models with non-Euclidean distance measurements (see, e.g., Wang and Ranalli, 2007) and for computational efficiency for large data sets (Cressie and Johannesson, 2008).

To make the model more flexible, following Cressie and Johannesson (2008), we considered radial basis functions at two scales. Let the "coarse" scale knots be $\{\boldsymbol{\kappa}_{C,j}; j = 1, \ldots, K_C\}$. Let the fine scale knots be $\{\boldsymbol{\kappa}_{F,j}; j = 1, \ldots, K_F\}$, where generally $K_F \geqslant 4K_C$. Note that Cressie and Johannesson (2008) use 3 scales with approximately 3 times as many knots at the next finer scale. Here, because we only have two scales, we use 4 times as many knots at the finer scale. The knots are generally spread out more or less regularly throughout the study area; more details on an algorithm for knot locations are given in Section 2.4.

Consider the log-linear model

$$\log(\boldsymbol{\lambda}) = \mathbf{X}\boldsymbol{\theta} = \mathbf{W}\boldsymbol{\beta} + \mathbf{Z}_C\boldsymbol{\gamma}_C + \mathbf{Z}_F\boldsymbol{\gamma}_F, \tag{5}$$

where $\mathbf{X} = [\mathbf{W}|\mathbf{Z}_C|\mathbf{Z}_F]$, $\boldsymbol{\theta} = [\boldsymbol{\beta}', \boldsymbol{\gamma}_C', \boldsymbol{\gamma}_F']'$, $\mathbf{Z} = [\mathbf{Z}_C|\mathbf{Z}_F]$, $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_C|\boldsymbol{\gamma}_F]$ and the jth column of $\mathbf{Z}_C$ has $C(\|\mathbf{s}_i - \boldsymbol{\kappa}_{C,j})\|; \rho_C)$ as the ith element, and the jth column of $\mathbf{Z}_F$ has $C(\|\mathbf{s}_i - \boldsymbol{\kappa}_{F,j})\|; \rho_F)$ as the ith element. We will not consider any covariates in our model, allowing all spatial variation to be modeled through the spatial basis functions, although future development could easily accommodate covariates here. From (4), we only consider an overall constant, $\mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} = \beta_0$. Also, we assume all plots are the same size, $|B_i| = |B|$. Then we can write,

$$\log(\mu(B_i)) = \beta_o + \log(|B|) + \mathbf{z}(\mathbf{s}_i)'\boldsymbol{\gamma}, \tag{6}$$

where $\mathbf{z}(\mathbf{s}_i)$ is the ith row of $\mathbf{Z}$. Model (6) will form the basis for estimation and prediction throughout the rest of this paper.

2.4 Knot Selection

To place coarse scale knots, a systematic grid of points was generated within A, and K-means clustering (MacQueen, 1967) on the coordinates was used to create K_C groups. Because K-means clustering minimizes within-group variance while maximizing among-group variance, the centroid of each group tends to be regularly spaced; i.e., it is a space-filling design that can work well when the region A has an irregular boundary, as in our example data set (Section 1.1). We also used K-means clustering placing K_F fine scale knots, but the systematic grid was generated within a minimum convex polygon that contained all non-zero counts intersected with A; this polygon was defined on the centroids of plots with nonzero counts, and an example can be seen in Figure 3. We found that this helped ensure convergence of the algorithm. If there are too many basis functions with a small range centered in a large area that is all zeros, the fitting algorithm that we describe next would fail to converge. The effect of knot numbers, both K_C and K_F, are examined in the simulation experiments in Section 3. Other methods for spatial knot placement could be used; for example see Nychka and Saltzman (1998). The software PROC GLIMMIX (SAS Institute Inc, 2008) generates spatial knots using vertexes of a k-d tree (Friedman et al., 1977). Regarding the number of spatial knots, Ruppert et al. (2003, pg. 255) recommend $K_C + K_F = n/4$, with no less than 20 and no more than 150.

2.5 Parameter Estimation

Recall that the ith plot B_i is very small in relation to A, and we let $\mathbf{s}_i$ be the centroid of the ith plot. The count in the ith plot is random, denoted $Y(B_i)$, and starting from an inhomogeneous Poisson process, from (6) we assume that $Y(B_i)$ has a Poisson distribution with mean $\mu(B_i) = \exp(\beta_o + \mathbf{z}(\mathbf{s}_i)'\boldsymbol{\gamma})$. This is Poisson regression, more generally formed as a generalized linear model (GLM) (McCullagh and Nelder, 1989),

$$E(\mathbf{Y}|\boldsymbol{\theta}) = g^{-1}(\mathbf{X}\boldsymbol{\theta}) = g^{-1}(\boldsymbol{\eta}) = \boldsymbol{\mu}, \tag{7}$$

where $\mathbf{Y} = (Y(B_1), \ldots, Y(B_n))$, and $\mathbf{X}$ and $\boldsymbol{\theta}$ were defined following (5). Conditional on fixed $\boldsymbol{\rho}$ values contained in the $\mathbf{Z}$ part of $\mathbf{X}$, iteratively weighted least squares (IWLS)

(Nelder and Wedderburn, 1972) provides maximum likelihood estimation for $\boldsymbol{\theta}$. Recall that the negative log likelihood for Poisson regression is

$$\ell(\boldsymbol{\rho}, \boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^{n} |B_i| \exp(\mathbf{x}_{\boldsymbol{\rho}}(\mathbf{s}_i)'\boldsymbol{\theta}) - y_i \log|B_i| - y_i \mathbf{x}_{\boldsymbol{\rho}}(\mathbf{s}_i)'\boldsymbol{\theta}, \tag{8}$$

where $\mathbf{x}_{\boldsymbol{\rho}}(\mathbf{s}_i)$ is the ith row of $\mathbf{X}$ in (5) with the specific case being (6). Here, we show the dependence of that row on $\boldsymbol{\rho}$ values. An iterative algorithm using block-wise coordinate descent for minimizing the negative likelihood is,

- condition on $\boldsymbol{\rho} = [\rho_C, \rho_F]'$ and use IWLS to estimate $\boldsymbol{\theta}$,
- embed the IWLS estimation in a numerical optimization of (8) for $\boldsymbol{\rho}$.

This optimization routine over just two parameters, $\boldsymbol{\rho}$, converges quickly and can use existing Poisson regression software for the IWLS update, so it satisfies the speed requirement of condition 1 in Section 1.3. To help ensure convergence, we constrained ρ_F to be between 0.5 and 3 times the minimum distance between any two knots in $\{\boldsymbol{\kappa}_F\}$, and constrained ρ_C to be greater than ρ_F but less than 3 times the minimum distance between any two knots in $\{\boldsymbol{\kappa}_C\}$. Optimization used the glm() and optim() functions in R (R Core Team, 2014), where optim() used the Nelder-Mead optimization algorithm (Nelder and Mead, 1965). To ensure boundary conditions, say a as a lower bound and b as an upper bound for one of the elements in $\boldsymbol{\rho}$, we used a transformation $\rho = a + (b-a)\exp(\rho^*)/(1+\exp(\rho^*))$, and then optimized for unconstrained ρ^* (note that a was a sliding lower boundary for ρ_C, but it would stabilize as ρ_F found its optimum).

Also, note the connection to the Janossy density for IPP (see, e.g., Cressie, 1993, p. 655). For some area B with $Y \in 1, 2, \ldots$ points at locations $\{\mathbf{s}_k; k = 1, 2, \ldots, Y\}$ within B, the Janossy likelihood is,

$$\mathcal{L}(\boldsymbol{\theta}; B) = \left\{ \prod_{k=1}^{Y} \lambda(\mathbf{s}_k|\boldsymbol{\theta}, \boldsymbol{\rho}) \right\} \exp\left\{ -\int_B \lambda(\mathbf{u}|\boldsymbol{\theta}, \boldsymbol{\rho}) d\mathbf{u} \right\}. \tag{9}$$

From Section 2.3, we are assuming that the plots are small enough so that the intensity function is approximately constant within plot, with the intensity value taken from the intensity surface at the centroid of the plot. Using this approximation, then from (9) the

negative loglikelihood for all plots is

$$\ell(\boldsymbol{\rho}, \boldsymbol{\theta}; \mathbf{y}) \approx \sum_{i=1}^{n} -y_i \log[\lambda(\mathbf{s}_i|\boldsymbol{\theta}, \boldsymbol{\rho})] + |B_i|\lambda(\mathbf{s}_i|\boldsymbol{\theta}, \boldsymbol{\rho}),$$

and, when using model (6) for $\lambda(\mathbf{s}_i|\boldsymbol{\theta}, \boldsymbol{\rho})$, this makes it apparent that minimizing (8) for $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ is an approximation to maximizing (9). This connection is important because, in Section 2.7, we use results from maximum likelihood estimation of the Janossy density in IPP literature to obtain variance estimates. Note that other approaches may be taken, including penalized splines (Ruppert et al., 2003) or Bayesian approaches (see Section 1.2).

2.6 Plug-in Abundance Estimator

Denote $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\rho}}$ as the maximum likelihood estimates from Section 2.5. Going back to our estimator, recall that $\widehat{T}(A) = T(\mathcal{B}) + \widehat{T}(\mathcal{U})$, and we will use our parameter estimates from Section 2.5 to obtain the predictor $\widehat{T}(\mathcal{U}) = \mu(\mathcal{U}) = \int_{\mathcal{U}} \lambda(\mathbf{u}|\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\theta}}) d\mathbf{u}$, where $\lambda(\mathbf{u}|\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\theta}}) = \exp(\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{u})'\hat{\boldsymbol{\theta}})$. The integral can be approximated with a dense grid of n_p points within $\mathbf{u}_j \in \mathcal{U}$,

$$\widehat{T}(A) = T(\mathcal{B}) + \sum_{j=1}^{n_p} |U_i| \exp(\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{u}_j)'\hat{\boldsymbol{\theta}}), \tag{10}$$

where $|U_i|$ is a small area around each $\mathbf{u}_j$. We generally assume all $|U_i|$ are equal to $|\mathcal{U}|/n_p$, yielding a 2-dimensional Riemann integral approximation, which is sufficient if n_p is large. Better approaches using numerical integration by quadrature could also be used.

2.7 Variance Estimation

The mean-squared prediction error of (1) is

$$\mathcal{M}(\hat{T}(A)) = E[(\hat{T}(A) - T(A))^2] = E[(\hat{T}(\mathcal{U}) - T(\mathcal{U}))^2] \tag{11}$$

Note that as $\mathcal{U} \cap A \to \varnothing$, then we count all animals in A, and $\mathcal{M}(\widehat{T}(A)) \to 0$, so that this estimator satisfies condition 4 in Section 1.3. Thus, a finite population correction factor is automatically embedded in the variance estimator. Also, $\hat{T}(\mathcal{U})$ depends on random counts in $\mathcal{B}$, while $T(\mathcal{U})$ depends on random counts in $\mathcal{U}$. Under the IPP assumption, these will be independent from each other. Further, assume that $\hat{T}(\mathcal{U})$ is an unbiased predictor, so

$E[\hat{T}(\mathcal{U})] = E[T(\mathcal{U})]$. Then $\mathcal{M}(\hat{T}(A)) = E[(\hat{T}(\mathcal{U})^2] - 2E[T(\mathcal{U})]^2 + E[T(\mathcal{U})^2]$, or

$$\mathcal{M}(\hat{T}(A)) = \text{var}[T(\mathcal{U})] + \text{var}[\hat{T}(\mathcal{U})].$$

For the IPP, $\text{var}[T(\mathcal{U})] = \mu(\mathcal{U})$, and this is estimated with

$$\hat{\mu}(\mathcal{U}) = \frac{|\mathcal{U}|}{n_p} \sum_{i=1}^{n_p} \exp[\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{s}_i)'\hat{\boldsymbol{\theta}}] \tag{12}$$

over the same fine grid of points used in (10). Recall that $\hat{T}(\mathcal{U}) = \int_{\mathcal{U}} \exp[\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{u})'\hat{\boldsymbol{\theta}}]d\mathbf{u}$. Define a vector $\mathbf{c}$ where the ith element of $\mathbf{c}$ is

$$\frac{\partial \hat{T}(\mathcal{U})}{\partial \theta_i} = \int_{\mathcal{U}} x_i(\mathbf{u}) \exp[\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{u})'\hat{\boldsymbol{\theta}}]d\mathbf{u}.$$

We approximate this integral with

$$\frac{\partial \hat{T}(\mathcal{U})}{\partial \theta_i} \approx \frac{|\mathcal{U}|}{n_p} \sum_{i=1}^{n_p} x_i(\mathbf{s}_i) \exp[\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{s}_i)'\hat{\boldsymbol{\theta}}], \tag{13}$$

where the sum is over a dense grid of n_p prediction points in the unsampled area. Using the delta method (Dorfman, 1938; Ver Hoef, 2012), $\text{var}[\hat{T}(\mathcal{U})] = \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c}$, where $\boldsymbol{\Sigma} = \text{var}(\hat{\boldsymbol{\theta}})$. A similar result is given by Johnson et al. (2010) in a distance sampling context. Then, as shown by Rathbun and Cressie (1994), if $\hat{\boldsymbol{\theta}}$ is a maximum likelihood estimator from the Janossy density for the IPP, then an estimator of $\boldsymbol{\Sigma}$ is

$$\hat{\boldsymbol{\Sigma}} = \left[\sum_{i=1}^{n} \int_{B_i} \mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{u}) \mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{u})' \exp[\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{u})'\hat{\boldsymbol{\theta}}]d\mathbf{u} \right]^{-1}.$$

Assuming that $|B_i| = |B| \ \forall \ i$ is small, this can be approximated as

$$\hat{\boldsymbol{\Sigma}} = \left[|B| \sum_{i=1}^{n} \mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{s}_i) \mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{s}_i)' \exp(\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{s}_i)'\hat{\boldsymbol{\theta}}) \right]^{-1}. \tag{14}$$

Note that, in (14), variances may become large if the dimension of $\boldsymbol{\theta}$ is too high (due to overfitting from too many knots). Through simulations, we will investigate the following

variance estimator,

$$\hat{\mathcal{M}}(\hat{T}(A)) = \hat{\mu}(\mathcal{U}) + \mathbf{c}'\hat{\mathbf{\Sigma}}\mathbf{c}. \tag{15}$$

where $\hat{\mu}(\mathcal{U})$ is given by (12), elements of $\mathbf{c}$ are given by (13), and $\hat{\mathbf{\Sigma}}$ is given by (14). Equation (15) has a nice interpretation by decomposing the variance into the prediction of the total due to fixed intensity surface $\hat{\mu}(\mathcal{U})$ (given the regression parameters $\boldsymbol{\theta}$), plus the variance in estimating the regression parameters $\boldsymbol{\theta}$. Note that we have not taken into account the estimation of $\boldsymbol{\rho}$. While this would be desirable, we use $\hat{\boldsymbol{\rho}}$ as plug-in estimators for now. This is similar to geostatistical models where covariance parameters are first estimated from the data, and then used for subsequent prediction (see, e.g., Schabenberger and Gotway, 2005, p. 263). While this is not ideal, and can be the subject of further research, our simulations show that it has little consequence for the type of data that we analyze.

2.8 Overdispersion

Animals (as well as other spatially patterned points) are often clustered at very fine spatial scales. For animals, this might occur as mother-offspring pairs, clustering around locally desirable habitats, etc. The inhomogeneous intensity surface estimated in the foregoing discussion will be unlikely to capture this fine scale clustering, which will contribute to the overall variance, and without considering it, the confidence intervals on abundance estimates will be too short. Various estimators of overdispersion for count models have been proposed, and the negative binomial and quasi-Poisson are commonly used (e.g., Ver Hoef and Boveng, 2007); see Hinde and Demétrio (1998) for an overview. Here, we consider quasi-type models, where, if the mean is ϕ, then the overdispersion is constant multiplier, ω, so the variance is $\omega\phi$. As we demonstrate next, some form of robust estimation or further modeling is required because overdispersion changes through space. In the negative binomial context, robust but nonspatial estimation of can be found in Moore and Tsiatis (1991), and nonparametric estimation is found in Gijbels et al. (2010). Our situation is different than general robustness because we want to either trim residuals based on data with low expected values, or downweight them. We describe several estimators next, and compare them in simulations.

Let $\phi_i = E(\mathbf{Y}_i|\hat{\boldsymbol{\beta}}) = g^{-1}(\mathbf{x}_i\hat{\boldsymbol{\beta}}) = \exp(\mathbf{x}_i\hat{\boldsymbol{\beta}})$ be the fitted intensity surface value for the ith plot, where $\mathbf{x}_i$ is the ith row of $\mathbf{X}$. Denoting y_i as the observed value for the ith plot, we considered four different ways to estimate overdispersion:

- The traditional estimator:

$$\omega_{OD} = \max\left(1, \frac{1}{n-q}\sum_{i=1}^{n}\frac{(y_i-\phi_i)^2}{\phi_i}\right),$$

where q is the rank of $\mathbf{X}$.

- A linear regression estimator. Under the Poisson model, the variance is equal to the mean. By regressing the squared residuals against the fitted value, any slope greater than one would be evidence of overdispersion. The linear regression is set up with a zero intercept, so the model is $(y_i - \phi_i)^2 = \omega\phi_i$. We used weighted least squares to obtain the estimator,

$$\omega_{WR} = \max\left(1, \arg\min_{\omega}\sum_{i=1}^{n}\sqrt{\phi_i}[(y_i-\phi_i)^2 - \omega\phi_i]^2\right),$$

where $\sqrt{\phi_i}$ were the weights. Notice that generally, this may not be a desirable estimator. Values with small expectations have virtually no effect on the slope, whereas values with larger expectations will have a great deal of leverage. In our case, this is a desirable feature, as discussed earlier. In fact, we create additional weight for values with large expectation by using $\sqrt{\phi_i}$.

- Estimator based on a trimmed mean of squared Pearson residuals from the upper quantile of fitted values. Let $\mathcal{F} = \{\phi_1, \phi_2, \ldots, \phi_n\}$ be an unordered set of expected values for the n observed counts, and $\{\phi_{(1)}, \phi_{(2)}, \ldots, \phi_{(n)}\}$ be the set of ordered values, from smallest to largest, where $\phi_{(1)} = \min(\mathcal{F})$ and $\phi_{(n)} = \max(\mathcal{F})$. Also, if $\phi_{(i)} = \phi_j$, then $y_{(i)} = y_j$; that is, the observed values are ordered by their fitted values as well. Let $0 \leqslant p < 1$ be some proportion, then

$$\omega_{TG}(p) = \max\left(1, \frac{1}{n-\lfloor np\rfloor}\sum_{i=\lfloor np\rfloor+1}^{n}\frac{(y_{(i)}-\phi_{(i)})^2}{\phi_{(i)}}\right),$$

where $\lfloor x\rfloor$ rounds x down to the nearest integer. That is, the proportion p of the squared Pearson residuals with the lowest fitted values are trimmed from the overdisperson computation.

Examples of the overdisperson estimators are shown in Figure 4, which were taken

from the data seen in Figure 3. The traditional estimator can be viewed as a constant fit (the average value) through all of the squared Pearson residuals for all fitted values, so this is shown as a horizontal solid line, the one that is below the short-dashed line (whose value is constant at one) in Figure 4A. Note especially the wide divergence in squared Pearson residuals for low expected values. This is not surprising because we are dividing by very small numbers, so any count greater than zero will have a very large residual. This instability, along with the fact that these values do not really contribute much to overall abundance, leads to estimator $\omega_{TG}(p)$. Here, we trim off the lowest expected values. Trimming off the lowest 75%, and averaging the rest, can be viewed as a constant fit (horizontal line) through the squared Pearson residuals for the upper 25% of fitted values, and is shown as the long-dashed horizontal line that is above the short-dashed line in Figure 4A. The other idea is to treat raw squared residuals, $(y_i - \phi_i)^2$ as a response variable in a zero-intercept regression, where the predictor variable is the fitted value ϕ_i. This is shown as the solid line in Figure 4B, which is above the one-to-one line. The estimated slope of this line is taken as the overdispersion estimate. Similar to trimming in $\omega_{TG}(p)$, the regression estimator $\omega_{WR}(p)$ downweights residuals with small expected values by forcing the line through zero, and it eliminates division by very small numbers. In fact, we considered weighted regression to add even more weight to higher fitted values. After some trail and error, we used weights $\sqrt{\phi_i}$, but this is clearly an area for further research.

With these three overdisperson estimators, we have several variance estimators of the abundance estimator (1) at our disposal,

$$\widehat{\text{var}}(\hat{T}(A))_k \equiv \omega_k \hat{\mathcal{M}}(\hat{T}(A)), \tag{16}$$

where $k = OD$, WR or, TG, and ω_{TG} has the additional trimming parameter p. We include one more estimator using the same logic applied to the IPP variance estimator as the trimmed overdispersion estimator $\omega_{TG}(p)$. If $\phi_{\lfloor np \rfloor}$ represents the smallest fitted value summed in ω_{TG}, then we computed (14) using only those i sites whose values satisfied $\exp(\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{s}_i)'\hat{\boldsymbol{\theta}}) > \phi_{\lfloor np \rfloor}$. Let us call this $\widetilde{\boldsymbol{\Sigma}}$, which when substituted into (15) and combined with ω_{TG} yields

$$\widehat{\text{var}}(\hat{T}(A))_{TL} \equiv \omega_{TG}(\hat{\mu}(\mathcal{U}) + \mathbf{c}'\widetilde{\boldsymbol{\Sigma}}\mathbf{c}). \tag{17}$$

For confidence intervals, note that from (10), the estimate is a sum of a large number of lognormal variates. That is, we can assume that $\hat{\boldsymbol{\theta}}$ are normal because they are maximum likelihood estimates. If each summand in (10) was independent, then (10) would converge

to normality because of the central limit theorem, but due to correlation, the distribution is unknown and may be asymmetric. We investigated this by simulating (10) using $\hat{\mathbf{\Sigma}}$ from (14) as estimated from various data sets. In all cases, (10) was skewed, and a log transformation made the distribution approximately normal. Thus, we recommend computing confidence intervals on the log scale, and then back-transforming. Using the delta method (Dorfman, 1938; Ver Hoef, 2012), an approximate $100(1-\alpha)\%$ level confidence interval is

$$\exp\left(\log(\hat{T}(A)) \pm \frac{z_{\alpha/2}\sqrt{\widehat{\text{var}}(\hat{T}(A))_k}}{\hat{T}(A)}\right), \tag{18}$$

for $k = OD$, WR, TG or TL, where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of a standard normal distribution. Note that this also yields the desirable property that the lower bound of the confidence interval is always greater than 0.

3 Simulation experiments

We simulated data under four different conditions to examine the performance of the abundance estimator and the variance estimators of abundance. In all experiments, data were simulated in the region $A = \{(x, y) : x \in [0, 10] \times y \in [0, 10]\}$. For each experiment, data sets were simulated 1000 times, with the number of points and their spatial locations changing, but the sample units were held fixed as shown in Figure 5.

3.1 Evaluating the experiments

For each experiment described below, the expected number of simulated points was near 1000. Let T_t be the actual number of points simulated in the tth simulation, let $\hat{T}_t$ be the estimator of the total from (10), and let $\hat{v}_{k,t}$ be a variance estimator given in (16) or (17) for $k = OD$, WR, TG or TL, for the tth simulation. The performance of the abundance estimator was evaluated in three ways:

- Bias for an experiment was computed as,

$$\frac{1}{1000}\sum_{t=1}^{1000}\hat{T}_t - T_t.$$

- Root-mean-squared prediction error (RMSPE) was computed as,

$$\sqrt{\frac{1}{1000}\sum_{t=1}^{1000}(\hat{T}_t - T_t)^2}.$$

- Coverage of 90% confidence interval (CI90) was computed as,

$$\frac{1}{1000}\sum_{t=1}^{1000} I\left(\exp(\log(\hat{T}_t) - 1.645\sqrt{\hat{v}_{k,t}}/\hat{T}_t) < T_t \;\&\; T_t < \exp(\log(\hat{T}_t) + 1.645\sqrt{\hat{v}_{k,t}}/\hat{T}_t)\right),$$

  where $I(\cdot)$ is the indicator function, equal to one if the argument is true, otherwise it is zero. Note that (CI90) can be computed for each k in (16) and (17), which we denote as CI90_k in the tables that summarize the experiments.

For all experiments we used $\omega_{TG}(0.75)$, but investigate the effect of p in Experiment 4. We also investigated knot density by changing K_C and K_F, but always using the algorithm described in Section 2.4. For each experiment, we included bias, RMSPE, and CI90 for simple random sampling (SRS), as described in the Introduction. We realize that SRS is inappropriate for these data, but it provides a convenient benchmark for comparison.

3.2 Experiment 1: Inhomogeneous spatial point process with regular sampling

The study area A was a square starting at (0,0) and 10 units on each side. Data were simulated using rejection sampling. Consider the intensity surface $\lambda(x, y) = x/20 + y/20$, which increases linearly from 0 at (0,0) to 1 at (10,10) within A. A location was simulated with x^*, y^*, and z^*, where each was drawn from Unif$(0, 1)$. If $z^* < \lambda(x^*, y^*)$, then the simulated location was retained. The set (x^*, y^*, z^*) was drawn 2000 times, so the expected number of locations retained was 1000 per simulated data set, but note that the actual number varied randomly among simulations. For each simulated data set, sample units as square plots that measured 0.3 on a side were systematically placed in a 16×16 grid as shown in Figure 5. The 256 plots covered 23.04% of the study area A.

The results of experiment 1 are given in Table 1. Note that for SRS and various knot proportions, there was little bias, which is $< 1\%$ of the average total. All methods had very similar RMSPE as well, and 90% confidence intervals generally had the appropriate

coverage, no matter which overdispersion method was used. Of course, the data were not simulated with overdispersion. The only exception occurred when using many knots for the trimmed overdispersion estimators CI90_{TL} and CI90_{TG}, where variance was overestimated.

3.3 Experiment 2: Inhomogeneous spatial point process with irregular sampling

For simulation experiment 2, locations were simulated exactly as they were in Experiment 1 (Section 3.2). However, we created unbalanced spatial sampling by removing one column and two rows of sample units, as shown in Figure 5. In this case, 210 plots covered 18.9% of the study area A.

The results of experiment 2 are given in Table 2, which should not be surprising for SRS. Indeed, the goal of this research was to find good estimators when high (or low) abundance areas were oversampled, and SRS makes no weighting adjustments for this. Consequently, it had a large bias, whereas $\widehat{T}(A)$ in (10) remained relatively unbiased with RMSPE only slightly larger than experiment 1 (note, too, that sample sizes were smaller here). The same basic patterns appeared for CI90, with generally valid confidence intervals (except for SRS), except when many knots are used for the trimmed overdispersion estimators CI90_{TL} and CI90_{TG}.

3.4 Experiment 3: Double spatial cluster process on inner rectangle

One difficult situation for estimating abundance from counts occurs when there is a large number of zeros and there is fine scale clustering, so we tested the abundance estimator under both of those conditions. For this experiment, seed points were simulated within a rectangle within the study area. Again, the study area A was a square starting at (0,0) and 10 units on each side. The inner rectangle itself had random boundaries, where the lower x-axis and y-axis boundaries were each randomly drawn from Unif$(3.5, 4.5)$, and the upper x-axis and y-axis boundaries were each drawn from Unif$(7.5, 8.5)$. Next 100 parent seed points were uniformly simulated over the inner rectangle (such a rectangle is shown with solid lines in Figure 3), and then each parent had a random number, Poi(15), of children that were uniformly distributed on a square with sides of length 2 centered on each parent. A second finer scale cluster process was added by creating 25 more parent points uniformly

distributed over the inner rectangle, where each parent had a random number, Poi(9), of children that were uniformly distributed on a square with sides 0.4 centered on each parent. After simulating all points, they were thinned using the same function as experiments 1 and 2, by simulating $z_i^* \sim \text{Unif}(0,1)$ at each simulated location with coordinates x_i and y_i, and keeping that location if $z_i^* < x_i/20 + y_i/20$. From 1000 simulated experiments, this yielded an average of 1034 points per simulation. An example of one simulation is given in Figure 5. The sample units were placed in the same positions as for Experiment 2 (Section 3.3).

The results of experiment 3 are given in Table 3. Once again, because of the unbalanced sampling, SRS had a large RMSPE and was highly biased, whereas $\widehat{T}(A)$ in (10) remained unbiased at generally less than 0.5% of the total. RMSPE was larger than experiments 1 and 2, but this is not surprising given the smaller area with positive count values. This experiment showed some poorer performance for CI90, especially for CI90_{OD}, which underestimated variance with coverage nearer 80% rather than 90%. Both CI90_{WR} and $CI90_{TG}$ performed more poorly with increasing numbers of knots. CI90_{TL} coverage is about 3% low for the fewest number of knots, but generally improves with the number of knots. Notice also that there is almost a 2% chance that the parameter estimation algorithm will fail when there are many knots.

3.5 Experiment 4: Double spatial cluster process on double inner rectangles

For this experiment, the seed points were simulated in a manner similar to experiment 3. However, the sample units were made smaller, with length 0.14 on a side, on a 26×26 grid in a study area, A, which was again a square starting at (0,0) and 10 units on each side. This time there were two inner rectangles. One inner rectangle had random boundaries where the lower x-axis and y-axis boundaries were each randomly drawn from $\text{Unif}(5.8, 6.2)$, and the upper x-axis and y-axis boundaries were each drawn from $\text{Unif}(7.8, 8.2)$. Next 75 parent seeds were uniformly simulated over this rectangle, and each parent seed had a random number, Poi(14), of children uniformly distributed in a box with sides of length 2 centered on each parent. A finer scale cluster process was added by creating 25 more parent points uniformly distributed over this inner rectangle, where each parent had a random number, Poi(8), children that were uniformly distributed on a square with sides 0.4 centered on each parent. A second inner rectangle had random boundaries where the lower x-axis boundary was randomly drawn from $\text{Unif}(0.8, 1.2)$ and the upper x-axis boundary was drawn from

Unif(3.8, 4.2), while the lower y-axis boundary was randomly drawn from Unif(4.8, 5.2), and the upper y-axis boundary was drawn from Unif(7.8, 8.2). Here, 25 parent seeds were uniformly simulated over this rectangle, and each parent seed had a random number, Poi(14), of children uniformly distributed in a box with sides of length 1 centered on each parent. A finer scale cluster process was achieved by creating 10 more parent points uniformly distributed over this inner rectangle, where each parent had a random number, Poi(8), of children that were uniformly distributed on a square with sides 0.4 centered on each parent. After simulating all points, they were thinned using the same function as experiments 1-3, by simulating $z_i^* \sim \text{Unif}(0, 1)$ at each simulated location with coordinates x_i and y_i, and keeping that location if $z_i^* < x_i/20 + y_i/20$. From 1000 simulated experiments, this yielded an average of 1012 points per simulation. An example of one simulation is given in Figure 5. We created unbalanced spatial sampling by removing one column and two rows of sample units, as shown in Figure 5. In this case, 600 plots covered 11.8% of the study area A.

The results of experiment 4 are given in Table 4. Once again, because of the unbalanced sampling, SRS had a large RMSPE and was highly biased, whereas $\widehat{T}(A)$ in (10) remained unbiased for smaller number of knots, but was $> 1\%$ of the total for the largest number of knots. RMSPE was less than experiments 3, likely due to a larger area with nonzero counts and smaller plots, leading to a larger sample size. Once again, CI90_{OD} underestimated variance with coverage nearer 83% rather than 90%. CI90_{WR} was about 4% high for small number of knots, but improved with numbers of knots. CI90_{TG} remains about 2% high for all combinations of knot numbers, while CI90_{TL} is 1% to 2% low for all combinations of knot numbers. Here, notice also that there is over 24% chance that the parameter estimation algorithm will fail when there are many knots.

The effects of varying p in CI90_{TG} and CI90_{TL} are shown in Figure 6, using 1000 simulated data sets as described for experiment 4. Notice that confidence coverage for CI90_{TG} was in the "valid" zone between $p = 0.4$ and $p = 0.8$, and CI90_{TL} was in the "valid" zone when $p \geqslant 0.5$. In a real setting, such a p value must be chosen by data examination (at least without further research). Looking at the simulated data in Figure 5, one would try to estimate the area dominated by zero counts, and then trim them. A p value anywhere from 0.6 to 0.8 seems reasonable, and would lead to nearly correct confidence intervals using either CI90_{TG} or CI90_{TL}. Clearly, CI90_{TG} is a more conservative strategy, but trimming aggressively with CI90_{TL} appears viable. Another consideration is that we expect that smaller p values would be more efficient by using more data.

4 Example: aerial surveys of harbor seals

The study area boundary, the locations of all 2080 plots, and the observed counts of seals within plots are shown in Figure 2, and the data were summarized in Section 1.1. We used $K_C = 4$ knots in the coarse grid and we used $K_F = 15$ knots in the fine grid shown in Figure 7. Notice that the fine grid of knots is contained in a bounding rectangle around only those plots with non-zero counts; with the reason explained in Section 3.1. The model fit took 17.56 seconds on a Intel Xeon 2.66GHz processor running under the linux operating system. The estimate range parameter ρ_F was 1.81 km, and ρ_C was 4.04 km. The fitted intensity surface is shown in Figure 7. The estimate obtained from integrating this surface, along with the observed count, using $\widehat{T}(A)$ in (10) was 4012. The standard error $\sqrt{\hat{\mathcal{M}}(\hat{T}(A))}$ in (15), without any corrections for overdispersion, was 111.86. Interestingly, the $\hat{\mu}(\mathcal{U})$ component of (15) was 3010, and the $\mathbf{c}'\hat{\mathbf{\Sigma}}\mathbf{c}$ component of (15) was 9504. The estimates of overdispersion given in Section 2.8 were $\omega_{OD} = 17.26$, $\omega_{WR} = 3.77$, $\omega_{TG} = 3.45$, and $\omega_{TL} = 2.79$. If we use ω_{TG} for overdispersion, that yields a standard error of 386. Then the 95% confidence interval, from (18), for the estimate of total abundance is (3322, 4845).

We tried different combinations of knots and p in ω_{TG} and ω_{TL}. No systematic attempt is made to present those results, and knot selection and overdispersion is a topic for future research. However, we did note that the abundance estimate was little changed for knots up to $K_C = 8$ and $K_F = 32$. However, there were changes in overdispersion estimates. Also, we noticed that for the knots that we selected, increasing p in ω_{TG} and ω_{TL} *decreased* overdispersion, rather than the increasing values seen in Figure 6. Conceptually, it is possible that an area with high counts could have less variability than a larger area that includes both high counts and low counts. Also, note that $\omega_{OD} = 17.26$, which is much larger than one, and larger than all other overdispersion estimators, in contrast to the example provided by Figure 4. However, the explanation is also provided by Figure 4 because for this data set there were a few non-zero counts with very low expected values that dominated ω_{OD}. We chose this example in part because it showed some exceptions for the overdispersion parameters. Based on dozens of similar real examples, most of them have $\omega_{OD} = 1$, and this appears to be an unstable estimator for our purposes.

5 Discussion and conclusions

Our objectives were to develop a model based estimator of a total by using counts from irregularly spaced plots. We wanted this estimator to have goals, properties, and performance similar to classical sampling: to estimate the realized total, not the mean of an assumed process, to have a finite population correction factor, to be unbiased with valid confidence intervals that must be robust to nonstationary mean and variance, and to be fast to compute. The problem was made more difficult by features of the data that were seen in the real example. First, the nonzero counts were highly clustered in space, and there were large areas of zeros. Secondly, the counts, where they occurred, showed overdispersion. Finally, sample sizes were quite large, so we needed to use computationally efficient methods. The general approach that we took was to assume that the data came from an inhomogeneous point process with overdispersion. We modeled the intensity surface of the inhomogeneous point process using spatial basis functions in a generalized linear mixed model framework.

To test the method, we started with fairly benign conditions in experiment 1. We then added complexity to the simulations that matched the complexity seen in the real data, by simulating spatially unbalanced sampling, data with overall trend in point density, several areas of clustered points, overdispersion within the cluster areas, and yet large areas with zeros, culminating in experiment 4. Overall, our method worked well, especially when using some of the newly introduced overdisperion estimators. One of the main contributions of this manuscript was the introduction of overdispersion estimators when the overdispersion appears to be varying spatially (a nonstationary overdispersion).

One of the interesting findings from our research was the effect of knot placement and proportion (Tables 1 - 4). Initially, we found a lack of convergence when fitting these models if there were too many knots, with short ranges, over areas containing all zeros. In retrospect, that may not seem surprising, but we have not found it reported in the literature. For that reason, we created spatially restricted knots over areas with non-zero values (dashed line in Figure 3; also see Figure 7). For placing knots, we used a K-means clustering algorithm on the spatial coordinates to create a space-filling design. Other approaches that could be used were mentioned in Section 2.4. Our results show that the bias does not depend very much on the number of knots, but the standard errors are quite sensitive. An encouraging result is that standard errors yield better confidence intervals when the number of knots is quite small, making the algorithm very fast. The whole issue of knot selection needs further research.

Besides more research on overdispersion estimators, and knot placement and number, there are numerous modifications that could be applied to our basic approach. We chose spatial basis functions that were Gaussian kernels at two scales, and there are many obvious modifications. Because we annually analyze dozens of data sets like our example for harbor seals, we took a maximum likelihood approach to estimating parameters and total abundance; however, Bayesian approaches could also be used (Wikle, 2002; Christensen and Waagepetersen, 2002). The ability to estimate the intensity function, essentially point-level information, from data at an aggregated level, i.e., counts from plots, depends on the plots being small in relation to changes in the intensity function. If plots are very large, then methods in this paper could still be used, but spatial points would need to be mapped within plots, and more traditional methods from the spatial point process literature could be used for estimating the intensity surface. For example, the R (R Core Team, 2014) package spatstat (Baddeley and Turner, 2005) can be used when the realized point patterns have a complicated "mask" comprised of many disjoint plots (Baddeley and Turner, 2006). Alternatively, area-to-point geostatistical methods (Kyriakidis, 2004) could be used. The main point is that once the intensity surface is estimated, (10) can be used to estimate total abundance. The variance of the total abundance can be estimated with (15) if maximum likelihood methods are used, along with one of the overdispersion factors (Section 2.8) that make sense for the problem under consideration.

ACKNOWLEDGMENTS

...

References

Baddeley, A. and Turner, R. (2005), "Spatstat: an R package for analyzing spatial point patterns," *Journal of Statistical Software*, 12, 1–42, URL: `www.jstatsoft.org`, ISSN: 1548-7660.

— (2006), "Modelling spatial point patterns in R," in *Case Studies in Spatial Point Pattern Modelling*, eds. Baddeley, A., Gregori, P., Mateu, J., Stoica, R., and Stoyan, D., New York: Springer-Verlag, no. 185 in Lecture Notes in Statistics, pp. 23–74, ISBN: 0-387-28311-0.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, New York, New York, USA: Chapman & Hall/CRC.

Barber, J. J. and Gelfand, A. E. (2007), "Hierarchical spatial modeling for estimation of population size," *Environmental and Ecological Statistics*, 14, 193–205.

Berman, M. and Turner, T. R. (1992), "Approximating Point Process Likelihoods with GLIM," *Journal of the Royal Statistical Society. Series C (Applied Statistis)*, 41, 31–38.

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion)," *Journal of the Royal Statistical Society, Series B*, 36, 192–236.

Brillinger, D. R. (1990), "Spatial-Temporal Modelling of Spatially Aggregate Birth Data," *Survey Methodology*, 16, 255–269.

— (1994), "Examples of Scientific Problems and Data Analysis in Demography, Neurophysiology, and Seismology," *Journal of Computational and Graphical Statistics*, 3, 1–22.

Christensen, O. F. and Waagepetersen, R. (2002), "Bayesian Prediction of Spatial Count Data Using Generalized Linear Mixed Models," *Biometrics*, 58, 280–286.

Cochran, W. G. (1977), *Sampling Techniques, Third Edition*, New York: John Wiley & Sons.

Cressie, N. and Johannesson, G. (2008), "Fixed rank kriging for very large spatial data sets," *Journal of the Royal Statistical Society, Series B*, 70, 209–226.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: John Wiley & Sons.

Dorfman, R. (1938), "A Note on the Delta-Method for Finding Variance Formulae," *The Biometric Bulletin*, 1, 129–137.

ESRI (2009), "ArcGIS Desktop: Release 9.3.1." Tech. rep., Environmental Systems Research Institute, Redlands, CA.

Friedman, J. H., Bentley, J. L., and Finkel, R. A. (1977), "An Algorithm for Finding Best Matches in Logarithmic Expected Time," *ACM Transactions on Mathematical Software*, 3, 209–226.

Gandin, L. S. (1959), "The Problem of Optimal Interpolation (In Russian)," *Trudy GGO*, 99, 67–75.

— (1960), "On Optimal Interpolation and Extrapolation of Meteorological Fields (In Russian)," *Trudy GGO*, 114, 75–89.

Gijbels, I., Prosdocimi, I., and Claeskens, G. (2010), "Nonparametric estimation of mean and dispersion functions in extended generalized linear models," *Test*, 19, 580–608.

Gotway, C. A. and Young, L. J. (2002), "Combining Incompatible Spatial Data," *Journal of the American Statistical Association*, 97, 632–648.

Hinde, J. and Demétrio, C. G. (1998), "Overdispersion: models and estimation," *Computational Statistics & Data Analysis*, 27, 151–170.

Johnson, D. S., Laake, J. L., and Ver Hoef, J. M. (2010), "A Model-based Approach for Making Ecological Inference from Distance Sampling Data," *Biometrics*, 66, 310–318.

Kyriakidis, P. C. (2004), "A Geostatistical Framework for Area-to-Point Spatial Interpolation," *Geographical Analysis*, 36, 259–289.

MacQueen, J. B. (1967), "Some Methods for Classification and Analysis of MultiVariate Observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, eds. Cam, L. M. L. and Neyman, J., University of California Press, vol. 1, pp. 281–297.

Matheron, G. (1963), "Principles of Geostatistics," *Economic Geology*, 58, 1246–1266.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models, 2nd Edition*, Chapman & Hall Ltd.

Monestiez, P., Dubroca, L., Bonnin, E., Durbec, J.-P., and Guinet, C. (2006), "Geostatistical Modelling of Spatial Distribution of *Balaenoptera physalus* in the Northwestern Mediterranean Sea from Sparse Count Data and Heterogeneous Observation Efforts," *Ecological Modelling*, 193, 615–628.

Moore, D. and Tsiatis, A. (1991), "Robust estimation of the variance in moment methods for extra-binomial and extra-Poisson variation," *Biometrics*, 47, 383–401.

Nelder, J. A. and Mead, R. (1965), "A Simplex Method for Function Minimization." *Computer Journal*, 7, 308–313.

Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society, Series A: General*, 135, 370–384.

Nychka, D. W. and Saltzman, N. (1998), "Design of air quality monitoring networks," in *Case Studies in Environmental Statistics*, eds. Nychka, D. e., Piegorsch, W. W. e., and Cox, L. H., Springer-Verlag Inc, pp. 229–234.

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rathbun, S. L. and Cressie, N. (1994), "Asymptotic Properties of Estimators for the Parameters of Spatial Inhomogeneous Poisson Point Processes," *Advances in Applied Probability*, 26, 122–154.

Royle, J. A., Kéry, M., Gautier, R., and Schmid, H. (2007), "Hierarchical Spatial Models of Abundance and Occurrence From Imperfect Survey Data," *Ecological Monographs*, 77, 465–481.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge University Press.

SAS Institute Inc (2008), *SAS/STAT® 9.2 Users Guide*, Cary, NC: SAS Institute Inc.

Schabenberger, O. and Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, Boca Raton, Florida: Chapman Hall/CRC.

Thogmartin, W. E., Sauer, J. R., and Knutson, M. G. (2004), "A Hierarchical Spatial Model of Avian Abundance With Application to Cerulean Warblers," *Ecological Applications*, 14, 1766–1779.

Thompson, S. K. (1992), *Sampling*, New York: John Wiley & Sons.

Ver Hoef, J. M. (2000), "Predicting Finite Populations from Spatially Correlated Data," in *ASA Proceedings of the Section on Statistics and the Environment*, American Statistical Association, pp. 93–98.

— (2002), "Sampling and Geostatistics for Spatial Data," *Ecoscience*, 9, 152–161.

— (2008), "Spatial Methods for Plot-based Sampling of Wildlife Populations," *Environmental and Ecological Statistics*, 15, 3–13.

— (2012), "Who Invented the Delta Method?" *The American Statistician*, 66, 124–127.

Ver Hoef, J. M. and Boveng, P. L. (2007), "Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data?" *Ecology*, 88, 2766–2772.

Wang, H. and Ranalli, M. G. (2007), "Low-rank Smoothing Splines on Complicated Domains," *Biometrics*, 63, 209–217.

Wikle, C. K. (2002), "Spatial modeling of count data: a case study in modelling breeding bird survey data on large spatial domains," in *Design and Analysis of Ecological Experiments*, eds. Lawson, A. and Denison, D., Chapman & Hall, London, pp. 199–209.

Wikle, C. K. and Berliner, L. M. (2005), "Combining information across spatial scales," *Technometrics*, 47.

Yaglom, A. M. (1962), *An Introduction to the Theory of Stationary Random Functions*, Mineola, New York: Dover Publications, Inc.

TABLES

Table 1: Results for bias, RMSPE, confidence interval coverage, and failure rate for simulation experiment 1. The number of coarse-scale knots used is given by K_C, and K_F is the number of fine-scale knots. An example of a single simulated data set is given in Figure 5.

		Knots			
	SRS	$K_C = 3$ $K_F = 8$	$K_C = 5$ $K_F = 16$	$K_C = 7$ $K_F = 24$	$K_C = 9$ $K_F = 32$
Bias	6.425	-1.277	-9.735	7.048	5.941
RMSPE	58.060	57.493	59.036	58.243	58.038
CI90[a]	0.914	0.892	0.886	0.886	0.893
CI90^b_{OD}		0.917	0.921	0.890	0.900
CI90^c_{WR}		0.895	0.890	0.886	0.894
CI90^d_{TG}		0.909	0.922	0.936	0.958
CI90^e_{TL}		0.901	0.898	0.911	0.928
Fail Rate[f]	0.000	0.000	0.000	0.000	0.000

[a] 90 % confidence interval coverage based on standard errors without overdispersion.
[b] 90 % confidence interval coverage using classical overdispersion
[c] 90 % confidence interval coverage using a weighted regression overdisperion estimator
[d] 90 % confidence interval coverage using a global trimmed mean overdisperion estimator
[e] 90 % confidence interval coverage using a local trimmed mean overdisperion estimator
[f] failure of the estimator due to lack of convergence or excessively large estimates or standard errors

Table 2: Results for bias, RMSPE, confidence interval coverage, and failure rate for simulation experiment 2. Details on column and row labels are given in Table 1 and the text. An example of a single simulated data set is given in Figure 5. Row names are described in Table 1.

		Knots			
	SRS	$K_C = 3$ $K_F = 8$	$K_C = 5$ $K_F = 16$	$K_C = 7$ $K_F = 24$	$K_C = 9$ $K_F = 32$
Bias	79.234	-1.333	-7.632	14.511	13.856
RMSPE	104.979	66.347	68.846	68.311	68.527
CI90[a]	0.726	0.883	0.864	0.883	0.883
CI90^b_{OD}		0.913	0.927	0.894	0.892
CI90^c_{WR}		0.889	0.874	0.888	0.886
CI90^d_{TG}		0.901	0.913	0.952	0.973
CI90^e_{TL}		0.893	0.887	0.919	0.939
Fail Rate[f]	0.000	0.000	0.000	0.000	0.001

Table 3: Results for bias, RMSPE, confidence interval coverage, and failure rate for simulation experiment 3. Details on column and row labels are given in Table 1 and the text. An example of a single simulated data set is given in Figure 5. Row names are described in Table 1.

		Knots			
	SRS	$K_C = 3$ $K_F = 8$	$K_C = 5$ $K_F = 16$	$K_C = 7$ $K_F = 24$	$K_C = 9$ $K_F = 32$
Bias	214.816	-2.389	-4.365	-2.919	-1.637
RMSPE	235.713	79.207	79.250	79.285	80.175
CI90[a]	0.774	0.780	0.777	0.780	0.783
CI90^b_{OD}		0.807	0.790	0.788	0.787
CI90^c_{WR}		0.913	0.903	0.864	0.848
CI90^d_{TG}		0.930	0.935	0.938	0.947
CI90^e_{TL}		0.877	0.877	0.872	0.901
Fail Rate[f]	0.000	0.000	0.000	0.000	0.018

Table 4: Results for bias, RMSPE, confidence interval coverage, and failure rate for simulation experiment 4. Details on column and row labels are given in Table 1 and the text. An example of a single simulated data set is given in Figure 5. Row names are described in Table 1.

		Knots			
	SRS	$K_C = 3$ $K_F = 8$	$K_C = 5$ $K_F = 16$	$K_C = 7$ $K_F = 24$	$K_C = 9$ $K_F = 32$
Bias	148.523	5.179	3.440	7.287	14.629
RMSPE	163.516	60.403	61.021	62.102	64.136
CI90[a]	0.834	0.837	0.828	0.831	0.827
CI90^b_{OD}		0.847	0.831	0.833	0.828
CI90^c_{WR}		0.937	0.926	0.917	0.910
CI90^d_{TG}		0.920	0.916	0.905	0.906
CI90^e_{TL}		0.892	0.892	0.878	0.867
Fail Rate[f]	0.000	0.000	0.000	0.000	0.242

FIGURES

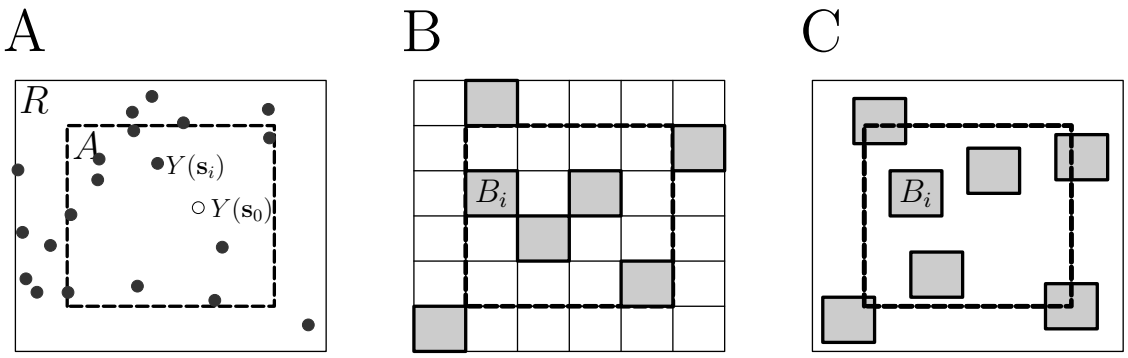

Figure 1: A. The domain of interest is R (thin solid line) with a spatial random field throughout, where the random variable at location $\mathbf{s}_i$ is denoted $Y(\mathbf{s}_i)$. Block kriging predicts the average or total for region A (heavy dashed line) B. A finite population version of block kriging. The samples are on a regular grid and a finite number of $\{Y(B_i)\}$ exhaustively fills both R and A, and the goal is to predict $Y(A) \equiv \sum_A Y(B_i)$ from a sample of $\{Y(B_i)\} \subseteq R$. C. The situation where the sample units have substantial area, but do not form a regular grid within R or A.

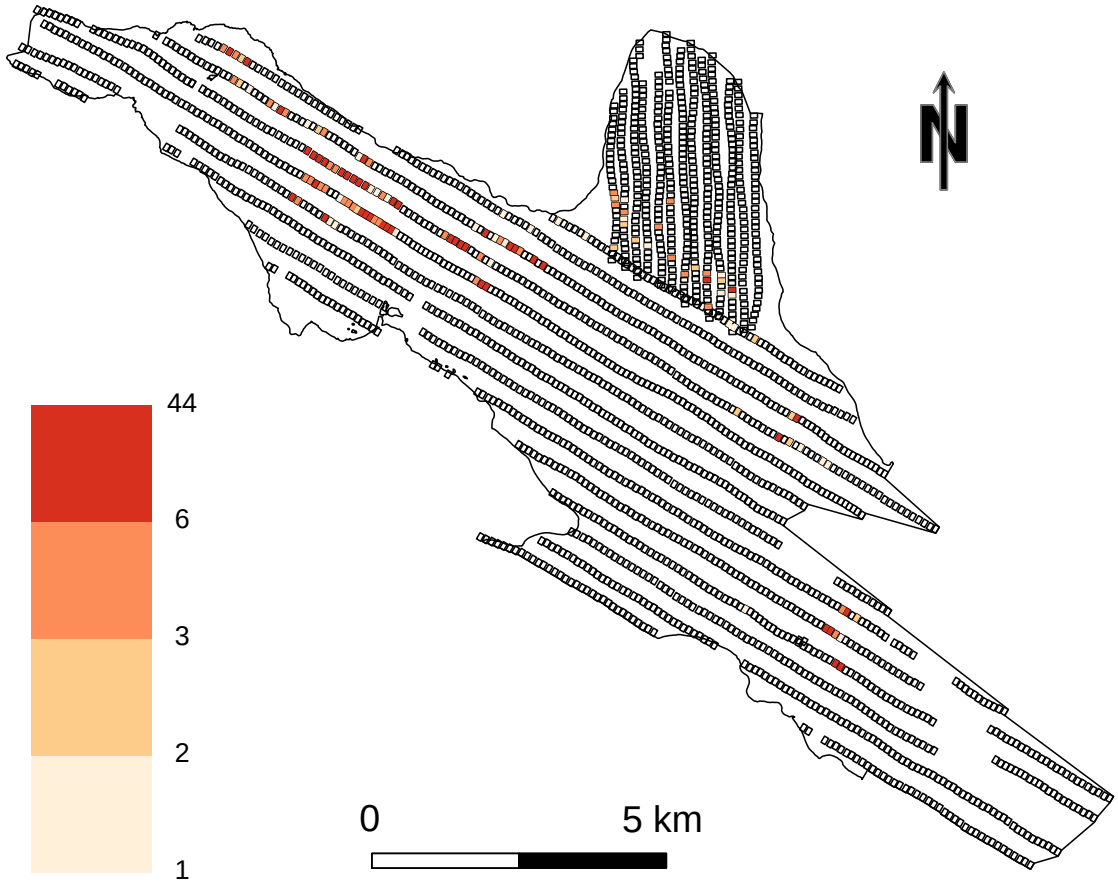

Figure 2: Example data set of aerial surveys for harbor seals conducted on 11 August 2008 in Icy Bay, Alaska. The outlines of aerial photographs are shown within the study area. Open plots have 0 seals, and darker shaded plots have more seals.

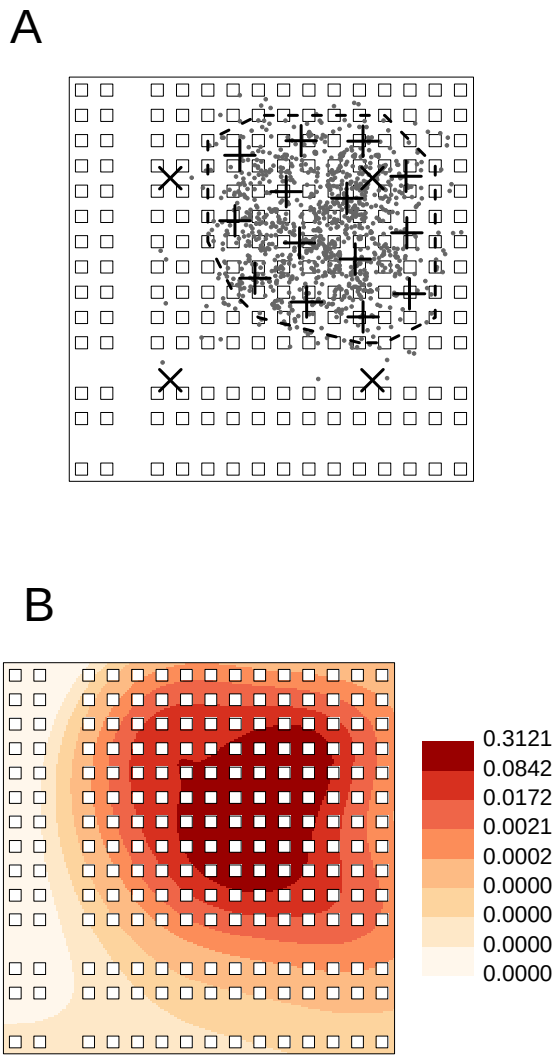

Figure 3: A. One simulated realization, where all simulated points are shown as grey dots, from simulation experiment 3, described in Section 3.4. The coarse-scale knot locations are shown with an "×" while the fine-scale knots are shown with a "+". The fine-scale knots are contained within the convex polygon given by the dashed lines, which bounds the centroids of plots containing nonzero counts. B. The fitted intensity surface throughout $\mathcal{U}$, scaled to the size of the prediction block, to yield the expected count per prediction block.

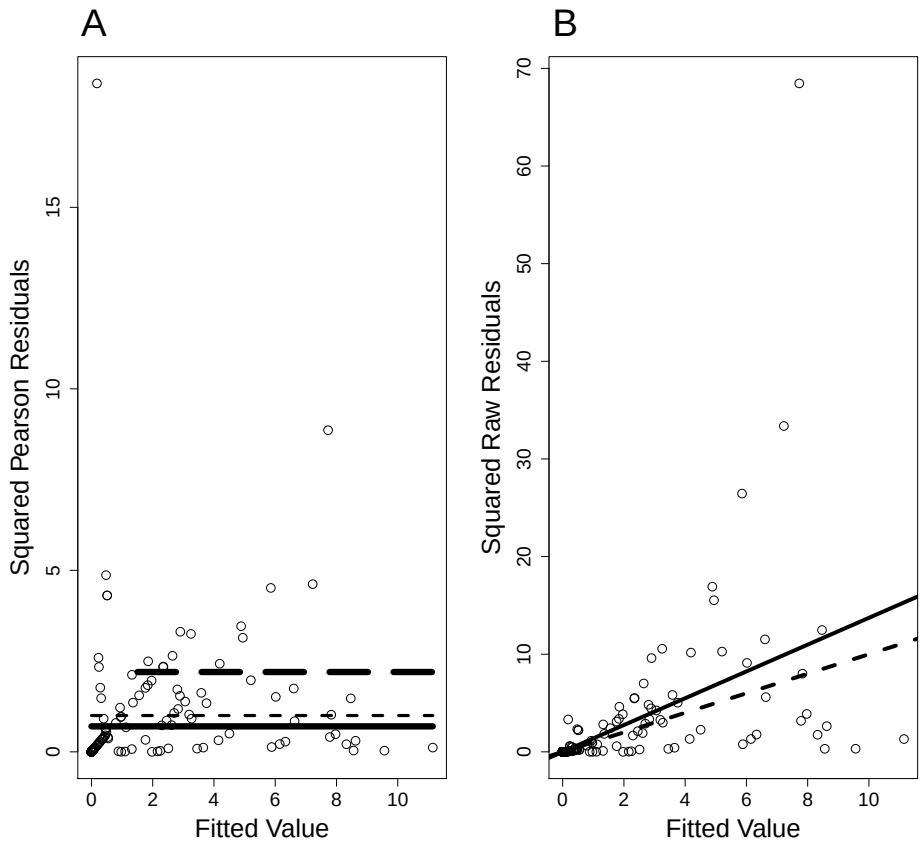

Figure 4: A. Squared Pearson residuals $(y_i - \phi_i)^2/\phi_i$ plotted against the fitted values ϕ_i for the example simulation in Figure 3. The short-dashed line is constant at 1, and the traditional overdispersion estimator is the solid line below the dashed line. The upper long-dashed line is the constant value of the trimmed overdispersion estimator, where only the upper 25% of the ordered values of the fits were used, and the line starts at the lowest of these fitted values. B. Squared raw residuals $(y_i - \phi_i)^2$ plotted against the fitted values ϕ_i. The regression estimator of overdispersion is the slope of the solid line, and the dashed line is the one-to-one line.

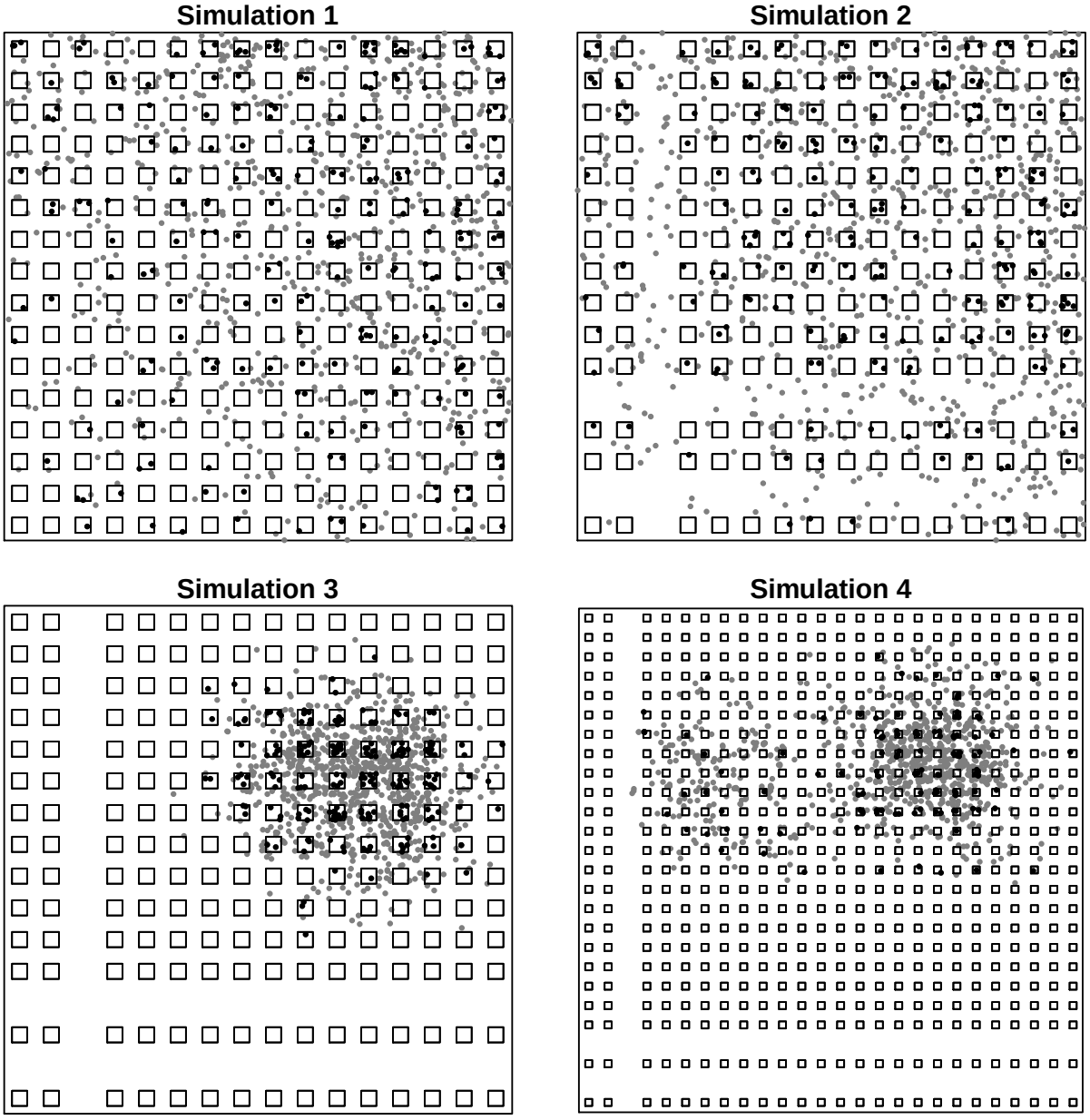

Figure 5: Examples of simulated data used to test methods. All simulated points are shown as grey and black dots. Sample units are shown as squares, and black dots were in sample units while the grey dots were out. The four types of simulations are described in the text.

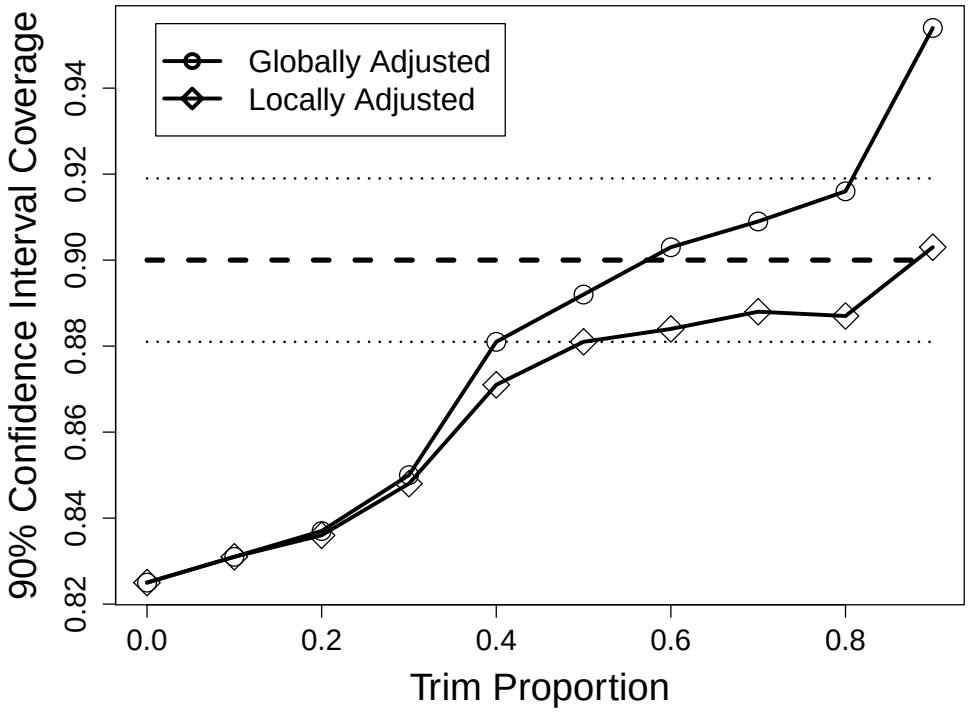

Figure 6: The 90% confidence interval coverage for various overdispersion trim proportions, using $K_C = 5$ and $K_F = 16$ for 1000 simulations from simulation experiment 4. The 90 % line is shown as a dashed horizontal line, and the dotted horizontal lines show the 95% bounds for an estimator that had a true coverage of 0.90.

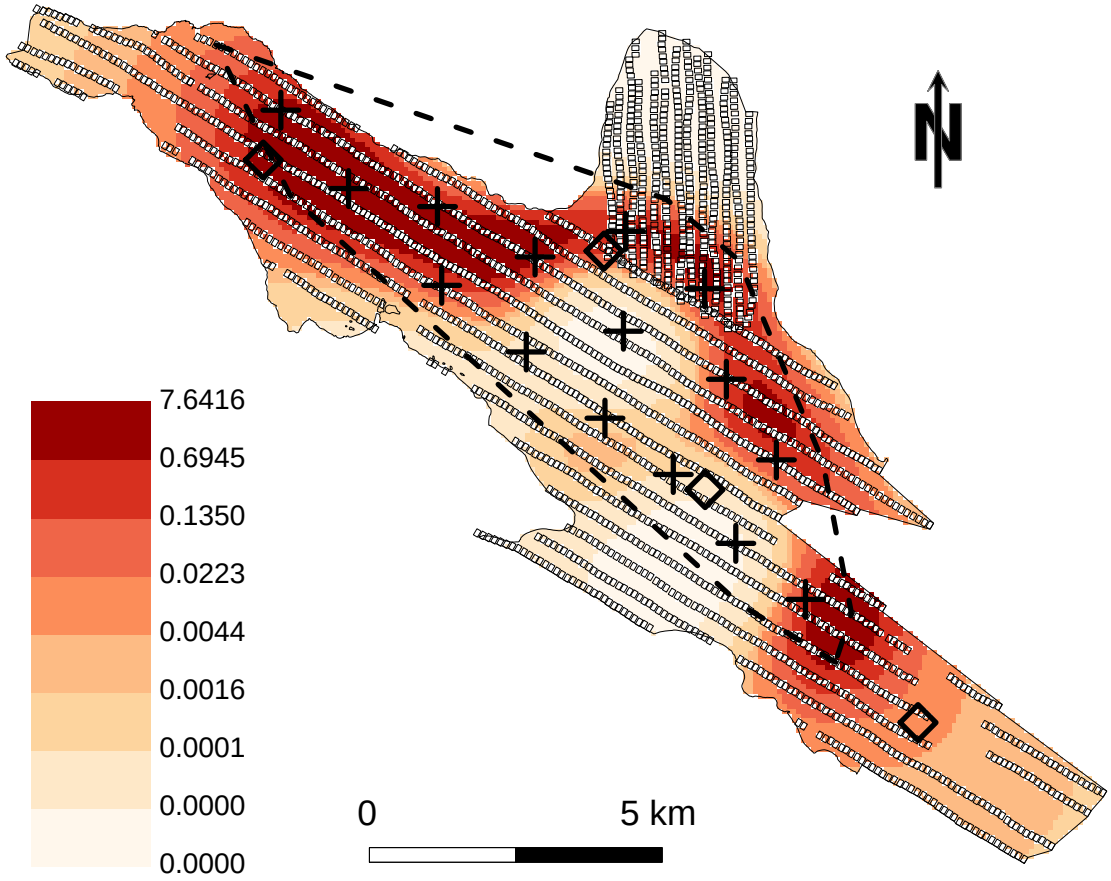

Figure 7: The study area for the harbor seal data with the fitted intensity surface throughout $\mathcal{U}$, scaled to yield the expected count per prediction grid block. The coarse scale knots are shown as open diamonds while the fine scale knots are shown as crosses. The minimum convex polygon enclosing all plots with nonzero counts is given by the dashed line.