Spatial logistic regression and change-of-support in Poisson point processes

A. Baddeley

CSIRO Mathematics, Informatics & Statistics
Leeuwin Centre, 65 Brockway Rd, Floreat WA 6014, Australia
e-mail: Adrian.Baddeley@csiro.au
and
School of Mathematics & Statistics, University of Western Australia

M. Berman

CSIRO Mathematics, Informatics & Statistics, Sydney, Australia

N.I. Fisher

University of Sydney, Australia

A. Hardegen

School of Mathematics & Statistics, University of Western Australia

R.K. Milne

School of Mathematics & Statistics, University of Western Australia

D. Schuhmacher

Institute of Mathematical Statistics and Actuarial Science, University of Bern, Switzerland

R. Shah

School of Mathematics & Statistics, University of Western Australia

and

R. Turner

Starpath Project, University of Auckland, New Zealand

Abstract: In Geographical Information Systems, spatial point pattern data are often analysed by dividing space into pixels, recording the presence or absence of points in each pixel, and fitting a logistic regression. We study weaknesses of this approach, propose improvements, and demonstrate an application to prospective geology in Western Australia. Models based on different pixel grids are incompatible (a 'change-of-support' problem) unless the pixels are very small. On a fine pixel grid, a spatial logistic

regression is approximately a Poisson point process with loglinear intensity; we give explicit distributional bounds. For a loglinear Poisson process, the optimal parameter estimator from pixel data is *not* spatial logistic regression, but complementary log-log regression with an offset depending on pixel area. If the pixel raster is randomly subsampled, logistic regression is conditionally optimal. Bias and efficiency depend strongly on the spatial regularity of the covariates. For discontinuous covariates, we propose a new algorithmic strategy in which pixels are subdivided, and demonstrate its efficiency.

Contents

1. Introduction

This paper investigates techniques used in Geographical Information Systems (GIS) to analyse spatial point pattern data. Figure 1 shows a motivating example, a map of the spatial locations of gold deposits and other geological features in a survey region. The goal of analysis is to predict the occurrence of gold deposits using other covariates. Further details are given in Section 3.

In statistical science, spatial point pattern data such as Figure 1 are usually analysed by applying spatial point process methods [29, 59, 43] to the exact spatial coordinates of the data points. However in GIS practice, the first step in analysing a spatial point pattern is usually to discretise it.

Logistic regression is commonly used in GIS to predict the probability of occurrence of mineral deposits [1, 20], archaeological finds [69, 49], landslides [21, 35] and other events which can be treated as points at the scale of interest. The study region is divided into pixels; in each pixel the presence or absence of any data points is recorded; then logistic regression [58, 31, 42] is used to predict the probability of the presence of a point as a function of predictor variables. Pixel-based spatial logistic regression for point events was developed in geology by Agterberg [1] on the suggestion of Tukey [74]. It was later independently rediscovered in archaeology [69, 39, 49, 50] and is now standard in GIS.

The implications of logistic regression are not well understood in the GIS literature; some writers describe it as a "nonparametric" technique [51, p. 24], and the interpretation of the fitted parameters is widely held to be obscure.
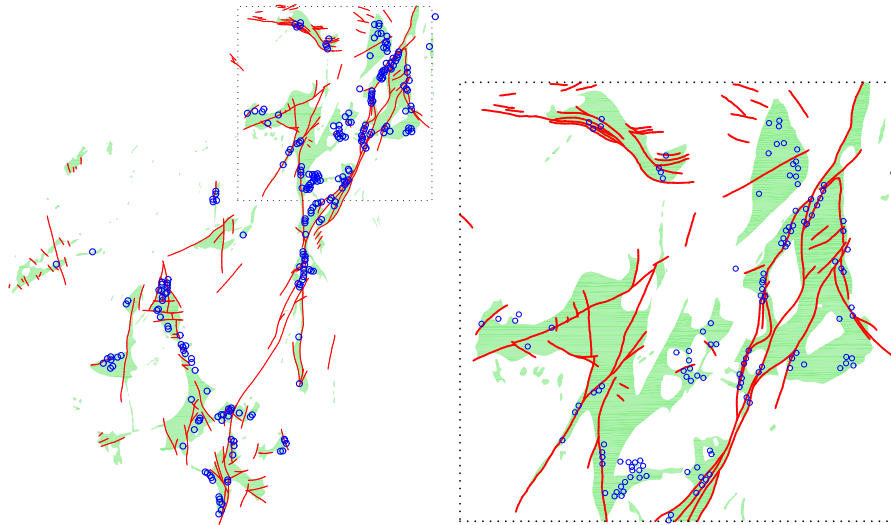
FIG 1. *Murchison data: 255 gold deposits (blue circles), geological faults (red solid lines) and greenstone outcrop (light green shading) in the Murchison area of Western Australia.* Left: *full survey region,* 330×400 *km.* Right: *magnified subregion,* 150×150 *km. Reproduced by kind permission of the Geological Survey of Western Australia.*

Nonetheless, it is clear that a pixel-based logistic regression is approximately equivalent to a Poisson point process model for the original data points.

Pixel-based logistic regression is not necessarily worse than other techniques, because all point process methods involve numerical approximations. The likelihood of a spatial Poisson point process is not generally available in closed form, so approximations are unavoidable. Spatial discretization has the advantage of converting the point process to a generalized linear model. Lewis [54] and Brillinger [16, 18, 17] showed that the likelihood of a general point process in one-dimensional time, or a Poisson point process in higher dimensions, can be usefully approximated by the likelihood of logistic regression for the discretised process. Asymptotic equivalence was established in [15]. This makes it practicable to fit spatial Poisson point process models of general form to point pattern data [14, 22, 8, 9] by enlisting the efficient and reliable software already developed for logistic regression. Approximation of a stochastic process by a generalized linear model is now commonplace in applied statistics [55, 56, 57].

Pixel discretisation is not fundamentally objectionable, but it tends to obscure the stochastic model. GIS literature is unclear about the precise meaning and implications of pixel-based logistic regression; the interpretation of a fitted model; prediction from the model; and the relationship between models fitted using different pixel sizes.

Hence, there are immediate benefits for GIS users in recognising that pixel-based logistic regression is essentially equivalent to a Poisson point process. A wide variety of statistical tools for inference, model selection, validation and prediction is available for Poisson processes. This will be outlined in a companion paper [7].

On the other hand, it appears to be unknown precisely how much efficiency is lost by discretising a point pattern onto a pixel grid. There are also unanswered questions about the relationship between different models and the optimal choice of fitting technique. Further practical algorithmic problems are encountered in fitting pixel-based logistic regressions; GIS manuals suggest pragmatic solutions, without casting much light on the problems.

Thus, what seems to be missing is a detailed study of pixel-based logistic regression from the standpoints of stochastic modelling and statistical methodology. This is the aim of the present paper. Specifically we examine the implications of pixel-based logistic regression as a stochastic model; the relevance of pixel size; the precise connection to Poisson point process models; and the performance of parameter estimators and algorithms. Our major findings are summarised in the next section.

2. Summary of findings

2.1. Pixel size

Of course, counting points inside pixels is a special case of aggregating spatial data into discrete geographical areas. This has been studied extensively

in epidemiology and in ecological and environmental statistics [32, 79, 78, 77]. However, most epidemiological research deals with geographical areas that are predetermined regions of appreciable size. In pixel-based regression, the pixel grid is chosen arbitrarily, and is usually very fine, which leads to a different set of methodological and practical problems.

When the pixel grid can be chosen arbitrarily, important questions include the equivalence of models fitted using different pixel grids (the "modifiable area unit problem" [61] or "change-of-support" [36, 11, 24]), the relation between discrete and continuous spatial models ("ecological fallacy" [66]) and bias due to aggregation over pixels ("ecological bias" [77, 78] or aggregation bias [27, 3]). GIS literature seems unclear on these questions, except to acknowledge that the interpretation of the fitted probabilities and model parameters clearly depends on the pixel size. The interpretation of fitted parameters is often said to be obscure [82, p. 175] and is typically based only on the sign of the slope parameters [35, pp. 405–407].

One of our unexpected findings is that it may be *impossible* to reconcile two spatial logistic regression models that were fitted to the same spatial point pattern data using different pixel grids (Section 5.1). Two such models are logically incompatible except in simple cases. Thus, there is no physical process which satisfies a logistic regression model whenever it is discretised on any pixel grid. The implication is that two research teams who apply spatial logistic regression to the same data, but using different pixel sizes, may obtain results that cannot be reconciled. A spatial logistic regression, fitted to data from a fully-explored survey region, cannot always be extrapolated to make predictions in a prospective exploration region.

2.2. *Poisson point process*

In the GIS community the accepted practical remedy for such inconsistency is to take a very fine pixel grid. As the pixel size tends to zero, spatial logistic regressions with the same form of linear predictor are asymptotically equivalent, provided the linear predictor includes an intercept term.

However, the limit of spatial logistic regressions as pixel size tends to zero is a Poisson point process. Essentially this follows from the assumption of independent responses in logistic regression. A more precise result is given in Section 6 using distributional approximation. Furthermore, the use of the logistic link implies that the intensity of the limiting Poisson point process is a *loglinear* function of the spatial covariates, termed a *modulated* Poisson process by Cox [23]. Thus, contrary to the general view in the GIS literature, there *is* a clear physical meaning for the model parameters of spatial logistic regression, when pixels are small (cf. [10]). The Poisson model allows prediction of several quantities of interest, as we discuss in a companion paper on applications [7].

Thus, spatial logistic regression is physically meaningful only when pixels are extremely small, in which case it is equivalent to a Poisson point process with loglinear intensity. In the remainder of the paper we assume a loglinear

Poisson point process model holds (for point events in the continuous spatial domain) and investigate optimal parameter estimation. In particular we compare the statistical performance of spatial logistic regression with the theoretically optimal method.

2.3. Estimation

We find that, in many cases, logistic regression using a fine grid of pixels is practically equivalent to the theoretical ideal of maximum likelihood estimation using the exact spatial coordinates of the data points. However, statistical performance and optimality depend very strongly on the regularity of the spatial covariates as functions of spatial location. If the covariates are piecewise constant inside each pixel, then the pixel presence/absence data conform exactly to a *complementary log–log regression,* i.e. binary regression with the link $g(p) = \log(-\log(1-p))$ instead of the logit link $g(p) = \log(p/(1-p))$. Complementary log–log regressions on different pixel grids are compatible. Logistic regression estimates of the parameters have bias of order $O(a)$ and efficiency of order $1 - O(a)$ where a is the pixel area. Alternatively if the covariates are smooth functions, then the pixel presence/absence data do not conform exactly to any generalized linear model. Logistic regression and complementary log-log regression estimates have bias of order $O(\delta)$ and efficiency of order $1 - O(\delta)$ where δ is pixel diameter. If the covariates are not smooth, logistic regression and complementary log-log regression estimates may converge poorly. All three scenarios are relevant in practice, but only the last one causes difficulty.

However, using a very fine grid of pixels may be computationally expensive and can cause numerical instability in model-fitting algorithms. In a fine pixel grid, almost all pixels do not contain a data point so that the overwhelming majority of responses are zeroes. Taylor series approximation of the loglikelihood may be inaccurate, leading to numerical instability and the Hauck-Donner effect [40]. This is likely if one covariate is highly influential [75, p. 198], intuitively because this approaches a scenario where the MLE does not exist [72].

To avoid such problems, a strategy popular in the GIS literature is to randomly sub-sample those pixels with value 0, while retaining all the pixels with value 1, and to apply logistic regression (with an offset term) to the selected pixels. Conditional on the sample, logistic regression is optimal (due to the relation between loglinear Poisson and logistic regressions) but subsampling inflates the bias and variance.

We propose a new algorithmic strategy for increasing efficiency and reducing computational load. This is applicable to the case where a covariate is piecewise smooth on each of several subdomains $D_1, \ldots, D_m$ whose boundaries are very irregular. Take a coarse pixel grid and subdivide each pixel S_j into sub-pixels $S_{jk} = S_j \cap D_k$. Treating the sub-pixels (S_{jk}) as a new pixel grid, apply logistic regression or complementary log–log regression with the appropriate offset. This strategy can greatly increase efficiency.

2.4. Plan of paper

The plan of the paper is as follows. The geological example in Figure 1 is discussed in detail in Section 3. Basic definitions are stated in Section 4. Incompatibility of logistic regressions using different pixel sizes is studied in Section 5. Poisson approximation results are given in Section 6. In the remainder of the paper we assume a loglinear Poisson point process model. Section 7 discusses ('exact') maximum likelihood for this model based on observation of the exact location of each point. Section 8 discusses maximum likelihood based on discretised observations on a pixel grid. In special cases, the discretised observations conform exactly to a generalized linear model, as explained in Section 9. Otherwise, spatial logistic regression is an approximation, as discussed in Section 10. The bias incurred by this approximation is calculated in Section 11. Two synthetic examples are analysed completely in Section 12. The GIS subsampling strategy is analysed in Section 13. Our new strategy of splitting pixels is presented in Section 14. The geological example (Section 3) is analysed in full in Section 15. Conclusions and recommendations are summarised in Section 16.

3. Motivating example

3.1. Data

Figure 1 shows the spatial locations of gold deposits and associated geological features in the Murchison area of Western Australia. The data were extracted from a large scale (1:500,000) study of the Murchison area by the Geological Survey of Western Australia [80]. The study region is contained in a 330×400 kilometre rectangle. At this scale, gold deposits are point-like. Gold deposits in this region should occur only in greenstone bedrock. 'Outcrop' is bedrock that is exposed at the surface and therefore observable in a geological survey. Geological faults can be observed reliably only within the greenstone outcrop region. However, some faults have been extrapolated (by geological "interpretation") beyond the outcrop.

Prospective geology aims to predict the occurrence of valuable deposits, or at least to identify areas of higher 'prospectivity', where a deposit is relatively likely. Surveys such as Figure 1 can be used as training data for learning about relationships between the abundance of deposits and geological covariates such as geochemistry and the proximity of other features. The appropriate relationship is effectively a point process model. A fitted model can then be applied in a new exploration area to predict the spatially varying intensity of deposits.

The ancient bedrock of Western Australia has been heavily eroded and leached, removing much geologically useful information from the surface. This militates against the use of geomagnetic or gravitational measurement, alluvial chemistry and other techniques. It is in this context that traditional surveys like Figure 1 are important.

The Murchison data have been analysed by various methods [19, 33, 37, 46]. The consensus is that there is strong evidence that the gold deposit pattern has higher intensity at locations closer to the faults.

3.2. Traditional analysis

To apply spatial logistic regression, we divided the study region into an $N \times N$ grid of pixels. For each pixel j, we recorded the binary response $y_j = 1$ if at least one gold deposit fell inside the pixel, and $y_j = 0$ otherwise. The two covariates were the greenstone indicator g_j (equal to 1 if the centre of pixel j is inside the greenstone area, and 0 otherwise) and the fault distance d_j (the distance in kilometres from the centre of pixel j to the nearest fault). We then fitted the two logistic regression models

$$\log \frac{p_j}{1 - p_j} \quad = \quad \beta_0 + \beta_1 g_j + \beta_2 d_j \tag{1}$$

$$\log \frac{p_j}{1 - p_j} \quad = \quad \gamma_0 + \gamma_1 d_j \tag{2}$$

where $p_j = \mathbb{P}\{Y_j = 1\}$. Table 1 shows the fitted coefficients obtained using different pixel resolutions N.

The standard iteratively-reweighted least squares fitting algorithm in the R system [63, 75] reported numerical exceptions when fitting model (2), but not with model (1).

Having fitted a model, the usual practice in geological applications is to display a pixel image of the fitted probabilities $\hat{p}_j$ as shown in Figure 2. Direct interpretation of the fitted parameters from spatial logistic regression is widely regarded as difficult in GIS literature [82, p. 175]. The intercept clearly depends strongly on pixel size; usually the other parameters are only given a qualitative interpretation based on the sign of the coefficient (cf. [35, pp. 405–407]).

The estimates of the slope parameter γ_1 in (2) are approximately independent of pixel size, but the estimates of factor effects β_1, β_2 in (1) change by up to 50% as pixel size is reduced. This effect is discussed in Section 5. The negative values of β_2, γ_1 are consistent with the interpretation that gold deposits are more abundant near a fault. Further analysis is reported in Section 15.

TABLE 1
Estimates of the parameters in the spatial logistic regression models (1) and (2) for the Murchison data, obtained using different pixel grids. The parameters $\beta_0, \beta_1, \gamma_0$ are dimensionless while β_2, γ_1 have units km^{-1}

grid	β_0	β_1	β_2	γ_0	γ_1
64×64	-2.64	2.19	-0.156	-1.16	-0.270
128×128	-3.96	2.30	-0.143	-2.27	-0.271
256×256	-5.45	2.53	-0.126	-3.60	-0.265
512×512	-7.31	2.91	-0.110	-4.98	-0.269
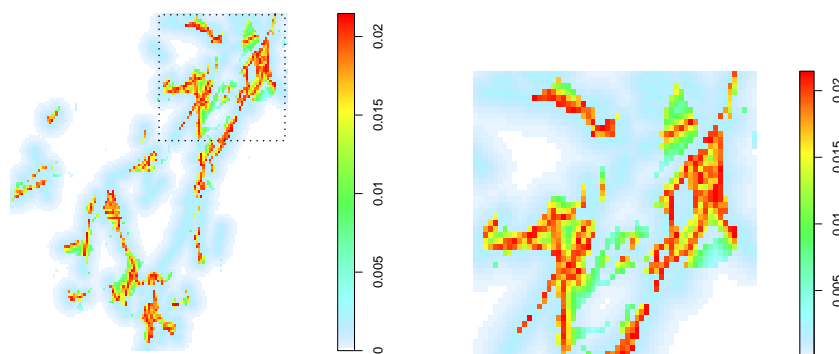
FIG 2. *Fitted pixel occupancy probabilities for the Murchison data of Figure 1 based on a logistic regression model (1) on a* $128{\times}128$ *pixel grid.* Left: *full survey region;* Right: *magnified subregion.*

4. Definitions

The following definitions will be used throughout the paper. The study area is a compact region $W \subset \mathbb{R}^d$ (usually $d = 2$) in which we observe a realisation of a spatial point process $\boldsymbol{X}$. A *pixel grid* $\mathcal{G}$ is a partition of W into disjoint measurable sets $S_j \subset W$ of area (or d-dimensional volume) a_j and diameter δ_j, for $j = 1, \dots, J$. The typical pixel grid is a rectangular array in which all pixels have the same area; pixels of unequal area may be necessary on a curved surface (such as the Earth or celestial sphere). Irregular geographical areas can be regarded as generalised pixels, so that our results also apply to aggregated point counts in spatial epidemiology. Let $N_j = N(\boldsymbol{X} \cap S_j)$ be the number of points of $\boldsymbol{X}$ falling in pixel S_j, and $Y_j = \mathbb{I}\{N_j > 0\}$ the indicator of the event that at least one point of $\boldsymbol{X}$ falls in S_j. Write $\mu_j = \mathbb{E}N_j$ for the expected number of points in S_j, and $p_j = \mathbb{E}Y_j = \mathbb{P}\{N_j > 0\}$ for the probability that at least one point of $\boldsymbol{X}$ falls in S_j.

A spatial logistic regression model for $\boldsymbol{X}$ with respect to $\mathcal{G}$ assumes that the random variables $Y_1, \dots, Y_n$ are independent with

$$\log \frac{p_j}{1 - p_j} = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j \tag{3}$$

for $j = 1, \dots, n$, where $\boldsymbol{\beta}$ is a k-dimensional vector parameter and $\boldsymbol{z}_j$ is a k-dimensional vector of covariates associated with pixel S_j.

The general theory does not specify how the pixel covariate values $\boldsymbol{z}_j$ should be determined when an artificial pixel grid $\mathcal{G}$ is superimposed on the domain W in continuous space. Suppose that $\boldsymbol{Z} : W \to \mathbb{R}^k$ is a vector-valued function providing the values $\boldsymbol{Z}(u)$ of a k-dimensional covariate at any location $u \in W$. The computationally simplest way to define $\boldsymbol{z}_j$ is to choose a representative or 'central' point $c_j \in S_j$ for each pixel S_j, and to read the covariate value at this

location,

$$\boldsymbol{z}_j = Z(c_j). \tag{4}$$

Alternatively, for suitable types of covariate $\boldsymbol{Z}$, the pixel value $\boldsymbol{z}_j$ may be defined as the mean value over the pixel,

$$\boldsymbol{z}_j = \frac{1}{a_j} \int_{S_j} \boldsymbol{Z}(u) \, \mathrm{d}u \tag{5}$$

where again a_j is the area (or d-dimensional volume) of the pixel.

Statistical properties are influenced by the spatial regularity of the function $\boldsymbol{Z}$. We will consider three different cases. In the first and most optimistic case $\boldsymbol{Z}$ is *constant within each pixel*,

$$\boldsymbol{Z}(u) = \boldsymbol{z}_j \text{ for } u \in S_j, \tag{6}$$

so that (4) and (5) agree. Covariates of this kind include indicators of broad geological classification, and national boundaries. In the second case $\boldsymbol{Z}$ is a Lipschitz function, i.e. satisfies $\|\boldsymbol{Z}(u) - \boldsymbol{Z}(v)\| \leq C\|u - v\|$ where $C < \infty$ is constant. Examples include distance transforms (e.g. $\boldsymbol{Z}(u)$ is the distance from u to the nearest geological fault), geographical coordinates, and kernel-smoothed geochemical assay values. In the third case, $\boldsymbol{Z}$ is a non-Lipschitz function, typified by indicators of very irregular spatial domains such as exposed rock areas.

5. Effect of pixel size

5.1. Incompatibility of models based on different pixel grids

Here we show that spatial logistic regressions at two different pixel scales can be mathematically inconsistent. Let $\mathcal{G} = \{S_j, j = 1, \ldots, J\}$ and $\mathcal{H} = \{T_k, k = 1, \ldots, K\}$ be two pixel grids on W. For convenience, assume that $\mathcal{H}$ is a refinement of $\mathcal{G}$, obtained by subdividing each pixel S_j in $\mathcal{G}$ into exactly r subpixels from $\mathcal{H}$,

$$S_j = \bigcup_{k \in K(j)} T_k \tag{7}$$

where $K(j)$ is the set of indices k such that $T_k \subset S_j$. Now let $\boldsymbol{X}$ be a point process on W, and assume that $\boldsymbol{X}$ satisfies a spatial logistic regression with respect to both grids $\mathcal{G}$ and $\mathcal{H}$, with linear predictors of the same form, say

$$\log \frac{p_j}{1 - p_j} = \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{z}(S_j) \tag{8}$$

$$\log \frac{q_k}{1 - q_k} = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}(T_k) \tag{9}$$

where $p_j = \mathbb{P}\{N(\boldsymbol{X} \cap S_j) > 0\}$ and $q_k = \mathbb{P}\{N(\boldsymbol{X} \cap T_k) > 0\}$ are the presence probabilities for the pixels in $\mathcal{G}$ and $\mathcal{H}$ respectively. The k-dimensional vector

covariate $\boldsymbol{z}(S_j)$ associated with a pixel S_j is assumed to be determined either by pixel averaging (5) or sampling (4) of a covariate function $\boldsymbol{Z} : W \to \mathbb{R}^k$. The canonical parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are also k-dimensional vectors.

The independence of pixel responses in $\mathcal{H}$, and the decomposition (7), imply that

$$1 - p_j = \prod_{k \in K(j)} (1 - q_k) \tag{10}$$

since both sides give the probability of having no points in S_j. This key equation identifies the scaling property of the pixel probabilities: the quantity $\log(1 - p_j)$ is additive over pixels. It also implies that the three equations (8), (9) and (10) could be *inconsistent* in general.

Lemma 1. *Assume that the covariate function $\boldsymbol{Z}$ is constant within pixels of $\mathcal{G}$, i.e. (6) holds.*

(a) *Suppose there are at least three pixels $S_j, S_{j'}, S_{j''}$ in $\mathcal{G}$ such that $p_j, p_{j'}, p_{j''}$ are distinct values and $\boldsymbol{z}_j, \boldsymbol{z}_{j'}, \boldsymbol{z}_{j''}$ are collinear vectors. Then the three equations (8), (9) and (10) are logically inconsistent.*

(b) *Suppose that $\{\boldsymbol{Z}(u) : u \in W\}$ is a basis for $\mathbb{R}^k$. Then (8), (9) and (10) are consistent.*

For the proof, see Appendix A.1. Lemma 1 has several implications in practice. First, two research teams who apply spatial logistic regression to the same data, but using different pixel sizes, may obtain results that cannot be reconciled. Second, there does not necessarily exist a random phenomenon in continuous space which we can rely upon to satisfy the assumptions of spatial logistic regression after it is discretised on a pixel raster of arbitrary scale. Third, a spatial logistic regression, fitted to data from a fully-explored survey region, can be extrapolated to make predictions in a prospective exploration region, only if the covariate data for the prospective region are converted to the same pixel grid size as the original survey data.

Example 1. *Consider spatial logistic regression with an intercept and a single slope parameter, $\log(p/(1 - p)) = \beta_0 + \beta_1 \zeta$. Thus $\boldsymbol{Z}(u) = (1, \zeta(u))$ where $\zeta : W \to \mathbb{R}$ is a real-valued spatial covariate. Assume ζ is constant within each pixel of $\mathcal{G}$, but takes at least three different values on W. Then Lemma 1(a) applies, and spatial logistic regressions on two grids $\mathcal{G}$ and $\mathcal{H}$ are logically incompatible.*

Inconsistency does not arise when the covariate $\boldsymbol{Z}$ is binary– or factor–valued, provided the model is not overdetermined.

Example 2. *Let the covariate $T(u)$ take categorical values (e.g. representing soil type or rock type) with possible categories $1, 2, \ldots, K$. Effectively this divides the domain W into subregions $W_k = \{u \in W : T(u) = k\}$. Assume that T is constant within pixels of $\mathcal{G}$. Then a spatial logistic regression (8) asserts that the presence probability is constant across pixels within a sub-domain W_k. The canonical covariate is $\boldsymbol{Z}(u) = (t_1(u), \ldots, t_K(u))$ where $t_k(u) = \mathbb{I}\{T(u) = k\}$ is the dummy variable associated with category k. The values of $\boldsymbol{Z}$ are the standard*

unit basis of $\mathbb{R}^K$*. Hence Lemma 1(b) applies, and two spatial logistic regressions* (8) *and* (9) *are logically consistent.*

Example 3. *Let* $\boldsymbol{Z}(u) = (1, a(u), b(u))$ *where* $a(u) = \mathbb{I}\{u \in A\}$ *and* $b(u) = \mathbb{I}\{u \in B\}$ *are the indicators of two sets* $A, B \subset W$ *with nonempty intersection. It can be shown [38] that logistic regressions on different pixel grids are incompatible.*

5.2. Approximate compatibility on fine grids

Equations (8), (9) and (10) become asymptotically equivalent when pixels are small, under certain conditions. Heuristically, as pixel size tends to zero, the presence probabilities p_j tend to zero, and the left side of (8) is $\log(p_j/(1-p_j)) \sim \log p_j$ where $\sim$ denotes asymptotic equivalence. The relation (69) for conversion between two pixel grids becomes $p_j = 1 - (1 - q_k)^r \sim rq_k$, so that

$$\log p_j \approx \log r + \log q_k, \qquad k \in K(j).$$

Here r is the number of pixels in grid $\mathcal{H}$ into which each pixel in $\mathcal{G}$ is subdivided. Equivalently $\log p_j - \log a_j \approx \log q_k - \log b_k$, where a_j, b_k are the areas (or d-dimensional volumes) of pixels $S_j \in \mathcal{G}$ and $T_k \in \mathcal{H}$ respectively. Hence the following adjusted versions of the spatial logistic regressions (8) and (9) are approximately equivalent, with the *same* coefficient vector $\boldsymbol{\beta}$:

$$\begin{aligned}
\log \frac{p_j}{1 - p_j} &= \log a_j + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}(S_j) \\
\log \frac{q_k}{1 - q_k} &= \log b_k + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}(T_k)
\end{aligned} \tag{11}$$

In practical applications, this requires that the linear predictor of the model include either an intercept term which can absorb the adjustment $\log a_j$, or a separate offset term $\log a_j$. A proof of asymptotic equivalence is given in Section 6.4.

6. Poisson process approximation

A spatial logistic regression on a fine pixel grid is approximately a Poisson point process. Intuitively this is clear, since spatial logistic regression assumes independence between pixels, and the presence probabilities p_j in a fine grid are small. The classical Poisson limit theorem is almost enough to establish this.

In Section 6.3 we obtain a stronger result giving a precise bound on the distance between the distribution of the two point processes in a quite general setting where we do not assume any particular model for the presence probabilities. A limit theorem for the special case of the logistic model (3) is then given as a consequence in Section 6.4.

6.1. Point processes

It is useful to represent a spatial point pattern $\boldsymbol{x}$ in $W \subset \mathbb{R}^d$ as a finite measure $\boldsymbol{x} = \sum_{i=1}^{n} \delta_{x_i}$ where $x_i \in W$ are the individual points $(i = 1, \ldots, n)$ and δ_u is the Dirac measure which puts mass 1 at location $u \in \mathbb{R}^d$. We shall allow a point pattern $\boldsymbol{x}$ to have multiple points at the same location. A point pattern without multiple points is called *simple*. The number of points of $\boldsymbol{x}$ falling in a set $A \subseteq W$ will be written as $\boldsymbol{x}(A)$ in the general case, and as $N(\boldsymbol{x} \cap A)$ if $\boldsymbol{x}$ is simple. Denote by $\mathfrak{N}$ the set of totally finite point measures (i.e. finite point patterns with multiplicity) on W, equipped with the (standard) vague topology and its Borel σ-algebra [44, Sections 4.1 and 15.7]. This is the minimal σ-algebra which renders $\boldsymbol{X} \mapsto \boldsymbol{X}(A)$ measurable for every compact $A \subseteq W$. Thus point processes are interpreted as random point measures.

6.2. A metric between distributions of point processes

Define the metric d_0 on W by $d_0(x, y) = \min(1, \|x - y\|)$ for all $x, y \in W$. Define the d_1-*distance* between two point measures $\boldsymbol{x} = \sum_{i=1}^{\boldsymbol{x}(W)} \delta_{x_i}$ and $\boldsymbol{y} = \sum_{i=1}^{\boldsymbol{y}(W)} \delta_{y_i}$ in $\mathfrak{N}$ by $d_1(\boldsymbol{x}, \boldsymbol{y}) = \min_{\tau \in \Sigma_n} \left[\frac{1}{n} \sum_{i=1}^{n} d_0(x_i, y_{\tau(i)}) \right]$ if $\boldsymbol{x}(W) = \boldsymbol{y}(W) = n > 0$, while $d_1(\boldsymbol{x}, \boldsymbol{y}) = 0$ if both patterns are empty $(\boldsymbol{x}(W) = \boldsymbol{y}(W) = 0)$ and $d_1(\boldsymbol{x}, \boldsymbol{y}) = 1$ if $\boldsymbol{x}(W) \neq \boldsymbol{y}(W)$, where Σ_n denotes the set of permutations of $\{1, 2, \ldots, n\}$. That is, in the most interesting case where $\boldsymbol{x}(W) = \boldsymbol{y}(W) > 0$, the d_1-distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is the average distance between points under an optimal pairing. It can be seen that $d_1 \leq 1$ is a metric that metrises the weak topology, which is the same as the vague topology, since W is compact.

Now let $\mathfrak{P}(\mathfrak{N})$ be the space of probability measures on $\mathfrak{N}$, that is, the space of point process distributions. Define the d_2-*distance* between two distributions $P, Q \in \mathfrak{P}(\mathfrak{N})$ by

$$d_2(P, Q) = \sup_{f \in \mathcal{F}_2} \left| \int f \, dP - \int f \, dQ \right| = \min_{\substack{\boldsymbol{X} \sim P \\ \boldsymbol{Y} \sim Q}} \mathbb{E}[d_1(\boldsymbol{X}, \boldsymbol{Y})] \qquad (12)$$

where $\mathcal{F}_2 = \{ f : \mathfrak{N} \to \mathbb{R}; \ |f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq d_1(\boldsymbol{x}, \boldsymbol{y}) \text{ for all } \boldsymbol{x}, \boldsymbol{y} \in \mathfrak{N} \}$.

Lemma 2. d_2 *is a metric on* $\mathfrak{P}(\mathfrak{N})$ *with values* ≤ 1 *that metrises convergence in distribution; that is, for point processes* $\boldsymbol{X}, \boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots$ *we have* $d_2\big(\mathscr{L}(\boldsymbol{X}^{(m)}), \mathscr{L}(\boldsymbol{X})\big) \to 0$ *if and only if* $\boldsymbol{X}^{(m)}$ *converges to* $\boldsymbol{X}$ *in distribution as* $m \to \infty$. $\qquad \square$

For the proof see [70] and [83]. See [71] for an equivalent metric $\bar{d}_2 \leq d_2$.

6.3. A general bound for Poisson process approximation of the observed process

Theorem 3. *Let* $\boldsymbol{X}$ *be a simple point process on a compact set* $W \subset \mathbb{R}^d$. *Suppose there is a pixel grid* $\mathcal{G}$ *with pixels* S_j *of diameter* δ_j, *for which* $Y_j = \mathbb{I}\{N_j > 0\}$

are independent, where $N_j = N(\boldsymbol{X} \cap S_j)$, for $j = 1, \ldots, J$. Let $\boldsymbol{\lambda}$ be any finite measure on W. Then

$$
\begin{aligned}
d_2\big(\mathscr{L}(\boldsymbol{X}), \mathrm{Pois}(\boldsymbol{\lambda})\big) &\leq 2 \max_j \delta_j + \mathbb{P}\{\exists j : N_j > 1\} \\
&+ \sum_{j=1}^{J} \big|\boldsymbol{\lambda}(S_j) - \mathbb{P}\{N_j > 0\}\big| \\
&+ 6.5 \max_j \mathbb{P}\{N_j > 0\};
\end{aligned}
\tag{13}
$$

if the expectation measure of $\boldsymbol{X}$ is $\boldsymbol{\lambda}$, this yields

$$
d_2\big(\mathscr{L}(\boldsymbol{X}), \mathrm{Pois}(\boldsymbol{\lambda})\big) \leq 2 \max_j \delta_j + \sum_{j=1}^{J} \big(\mathbb{E}N_j - \mathbb{P}\{N_j = 1\}\big) + 6.5 \max_j \mathbb{E}N_j. \tag{14}
$$

See Appendix A.2 for the proof. Note that we do not assume any particular model for the pixel probabilities $p_j = \mathbb{P}\{N_j > 0\}$, only that the Y_j are independent. This result applies equally to logistic regression, complementary log–log regresssion and other models mooted in the GIS literature. Moreover, the proof technique is flexible enough to allow for pixel values that are weakly dependent at the cost of additional small terms in the upper bound.

From the above result we obtain a Poisson limit theorem for a quite general sequence of point processes whose discretisations exhibit independent pixel values on successively finer pixel grids.

Definition 1. *A sequence $\big(\mathcal{G}^{(m)}\big)_{m \in \mathbb{N}}$ of partitions $\mathcal{G}^{(m)} = \{S_{m,j}; 1 \leq j \leq J_m\}$ of W is called* **nested** *if $S_{m-1,i} \cap S_{m,j} \in \{S_{m,j}, \emptyset\}$ for all $m \in \mathbb{N}$, $1 \leq i \leq J_{m-1}$, and $1 \leq j \leq J_m$. It is* **asymptotically infinitesimal** *if the pixel diameters $\delta_{m,j} = \mathrm{diam}(S_{m,j})$ and d-dimensional volumes $a_{m,j} = |S_{m,j}|$ satisfy $\delta_m = \max_{1 \leq j \leq J_m} \delta_{m,j} \to 0$ and hence $a_m = \max_{1 \leq j \leq J_m} a_{m,j} \to 0$ as $m \to \infty$.*

Corollary 4. *Let $(\boldsymbol{X}^{(m)})_{m \in \mathbb{N}}$ be a sequence of simple point processes on the compact set $W \subset \mathbb{R}^d$ with intensities $(\lambda^{(m)})_m$, such that $\lambda^{(m)} \to \lambda$ in L^1 as $m \to \infty$ for some integrable $\lambda : W \to \mathbb{R}_+$. Let $\big(\mathcal{G}^{(m)}\big)_{m \in \mathbb{N}}$ be an asymptotically infinitesimal, nested sequence of partitions of W. Assume that the random variables $N_{m,j} = N(\boldsymbol{X}^{(m)} \cap S_{m,j})$ satisfy*

(i) $\big(\mathbb{I}\{N_{m,j} > 0\}\big)_{1 \leq j \leq J_m}$ is independent for every m;
(ii) $\sum_{j=1}^{J_m} \big(\mathbb{E}N_{m,j} - \mathbb{P}\{N_{m,j} = 1\}\big) \to 0$ as $m \to \infty$.

Then $\boldsymbol{X}^{(m)}$ converges in distribution as $m \to \infty$ to a Poisson process with intensity λ.

See Appendix A.2 for the proof.

Definition 2. *In the notation of Corollary 4, the sequence of pixel grids $\big(\mathcal{G}^{(m)}\big)_m$ is an* **asymptotically exact** *discretisation of the sequence of point processes $(\boldsymbol{X}^{(m)})_m$ if*

$$
\mathbb{P}\{\exists j : N_{m,j} > 1\} \to 0 \quad as \quad m \to \infty.
$$

Asymptotic exactness is slightly weaker than condition (ii) of Corollary 4. By Hölder's inequality, for any $\alpha > 0$,

$$\sum_{j=1}^{J_m} \left(\mathbb{E} N_{m,j} - \mathbb{P}\{N_{m,j} = 1\} \right)$$

$$= \sum_{j=1}^{J_m} \mathbb{E}\left[N_{m,j} \, \mathbb{I}\{N_{m,j} > 1\} \right]$$

$$\leq \mathbb{E}\left[\boldsymbol{X}^{(m)}(W) \max_{1 \leq j \leq J_m} \mathbb{I}\{N_{m,j} > 1\} \right]$$

$$\leq \left(\mathbb{E}\left[\boldsymbol{X}^{(m)}(W)^{1+\alpha} \right] \right)^{1/(1+\alpha)} \left(\mathbb{E}\left[\max_{1 \leq j \leq J_m} \mathbb{I}\{N_{m,j} > 1\} \right] \right)^{\alpha/(1+\alpha)}.$$

Thus (ii) holds if the discretisation of $\boldsymbol{X}^{(m)}$ is asymptotically exact and the sequence $\left(\mathbb{E}\left[\boldsymbol{X}^{(m)}(W)^{1+\alpha} \right] \right)_m$ is bounded for some $\alpha > 0$.

6.4. Poisson process approximation for spatial logistic regression

In spatial logistic regression, the relationship between the presence probabilities p and the explanatory covariates $\boldsymbol{Z}$ is controlled by the form of the logistic link function. In the fine-pixel limit, this also controls the relationship between the intensity of the limiting Poisson process and the covariates. As pixel size tends to zero we have $p \to 0$ and $\log(p/(1-p)) \sim \log p$, so that the limiting Poisson intensity is a loglinear function of the covariates.

Theorem 5. *Assume that the observation window W is compact and the covariate function $\boldsymbol{Z} : W \to \mathbb{R}^k$ is continuous. Let $\left(\mathcal{G}^{(m)} \right)_{m \in \mathbb{N}}$ be an asymptotically infinitesimal, nested sequence of partitions of W. For each m, let $\boldsymbol{X}^{(m)}$ be a simple point process which, when discretised, satisfies the assumptions of spatial logistic regression; that is, writing $N_{m,j} = N(\boldsymbol{X}^{(m)} \cap S_{m,j})$, the indicators $\mathbb{I}\{N_{m,j} > 0\}$ are independent for $j = 1, 2, \ldots, J_m$ and the presence probabilities $p_{m,j} = \mathbb{P}\{N_{m,j} > 0\}$ satisfy*

$$\log \frac{p_{m,j}}{1 - p_{m,j}} = \log a_{m,j} + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_{m,j} \ .$$

The covariate value $\boldsymbol{z}_{m,j}$ is either a sampled value (4) or the average value (5) of $\boldsymbol{Z}$ over the pixel $S_{m,j}$. Suppose furthermore that the discretisation of $\boldsymbol{X}^{(m)}$ is asymptotically exact.

Then $\boldsymbol{X}^{(m)}$ converges in distribution as $m \to \infty$ to a Poisson process with loglinear intensity λ, where $\log \lambda(u) = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u)$.

7. Maximum likelihood for loglinear Poisson point process

In the remainder of this paper we assume a loglinear Poisson model holds, and investigate the optimal way to fit such a model. We begin with a brief study of 'exact' maximum likelihood for the point process.

7.1. Assumptions

Assume $\boldsymbol{X}$ is a Poisson point process with log intensity

$$\log \lambda_{\boldsymbol{\beta}}(u) = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u) \tag{15}$$

on a bounded set $W \subset \mathbb{R}^d$, where $\boldsymbol{\beta} \in \mathbb{R}^k$ is the parameter, and $\boldsymbol{Z} : W \to \mathbb{R}^k$ is such that $\lambda_{\boldsymbol{\beta}}$ is integrable on W. Let $\mathcal{G}$ be a pixel grid consisting of pixels S_j of area a_j, $j = 1, \ldots, J$. The corresponding pixel counts N_j and presence/absence indicators Y_j were defined in Section 4. Under the Poisson model, the N_j are independent Poisson random variables with means

$$\mathbb{E}N_j = \mu_j = \int_{S_j} \lambda_{\boldsymbol{\beta}}(u)\,\mathrm{d}u = \int_{S_j} \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u))\,\mathrm{d}u \tag{16}$$

and the presence-absence indicators are independent Bernoulli variables with

$$\mathbb{P}\{Y_j = 1\} = p_j = 1 - \exp(-\mu_j). \tag{17}$$

7.2. Point process likelihood

First consider the likelihood of the point process $\boldsymbol{X}$ based on complete information about the realisation

$$\boldsymbol{x} = \{x_1, \ldots, x_n\}, \quad x_i \in W. \tag{18}$$

The loglikelihood of the Poisson point process with intensity $\lambda_{\boldsymbol{\beta}}(u)$, of general form, is

$$\log L(\boldsymbol{\beta}) = \sum_i \log \lambda_{\boldsymbol{\beta}}(x_i) - \int_W \lambda_{\boldsymbol{\beta}}(u)\,\mathrm{d}u \tag{19}$$

as shown in [26, pp. 23, 213]. For the loglinear intensity (15) the loglikelihood takes the form

$$\ell(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) = \boldsymbol{\beta}^{\mathrm{T}} \sum_i \boldsymbol{Z}(x_i) - \int_W \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u))\,\mathrm{d}u. \tag{20}$$

This model is a canonically parametrised exponential family [13, p. 113], [53, pp. 23–24] with concave loglikelihood, with efficient score

$$U(\boldsymbol{\beta}) = \sum_i \boldsymbol{Z}(x_i) - \int_W \boldsymbol{Z}(u)\lambda_{\boldsymbol{\beta}}(u)\,\mathrm{d}u \tag{21}$$

and Fisher information

$$\begin{aligned} I(\boldsymbol{\beta}) &= \int_W \left[\frac{\partial}{\partial \boldsymbol{\beta}} \log \lambda_{\boldsymbol{\beta}}(u)\right]\left[\frac{\partial}{\partial \boldsymbol{\beta}} \log \lambda_{\boldsymbol{\beta}}(u)\right]^{\mathrm{T}} \lambda_{\boldsymbol{\beta}}(u)\,\mathrm{d}u \\ &= \int_W \boldsymbol{Z}(u)\boldsymbol{Z}(u)^{\mathrm{T}} \lambda_{\boldsymbol{\beta}}(u)\,\mathrm{d}u. \end{aligned} \tag{22}$$

For a general concave loglikelihood, with arbitrary data, the existence and uniqueness of the maximum likelihood estimate (MLE) are not guaranteed, and depend on the absence of directions of recession [67, 81, 72, 2, 34]. For the loglikelihood (20) of the loglinear Poisson process, with data obtained from a realization of the Poisson process, the MLE exists uniquely except in trivial cases.

Lemma 6. *Let $\boldsymbol{X}$ be a Poisson point process with loglinear intensity* (15). *Assume the matrix*

$$M = \int_W \boldsymbol{Z}(u)\,\boldsymbol{Z}(u)^{\mathrm{T}}\,\mathrm{d}u$$

is positive definite. Then the model is identifiable. Conditional on the event that $\sum_i \boldsymbol{Z}(x_i) \neq 0$, the MLE $\hat{\boldsymbol{\beta}} = \mathsf{argmax}\,\ell(\boldsymbol{\beta})$ exists and is unique.

For a sketch proof, see Appendix A.4. The condition on M is analogous to requiring the design matrix of a GLM to have full rank. Note that the Lemma implicitly assumes that the data $\boldsymbol{Z}(x_i)$ are consistent with the model, since they are obtained from a realization of $\boldsymbol{X}$. The condition that $\sum_i \boldsymbol{Z}(x_i) \neq 0$ is satisfied if, for example, the model contains an intercept term and the realization $\boldsymbol{x}$ is not empty. Explicit examples are studied in Section 12.

7.3. Asymptotic normality

For very general Poisson point process models, Rathbun and Cressie [64, Theorem 11, pp. 135–136] and Kutoyants [48, Thm. 2.4, p. 51] proved that the MLE is consistent, asymptotically normal and asymptotically efficient in a 'large domain' limiting regime. See also [47]. The following simpler conditions (adapted from [64]) are sufficient here. We use $||\cdot||_p$ to denote the usual L^p norm in $\mathbb{R}^d$.

Theorem 7. *Let $W_n, n = 1, 2, \ldots$ be an increasing sequence of compact subsets of $\mathbb{R}^d$ whose union is $\mathbb{R}^d$. Let $\boldsymbol{X}$ be a Poisson point process on $\mathbb{R}^d$ with loglinear intensity* (15) *where $\boldsymbol{\beta} \in \mathbb{R}^k$ is a vector parameter and $\boldsymbol{Z} : \mathbb{R}^d \to \mathbb{R}^k$ a vector covariate. Assume*

1. *$\boldsymbol{Z}$ is uniformly bounded, i.e. there is $K < \infty$ such that $||\boldsymbol{Z}(u)||_2 < K$ for all $u \in \mathbb{R}^d$;*
2. *the matrix*

$$M_n = \int_{W_n} \boldsymbol{Z}(u)\,\boldsymbol{Z}(u)^{\mathrm{T}}\,\mathrm{d}u$$

   is positive definite for all n; and
3. *$|M_n| \to \infty$, where $|\cdot|$ is the matrix norm $|H| = \sup\{||Hx||_1 : ||x||_2 = 1\}$.*

Let $\hat{\boldsymbol{\beta}}_n$ be the MLE of $\boldsymbol{\beta}$ based on $\boldsymbol{X} \cap W_n$. Then $\hat{\boldsymbol{\beta}}_n$ is a consistent, asymptotically normal and asymptotically efficient estimator of $\boldsymbol{\beta}$ with asymptotic variance $I(\boldsymbol{\beta})^{-1}$, where $I(\boldsymbol{\beta})$ is defined in (22) *with W replaced by W_n.*

The integral in the loglinear Poisson process likelihood (20) is the Laplace transform of the covariate function $\boldsymbol{Z}$. Consequently the likelihood is not generally available in closed form, and the maximum likelihood estimate does not

generally permit exact analytic solution. Some form of numerical approximation of the integrals appearing in (20)–(22) is required. Approximations relevant to the spatial Poisson point process have been described in [14, 8, 22].

8. Maximum likelihood for discretised observations

In this section we again assume the data points are generated by a loglinear Poisson point process model (15) but now suppose that observations are discretised on a pixel grid. The available observations are either the counts N_j or the indicators Y_j. Various pathologies can now occur.

8.1. Counts

The loglikelihood based on the pixel counts N_j is

$$\log L_d(\boldsymbol{\beta}) = \sum_j (N_j \log \mu_j - \mu_j) \tag{23}$$

where the means μ_j depend on $\boldsymbol{\beta}$ through (16). The score and Fisher information are respectively

$$U_d(\boldsymbol{\beta}) = \sum_j \left(\frac{N_j}{\mu_j} - 1\right) \frac{\partial \mu_j}{\partial \boldsymbol{\beta}} \tag{24}$$

$$I_d(\boldsymbol{\beta}) = \sum_j \frac{1}{\mu_j} \left[\frac{\partial \mu_j}{\partial \boldsymbol{\beta}}\right] \left[\frac{\partial \mu_j}{\partial \boldsymbol{\beta}}\right]^{\mathrm{T}} \tag{25}$$

where

$$\frac{\partial \mu_j}{\partial \boldsymbol{\beta}} = \int_{S_j} \boldsymbol{Z}(u) \lambda_{\boldsymbol{\beta}}(u) \, \mathrm{d}u. \tag{26}$$

The score could also be derived from the missing data principle [73]. Jensen's inequality (or the missing data principle for exponential families [62]) shows that $I(\boldsymbol{\beta}) - I_d(\boldsymbol{\beta})$ is nonnegative definite.

Note that (23) is an incomplete data loglikelihood [28]. It is generally not an exponential family, and may exhibit pathological behaviour. The Hessian is

$$J_d(\boldsymbol{\beta}) = \int_W \boldsymbol{Z}(u)\,\boldsymbol{Z}(u)^{\mathrm{T}} \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}(u)) \, \mathrm{d}u - \sum_j N_j v_j(\boldsymbol{\beta}) \tag{27}$$

where

$$v_j(\boldsymbol{\beta}) = \frac{1}{\mu_j} \frac{\partial^2 \mu_j}{\partial \boldsymbol{\beta}^2} - \frac{1}{\mu_j^2} \left(\frac{\partial \mu_j}{\partial \boldsymbol{\beta}}\right)^2. \tag{28}$$

In view of (26) and

$$\frac{\partial^2 \mu_j}{\partial \boldsymbol{\beta}^2} = \int_{S_j} \boldsymbol{Z}(u)\,\boldsymbol{Z}(u)^{\mathrm{T}} \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}(u)) \, \mathrm{d}u, \tag{29}$$

we recognise $v_j(\boldsymbol{\beta})$ as the variance-covariance matrix of $\boldsymbol{Z}(U)$ where U is a random point in S_j with probability density $f(u) = \lambda_{\boldsymbol{\beta}}(u)/\mu_j$. Hence (27) is the difference between two nonnegative definite matrices. For some realisations (N_j) and parameter values $\boldsymbol{\beta}$ the Hessian is not positive definite, and the log-likelihood may have multiple local maxima. Indeed the discretised model may be unidentifiable.

Example 4. *In the unit square $D = [0,1] \times [0,1]$ define the chessboard indicator function $h(x,y) = r([8x] + [8y])$ where $[\cdot]$ denotes integer part, and $r(n) = \mathbb{I}\{n \text{ is odd}\}$. The Poisson process with intensity $\lambda_{\boldsymbol{\beta}}(x,y) = \exp(\beta_0 + \beta_1 h(x,y))$ is identifiable. Now let $\mathcal{G}$ be the 4×4 grid of square pixels in D. We have $\mu_j = \exp(\beta_0)(1 + \exp(\beta_1))/32$ for all j. So the discretised model on $\mathcal{G}$ is not identifiable.*

The only case in which (23) is guaranteed to have a unique local and global maximum is when $v_j(\boldsymbol{\beta}) = 0$ in (27)–(28), which implies that the covariate $\boldsymbol{Z}(u)$ is constant within each pixel. This case is considered in Section 9.

8.2. Indicators

The presence/absence indicators Y_j are Bernoulli with success probabilities

$$p_j = p_j(\boldsymbol{\beta}) = 1 - \exp(-\mu_j) = 1 - \exp\left(-\int_{S_j} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u)} \, \mathrm{d}u\right). \tag{30}$$

We have

$$\frac{\partial p_j}{\partial \boldsymbol{\beta}} = (1 - p_j)\frac{\partial \mu_j}{\partial \boldsymbol{\beta}} \tag{31}$$

so the loglikelihood for Bernoulli responses

$$\log L_b(\boldsymbol{\beta}) = \sum_j (Y_j \log \frac{p_j}{1 - p_j} + \log(1 - p_j)) \tag{32}$$

has score and Fisher information

$$U_b(\boldsymbol{\beta}) = \sum_j \frac{\partial \mu_j}{\partial \boldsymbol{\beta}} \left(\frac{Y_j}{p_j} - 1\right) \tag{33}$$

$$I_b(\boldsymbol{\beta}) = \sum_j \frac{1 - p_j}{p_j} \left[\frac{\partial \mu_j}{\partial \boldsymbol{\beta}}\right] \left[\frac{\partial \mu_j}{\partial \boldsymbol{\beta}}\right]^{\mathrm{T}}. \tag{34}$$

Jensen's inequality or the missing information principle can again be used to show that $I_d(\boldsymbol{\beta}) - I_b(\boldsymbol{\beta})$ is nonnegative definite. Pathological behaviour may occur here too, as the Hessian is not necessarily positive definite. In Example 4 the discretised model is again unidentifiable.

9. Reduction to generalized linear models

This section considers the special case (6) where the covariate $\boldsymbol{Z}$ is *constant within each pixel*. As shown in Section 8, this is the only case in which the discretised observations conform exactly to a generalized linear model.

9.1. Counts

If the covariate is constant within each pixel (6), the mean count (16) simplifies to

$$\mu_j = a_j \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j) \tag{35}$$

and (23) is the loglikelihood of *loglinear Poisson regression* with an offset term:

$$\log \mu_j = \log a_j + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j. \tag{36}$$

Discretisation causes no loss of information in this case: the counts N_j are sufficient for the original model, and $I(\boldsymbol{\beta}) = I_d(\boldsymbol{\beta})$. The discretised model is a canonically parametrised exponential family.

9.2. Indicators

If the covariate is constant within each pixel (6), the presence probability (30) reduces to

$$p_j = 1 - \exp(-a_j \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j)) \tag{37}$$

so that (32) is the loglikelihood of *complementary log–log regression* with an offset:

$$\log(-\log(1 - p_j)) = \log a_j + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j, \tag{38}$$

rather than logistic regression. That is, the exact likelihood is that of binary regression with the complementary log-log link

$$g(p) = \log(-\log(1 - p)) \tag{39}$$

rather than the logistic link. Information is lost by replacing N_j by Y_j. The Fisher information can be calculated from (34). This model is an exponential family, but is not canonically parametrised. The conditions for uniqueness of the MLE are the standard conditions for this discrete model [58].

Note that complementary log–log regressions on different pixel grids are compatible. Under the assumptions of Section 5.1, complementary log–log regressions on different pixel grids satisfy the multiplicative property (10) exactly.

9.3. A rule of thumb

Comparing (34) with (25) gives a rule-of-thumb for the relative efficiency of the MLE of $\boldsymbol{\beta}$ based on the presence/absence indicators Y_j, relative to the MLE based on the pixel counts N_j. If there is only one real covariate, the relative

efficiency is

$$\frac{I_b(\boldsymbol{\beta})}{I_d(\boldsymbol{\beta})} = \frac{\sum_j (\frac{\partial}{\partial \boldsymbol{\beta}} \mu_j)^2 (1 - p_j)/p_j}{\sum_j (\frac{\partial}{\partial \boldsymbol{\beta}} \mu_j)^2 / \mu_j} = \frac{\sum_j w_j / (e^{\mu_j} - 1)}{\sum_j w_j / \mu_j}$$

where $w_j = (\partial \mu_j / \partial \boldsymbol{\beta})^2$. A rule of thumb (derived by assuming the w_j are all equal) is to approximate

$$\frac{I_b(\boldsymbol{\beta})}{I_d(\boldsymbol{\beta})} \approx \frac{\bar{\mu}}{e^{\bar{\mu}} - 1}$$

where $\bar{\mu} = \frac{1}{N} \sum \mu_j$ is the average expected number of points per pixel. Thus the loss of efficiency should be tolerable as long as $\bar{\mu} < 0.2$, say. For pixels of equal area, $\bar{\mu} = \frac{1}{N} \int_W \lambda_{\boldsymbol{\beta}}(u)\, \mathrm{d}u$ can be estimated by $N(\boldsymbol{X})/N$, the average number of data points per pixel.

9.4. Relations between regression models

The following fundamental relationships give some insight into the preceding results. Let $N_1, \ldots, N_J$ be independent Poisson variables satisfying a loglinear regression $\mu_j = \mathbb{E}N_j = \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j)$. Let $Y_j = \mathbb{I}\{N_j > 0\}$. Then $Y_1, \ldots, Y_J$ satisfy a complementary log–log regression with the same linear predictor, $\log(-\log(1 - p_j)) = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j$ where $p_j = \mathbb{P}\{Y_j = 1\}$. This is well known [58, p. 212]. Now let

$$q_j = \mathbb{P}\{Y_j = 1 \mid N_j \le 1\} = \mathbb{P}\{N_j > 0 \mid N_j \le 1\} = \frac{\mu_j}{1 + \mu_j};$$

then $\log(q_j/(1 - q_j)) = \log \mu_j = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j$. Thus, *conditional on $N_j \le 1$ for all j, the responses Y_j satisfy a logistic regression with the same linear predictor.* One consequence is that spatial logistic regression and complementary log–log regression will be approximately equivalent if $\mathbb{P}\{N_j \ge 2 \text{ for some } j\}$ is negligible.

10. Approximation by generalized linear models

In general, the discretised responses Y_j do not conform exactly to any generalized linear model. This reflects a familiar difficulty with incomplete data likelihoods, in particular those from an exponential family. In the spatial context, it is an instance of the 'ecological fallacy' or 'modifiable area unit problem', closely related to the geostatistical problem of 'change of support' [36, 11, 66, 61, 24].

Both logistic regression and complementary log–log regression are practical approximations. The approximate loglikelihoods can be maximised using standard software, and may have better properties than the true loglikelihoods (23) and (32) of the discretised responses.

10.1. Approximate likelihoods

Section 10.1.1 describes the spatial logistic regression approximation, and Section 10.1.2 investigates spatial binary regression with a complementary log–log link.

10.1.1. Logistic regression

In spatial logistic regression, the true binary loglikelihood (32) is approximated by the loglikelihood of logistic regression

$$
\begin{aligned}
\log L_{\mathrm{LR}}(\boldsymbol{\beta}) &= \sum_j \left(Y_j \log \frac{p_{\mathrm{LR}}(\boldsymbol{\beta}, j)}{1 - p_{\mathrm{LR}}(\boldsymbol{\beta}, j))} + \log(1 - p_{\mathrm{LR}}(\boldsymbol{\beta}, j)) \right) \\
&= \boldsymbol{\beta}^{\mathrm{T}} \sum_j Y_j \boldsymbol{z}_j + \sum_j Y_j \log a_j \\
&\quad - \sum_j \log(1 + \exp(\log a_j + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j))
\end{aligned}
\tag{40}
$$

where $\boldsymbol{z}_j$ are covariate values associated with pixels S_j, and

$$
p_{\mathrm{LR}}(\boldsymbol{\beta}, j) = \frac{\exp(\log a_j + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j)}{1 + \exp(\log a_j + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j)} = \left[1 + \exp(- \log a_j - \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j) \right]^{-1} \tag{41}
$$

is the probability predicted by spatial logistic regression (11) in which the linear predictor includes an offset equal to the log of pixel area. The approximate MLE exists uniquely under usual conditions [58].

Note that the true loglikelihood (32) is not equivalent to (40), and indeed (32) is usually not equivalent to the loglikelihood of any GLM. The mean response $\mathbb{E}Y_j = p_j$ depends on $\boldsymbol{\beta}$ through (30). Except in some extreme cases, this cannot be cast in the form $g(p_j) = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j$ required for a binary GLM. Similarly, the loglikelihood (23) of the counts N_j is generally not equivalent to the loglikelihood of any Poisson GLM.

10.1.2. Complementary log–log regression

The true binary loglikelihood (32) may alternatively be approximated by the loglikelihood of complementary log-log regression (38),

$$
\begin{aligned}
\log L_{\mathrm{CL}}(\boldsymbol{\beta}) &= \sum_j \left(Y_j \log \frac{p_{\mathrm{CL}}(\boldsymbol{\beta}, j)}{1 - p_{\mathrm{CL}}(\boldsymbol{\beta}, j))} + \log(1 - p_{\mathrm{CL}}(\boldsymbol{\beta}, j)) \right) \\
&= \sum_j Y_j \log(1 - \exp(-a_j \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j))) \\
&\quad + \sum_j (1 - Y_j) a_j \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j)
\end{aligned}
\tag{42}
$$

where

$$
p_{\mathrm{CL}}(\boldsymbol{\beta}, j) = 1 - \exp(-a_j \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j)) \tag{43}
$$

is the probability predicted by complementary log-log regression, with surrogate covariate values $\boldsymbol{z}_j$ determined by sampling (4) or averaging (5) the covariate function $\boldsymbol{Z}$.

10.2. Fine-pixel limit

When pixel size reduces to zero (the "fine-pixel limit") we shall prove that the likelihood of spatial logistic regression converges to the exact likelihood, and that the spatial logistic regression estimator $\widetilde{\boldsymbol{\beta}}$ converges to the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$.

To obtain convergence of the loglikelihoods as well as the estimators, the loglikelihood of spatial logistic regression (40) will be adjusted by subtracting $\sum_j Y_j \log a_j$ (where a_j is the area of pixel j) to yield the adjusted loglikelihood

$$\log L_{\mathrm{adj}} = \boldsymbol{\beta}^{\mathrm{T}} \sum_j Y_j \boldsymbol{z}_j - \sum_j \log(1 + a_j \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j)) \tag{44}$$

where Y_j is the presence indicator for pixel j.

Theorem 8. *Consider the Poisson point process model with intensity of the form $\lambda_{\boldsymbol{\beta}}(u) = \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u))$ in a fixed bounded region $W \subset \mathbb{R}^d$, where $\boldsymbol{\beta} \in \Theta = \mathbb{R}^k$ and $\boldsymbol{Z} : W \to \mathbb{R}^k$. Assume the function $\boldsymbol{Z}$ is Lipschitz with constant C, and that $\int_W \boldsymbol{Z}(u) \boldsymbol{Z}(u)^{\mathrm{T}} \mathrm{d}u$ is positive definite. Let $\boldsymbol{\beta}_0$ be the true value of the parameter. Let $\mathcal{G}_n$, $n = 1, 2, \ldots$ be a nested, asymptotically infinitesimal sequence of pixel grids on W satisfying $a_{n+1} \le a_n/2$ where n is the maximum pixel area in $\mathcal{G}_n$.*

Consider the adjusted loglikelihood (44) of spatial logistic regression, in which the covariate value $\boldsymbol{z}_j$ associated with a pixel S_j is taken to be the average (5) of $\boldsymbol{Z}$ over the pixel. With probability 1,

1. *$\log L_{adj}(\boldsymbol{\beta})$ converges pointwise to the exact loglikelihood (20),*
2. *the first and second derivatives of $\log L_{adj}(\boldsymbol{\beta})$ converge pointwise to the corresponding derivatives of the exact loglikelihood, and*
3. *the spatial logistic regression estimator $\widetilde{\beta} = arg\ max\ L_{adj}(\boldsymbol{\beta})$ converges to the maximum likelihood estimator $\widehat{\boldsymbol{\beta}} = arg\ max\ L(\boldsymbol{\beta})$.*

For the proof, see Appendix A.5. A similar result holds for the adjusted loglikelihood of complementary log–log regression.

11. Bias in regression estimates

For practical applications it is important to assess the bias that arises from using logistic regression or complementary log–log regression in a pixel grid of a given resolution.

11.1. Approximate bias of logistic regression

Approximations for the bias can be obtained using a standard linearisation argument. Let $\widetilde{\boldsymbol{\beta}}$ denote the estimate of $\boldsymbol{\beta}$ obtained using logistic regression. This is the zero of the logistic score

$$U_{\mathrm{LR}}(\boldsymbol{\beta}) = \sum_j (Y_j - p_{\mathrm{LR}}(\boldsymbol{\beta}, j)) \boldsymbol{z}_j. \tag{45}$$

The derivative of the logistic score

$$
\begin{aligned}
U'_{\mathrm{LR}}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} U_{\mathrm{LR}}(\boldsymbol{\beta}) & = -\sum_j \frac{a_j e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j}}{(1 + a_j e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j})^2} \boldsymbol{z}_j \, \boldsymbol{z}_j{}^{\mathrm{T}} \\
& = -\sum_j p_{\mathrm{LR}}(\boldsymbol{\beta}, j)(1 - p_{\mathrm{LR}}(\boldsymbol{\beta}, j)) \boldsymbol{z}_j \, \boldsymbol{z}_j{}^{\mathrm{T}} \quad (46)
\end{aligned}
$$

is nonrandom (a consequence of the fact that the logit link is the canonical link for binomial likelihoods) so equals $-I_{\mathrm{LR}}(\boldsymbol{\beta})$ where $I_{\mathrm{LR}}(\boldsymbol{\beta})$ is the Fisher information of logistic regression.

Assume a loglinear Poisson process (15) and denote by $\boldsymbol{\beta}_0$ the true value of $\boldsymbol{\beta}$. Then the true expectation of the logistic score is

$$
\mathbb{E}_0 U_{\mathrm{LR}}(\boldsymbol{\beta}) = \sum_j (p_j - p_{\mathrm{LR}}(\boldsymbol{\beta}, j)) \boldsymbol{z}_j \tag{47}
$$

where $p_j = p_j(\boldsymbol{\beta}_0)$ is the true value of the success probability (30) for the loglinear Poisson process. Expanding $U_{\mathrm{LR}}(\boldsymbol{\beta})$ about $\boldsymbol{\beta}_0$ to first order, $U_{\mathrm{LR}}(\boldsymbol{\beta}) \approx U_{\mathrm{LR}}(\boldsymbol{\beta}_0) - (\boldsymbol{\beta} - \boldsymbol{\beta}_0) I_{\mathrm{LR}}(\boldsymbol{\beta}_0)$. Since $U_{\mathrm{LR}}(\widetilde{\boldsymbol{\beta}}) = 0$, we have $\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \approx I_{\mathrm{LR}}(\boldsymbol{\beta}_0)^{-1} U(\boldsymbol{\beta}_0)$ and heuristically

$$
\mathbb{E}_0[\widetilde{\boldsymbol{\beta}}] - \boldsymbol{\beta}_0 \approx I_{\mathrm{LR}}(\boldsymbol{\beta}_0))^{-1} \mathbb{E}_0 U_{\mathrm{LR}}(\boldsymbol{\beta}_0) \tag{48}
$$

where the right side is equal to

$$
\left[\sum_j p_{\mathrm{LR}}(\boldsymbol{\beta}_0, j)(1 - p_{\mathrm{LR}}(\boldsymbol{\beta}_0, j)) \boldsymbol{z}_j \, \boldsymbol{z}_j{}^{\mathrm{T}} \right]^{-1} \sum_j (p_j - p_{\mathrm{LR}}(\boldsymbol{\beta}_0, j)) \boldsymbol{z}_j \tag{49}
$$

with p_j given by (30) with $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. In test examples (Section 12.2) this is a tolerable approximation to the bias. An asymptotic justification for (48) is discussed in Section 11.2.

In real applications, the approximate bias incurred in spatial logistic regression may be estimated by re-fitting the model using the complementary log-log link, then either (a) evaluating the difference between the parameter estimates obtained using the two link functions, or (b) maximum likelihood estimation, using (49) where p_j and $\boldsymbol{\beta}_0$ are replaced by the fitted probabilities and estimated parameter vector obtained using the complementary log-log link.

11.2. Asymptotics of bias in spatial logistic regression

Asymptotic analysis of the bias approximation (49) is complicated because it involves two asymptotic regimes. Under "large-domain asymptotics" (Section 7.3) the exact maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$ is close to $\boldsymbol{\beta}_0$. Under "fine-pixel asymptotics" (Section 10.2) the spatial logistic regression estimator $\widetilde{\boldsymbol{\beta}}$ converges to the exact maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$. For a suitable sequence of models

with increasing domains W_n and increasingly fine pixel grids $\mathcal{G}_n$, both statements apply and (49) is an asymptotic approximation to the bias. Convergence involves a balance between the effects of reducing pixel size, which tends to reduce discretisation bias, and expanding the domain W_n, which tends to increase bias. The asymptotic order of the expression on the right hand side of (49) depends on the rates of convergence in these two regimes and on the behaviour of the function $\boldsymbol{Z}$.

Lemma 9. *Let $\boldsymbol{X}$ be a Poisson point process with loglinear intensity (15) observed on an increasing sequence of domains $W_n \uparrow \mathbb{R}^d$. Assume the conditions of Theorem 7 hold. Let $\mathcal{G}_n$, $n = 1, 2, \dots$ be a nested sequence of pixel grids on $\mathbb{R}^d$ satisfying the assumptions of Theorem 8 and set $a_n = \max_j a_{n,j}$. Then*

(a) if $\boldsymbol{Z}$ is Lipschitz with constant C on each pixel of $\mathcal{G}_n$, for sufficiently large n, then

$$\mathbb{E}_0(U_{LR}(\boldsymbol{\beta}_0)) = \sum_j \left(\int_{S_{n,j}} \lambda_{\boldsymbol{\beta}_0}(u)\,\mathrm{d}u - a_{n,j} e^{\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{z}_j} \right) \boldsymbol{z}_j + R_n \qquad (50)$$

with $||R_n|| = O(a_n|W_n|)$, and the leading term is

$$||\mathbb{E}_0(U_{LR}(\boldsymbol{\beta}_0))|| \leq O(\delta_n|W_n|); \qquad (51)$$

(b) if $\boldsymbol{Z}$ is constant within pixels of $\mathcal{G}_n$, for all sufficiently large n,

$$\mathbb{E}_0[U_{LR}(\boldsymbol{\beta}_0)] = \frac{1}{2} \sum_j a_{n,j}^2 e^{2\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{z}_j} \boldsymbol{z}_j + R_n \qquad (52)$$

with $||R_n|| \leq O(a_n^2|W_n|)$, and

$$||\mathbb{E}_0[U_{LR}(\boldsymbol{\beta}_0)]|| \leq O(a_n|W_n|). \qquad (53)$$

For a proof see Appendix A.6. The lemma implies that the value of the large-domain asymptotic bias term $I(\boldsymbol{\beta}_0)^{-1}\mathbb{E}_0 U_{\mathrm{LR}}(\boldsymbol{\beta}_0)$ is at most of order $O(\delta_n)$ and at best of order $O(a_n)$, depending on the behaviour of the function $\boldsymbol{Z}$.

Theorem 10 (Large domain, fine-pixel limit). *Assume the conditions of Lemma 9. Let $\widetilde{\boldsymbol{\beta}}_n$ be the estimator of $\boldsymbol{\beta}$ obtained by logistic regression using the pixels of $\mathcal{G}_n \cap W_n$. Write*

$$
\begin{aligned}
I_n &= \frac{1}{|W_n|} \int_{W_n} \lambda(u) \boldsymbol{Z}(u)\,\boldsymbol{Z}(u)^{\mathrm{T}}\,\mathrm{d}u \\
B_n &= \frac{1}{|W_n|} \int_{W_n} \lambda(u)^2 \boldsymbol{Z}(u)\,\mathrm{d}u.
\end{aligned}
$$

Assume that $I_n \to \mathcal{I}$ and $B_n \to B$ where $\mathcal{I}$ is a finite, positive definite matrix, and B is a finite vector. If either

(a) $\boldsymbol{Z}$ *is Lipschitz*, $\delta_n = o(|W_n|^{-1})$, *and*

$$b_n = \mathcal{I}^{-1} \frac{1}{|W_n|} \sum_j \left(\int_{S_{jn}} \lambda(u) \, \mathrm{d}u - a_{jn} e^{\boldsymbol{\beta_0}^{\mathrm{T}} \boldsymbol{z}_j} \right) \boldsymbol{z}_j;$$

or

(b) $\boldsymbol{Z}$ *is constant within pixels of* $\mathcal{G}_n$, *for all sufficiently large* n, *and* $a_n = o(|W_n|^{-1})$, *and*

$$b_n = \mathcal{I}^{-1} \frac{1}{|W_n|} \frac{1}{2} \sum_j a_{jn}^2 e^{2\boldsymbol{\beta_0}^{\mathrm{T}} \boldsymbol{z}_j} \boldsymbol{z}_j,$$

then $\widetilde{\boldsymbol{\beta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\beta}_0$ *and* $|W_n|^{1/2}(\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 - b_n) \to N(0, \mathcal{I}^{-1})$.

The proof of Theorem 10 is standard and is omitted. Note that pixel size must converge to zero at a faster rate in case (a) than allowed in case (b).

11.3. Bias in complementary log-log regression

Similar arguments apply to complementary log-log regression. The surrogate score is

$$U_{\mathrm{CL}}(\boldsymbol{\beta}) = \sum_j \left(\frac{Y_j}{p_{\mathrm{CL}}(\boldsymbol{\beta}, j)} - 1 \right) e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j} \, \boldsymbol{z}_j$$

where $p_{\mathrm{CL}}(\boldsymbol{\beta}, j)$ is given by (43). The derivative of the surrogate score is

$$U'_{\mathrm{CL}}(\boldsymbol{\beta}) = \sum_j \left\{ Y_j \left[\frac{1}{p_{\mathrm{CL}}(\boldsymbol{\beta}, j)} - \frac{1 - p_{\mathrm{CL}}(\boldsymbol{\beta}, j)}{p_{\mathrm{CL}}(\boldsymbol{\beta}, j)^2} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j} \right] - 1 \right\} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j} \, \boldsymbol{z}_j \boldsymbol{z}_j^{\mathrm{T}}.$$

Under the loglinear Poisson point process, the surrogate score and its derivative have ('true') expected values

$$\mathbb{E}_0 U_{\mathrm{CL}}(\boldsymbol{\beta}) = \sum_j \left(\frac{p_j}{p_{\mathrm{CL}}(\boldsymbol{\beta}, j)} - 1 \right) e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j} \boldsymbol{z}_j$$

$$\mathbb{E}_0 U'_{\mathrm{CL}}(\boldsymbol{\beta}) = \sum_j \left\{ p_j \left[\frac{1}{p_{\mathrm{CL}}(\boldsymbol{\beta}, j)} - \frac{1 - p_{\mathrm{CL}}(\boldsymbol{\beta}, j)}{p_{\mathrm{CL}}(\boldsymbol{\beta}, j)^2} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j} \right] - 1 \right\} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j} \, \boldsymbol{z}_j \boldsymbol{z}_j^{\mathrm{T}}$$

where p_j is the true probability (30). So the bias is approximately

$$- \left(\mathbb{E}_0 U'_{\mathrm{CL}}(\boldsymbol{\beta}_0) \right)^{-1} \mathbb{E}_0 U_{\mathrm{CL}}(\boldsymbol{\beta}_0) \tag{54}$$

and a similar analysis can be performed.

12. Theoretical examples

It is useful to study several examples which are computable analytically or with simple numerical approximations.

12.1. Factor covariate

Recall the setting of Example 2 where $T(u)$ is a categorical covariate with possible values $1, 2, \ldots, K$ effectively dividing W into subdomains $W_k = \{u \in W : T(u) = k\}$. Consider a Poisson point process on W with different intensities λ_k on the subdomains W_k, where $\lambda_1, \ldots, \lambda_K$ are parameters. Thus $\lambda(u) = \lambda_{T(u)}$, which can be recast in the form (15) by taking $\beta_k = \log \lambda_k$ and $\boldsymbol{Z}(u) = (t_1(u), \ldots, t_K(u))$ where $t_k(u) = \mathbb{I}\{T(u) = k\}$ is the indicator of W_k.

The exact maximum likelihood estimate of β_k is $\hat{\beta}_k = \log \hat{\lambda}_k$ where $\hat{\lambda}_k$ is the observed intensity of points $\hat{\lambda}_k = N(\boldsymbol{X} \cap W_k)/|W_k|$ on the subregion W_k. The Fisher information matrix (22) is diagonal with entries $\lambda_k |W_k|$.

Divide W into a grid of pixels of equal area a, and suppose $T(u)$ is constant within each pixel. For all pixels S_j contained in W_k, we have $\mu_j = a\lambda_k$ and $\partial \mu_j / \partial \beta_m = \mathbb{I}\{k = m\}\mu_j$. By straightforward calculation of (34) the Fisher information $I_b(\boldsymbol{\beta})$ is the diagonal matrix with entries $\lambda_k |W_k| a\lambda_k/(e^{a\lambda_k} - 1)$. Thus, replacing the pixel counts N_j by the indicators Y_j has a relative efficiency of $a\lambda_k/(e^{a\lambda_k} - 1)$ for each parameter β_k.

12.2. Intensity loglinear in Cartesian coordinates

Consider the Poisson point process in the unit square $D = [0, 1] \times [0, 1]$ with intensity

$$\lambda(x, y) = \exp(\beta_0 + \beta_1 x + \beta_2 y), \tag{55}$$

where $\beta_0, \beta_1, \beta_2$ are parameters. Figure 3 shows one simulated realisation of the model with $\boldsymbol{\beta} = (2, 3, 2)$.

12.2.1. Exact MLE

Suppose the data points are at locations (x_i, y_i) for $i = 1, \ldots, n$. For the model (55), the score is (21) where $\boldsymbol{Z}((x, y)) = (1, x, y)$. The integrals in (21) are separable so that we can simplify the score components to $U_0 = n - e^{\beta_0} s_0(\beta_1) s_0(\beta_2)$,
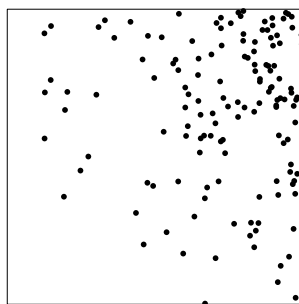

FIG 3. *Simulated realisation of inhomogeneous Poisson process with intensity* (55). *Simulated pattern has 141 points.*

$U_1 = \sum x_i - e^{\beta_0} s_1(\beta_1) s_0(\beta_2)$ and $U_2 = \sum y_i - e^{\beta_0} s_0(\beta_1) s_1(\beta_2)$, where $s_k(\beta) = \int_0^1 x^k e^{\beta x} \, dx$ if $\beta \neq 0$. We have $s_k(0) = 1/k$ and if $\beta > 0$ then $s_0(\beta) = \frac{1}{\beta} \left[e^\beta - 1 \right]$, $s_1(\beta) = \frac{1}{\beta^2} \left[e^\beta (\beta - 1) + 1 \right]$ and $s_2(\beta) = \frac{1}{\beta^3} \left[e^\beta (\beta^2 - 2\beta + 2) - 2 \right]$. Define $t_k(\beta) = s_k(\beta)/s_0(\beta)$. The score equations reduce to

$$t_1(\beta_1) = \bar{x}, \qquad t_1(\beta_2) = \bar{y}, \qquad \beta_0 = \log n - \log s_0(\beta_1) - \log s_0(\beta_2) \qquad (56)$$

which can be solved numerically. By elementary analysis of the Poisson process (using separability), the Fisher information matrix is

$$I = e^{\beta_0} s_0(\beta_1) s_0(\beta_2) \begin{bmatrix} 1 & t_1(\beta_1) & t_1(\beta_2) \\ t_1(\beta_1) & t_2(\beta_1) & t_1(\beta_1) t_1(\beta_2) \\ t_1(\beta_2) & t_1(\beta_1) t_1(\beta_2) & t_2(\beta_2) \end{bmatrix}. \qquad (57)$$

12.2.2. Exact MLE of discretised observations

Divide the unit square into an $N \times N$ array of square pixels

$$S_{ij} = \left[\frac{i-1}{N}, \frac{i}{N} \right) \times \left[\frac{j-1}{N}, \frac{j}{N} \right)$$

for $i, j = 1, \ldots, N$. Then

$$\mu_{ij} = \int_{S_{ij}} \lambda_{\boldsymbol{\beta}}(x, y) \, dx \, dy = e^{\beta_0} \frac{\gamma_1^{i-1}(\gamma_1 - 1)}{\beta_1} \frac{\gamma_2^{j-1}(\gamma_2 - 1)}{\beta_2} \qquad (58)$$

where $\gamma_1 = \exp(\beta_1/N)$ and $\gamma_2 = \exp(\beta_2/N)$. The partial derivatives are $(\partial/\partial\beta_0)\mu_{ij} = \mu_{ij}$ and

$$\frac{\partial}{\partial\beta_1} \mu_{ij} = \mu_{ij} \left[\left(\frac{i}{N} - \frac{1}{\beta_1} \right) + \frac{1}{N(\gamma_1 - 1)} \right]$$

$$\frac{\partial}{\partial\beta_2} \mu_{ij} = \mu_{ij} \left[\left(\frac{j}{N} - \frac{1}{\beta_2} \right) + \frac{1}{N(\gamma_2 - 1)} \right]$$

Substituting into (25) and (34) yields the Fisher information for the count and indicator data respectively.

Table 2 shows the asymptotic standard errors of the MLE $\hat{\boldsymbol{\beta}}$ (obtained as the square roots of the diagonal entries of the inverse of the Fisher information matrix) calculated for the model (55) with $(\beta_0, \beta_1, \beta_2) = (2, 3, 2)$ discretised on a 32×32 pixel grid. We have $\bar{\mu} = 150.2/32^2 = 0.147$ so the rule-of-thumb (Section 9.3) estimate of relative efficiency is $0.147/(\exp(0.147) - 1) = 0.93$. From Table 2 the actual relative efficiency (in replacing pixel counts by presence/absence) for the estimator of β_0 is $(0.3315/0.3494)^2 = 0.90$.

TABLE 2
Asymptotic standard errors of MLE components for the model (55) with
$(\beta_0, \beta_1, \beta_2) = (2, 3, 2)$ on the unit square, discretised on a 32×32 pixel grid

Observations	β_0	β_1	β_2
exact points	0.3313	0.3449	0.3107
discrete counts	0.3315	0.3452	0.3109
binary	0.3494	0.3655	0.3322

TABLE 3

Approximate bias in spatial logistic regression or complementary log-log regression for the model (55) at different pixel resolutions. Values indicated as "0" are smaller than 10^{-10}. Other values are rounded to 4 decimal places

grid	logistic			cloglog		
	β_0	β_1	β_2	β_0	β_1	β_2
32×32	-0.2141	0.3085	0.2363	0.0005	0.0000	0.0000
64×64	-0.0611	0.0828	0.0644	-0.0001	0	0
128×128	-0.0156	0.0208	0.0162	0.0000	0	0
256×256	-0.0039	0.0052	0.0041	0.0000	0	0
512×512	-0.0010	0.0013	0.0010	0.0000	0	0

TABLE 4

Simulation estimates of bias, standard deviation and root-mean-square error for different estimators of the parameter vector $\boldsymbol{\beta}$ in model (55) using exact coordinates or a 32×32 pixel grid. See text for explanation

	Binary (logistic)			Binary (cloglog)			Continuous		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
bias	-0.245	0.325	0.247	-0.026	0.016	0.011	-0.020	0.008	0.006
sd	0.397	0.415	0.376	0.349	0.367	0.331	0.332	0.347	0.311
rmse	0.466	0.527	0.450	0.350	0.367	0.331	0.332	0.347	0.311

12.2.3. Approximation by GLM

Table 3 shows the approximate bias (according to equations (49) and (54)) incurred by approximating this model by either a logistic regression or a complementary log-log regression, for various pixel grids. Note that the bias in logistic regression is consistently much larger than in complementary log-log regression.

12.2.4. Simulation experiments

We generated 10,000 realisations of the Poisson point process with intensity (55) in the unit square with parameters $\boldsymbol{\beta} = (2, 3, 2)$. For each realisation we fitted the model by various methods. Table 4 shows simulation estimates of bias, standard deviation and root-mean-square error for the different estimators of $\boldsymbol{\beta}$. "Binary (logistic)" is the maximum likelihood estimator for spatial logistic regression, i.e. the maximiser of (40), where the responses Y_j are pixel presence/absence indicators on a 32×32 pixel grid, and the linear predictor is of the form $\eta = \beta_0 + \beta_1 x + \beta_2 y$. "Binary (cloglog)" is the corresponding MLE for binary regression with the complementary log-log link. "Continuous" is the exact maximum likelihood estimate of $\boldsymbol{\beta}$ based on exact coordinates of all points in the Poisson process, obtained as the solution of (56). The bias and variance were estimated using the sample moments from the simulation.

In this model the expected number of data points is $\mathbb{E}[N] = e^{\beta_0} s_0(\beta_1) s_0(\beta_2) = $ 150.2. For a 32×32 grid, the mean expected number of points per pixel is $\bar{\mu} = \mathbb{E}[N]/32^2 = 0.15$, the maximum presence probability p_j is 0.66 and the maximum value of $\mathbb{P}\{N_j \geq 2\}$ is 0.29. Comparison with Table 2 shows that the predictions of asymptotic theory are quite accurate.

13. Sampling zero pixels

The use of fine discretisations of the spatial domain increases the computational load, and may cause numerical instability due to the overwhelming dominance of zero values of the indicators Y_j. A practical strategy that is common in the GIS literature (e.g. [6, p. 378], [35, §3.3.1, p. 403]) is to take a *sample* of the zero values. Then logistic regression is applied to the data consisting of the '1' pixels (all pixels containing a point of the point pattern) and the sampled '0' pixels. The validity of this procedure depends on the random sampling design.

13.1. Simple random sampling

Logistic regression for sampled data is valid if pixels are selected independently.

Lemma 11. *For a particular pixel grid $\mathcal{G}$ let $Y_1, \ldots, Y_n$ be independent Bernoulli variables satisfying a spatial logistic regression* (3). *Conditionally on $Y_1, \ldots, Y_n$, let S be a random sample of pixels, such that (a) if $Y_j = 1$ then pixel j is sampled, (b) if $Y_j = 0$, then pixel j is sampled with probability $s_j > 0$ independently of other pixels. Then the conditional likelihood of $Y_1, \ldots, Y_n$ given S is the likelihood of the logistic regression*

$$\log \frac{p_j}{1 - p_j} = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j - \log s_j$$

for $j \in S$.

Similarly, for spatial logistic regression with an offset (11) the conditional likelihood is that of logistic regression with an *additional* offset $-\log s_j$. The proof is straightforward. This is a subsampling property of logistic regression that is not shared by complementary log–log regression.

Since the discretisation is assumed to be very fine, the sampling procedure is virtually equivalent to generating a point process $\boldsymbol{S}$ of sample points. Logistic regression is optimal if the sample process $\boldsymbol{S}$ is a Poisson process.

Lemma 12. *Suppose $\boldsymbol{X}$ is a Poisson point process in W with loglinear intensity* (15). *Let $\boldsymbol{S}$ be a Poisson point process in W with intensity function $\rho(u)$, independent of $\boldsymbol{X}$. Then the conditional likelihood of $\boldsymbol{X}$ given the superposition $\boldsymbol{U} = \boldsymbol{X} \cup \boldsymbol{S}$ is the likelihood of the logistic regression*

$$\log \frac{p(u)}{1 - p(u)} = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u) - \log \rho(u)$$

where $p(u)$ is the conditional probability that the point u belongs to $\boldsymbol{X}$ given that $u \in \boldsymbol{U}$.

The connection between the Poisson and logistic models is of course very well known [58, p. 212]. The particular result above is well-known in the context of spatial epidemiology, e.g. [30]; however in that context $\boldsymbol{X}$ and $\boldsymbol{S}$ are observed

point processes of cases and controls respectively, while in our context $\boldsymbol{S}$ is synthetic.

Subsampling entails a loss of efficiency, and this is easily calculated in the point process case. In Lemma 12 the conditional loglikelihood is

$$
\begin{aligned}
\log L(\boldsymbol{\beta}; \boldsymbol{X}, \boldsymbol{S}) &= \sum_{x_i \in \boldsymbol{X}} \log \frac{p(x_i)}{1 - p(x_i)} + \sum_{u_j \in \boldsymbol{X} \cup \boldsymbol{S}} \log(1 - p(u_j)) \\
&= \sum_{x_i \in \boldsymbol{X}} \left[\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(x_i) - \log \rho(x_i) \right] \\
&\qquad - \sum_{u_j \in \boldsymbol{X} \cup \boldsymbol{S}} \log\left(1 + \frac{\lambda_{\boldsymbol{\beta}}(u_j)}{\rho(u_j)} \right)
\end{aligned}
\tag{59}
$$

and the score is

$$
U(\boldsymbol{\beta}; \boldsymbol{X}, \boldsymbol{S}) = \sum_{x_i \in \boldsymbol{X}} \boldsymbol{Z}(x_i) \frac{\rho(x_i)}{\rho(x_i) + \lambda_{\boldsymbol{\beta}}(x_i)} - \sum_{u_j \in \boldsymbol{S}} \boldsymbol{Z}(u_j) \frac{\lambda_{\boldsymbol{\beta}}(u_j)}{\rho(u_j) + \lambda_{\boldsymbol{\beta}}(u_j)}.
\tag{60}
$$

Since $\boldsymbol{X}$ and $\boldsymbol{S}$ are independent Poisson processes, the Fisher information is

$$
\begin{aligned}
I(\boldsymbol{\beta}) &= \int_W \boldsymbol{Z}(u)\,\boldsymbol{Z}(u)^{\mathrm{T}} \left(\frac{\rho(u)}{\rho(u) + \lambda_{\boldsymbol{\beta}}(u)} \right)^2 \lambda_{\boldsymbol{\beta}}(u)\,\mathrm{d}u \\
&\qquad + \int_W \boldsymbol{Z}(u)\,\boldsymbol{Z}(u)^{\mathrm{T}} \left(\frac{\lambda_{\boldsymbol{\beta}}(u)}{\rho(u) + \lambda_{\boldsymbol{\beta}}(u)} \right)^2 \rho(u)\,\mathrm{d}u \\
&= \int_W \boldsymbol{Z}(u)\,\boldsymbol{Z}(u)^{\mathrm{T}} \left(\frac{\rho(u)}{\rho(u) + \lambda_{\boldsymbol{\beta}}(u)} \right) \lambda_{\boldsymbol{\beta}}(u)\,\mathrm{d}u.
\end{aligned}
\tag{61}
$$

Comparison with (22) shows the loss of efficiency relative to maximum likelihood. Note that subsampling avoids bias associated with spatial irregularity of covariates, assuming the exact values of the covariates are evaluated at each data point and sample point.

13.2. *Other sampling patterns*

If pixels are not sampled independently, or in continuous space if the sample points are not a Poisson process, then logistic regression is not appropriate. The conditional likelihood of $\boldsymbol{X}$ given $\boldsymbol{X} \cup \boldsymbol{S}$ is generally intractable.

Berman and Turner [14] proposed approximating the integral in the Poisson loglikelihood (19) by a finite sum over arbitrary sample points. Generalising, Rathbun et al [65] considered estimating functions of the form

$$
U_R(\boldsymbol{\beta}) = \sum_{x_i \in \boldsymbol{X}} \boldsymbol{Z}(x_i) - \sum_{u \in \boldsymbol{S}} \boldsymbol{Z}(u) \frac{\lambda_{\boldsymbol{\beta}}(u)}{\rho(u)}
\tag{62}
$$

where $\rho(u)$ is the intensity of $\boldsymbol{S}$. The conditional expectation of $U_R(\boldsymbol{\beta})$ given $\boldsymbol{X}$ (i.e. with respect to the distribution of $\boldsymbol{S}$) is equal to the likelihood score:

$\mathbb{E}[U_R(\boldsymbol{\beta}) \mid \boldsymbol{X}] = U(\boldsymbol{\beta})$. In particular U_R is an unbiased estimating function. More generally Waagepetersen [76] studied estimating functions of the form

$$U_W(\boldsymbol{\beta}) = \sum_{x_i \in \boldsymbol{X}} \boldsymbol{Z}(x_i) - \sum_{u_j \in \boldsymbol{X} \cup \boldsymbol{S}} \boldsymbol{Z}(u_j) \frac{\lambda_{\boldsymbol{\beta}}(u_j)}{s(u_j)} \qquad (63)$$

where $s(u_j)$ is a suitable function. In particular if $s(u_j)$ is the intensity of $\boldsymbol{X} \cup \boldsymbol{S}$ then this equation also has the property that $\mathbb{E}[U_W(\boldsymbol{\beta}) \mid \boldsymbol{X}] = U(\boldsymbol{\beta})$ and hence is an unbiased estimating equation.

14. Splitting pixels

In this section we propose an alternative algorithmic strategy for avoiding the computational and numerical problems inherent in using very fine pixel grids, while reducing bias due to spatial approximation of irregular covariates.

The most common type of spatial irregularity in the covariate function arises when $\boldsymbol{Z}(u)$ is the indicator function of a spatial domain D with a very irregular boundary. The greenstone outcrop in the Murchison data (Section 3) is an example of such a domain. The indicator covariate causes a fundamental difficulty for pixel-based spatial logistic regression which is illustrated in Figure 4. A single square pixel S_j is shown, containing one data point x_i. The irregular spatial domain D covers part of the pixel, including the pixel centre c_j, but excluding the data point x_i.

Consider the Poisson process with intensity $\lambda(u) = \exp(\beta_0 + \beta_1 I(u))$ where $I(u) = \mathbb{I}\{u \in D\}$ is the indicator covariate. In exact maximum likelihood estimation for the Poisson process (Section 7) the contribution to the score (21) from the pixel S_j sketched in Figure 4 has components

$$
\begin{array}{llll}
U_{0j} & = & 1 - \int_{S_j} e^{\beta_0 + \beta_1 I(u)} \, \mathrm{d}u & = & 1 - e^{\beta_0}(a_j - A_j + e^{\beta_1} A_j) \\
U_{1j} & = & I(x_j) - \int_{S_j} I(u) e^{\beta_0 + \beta_1 I(u)} \, \mathrm{d}u & = & I(x_j) - e^{\beta_0 + \beta_1} A_j
\end{array}
$$

where $a_j = |S_j|$ and $A_j = \int_{S_j} I(u) \, \mathrm{d}u = |D \cap S_j|$. In the GLM approximation to the likelihood of the pixel counts (Section 10) we effectively 'approximate' A_j
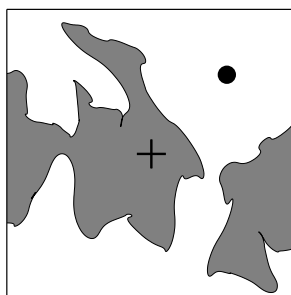

FIG 4. *Illustration of the problem of irregular discontinuities. A single square pixel is shown, containing one data point (●). A spatial domain (grey shading) with irregular boundary covers part of the pixel, including the pixel centre (+), but excluding the data point.*

by $a_j I(c_j)$ and 'approximate' $I(x_i)$ by $I(c_j)$. In the case illustrated in Figure 4 this means replacing $I(x_i) = 0$ by $I(c_j) = 1$. Summing over all j we will have $\sum_j a_j I(c_j) \approx \sum_j A_j$ when pixel size is small, assuming the boundary of D is rectifiable, but $\sum_i I(x_i)$ could be very different from $\sum_j I(c_j)$ even for very fine pixel grids. This is a counterpart of the biostatistical principle that it is important to have accurate estimates of the hazard exposure for the cases in a case-control study.

A simple strategy for reducing bias is to subdivide each pixel into sub-pixels where the covariate is smooth. Suppose the study region W can be partitioned into subdomains $D_1, \ldots, D_M$ such that the restriction of $\boldsymbol{Z}$ to D_m is a Lipschitz function, for each $m = 1, \ldots, M$. This embraces the situation where we have several indicator covariates and several smooth covariates, from which we construct new covariates by algebraic combination.

The strategy is to split each pixel S_j into sub-pixels $T_{jm} = S_j \cap D_m$ for $j = 1, \ldots, J$ and $m = 1, \ldots, M$. Then the partition $(T_{jm})_{j,m}$ is treated as the pixel grid. Pixel covariate values $\boldsymbol{z}_{jm}$ may be assigned either by averaging $\boldsymbol{Z}$ over T_{jm}, or by evaluating $\boldsymbol{Z}$ at an arbitrary point $c_{jm} \in T_{jm}$. Then spatial logistic regression or complementary log–log regression is applied, with an offset $\log a_{jm}$ where $a_{jm} = |T_{jm}|$. The split pixel grid satisfies the conditions of Lemma 9 and Theorem 10, which predict that the parameter estimates should converge at rate $O(\delta_n)$ or faster.

Split pixels T_{jm} are somewhat analogous to 'mixels' (mixed pixels) as described in the remote sensing literature (cf. [41, 45, 52]), except that in our context the exact geometry of the subpixels is assumed known and does not need to be estimated.

In practice the split pixel strategy should be most efficient when the covariate discontinuities are covered by a relatively small number of pixels S_j, so that the total number of nonempty sub-pixels T_{jm} is small. Computational cost depends on the complexity of computing the areas $|T_{jm}|$ and of evaluating covariates.

15. Analysis of Murchison data

The various alternatives to spatial logistic regression developed in this paper are now demonstrated on the Murchison data introduced in Section 3. Software for performing the analysis was developed in the R system [63] as part of the `spatstat` library [9].

15.1. Pixel regression models

Table 5 shows the results of fitting spatial logistic regressions with an offset,

$$\log \frac{p_j}{1 - p_j} \quad = \quad \log a + \beta_0 + \beta_1 g_j + \beta_2 d_j \tag{64}$$

$$\log \frac{p_j}{1 - p_j} \quad = \quad \log a + \gamma_0 + \gamma_1 d_j \tag{65}$$

TABLE 5

Estimates of the parameters in the offset spatial logistic regression models (64) and (65) for the Murchison data, obtained using different pixel grids

grid	β_0	β_1	β_2	γ_0	γ_1
64×64	-6.11	2.19	-0.156	-4.64	-0.270
128×128	-6.04	2.30	-0.143	-4.36	-0.271
256×256	-6.25	2.53	-0.126	-4.30	-0.265
512×512	-6.63	2.91	-0.110	-4.29	-0.269

TABLE 6

Estimates of the parameters in the offset complementary log–log regression models (66) and (67) for the Murchison data, obtained using different pixel grids

grid	β_0	β_1	β_2	γ_0	γ_1
64×64	-6.18	2.03	-0.146	-4.74	-0.260
128×128	-6.07	2.24	-0.140	-4.40	-0.267
256×256	-6.26	2.52	-0.125	-4.31	-0.264
512×512	-6.63	2.90	-0.110	-4.30	-0.269

to the Murchison data, where $a = |W|/N^2$ is the pixel area. This could alternatively have been derived from Table 1 by adding $-\log a$ to the estimates of β_0 and γ_0 in Table 1.

Table 6 shows the results of fitting the analogous complementary log–log regressions

$$\log(-\log(1 - p_j)) = \log a + \beta_0 + \beta_1 g_j + \beta_2 d_j \tag{66}$$
$$\log(-\log(1 - p_j)) = \log a + \gamma_0 + \gamma_1 d_j \tag{67}$$

Numerical exceptions were again encountered in fitting model (67) but not (66).

The estimates of the slope parameter γ_1 in (67) are approximately independent of pixel size, but the estimates of β_1, β_2 in (66) change by up to 40% as pixel size is reduced.

These findings were broadly anticipated by our analysis in Section 11. For any spatial location u let $d(u)$ be the distance from u to the nearest fault, and $g(u)$ the indicator which equals 1 if u lies inside the greenstone outcrop region. The function d is Lipschitz with constant 1, by the triangle inequality. Lemma 9(a) applies to model (67) so the bias in the spatial logistic regression estimates of γ_0, γ_1 should be at most of order $O(\delta_n)$. However, g is not Lipschitz and has jump discontinuities along a very irregular boundary. Hence, Lemma 9 does not apply to model (66) and the estimates could depend strongly on the pixel grid geometry.

15.2. Subsampling

Table 7 summarises the results of applying the subsampling strategy of Section 13, Lemma 12, to the Murchison data. The objective is to fit the loglinear Poisson point process model

$$\lambda(u) = \exp(\beta_0 + \beta_1 g(u) + \beta_2 d(u)). \tag{68}$$

TABLE 7

Sample means and standard deviations of the parameter estimates for the loglinear Poisson model (68) *obtained by logistic regression using a Poisson process of sampling points with intensity ρ, expected number of sampling points $s = \rho|W|$*

ρ	s	β_0	β_1	β_2
10λ	2,550	-6.53 (0.06)	2.85 (0.07)	-0.121 (0.008)
100λ	25,500	-6.59 (0.03)	2.86 (0.02)	-0.112 (0.003)
500λ	127,500	-6.59 (0.01)	2.85 (0.01)	-0.112 (0.001)

TABLE 8

Sample means and standard deviations of parameter estimates for the loglinear Poisson model (68) *obtained by logistic regression on a random subset of the 512×512 pixel grid with selection probability q, expected number of sample points $s = 512^2 q$. Last line obtained using all pixels*

K	q	s	β_0	β_1	β_2
10	0.01	2,540	-6.58 (0.07)	2.90 (0.08)	-0.117 (0.007)
100	0.10	25,400	-6.63 (0.02)	2.90 (0.02)	-0.111 (0.003)
500	0.49	127,400	-6.63 (0.01)	2.91 (0.01)	-0.109 (0.001)
	1	262,144	-6.63	2.91	-0.110

A Poisson process of sample points with constant intensity ρ was generated; the covariates $d(u)$ and $g(u)$ were evaluated *exactly* at the data points and sample points; and logistic regression with offset $-\log \rho$ was fitted to the indicators ($Y = 1$ for data points, $Y = 0$ for sample points). Sampling intensity was $\rho = K\hat{\lambda}$ where $K \in \{10, 100, 500\}$ and $\hat{\lambda} = n/|W|$ was the empirical mean intensity of the data points. There are approximately $s = nK$ sample points, so the computational load is roughly equivalent to pixel-based logistic regression with an $m \times m$ pixel grid, where $m = \sqrt{n(K+1)} \approx 50, 160$ and 360 respectively. Estimates for $K = 10$ and 100 are based on 100 simulations while estimates for $K = 500$ are based on 30 simulations.

Comparing Tables 7 and 9 we find that, for approximately equal computational effort, the mean square error is approximately equivalent.

Note that Table 7 refers to the strategy described in Lemma 12 of taking Poisson sample points in *continuous* space, and using the *exact* values of the covariates at the sample points.

Table 8 shows the results of subsampling the pixels in spatial logistic regression as described in Lemma 11. For a 512×512 grid of pixels, the covariates were evaluated at the centre of each pixel. Pixels that did not contain a data point were randomly retained or deleted, with retention probability q, independently of other pixels. The retention probability q was taken to be K times the fraction of pixels containing a data point, with K chosen to be 10, 100 or 500. Logistic regression with offset $\log a - \log q$, where a is pixel area, was fitted to the data from the retained pixels. Table 8 shows the sample means and standard deviations of the resulting parameter estimates. For comparison, the last line of Table 8 gives the fitted coefficients obtained when all pixel data are used (obtained from Table 5). There are some persistent discrepancies between the outcomes of Tables 7 and 8 which we attribute to discretisation effects in the latter.

15.3. Split pixels

Table 9 shows the results of using the split-pixel strategy proposed in Section 14 to fit the spatial logistic regression model (64) and the complementary log–log regression (66). Any pixel S_j that was intersected by the irregular polygonal boundary of the greenstone outcrop was divided into two sub-pixels T_{j0} and T_{j1} by this boundary. The areas of the two sub-pixels were computed using the discrete version of Green's formula. The distance covariate d_{jm} for subpixel T_{jm} was simply taken to be the value of d_j for S_j, computed as the distance from the centre of S_j to the nearest geological fault. The greenstone indicator was assigned $g_{jm} = m$ for $m = 0, 1$.

Complementary log–log regression with pixel splitting produces very stable results. The parameter estimates obtained with a 64×64 pixel grid are within 5% of the estimates for the 512×512 grid. For a graphical comparison between the different methods, Figure 5 is a plot of estimates of the slope parameter β_2 in the Poisson loglinear model (68) using logistic regression or complementary log–log regression, with or without pixel splitting. These results confirm the high efficiency of the pixel-splitting strategy.

TABLE 9
Parameter estimates using the split pixel strategy for the offset spatial logistic regression model (64) *and the offset complementary log–log regression model* (66) *for the Murchison data*

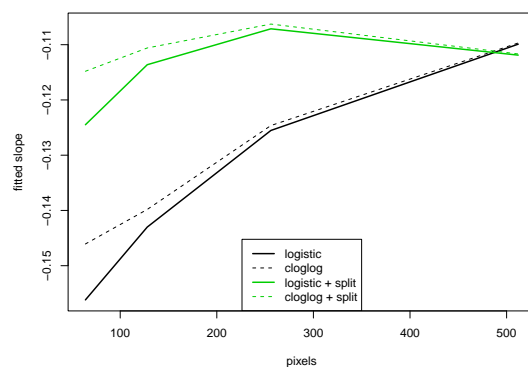| grid	logistic			cloglog		
	β_0	β_1	β_2	β_0	β_1	β_2
64×64	-6.70	2.86	-0.125	-6.78	2.74	-0.115
128×128	-6.69	2.97	-0.114	-6.71	2.92	-0.111
256×256	-6.65	2.94	-0.107	-6.66	2.92	-0.106
512×512	-6.59	2.86	-0.112	-6.59	2.86	-0.112

FIG 5. *Estimates of slope parameter β_2 in the loglinear model* (68) *for the Murchison data using different algorithms.*

15.4. Discussion of analysis

The model fitted by this analysis is a Poisson point process with intensity (68). A colour image plot of the fitted intensity function (not shown) is visually very similar to a plot of the fitted pixel occupancy probabilities (Figure 2 of Section 3.2). However the values of the fitted intensity function have a physical meaning independent of the pixel size.

The foregoing data analysis is designed only to demonstrate features of the method. In particular it should be understood that other loglinear Poisson models could be fitted to the Murchison data using the same covariates (i.e. the original covariates can be transformed to make new canonical covariates). For example the intensity could be a log-quadratic function of distance to the nearest fault.

We have not undertaken a full analysis, which would require exploratory consideration of different models, inspection of model diagnostics, and model selection. Model diagnostics will be discussed in a sequel paper. Formal techniques for model selection are available since the Poisson likelihood can be evaluated explicitly; one may apply the likelihood ratio test based on an analysis of deviance, or criteria such as AIC.

We have avoided discussing the validity of the Poisson process model for the Murchison data. Gold deposits might well be spatially clustered (positively associated) due to random effects in the formation process. Landslides and avalanches are likely to be a regular (negatively associated) point process because each event requires the buildup of material [60, sec. 4]. Such models can be dealt with using established techniques of spatial statistics [25, 29, 59]. This is a general issue for GIS data analysis that is brought into focus by the current work.

We implicitly assumed that the point pattern dataset is a complete and accurate record of the realisation of the point process. That is, the map of gold deposits is complete and accurate within the study region. In reality such maps typically show only the *known* deposits. At the very least, one should allow for spatial variation in the intensity or reliability of the survey. To a first approximation, everything is observable inside an outcrop (rock not covered by soil) while nothing is directly observable elsewhere.

16. Conclusions and recommendations

In applications where a study region is divided into pixels of arbitrary size, we have shown that spatial logistic regression models do not necessarily have a consistent meaning, independent of the choice of pixel size. A consistent meaning is obtained only when pixels are very small; in this case, the model is a Poisson point process, whose intensity is a loglinear function of the (canonical) covariates.

Estimation of the parameters of a loglinear Poisson process can be performed to a good approximation using presence/absence data on a coarse pixel grid:

the preferred algorithm is binary regression with the complementary log–log link (instead of the logistic link) and with an offset term equal to the logarithm of pixel area. Our theoretical and empirical results show that this technique is typically much more accurate than logistic regression for a given grid size.

The assignment of values to the pixel covariates z_j, particularly for the pixels that contain data points, can have substantial influence on statistical performance. This is analogous to the importance of obtaining accurate estimates of hazard exposure for the cases in an epidemiological case–control study. An important instance of this problem occurs when the covariate has jump discontinuities along a highly irregular edge. The split pixel strategy is a very effective solution, and is recommended especially in very large datasets.

The MLE is not necessarily appropriate; other possibilities for fitting Poisson point process models include robust estimation [4, 5, 84]. However, robust estimating equations have a structure similar to the score of the loglinear Poisson model [5], so we expect similar issues to arise in robust estimation.

Acknowledgements

Appendix A: Proofs

A.1. Incompatibility Lemma

Proof of Lemma 1. In the subdivision (7) the component pixels T_k and their union S_j all share the same covariate value $z(T_k) = z(S_j) = z_j$ for $k \in K(j)$. Through (9) this implies that each of the component pixels T_k has the same presence probability q_k, so that

$$1 - p_j = (1 - q_k)^r, \qquad k \in K(j). \tag{69}$$

Assume that equations (8), (9) and (69) all hold. Then

$$1 + \exp(\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{z}_j) = \left[1 + \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j)\right]^r$$

and hence

$$\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{z}_j = \log \left\{ \left[1 + \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j)\right]^r - 1 \right\} \tag{70}$$

for all pixels S_j in $\mathcal{G}$. The function $f(x) = \log((1 + e^x)^r - 1)$ appearing on the right hand side of (70) is strictly increasing and strictly convex.

Assumption (a) implies that we can write $\boldsymbol{z}_j = t\boldsymbol{z}_{j'} + (1-t)\boldsymbol{z}_{j''}$ with $0 < t < 1$. Substituting in (70) yields

$$t\,\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{z}_{j'} + (1-t)\,\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{z}_{j''} = f\left(t\,\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{z}_{j'} + (1-t)\,\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{z}_{j''}\right).$$

Applying (70) to each term on the left side and writing $v_j = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{z}_j$ for any j, we obtain $tf(v_{j'}) + (1-t)f(v_{j''}) = f(tv_{j'} + (1-t)v_{j''})$. But this is impossible, since f is strictly convex and the values $v_{j'}, v_{j''}$ are distinct. The contradiction implies that (10), (8), (9) are inconsistent.

Assumption (b) implies that $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{z}_j$, and hence $f(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{z}_j)$, takes at most k distinct values. Since the vectors $\boldsymbol{z}_j$ form a basis, equation (70) has solutions. $\qquad\square$

A.2. Poisson process approximation for general models

This section contains proofs for Theorem 3 and Corollary 4 from Section 6.3.

Proof of Theorem 3. Choose a fixed point $\gamma_j \in S_j$ for every $j \in \{1, \ldots, J\}$ and define the function $\gamma : W \to \{\gamma_j; 1 \le j \le J\}$ by $\gamma(x) = \gamma_j$ if $x \in S_j$. Denote by $\widetilde{\boldsymbol{X}} = \sum_{j=1}^J Y_j \delta_{\gamma_j}$ the pixel point process, which has expectation measure $\boldsymbol{\mu} = \sum_{j=1}^J p_j \delta_{\gamma_j}$. Let $\boldsymbol{\Pi}$ be a Poisson process on W with expectation measure $\boldsymbol{\lambda}$ and $\widetilde{\boldsymbol{\Pi}}$ a Poisson process with expectation measure $\boldsymbol{\mu}$. Write $r_j = \sup_{u \in S_j} \|u - \gamma_i\|$ and $\bar{r} = \max_{1 \le j \le J} r_j$. We can split up the initial distance as

$$d_2\big(\mathscr{L}(\boldsymbol{X}), \mathrm{Pois}(\boldsymbol{\lambda})\big) \le d_2\big(\mathscr{L}(\boldsymbol{X}), \mathscr{L}(\widetilde{\boldsymbol{X}})\big) + d_2\big(\mathscr{L}(\widetilde{\boldsymbol{X}}), \mathscr{L}(\widetilde{\boldsymbol{\Pi}})\big) + d_2\big(\mathscr{L}(\widetilde{\boldsymbol{\Pi}}), \mathscr{L}(\boldsymbol{\Pi})\big) \tag{71}$$

and examine each of the summands on the right hand side separately. For the first one, we obtain

$$\begin{aligned}
d_2\big(\mathscr{L}(\boldsymbol{X}), \mathscr{L}(\widetilde{\boldsymbol{X}})\big) &\le \mathbb{E}d_1(\boldsymbol{X}, \widetilde{\boldsymbol{X}}) \\
&= \mathbb{E}\big[d_1(\boldsymbol{X}, \widetilde{\boldsymbol{X}})\,\mathbb{I}\{N_j = Y_j \text{ for every } j\}\big] \\
&\quad + \mathbb{E}\big[d_1(\boldsymbol{X}, \widetilde{\boldsymbol{X}})\,\mathbb{I}\{N_j > 1 \text{ for some } j\}\big] \\
&= \mathbb{E}\Big[\big(\tfrac{1}{\boldsymbol{X}(W)}\textstyle\sum_{x \in \boldsymbol{X}}\|x - \gamma(x)\|\big)\,\mathbb{I}\{N_j = Y_j \text{ for every } j\}\Big] \\
&\quad + \mathbb{P}\{N_j > 1 \text{ for some } j\} \\
&\le \bar{r} + \mathbb{P}\{N_j > 1 \text{ for some } j\}.
\end{aligned} \tag{72}$$

We write $\boldsymbol{\Pi}^* = \sum_{j=1}^J \boldsymbol{\Pi}(S_j)\delta_{\gamma_j}$ and use a somewhat similar argument for the third summand. Employing two well-known facts about the total variation distance and Poisson distributions (see [12], Proposition A.1.1 (1.15) and Theorem 1.C), we have

$$d_2\big(\mathscr{L}(\mathbf{\Pi}),\mathscr{L}(\widetilde{\mathbf{\Pi}})\big) \le d_2\big(\mathscr{L}(\mathbf{\Pi}),\mathscr{L}(\mathbf{\Pi}^*)\big) + d_2\big(\mathscr{L}(\mathbf{\Pi}^*),\mathscr{L}(\widetilde{\mathbf{\Pi}})\big)$$

$$\le \mathbb{E}\big[d_1(\mathbf{\Pi},\mathbf{\Pi}^*)\big] + d_{\mathrm{TV}}\big(\mathscr{L}(\mathbf{\Pi}^*),\mathscr{L}(\widetilde{\mathbf{\Pi}})\big)$$

$$\le \mathbb{E}\big[\tfrac{1}{\mathbf{\Pi}(W)}\textstyle\sum_{x\in\mathbf{\Pi}}\|x-\gamma(x)\|\big] + \sum_{j=1}^{J} d_{\mathrm{TV}}\big(\mathrm{Pois}(\boldsymbol{\lambda}(S_j)),\mathrm{Pois}(p_j)\big)$$

$$\le \bar{r} + \sum_{j=1}^{J}\big|\boldsymbol{\lambda}(S_j) - \mathbb{P}\{N_j > 0\}\big|.$$

$$(73)$$

For the second summand on the right hand side of inequality (71), we use a result obtained by Stein's method for Poisson process approximation. Theorem 4.3.C in [70], which is derived by a slight modification of the proof of Equation (6.3) of Theorem 6.1 in [83], yields that

$$d_2\big(\mathscr{L}(\widetilde{\mathbf{X}}),\mathscr{L}(\widetilde{\mathbf{\Pi}})\big) = d_2\Big(\mathscr{L}\big(\textstyle\sum_{j=1}^{J}Y_j\delta_{\gamma_j}\big),\mathrm{Pois}\big(\textstyle\sum_{j=1}^{J}(\mathbb{E}Y_j)\delta_{\gamma_j}\big)\Big)$$

$$\le \Big(1 \wedge \frac{6.5}{\boldsymbol{\mu}(W)}\Big)\sum_{j=1}^{J}p_j^2 \le 6.5 \max_{1\le j\le J} p_j.$$

$$(74)$$

Plugging the inequalities obtained for the three summands into (71) leads to the upper bound claimed in (13). Inequality (14) is a simple consequence. □

The same proof technique can be used to deal with (weakly) dependent pixel values Y_j. In this case the inequality (74) is replaced by that obtained in [12, Theorem 10.F] or, more generally, by [83, (5.51)]. This leads to a reasonable but more complicated upper bound which contains extra summands to control the amount and structure of dependence amongst the Y_j.

Proof of Corollary 4. Write $\boldsymbol{\lambda}(du) = \lambda(u)du$. By Condition (i) we can apply inequality (13) of Theorem 3 for each of the partitions $\mathcal{G}^{(m)}$, $m \in \mathbb{N}$. Since $\mathbb{P}\{N_{m,j} > 0\} \le \boldsymbol{\lambda}(S_{m,j}) + \sum_{j=1}^{J_m}|\mathbb{E}N_{m,j} - \boldsymbol{\lambda}(S_{m,j})|$, this yields

$$d_2\big(\mathscr{L}(\mathbf{X}^{(m)}),\mathrm{Pois}(\boldsymbol{\lambda})\big) \le 2\delta_m + \sum_{j=1}^{J_m}\big(\mathbb{E}N_{m,j} - \mathbb{P}\{N_{m,j} = 1\}\big)$$

$$+ 7.5\sum_{j=1}^{J_m}\big|\boldsymbol{\lambda}(S_{m,j}) - \mathbb{E}N_{m,j}\big| + 6.5 \max_j \boldsymbol{\lambda}(S_{m,j}).$$

$$(75)$$

By the prerequisites the first two summands converge to zero as $m \to \infty$. For the third summand, we have

$$\sum_{j=1}^{J_m}\big|\boldsymbol{\lambda}(S_{m,j}) - \mathbb{E}N_{m,j}\big| \le \sum_{j=1}^{J_m}\bigg|\int_{S_{m,j}}\big(\lambda(u) - \lambda^{(m)}(u)\big)\,du\bigg| \le \int_W\big|\lambda(u) - \lambda^{(m)}(u)\big|\,du,$$

which goes to zero as well. Finally, it can be shown from the fact that $(\mathcal{G}^{(m)})_m$ is nested and asymptotically infinitesimal that $\max_j \boldsymbol{\lambda}(S_{m,j}) \to 0$ as $m \to \infty$; compare Lemma A1.6.II in [26] (noting that $(\mathcal{G}^{(m)})_m$ is a dissecting system for W in the sense defined there). Thus inequality (75) implies that $d_2\big(\mathscr{L}(\boldsymbol{X}^{(m)}),$ $\mathrm{Pois}(\boldsymbol{\lambda})\big) \to 0$, and hence $\boldsymbol{X}^{(m)}$ converges in distribution to $\mathrm{Pois}(\boldsymbol{\lambda})$ by Lemma 2. $\qquad\square$

A.3. Poisson process approximation for spatial logistic regression

This section contains a proof of Theorem 5 from Section 6.4.

Write $\boldsymbol{\lambda}(du) = \lambda(u)du$ for the expectation measure of the Poisson process. Note that $p_{m,j} \leq p_{m,j}/(1 - p_{m,j}) \leq Ka_{m,j}$, where $K = \max_{u \in W} \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u))$. For each m the conditions of Theorem 3 are satisfied, so that we can employ inequality (13). The third summand in the upper bound can then be estimated further as

$$\sum_{j=1}^{J_m} \big|\boldsymbol{\lambda}(S_{m,j}) - \mathbb{P}\{N_{m,j} > 0\}\big| = \sum_{j=1}^{J_m} \left| \int_{S_{m,j}} \exp\big(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u)\big)\,\mathrm{d}u - p_{m,j} \right|$$

$$\leq |W|\varepsilon_m + \sum_{j=1}^{J_m} \left|\exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_{m,j})a_{m,j} - p_{m,j}\right|$$

$$= |W|\varepsilon_m + \sum_{j=1}^{J_m} \frac{p_{m,j}^2}{1 - p_{m,j}}$$

$$\leq |W|\varepsilon_m + K^2|W|\,a_m,$$

where $\varepsilon_m = \max_{1 \leq j \leq J_m} \sup_{u \in S_{m,j}} \big|\exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_{m,j}) - \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u))\big|$, which goes to zero by the choice of $\boldsymbol{z}_{m,j}$, the uniform continuity of Z on W, and $\delta_m \to 0$. Hence we have for the total d_2-error

$$d_2\big(\mathscr{L}(\boldsymbol{X}^{(m)}), \mathrm{Pois}(\boldsymbol{\lambda})\big) \leq 2\delta_m + \mathbb{P}\{\exists j : N_{m,j} > 1\} + |W|\varepsilon_m + K\big(K|W| + 6.5\big)a_m. \tag{76}$$

By the prerequisites all the summands on the right hand side of (76) tend to zero, so that, by Lemma 2, $\boldsymbol{X}^{(m)}$ converges in distribution to a Poisson process with expectation measure $\boldsymbol{\lambda}$ as $m \to \infty$.

A.4. Existence of MLE for loglinear Poisson process

Following is a sketch proof of Lemma 6 in Section 7.

Let $D = \{\sum_i t_i \boldsymbol{Z}(x_i) : t_i > 0\}$ be the relative interior of the convex cone generated by the covariate values $\boldsymbol{Z}(x_i)$ for the data points. Let C be the relative interior of the convex cone generated by the image of the covariate function $\boldsymbol{Z}$, defined as follows. First the image measure $\boldsymbol{Z}^{\sharp}$ is a finite measure on $\mathbb{R}^k$ defined by $\boldsymbol{Z}^{\sharp}(B) = \int_W \mathbb{I}\{\boldsymbol{Z}(u) \in B\}\,\mathrm{d}u$ for measurable $B \subseteq \mathbb{R}^k$. Its support $\mathrm{spt}(\boldsymbol{Z}^{\sharp})$ is

the smallest closed set S whose complement has measure zero under $\boldsymbol{Z}^\sharp$. Finally C is the relative interior of the convex cone generated by $\mathrm{spt}(\boldsymbol{Z}^\sharp)$.

By arguments similar to [72] the negative loglikelihood has no directions of recession iff $D \cap C \neq \emptyset$. For a realization of the Poisson process, the values $\boldsymbol{Z}(x_i)$ are almost surely in $\mathrm{spt}(\boldsymbol{Z}^\sharp)$, so $D \subseteq C$. Hence the MLE exists uniquely iff $D \neq \emptyset$, which is equivalent to requiring $\boldsymbol{Z}(x_i) \neq 0$ for at least one i.

A.5. Fine-pixel limit for logistic regression

This section contains a proof of Theorem 8 from Section 10.2.

For any fixed $\boldsymbol{\beta}$, the function $h(u) = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u)$ is uniformly bounded above and below, since $\boldsymbol{Z}$ is Lipschitz and W is bounded. Hence the values $\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j$ are uniformly bounded (independent of pixel size). Using the first order Taylor expansion of $\log(1 + x)$ at $x = 0$, we get

$$\sum_j \log(1 + a_{n,j} \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j)) = \sum_j a_{n,j} \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j) + R_1$$

where the remainder satisfies

$$|R_1| \leq \frac{1}{2} \sum_j a_{n,j}^2 \exp(2\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j) \leq \frac{1}{2} a_n \sum_j a_{n,j} \exp(2\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j).$$

Accordingly

$$\sum_j \log(1 + a_{n,j} \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j)) \to \int_W \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u)) \, \mathrm{d}u$$

for each fixed $\boldsymbol{\beta}$, since $g(u) = \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u))$ is clearly Riemann integrable.

Let $N_{n,j}$ denote the number of points of $\boldsymbol{X}$ in pixel $S_{n,j}$ from grid $\mathcal{G}_n$. Then $N_{n,j}$ is Poisson distributed with mean $\mu_{n,j} = \int_{S_{n,j}} \lambda_{\boldsymbol{\beta}_0}(u) \, \mathrm{d}u$. Consider the event $A_n = \{\exists j : N_{n,j} > 1\}$ for grid $\mathcal{G}_n$. We have

$$\mathbb{P}(A_n) = 1 - \prod_j e^{-\mu_{n,j}} (1 + \mu_{n,j}).$$

By Taylor expansion

$$\mathbb{P}(A_n) \leq 1 - \exp(-\frac{1}{2} a_n \sum_j a_{n,j} e^{2\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_j})$$

so that $\mathbb{P}(A_n) \to 0$. Furthermore, since $a_{n+1} \leq a_n/2$, we have $\sum_n \mathbb{P}(A_n) < \infty$ and by the Borel-Cantelli Lemma $\mathbb{P}(A) = 0$ where $A = \{A_n \text{ occurs infinitely often}\}$. For realisations of $\boldsymbol{X}$ outside the null set A, Lipschitz continuity of $\boldsymbol{Z}$ gives

$$|\sum_j Y_j \boldsymbol{z}_j - \sum_{x_i \in \boldsymbol{X}} \boldsymbol{Z}(x_i)| \leq NC\delta_n$$

where $N = N(\boldsymbol{X} \cap W)$ is the total number of points in the realisation of the process. Thus

$$\sum_j Y_j \boldsymbol{z}_j \xrightarrow{\text{a.s.}} \sum_{x_i \in \boldsymbol{X}} \boldsymbol{Z}(x_i).$$

Hence $\log L_{\text{adj}}(\boldsymbol{\beta})$ converges pointwise to the exact log likelihood, almost surely.

By similar arguments, the first and second derivatives of $\log L_{\text{adj}}(\boldsymbol{\beta})$ converge pointwise to the corresponding derivatives of the exact loglikelihood, almost surely.

The functions $\ell(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$ and $\ell_n(\boldsymbol{\beta}) = \log L_{\text{adj}}(\boldsymbol{\beta})$ are convex, with negative definite second derivative. Hence the respective maximisers $\widetilde{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\beta}}$ are uniquely determined. Since ℓ_n converges pointwise to ℓ, and the first derivative of ℓ_n converges pointwise to the first derivative of ℓ, we have that $\ell_n(\boldsymbol{\beta})$ converges uniformly to $\ell(\boldsymbol{\beta})$ in any bounded region. Hence ℓ_n hypoconverges to ℓ by [68, Thm 7.17, p. 252]. The upper level sets $u_n(t) = \{\boldsymbol{\beta} : \ell_n(\boldsymbol{\beta}) \geq t\}$ are bounded convex sets, so hypoconvergence implies [68, Exercise 7.32, p. 266] that for fixed t there is a bounded set B_t such that $u_n(t) \subset B_t$ for sufficiently large n. Since ℓ_n and ℓ have unique maxima, by Theorem 7.33 of [68, p. 266] we have $\widetilde{\boldsymbol{\beta}}_n \to \widehat{\boldsymbol{\beta}}$.

A.6. Asymptotics of bias in spatial logistic regression

This section contains a proof of Lemma 9 in Section 11.2.

For a pixel $S_{n,j}$ in the grid $\mathcal{G}_n$

$$p_{n,j} - p_{\text{LR}}(\boldsymbol{\beta}_0, j) = 1 - \exp\left(-\int_{S_{n,j}} e^{\boldsymbol{\beta}_0^{\text{T}} \boldsymbol{Z}(u)} \, du\right) - \frac{1}{1 + \exp(-\boldsymbol{\beta}_0^{\text{T}} \boldsymbol{z}_j)/a_{n,j}}.$$

Applying separate Taylor expansions to $1 - \exp(-x)$ and $x/(1+x)$

$$p_{n,j} - p_{\text{LR}}(\boldsymbol{\beta}_0, j) = \int_{S_{n,j}} \exp(\boldsymbol{\beta}_0^{\text{T}} \boldsymbol{Z}(u)) \, du - a_{n,j} \exp(\boldsymbol{\beta}_0^{\text{T}} \boldsymbol{z}_j) + R_j + R_j'$$

where the remainders satisfy

$$R_j \;\leq\; \frac{1}{2}\left(\int_{S_{n,j}} e^{\boldsymbol{\beta}_0^{\text{T}} \boldsymbol{Z}(u)} \, du\right)^2 \leq \frac{1}{2} a_n a_{n,j} e^{2C\|\boldsymbol{\beta}_0\|}$$

$$R_j' \;\leq\; \left(a_{n,j} \exp(\boldsymbol{\beta}_0^{\text{T}} \boldsymbol{z}_j)\right)^2 \leq a_n a_{n,j} e^{2C\|\boldsymbol{\beta}_0\|}.$$

Summing over j and using (47) we have

$$\mathbb{E}_0 U_{\text{LR}}(\boldsymbol{\beta}_0) = \sum_j \left(\int_{S_{n,j}} \exp(\boldsymbol{\beta}_0^{\text{T}} \boldsymbol{Z}(u)) \, du - a_{n,j} \exp(\boldsymbol{\beta}_0^{\text{T}} \boldsymbol{z}_j) + R_j + R_j'\right) \boldsymbol{z}_j$$

where $\sum_j R_j \leq O(a_n|W_n|)$ and similarly for $\sum_j R'_j$. Since $||\boldsymbol{z}_j||$ is uniformly bounded,

$$|| \sum_j (R_j + R'_j) \boldsymbol{z}_j || = O(a_n|W_n|)$$

which proves (50). By Lipschitz continuity in each pixel, for all $u \in S_{n,j}$ we have $|| \exp(\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{Z}(u)) - \exp(\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{z}_j) || \leq C ||\boldsymbol{z}_j|| \delta_{n,j}$, so that

$$|| \int_{S_j} \exp(\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{Z}(u)) \, \mathrm{d}u - a_{n,j} \exp(\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{z}_j) || \leq a_{n,j} C ||\boldsymbol{z}_j|| \delta_n$$

and

$$|| \sum_j \left(\int_{S_j} \exp(\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{Z}(u)) \, \mathrm{d}u - a_n \exp(\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{z}_j) \right) \boldsymbol{z}_j || \leq |W_n| C \delta_n \sup_{u \in W} ||\boldsymbol{Z}(u)||^2.$$

This proves (51). For part (b) of Lemma 9, when $\boldsymbol{Z}$ is constant within pixels, the true presence probability is

$$p_{n,j} = 1 - \exp(-\mu_{n,j}) = 1 - \exp(-a_{n,j} \exp(\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{z}_j))$$

where $\boldsymbol{z}_j = Z(c_j)$ for arbitrary $c_j \in S_{n,j}$. Thus $p_{n,j} - p_{\mathrm{LR}}(\boldsymbol{\beta}_0, j) = f(a_{n,j} \times \exp(\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{z}_j))$ where

$$f(x) = 1 - e^{-x} - \frac{x}{1+x} = \frac{1}{2}x^2 + R(x)$$

with remainder $|R(x)| \leq x^3$. Hence

$$\sum_j (p_{n,j} - p_{\mathrm{LR}}(\boldsymbol{\beta}_0, j)) \boldsymbol{z}_j = \frac{1}{2} \sum_j a_{n,j}^2 \exp(2\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{z}_j) \boldsymbol{z}_j + \sum_j R_j \boldsymbol{z}_j$$

where $|R_j| \leq a_{n,j}^3 \exp(3\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{z}_j)$. Since $\boldsymbol{Z}$ is constant within pixels,

$$a_{n,j} \exp(3\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{z}_j) \boldsymbol{z}_j = \int_{S_{n,j}} \lambda_{\boldsymbol{\beta}_0}(u)^3 \boldsymbol{Z}(u) \, \mathrm{d}u$$

so that

$$|| \sum_j R_j \boldsymbol{z}_j || \leq a_n^2 \int_{W_n} \lambda_{\boldsymbol{\beta}_0}(u)^3 ||\boldsymbol{Z}(u)|| \, \mathrm{d}u \leq O(a_n^2|W_n|)$$

which establishes (52). Similarly for the leading term

$$|| \frac{1}{2} \sum_j a_{n,j}^2 \exp(2\boldsymbol{\beta}_0{}^{\mathrm{T}} \boldsymbol{z}_j) \boldsymbol{z}_j || \leq \frac{a_n}{2} \int_{W_n} \lambda_{\boldsymbol{\beta}_0}(u)^2 ||\boldsymbol{Z}(u)|| \, \mathrm{d}u = O(a_n|W_n|)$$

which proves (53).

References

[1] F.P. AGTERBERG. Automatic contouring of geological maps to detect target areas for mineral exploration. *Journal of the International Association for Mathematical Geology*, 6:373–395, 1974.

[2] A. ALBERT and J.A. ANDERSON. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71:1–10, 1984. MR0738319

[3] J.E. ALT, G. KING, and C.S. SIGNORINO. Aggregation among binary, count and duration models: estimating the same quantities from different levels of data. *Political Analysis*, 9:21–44, 2001.

[4] R. ASSUNÇÃO. *Robust estimation in point processes*. PhD thesis, University of Washington, Seattle, 1994. MR2691711

[5] R. ASSUNÇÃO and P. GUTTORP. Robustness for inhomogeneous Poisson point processes. *Annals of the Institute of Statistical Mathematics*, 51:657–678, 1999. MR1735320

[6] P.M. ATKINSON and R. MASSARI. Generalised linear modelling of susceptibility to landsliding in the central Appenines, Italy. *Computers and Geosciences*, 24:373–385, 1998.

[7] A. BADDELEY. Interpreting results of spatial logistic regression. In preparation.

[8] A. BADDELEY and R. TURNER. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics*, 42(3):283–322, 2000. MR1794056

[9] A. BADDELEY and R. TURNER. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42, 2005. URL: www.jstatsoft.org, ISSN: 1548-7660.

[10] A. BADDELEY and R. TURNER. Modelling spatial point patterns in R. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, editors, *Case Studies in Spatial Point Pattern Modelling*, number 185 in Lecture Notes in Statistics, pages 23–74. Springer-Verlag, New York, 2006. ISBN: 0-387-28311-0. MR2232122

[11] S. BANERJEE and A.E. GELFAND. Prediction, interpolation and regression for spatially misaligned data. *Sanhkya A*, 64:227–245, 2002. MR1981755

[12] A.D. BARBOUR, L. HOLST, and S. JANSON. *Poisson Approximation*. Number 2 in Oxford Studies in Probability. Oxford University Press, 1992. MR1163825

[13] O.E. BARNDORFF-NIELSEN. *Information and Exponential families in statistical theory*. Wiley, Chichester, New York, Brisbane, Toronto, 1978. MR0489333

[14] M. BERMAN and T.R. TURNER. Approximating point process likelihoods with GLIM. *Applied Statistics*, 41:31–38, 1992.

[15] J. BESAG, R. MILNE, and S. ZACHARY. Point process limits of lattice processes. *Journal of Applied Probability*, 19:210–216, 1982. MR0644433

[16] D.R. BRILLINGER. Comparative aspects of the study of ordinary time series and of point processes. In P.R. Krishnaiah, editor, *Developments in Statistics*, pages 33–133. Academic Press, 1978. MR0501668

[17] D.R. BRILLINGER and H.K. PREISLER. Two examples of quantal data analysis: a) multivariate point process, b) pure death process in an experimental design. In *Proceedings, XIII International Biometric Conference, Seattle*, pages 94–113. International Biometric Society, 1986.

[18] D.R. BRILLINGER and J.P. SEGUNDO. Empirical examination of the threshold model of neuron firing. *Biological Cybernetics*, 35:213–220, 1979.

[19] W.M. BROWN, T.D. GEDEON, A.J. BADDELEY, and D.I. GROVES. Bivariate J-function and other graphical statistical methods help select the best predictor variables as inputs for a neural network method of mineral prospectivity mapping. In U. Bayer, H. Burger, and W. Skala, editors, *IAMG 2002: 8th Annual Conference of the International Association for Mathematical Geology*, volume 1, pages 257–268. International Association of Mathematical Geology, 2002.

[20] C.F. CHUNG and F.P. AGTERBERG. Regression models for estimating mineral resources from geological map data. *Mathematical Geology*, 12:473–488, 1980.

[21] C.F. CHUNG and A.G. FABBRI. Probabilistic prediction models for landslide hazard mapping. *Photogrammetric engineering and remote sensing*, 62(12):1389–1399, 1999.

[22] M. CLYDE and D. STRAUSS. Logistic regression for spatial pair-potential models. In A. Possolo, editor, *Spatial Statistics and Imaging*, volume 20 of *Lecture Notes - Monograph series*, chapter II, pages 14–30. Institute of Mathematical Statistics, 1991. ISBN 0-940600-27-7. MR1195558

[23] D.R. COX. The statistical analysis of dependencies in point processes. In P.A.W. Lewis, editor, *Stochastic Point Processes*, pages 55–66. Wiley, New York, 1972. MR0375705

[24] N. CRESSIE. Change of support and the modifiable area unit problem. *Geographical Systems*, 3:159–180, 1996.

[25] N.A.C. CRESSIE. *Statistics for Spatial Data.* John Wiley and Sons, New York, 1991. MR1127423

[26] D.J. DALEY and D. VERE-JONES. *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods.* Springer Verlag, New York, second edition, 2003. MR1950431

[27] C.B. DEAN and R. BALSHAW. Efficiency lost by analyzing counts rather than exact times in Poisson and overdispersed Poisson regression. *Journal of the American Statistical Association*, 92:1387–1398, 1997. MR1615249

[28] A.P. DEMPSTER, N.M. LAIRD, and D.B. RUBIN. Maximum likelihood from incomplete data via the E–M algorithm. *Journal of the Royal Statistical Society B*, 39:1–22, 1977. MR0501537

[29] P.J. DIGGLE. *Statistical Analysis of Spatial Point Patterns.* Hodder Arnold, London, second edition, 2003. MR0743593

[30] P.J. DIGGLE and B. ROWLINGSON. A conditional approach to point process modelling of elevated risk. *Journal of the Royal Statistical Society, series A (Statistics in Society)*, 157(3):433–440, 1994.

[31] A.J. DOBSON and A.G. BARNETT. *An introduction to generalized linear models.* CRC Press, third edition, 2008. MR2459739

[32] P. ELLIOTT, J. WAKEFIELD, N. BEST, and D. BRIGGS, editors. *Spatial Epidemiology: Methods and Applications.* Oxford University Press, Oxford, 2000.

[33] R. FOXALL and A. BADDELEY. Nonparametric measures of association between a spatial point process and a random set, with geological applications. *Applied Statistics*, 51(2):165–182, 2002. MR1900164

[34] C.J. GEYER. Likelihood inference for spatial point processes. In O.E. Barndorff-Nielsen, W.S. Kendall, and M.N.M. van Lieshout, editors, *Stochastic Geometry: Likelihood and Computation*, number 80 in Monographs on Statistics and Applied Probability, chapter 3, pages 79–140. Chapman and Hall / CRC, Boca Raton, Florida, 1999. MR1673118

[35] P.V. GORSEVSKI, P.E. GESSLER, R.B. FOLZ, and W.J. ELLIOTT. Spatial prediction of landslide hazard using logistic regression and ROC analysis. *Transactions in GIS*, 10:395–415, 2006.

[36] C.A. GOTWAY and L.J. YOUNG. Combining incompatible spatial data. *Journal of the American Statistical Association*, 97:632–648, 2002. MR1951636

[37] D.I. GROVES, R.J. GOLDFARB, C.M. KNOX-ROBINSON, J. OJALA, S. GARDOLL, G.Y. YUN, and P. HOLYLAND. Late-kinematic timing of orogenic gold deposits and significance for computer-based exploration techniques with emphasis on the Yilgarn Block, Western Australia. *Ore Geology Reviews*, 17:1–38, 2000.

[38] A. HARDEGEN. *Efficient parameter estimation for spatial point processes.* PhD thesis, University of Western Australia, 2010. In preparation.

[39] R.J. HASENSTAB. A preliminary cultural resource sensitivity analysis for the proposed flood control facilities construction in the Passaic River basin of New Jersey. Technical report, Soil Systems, Inc, New York, 1983. Submitted to the Passaic River Basin Special Studies Branch, Department of the Army, USA. New York District Army Corps of Engineers.

[40] W.W. HAUCK, JR. and A. DONNER. Wald's test as applied to hypotheses in logit analysis. *J. Amer. Statist. Assoc.*, 72(360, part 1):851–853, 1977. MR0471146

[41] H.J. HERZOG. Change detection metrics and mixels (mixed picture elements) for computer analysis of low resolution remote sensing imagery. In *IEEE Region Six Conference Record*, pages 17–20. IEEE, May 1977. Persistent link: http://ieeexplore.ieee.org/servlet/opac?punumber=5782.

[42] D.W. HOSMER and S. LEMESHOW. *Applied logistic regression.* Wiley, second edition, 2000.

[43] J. ILLIAN, A. PENTTINEN, H. STOYAN, and D. STOYAN. *Statistical Analysis and Modelling of Spatial Point Patterns.* John Wiley and Sons, Chichester, 2008. MR2384630

[44] O. KALLENBERG. *Random measures.* Akademie-Verlag, Berlin, fourth edition, 1986. MR0854102

[45] A. KITAMOTO and M. TAKAGI. Image classification using probabilistic models that reflect the internal structure of mixels. *Pattern Analysis and Applications*, 2:31–43, 1999.

[46] C.M. KNOX-ROBINSON and D.I. GROVES. Gold prospectivity mapping using a geographic information system (GIS), with examples from the Yilgarn Block of Western Australia. *Chronique de la Recherche Minière*, 529:127–138, 1997.

[47] K. KRICKEBERG. Processus ponctuels en statistique. In P.L. Hennequin, editor, *Ecole d'Eté de Probabilités de Saint-Flour X*, volume 929 of *Lecture Notes in Mathematics*, pages 205–313. Springer, 1982. MR0665597

[48] Y.A. KUTOYANTS. *Statistical Inference for Spatial Poisson Processes.* Number 134 in Lecture Notes in Statistics. Springer, New York, 1998. MR1644620

[49] K.L. KVAMME. Computer processing techniques for regional modeling of archaeological site locations. *Advances in Computer Archaeology*, 1:26–52, 1983.

[50] K.L. KVAMME. A view from across the water: the North American experience in archeological GIS. In G.R. Lock and Z. Stančič, editors, *Archaeology and Geographical Information Systems: a European Perspective*, pages 1–14. CRC Press, 1995.

[51] K.L. KVAMME. There and back again: revisiting archeological locational modeling. In M.W. Mehrer and K.L.Wescott, editors, *GIS and archaeological site modelling*, pages 3–40. CRC Press, 2006.

[52] C.R.O. LAWOKO and G.J. MCLACHLAN. Bias associated with the discriminant analysis approach to the estimation of mixing proportions. *Pattern Recognition*, 22:763–766, 1989.

[53] E.L. LEHMANN. *Theory of point estimation.* John Wiley and Sons, New York, 1983. MR0702834

[54] P.A.W. LEWIS. Recent results in the statistical analysis of univariate point processes. In P.A.W. Lewis, editor, *Stochastic point processes*, pages 1–54. Wiley, New York, 1972. MR0359226

[55] J.K. LINDSEY. *The analysis of stochastic processes using GLIM.* Springer, Berlin, 1992.

[56] J.K. LINDSEY. *Modelling frequency and count data.* Oxford University Press, 1995.

[57] J.K. LINDSEY. *Applying generalized linear models.* Springer, 1997.

[58] P. MCCULLAGH and J.A. NELDER. *Generalized Linear Models.* Chapman and Hall, second edition, 1989. MR0727836

[59] J. MØLLER and R.P. WAAGEPETERSEN. *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall/CRC, Boca Raton, 2004. MR2004226

[60] G.C. OHLMACHER and J.C. DAVIS. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering Geology*, 69:331–343, 2003.

[61] S. OPENSHAW. *The modifiable area unit problem.* Geo Books, Norwich, 1984.

[62] T. ORCHARD and M.A. WOODBURY. A missing information principle: theory and applications. In L.M. Le Cam, J. Neyman, and E.L. Scott, editors, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 697–715, Berkeley, Calif., 1972. University of California Press. MR0400516

[63] R Development Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2004. ISBN 3-900051-07-0.

[64] S.L. RATHBUN and N. CRESSIE. Asymptotic properties of estimators of the parameters of spatial inhomogeneous Poisson point processes. *Advances in Applied Probability*, 26:122–154, 1994. MR1260307

[65] S.L. RATHBUN, S. SHIFFMAN, and C.J. GWALTNEY. Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika*, 94:153–165, 2007. MR2367829

[66] W.S. ROBINSON. Ecological correlations and the behavior of individuals. *American Sociological Review*, 15:351–357, 1950.

[67] R.T. ROCKAFELLAR. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1972. MR1451876

[68] R.T. ROCKAFELLAR and R.J.-B. WETS. *Variational analysis.* Springer, Berlin, 1998. MR1491362

[69] S.C. SCHOLTZ. Location choice models in Sparta. In R. Lafferty III, J.L. Ottinger, S.C. Scholtz, W.F. Limp, B. Watkins, and R. D. Jones, editors, *Settlement Predictions in Sparta: A Locational Analysis and Cultural Resource Assessment on the Uplands of Calhoun County, Arkansas*, number 14 in Arkansas Archaeological Survey Research Series, pages 207–222. Arkansas Archaeological Survey, Fayetteville, Arkansas, USA, 1981.

[70] D. SCHUHMACHER. *Estimation of distances between point process distributions.* PhD thesis, University of Zurich, Switzerland, 2005.

[71] D. SCHUHMACHER and A. XIA. A new metric between distributions of point processes. *Advances in Applied Probability*, 40:651–672, 2008. MR2454027

[72] M. SILVAPULLE. On the existence and uniqueness of the maximum likelihood estimates for the binomial response models. *Journal of the Royal Statistical Society, series B*, 43:310–313, 1981. MR0637943

[73] R. SUNDBERG. Maximum likelihood theory for incomplete data from an exponential family. *Scandinavian Journal of Statistics*, 1:49–58, 1974. MR0381110

[74] J.W. TUKEY. Discussion of paper by F.P. Agterberg and S.C. Robinson. *Bulletin of the International Statistical Institute*, 44(1):596, 1972. Proceedings, 38th Congress, International Statistical Institute.

[75] W.N. VENABLES and B.D. RIPLEY. *Modern Applied Statistics with S-Plus.* Springer, fourth edition, 2002. MR1337030

[76] R. WAAGEPETERSEN. Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika*, 95:351–363, 2008. MR2422695

[77] J. WAKEFIELD. A critique of statistical aspects of ecological studies in spatial epidemiology. *Environmental and Ecological Statistics*, 11:31–54, 2004. MR2019862

[78] J. WAKEFIELD. Disease mapping and spatial regression with count data. *Biostatistics*, 8:158–183, 2007.

[79] L.A. WALLER and C.A. GOTWAY. *Applied spatial statistics for public health data*. Wiley, 2004. MR2075123

[80] K.P. WATKINS and A.H. HICKMAN. Geological evolution and mineralization of the Murchison Province, Western Australia. Bulletin 137, Geological Survey of Western Australia, 1990. Published by Department of Mines, Western Australia, 1990. Available online from Department of Industry and Resources, State Government of Western Australia, `www.doir.wa.gov.au`.

[81] R.W.M. WEDDERBURN. On the existence and uniqueness of maximum likelihood estimates for certain generalized linear models. *Biometrika*, 63:27–32, 1976. MR0408092

[82] D. WHEATLEY and M. GILLINGS. *Spatial technology and archaeology: the archaeological applications of GIS*. Taylor and Francis, 2002.

[83] A. XIA. Stein's method and Poisson process approximation. In *An introduction to Stein's method*, volume 4 of *Lecture Notes Series, Institute of Mathematical Sciences, National University of Singapore*, pages 115–181. Singapore University Press, Singapore, 2005. MR2235450

[84] T. YOSHIDA and T. HAYASHI. On the robust estimation in Poisson processes with periodic intensities. *Annals of the Institute of Statistical Mathematics*, 42:489–507, 1990. MR1093965