

Chapter 2

A Statistical Commentary on Mineral Prospectivity Analysis



Adrian Baddeley

Abstract We compare and contrast several statistical methods for predicting the occurrence of mineral deposits on a regional scale. Methods include logistic regression, Poisson point process modelling, maximum entropy, monotone regression, nonparametric curve estimation, recursive partitioning, and ROC (Receiver Operating Characteristic) curves. We discuss the use and interpretation of these methods, the relationships between them, their strengths and weaknesses from a statistical standpoint, and fallacies about them. Potential improvements and extensions include models with a flexible functional form; techniques which take account of sampling effort, deposit endowment and spatial association between deposits; conditional simulation and prediction; and diagnostics for validating the analysis.

2.1 Introduction

The pioneering work of Agterberg (1974) developed a statistical strategy for predicting the likely occurrence of mineral deposits. In essence, the observed association between known deposits and other known geostructural or geochemical information is used to predict the spatially-varying abundance of unknown deposits. The association between predictors and deposits is modelled by logistic regression.

This general approach to prospectivity analysis has been extended and adopted across a wide range of applications, for predicting mineral deposits (Chung and Agterberg 1980; Bonham-Carter 1995), archaeological finds (Scholtz 1981; Kvamme 1983), landslides (Chung and Fabbri 1999; Gorsevski et al. 2006), animal and plant species (Franklin 2009) and other features which can be treated as points at the scale of interest. Extensions and modifications include logistic regression for sampled data, maximum entropy, and weights-of-evidence modelling.

However, the scientific literature contains many conflicting statements about the interpretation of these methods. For example, there are different understandings of

A. Baddeley (✉)
Department of Mathematics and Statistics, Curtin University, GPO Box U1987,
Perth, WA 6845, Australia
e-mail: adrian.baddeley@curtin.edu.au

the fundamental scope and validity of logistic regression, about the degree of flexibility inherent in the assumptions, and about the interpretation of the results. This is a concern, because misunderstanding of a statistical technique poses the obvious risk that it may be mis-applied, its results misinterpreted, or its performance incorrectly evaluated.

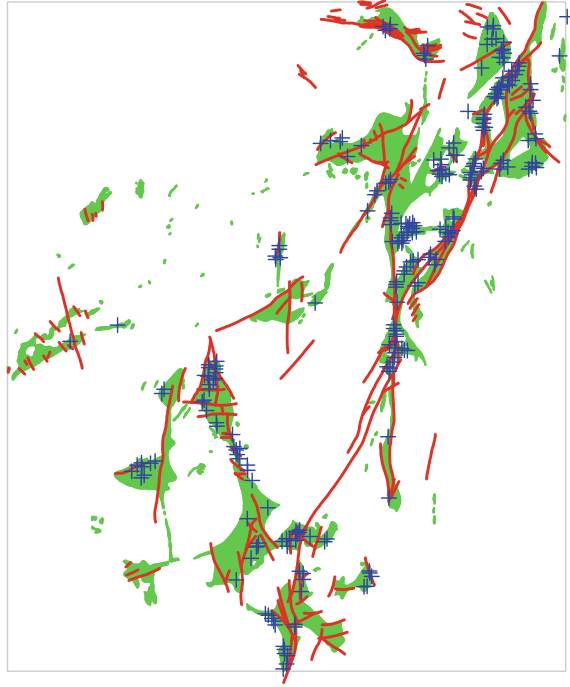
In statistical science the understanding of these techniques has also changed dramatically over the last four decades. The modern synthesis of statistical modelling permits a new and deeper appreciation of prospectivity methods. New tools from statistical science may enable exploration geologists to perform a more searching analysis of their survey data.

Accordingly, this article offers a commentary and critique of prospectivity analysis from the standpoint of modern statistical methodology. We begin by examining the fundamentals of logistic regression, explaining the interpretation of the results, and discussing its strengths and weaknesses. We explain the close relationship between logistic regression, point process modelling, and maximum entropy methods. We canvas some alternative methods which are less well known, including monotone regression, nonparametric regression, recursive partitioning models, and ROC curves. (The popular weights-of-evidence method is not discussed here, but will be treated in detail in another article.) New tools include robust estimation, model selection and variable selection, conditional prediction and model diagnostics. Several unanswered questions in prospectivity analysis are identified as topics for future research in statistical methodology.

2.2 Example Data

For the sake of demonstration and discussion, we shall use a vastly oversimplified example. The Murchison geological survey data shown in Fig. 2.1 record the spatial locations of gold deposits and associated geological features in the Murchison area of Western Australia. They are extracted from a regional survey (scale 1:500,000) of the Murchison area made by the Geological Survey of Western Australia (Watkins and Hickman 1990). The features shown in the Figure are the known locations of gold deposits, the known or inferred locations of geological faults, and greenstone outcrop. The study region is contained in a 330×400 km rectangle. At this scale, gold deposits are point-like, i.e. their spatial extent is negligible. These data were previously analysed in Foxall and Baddeley (2002), Brown et al. (2002); see also Groves et al. (2000), Knox-Robinson and Groves (1997). Data were kindly provided by Dr. Carl Knox-Robinson, and permission granted by Dr. Tim Griffin, Geological Survey of Western Australia and by Dr. Knox-Robinson.

Fig. 2.1 Murchison geological survey data. Known gold deposits (blue crosses), major geological faults (red lines) and greenstone outcrop (green shading) in a survey region about 200 by 300 km across



Evidently, both the geological fault pattern and the greenstone outcrop have “predictive” value for gold prospectivity, because gold deposits are strongly associated with proximity to both features. For the purposes of analysis in this article, we require predictors to be spatial variables. A predictor Z should be a function $Z(u)$ defined at any spatial location u . For a map of rock type such as the greenstone outcrop, the simplest choice for the predictor value $Z(u)$ at location u is the “indicator” equal to 1 if the location u falls inside the greenstone, and 0 if it falls outside. For a map of linear features such as geological faults, a common choice for the predictor value $Z(u)$ is the distance from u to the nearest fault. Figure 2.2 shows contours of this distance function for the Murchison data.

It is important to note that our choice of spatial predictor $Z(u)$ will affect the results of the analysis: the results would usually be different if we replace the distance function in Fig. 2.2 by the squared distance or the square root of distance, etc. Several other choices of spatial predictor derived from geological fault data are canvassed in Berman and Turner (1992). Likewise for the greenstone outcrop we could have chosen another predictor, such as the distance function of the greenstone. The choice of predictor can be revisited after the analysis, as discussed in Sect. 2.4.6.

Fig. 2.2 Contours of distance to the nearest fault in the Murchison survey



2.3 Logistic Regression

Here we recapitulate and re-examine some details of the logistic regression technique, for the purposes of discussion.

2.3.1 Basics of Logistic Regression

Logistic regression is a general statistical technique for modelling the relationship between a binary response variable and a numerical explanatory variable (Berkson 1955; McCullagh and Nelder 1989; Dobson and Barnett 2008; Hosmer and Lemeshow 2000). The use of logistic regression to predict the presence/absence of point events was pioneered in geology by Agterberg (1974, 1980), apparently on the suggestion of the statistician Tukey (1972): see Agterberg (2001). The study region is divided into pixels; in each pixel the presence or absence of any deposits is recorded; then logistic regression is used to predict the probability of the presence of a deposit as a function of predictor variables. This was later independently rediscovered in archaeology (Scholtz 1981; Hasenstab 1983; Kvamme 1983, 1995) and

is now a standard technique in GIS applications (Bonham-Carter 1995) including spatial ecology (Franklin 2009).

The study region is divided into pixels of equal area. For each pixel, we record whether mineral deposits are present or absent. We then fit the *logistic regression* relationship

$$\log \frac{p}{1-p} = \alpha + \beta z \quad (2.1)$$

where p is the probability of presence of a deposit (or deposits) in a given pixel, and z is the corresponding value of the predictor variable.

Here α and β are model parameters which are estimated from the data. Some writers state that the interpretation of α and β is “obscure” (Wheatley and Gillings 2002, p. 175), perhaps because of the unfamiliar form of the left hand side of (2.1). The quantity $p/(1-p)$ is the *odds* of presence against absence, that is, the probability p of presence of a deposit, divided by the probability $1-p$ of absence. The left hand side of (2.1) is the logarithm of the odds of presence. (In this paper ‘log’ always refers to the natural logarithm, with base e .) The logistic regression relationship (2.1) states that the *log odds* of presence is a linear function of the predictor z . The straight line has slope β and intercept α . The transformation $\log(p/(1-p))$ ensures that

$$p = \frac{e^{\alpha+\beta z}}{1 + e^{\alpha+\beta z}} \quad (2.2)$$

is a well-defined probability value (between 0 and 1) for any possible values of α , β and z . The log odds is the “canonical” choice of transformation in order to satisfy some desirable statistical properties (McCullagh and Nelder 1989), and arises naturally in many applications. Bookmakers often quote gambling odds that are equally spaced on a logarithmic scale, such as the sequence 2:1, 4:1, 8:1, 16:1. Since logistic regression is widely used in medical and public health research, standard statistical textbooks contain many useful ways to interpret and explain these quantities (Hosmer and Lemeshow 2000).

Once the parameters α , β have been estimated from data (as detailed in Sect. 2.3.3), the predicted probabilities p_j can be computed using (2.2) and displayed as colours or greyscales in a pixel image, as shown in Fig. 2.3. Qualitative interpretation of the map seems to be adequate for many purposes, while many writers recommend using only the *sign* of the slope parameter β (Gorsevski et al. 2006, pp. 405–407). However, much more can be done with the fitted logistic regression, as we discuss below.

The general appearance of Fig. 2.3 is very similar to the contour plot of distance to nearest fault in Fig. 2.2. This is a foreseeable consequence of the simple model (2.1) which implies that contours of probability are contours of distance to nearest fault. This is not true of more complicated models involving several predictors.

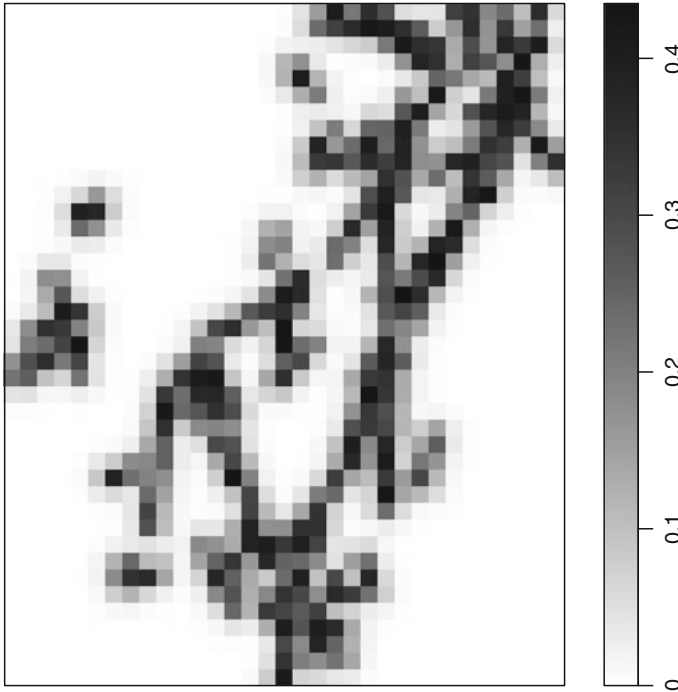


Fig. 2.3 Fitted probability of a gold deposit in each 10-km-square pixel in the Murchison survey, estimated by logistic regression. Pixel values are probabilities (between 0 and 1)

2.3.2 Flexibility and Validity

Some writers describe logistic regression as a ‘nonparametric’ technique (Kvamme 2006, p. 24), which would suggest that it is able to detect and respond to any kind of relationship, not specified in advance, between the predictor z and the presence probability p . On the contrary, logistic regression is a parametric model of a very simple kind. The relationship z and p is rigidly defined by Eqs. (2.1) and (2.2): *the relationship is linear* on the scale of the log odds. The position of the line is determined by the two parameters α and β . Logistic regression could be *false* for a particular application: that is, the model assumptions could be incorrect.

Logistic regression is an example of a “generalised linear model” (McCullagh and Nelder 1989; Dobson and Barnett 2008), essentially a linear regression of the transformed probabilities against the predictor. In the analysis of the Murchison data shown here, if we replace the distance function $Z(u)$ by its square $Z(u)^2$, or square root $\sqrt{Z(u)}$, etc. in the logistic regression, we obtain a different model, which is incompatible with the original model. If the log odds are a linear function of squared distance, then they are *not* a linear function of distance. Consequently, the choice of predictor variable is very important, and it involves an implicit *model assumption*

about the relationship between presence probability p and predictor z . Even the sign of the fitted slope parameter β could be misleading if the predictor was chosen incorrectly.

Such freedom as does exist in the logistic regression model is the freedom to choose the predictor or predictors $Z(u)$. Once the predictor is chosen, the model becomes rigid. If there is concern about the form of relationship between p and Z , one simple strategy is to fit a polynomial, instead of linear, relationship between the log odds and the predictor variable.

Statistical science has developed an armory of techniques for “validating” a regression analysis (Harrell 2001; Hosmer and Lemeshow 2000). These include diagnostics for checking the validity of the logistic regression relationship (2.1), measures of sensitivity of the fitted model to the data, techniques for selecting the most important variables and the most informative models, and measures of goodness-of-fit. As far as the author is aware, these techniques are rarely used in geoscience. This presents the risk of failing to detect situations where logistic regression analysis is not appropriate. Model validation is a kind of “due diligence” for data analysts.

A weakness of all parametric modelling is that, because of its “low degrees of freedom”, the model predictions at a given location are heavily influenced by the entire dataset, including data observed under very different conditions. In the Murchison example, the predicted probability of presence of a gold deposit declines dramatically between distances 0, 1 and 2 km from the nearest fault. This is not necessarily a reflection of the observed frequency of occurrence of gold deposits at these distances: rather, it is a consequence of the large negative value of the estimated slope parameter β , which arises because of the scarcity of gold deposits at much larger distances.

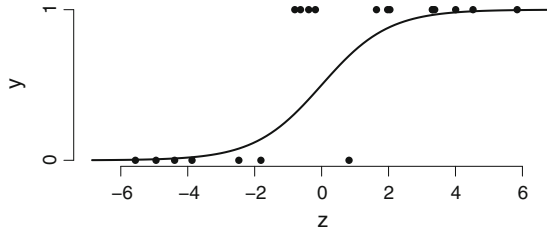
Extension of the logistic regression technique to account for characteristics of the mineral deposits, such as total endowment of gold, would be problematic because it would effectively require a model for the probability distribution of the endowment (and this might also be spatially-varying). However, it is straightforward to apply logistic regression to different subsets of the deposits, for example to predict the occurrence of deposits with endowment exceeding a specified threshold.

The logistic regression technique described here assumes that the relationship (2.1) holds throughout the study region, with the same parameter values α, β throughout. This assumption can be avoided using geographically-weighted logistic regression (Lloyd 2011) or local likelihood estimation (Loader 1999; Baddeley 2017) which allow the parameters to be spatially-varying.

2.3.3 Fitting Procedure and Implicit Assumptions

For the discussion it will be important to know a few details about the procedure that is used to fit the logistic regression relationship.

Fig. 2.4 Illustrative example of binary responses y (filled circles) and fitted probabilities (solid curve) plotted against predictor value z



Suppose there are N pixels, with covariate values z_1, \dots, z_N respectively, and pixel presence/absence indicators y_1, \dots, y_N respectively, where $y_j = 1$ if the j -th pixel contains a mineral deposit, and $y_j = 0$ if not. The goal is to fit a relationship of the form (2.2). This is not a simple matter of curve-fitting, because the data (z_j, y_j) do not lie “along” or “near” the curve in any sense. See Fig. 2.4. Instead, it is necessary to specify a measure of closeness or agreement between the curve and the observed data: the model is fitted by choosing the parameter values α, β which make this agreement as close as possible.

The classical fitting method is *maximum likelihood*. Given the data y_1, \dots, y_N and z_1, \dots, z_N , define the *likelihood* $L(\alpha, \beta)$ to be the theoretical probability of obtaining the observed pattern of outcomes (y_1, \dots, y_N) , as a function of the unknown parameter values α and β . The likelihood is a measure of agreement between the logistic regression curve and the observed data.

To find the likelihood, first consider a single pixel j where $j = 1, 2, \dots, N$. The probability of obtaining a presence ($y_j = 1$) in this pixel is

$$p_j = \frac{e^{\alpha + \beta z_j}}{1 + e^{\alpha + \beta z_j}} \quad (2.3)$$

and the probability of an absence ($y_j = 0$) is $1 - p_j$. The likelihood for pixel j is the probability of obtaining the *observed* outcome y_j ,

$$L_j = \mathbb{P}\{Y_j = y_j\} = \begin{cases} p_j & \text{if } y_j = 1 \\ 1 - p_j & \text{if } y_j = 0 \end{cases}$$

or more compactly

$$L_j = p_j^{y_j} (1 - p_j)^{1 - y_j} = \left(\frac{p_j}{1 - p_j} \right)^{y_j} (1 - p_j)$$

which is a function $L_j = L_j(\alpha, \beta)$ of the unknown values of the parameters. Then the full likelihood is the predicted probability of the entire observed pattern of presences and absences (y_1, \dots, y_N) ,

$$L = L_1 L_2 \dots L_N, \quad (2.4)$$

and is a function $L = L(\alpha, \beta)$ of the unknown parameter values α, β . Equation (2.4) assumes that the outcomes in different pixels are statistically independent of each other, because the likelihood is obtained by multiplying likelihood contributions from each pixel. That is, the logistic regression technique, as it is commonly applied to presence/absence data, makes two assumptions:

1. the probability of presence p is related to the predictor variable z by a logistic regression relationship (2.1);
2. presence/absence outcomes in different pixels are statistically independent of each other.

The (parametric) *maximum likelihood* fitting rule is to choose the values of the parameters α, β which maximise the likelihood $L(\alpha, \beta)$. This is a standard procedure in classical statistics, carrying with it many useful additional tools such as standard errors, confidence intervals, and significance tests (Hogg and Craig 1970; Freedman et al. 2007).

Ignoring some pathological cases (e.g. where no deposits are observed), the likelihood is maximised by setting its partial derivatives to zero. Equivalently we may work with the derivatives of $\log L$. This yields the *score equations* for logistic regression

$$\sum_{j=1}^N p_j = \sum_{j=1}^N y_j \quad (2.5)$$

$$\sum_{j=1}^N p_j z_j = \sum_{j=1}^N y_j z_j \quad (2.6)$$

obtained by setting $\partial \log L / \partial \alpha = 0$ and $\partial \log L / \partial \beta = 0$ respectively. Typically the score equations have a unique solution in (α, β) , giving the maximum likelihood estimates $\hat{\alpha}, \hat{\beta}$ of the parameters. There are no explicit formulae for $\hat{\alpha}, \hat{\beta}$ and the score equations must be solved numerically.

The score equations (2.5)–(2.6) have a commonsense interpretation in their own right. In (2.5) the right hand side is the observed number of deposits, while the left hand side is the expected (mean) number of deposits according to the model. In (2.6) the right hand side is the sum of the predictor values at the observed deposits, while the left hand side is the expected (mean) value of this sum according to the model. In this case maximum likelihood is equivalent to the “method of moments” in which parameters are estimated by equating the observed value of a statistic to its theoretical mean value.

Logistic regression is a simple two-parameter model, equivalent to linear regression on a transformed scale. The parameters are estimated using the entire dataset, as shown by Eq. (2.4) or (2.5)–(2.6). Consequently, the presence probability predicted by logistic regression, for a pixel with predictor value z , is influenced by data where the predictor value is very different from z , as discussed above.

It is not obligatory to use maximum likelihood estimation to fit the logistic regression model. Although maximum likelihood is theoretically optimal if the logistic regression model is true, it may fail if the model is false (“non-robust to mis-specification”) and it is sensitive to anomalies in the data (“non-robust against outliers”). Robustness against outliers can be improved using *penalised likelihood* in which the likelihood L is multiplied by a term $b(\alpha, \beta)$ which penalises large parameter values.

2.3.4 Pixel Size and Model Consistency

Dependence on Pixel Size

The results of a logistic regression analysis clearly depend on the size of the pixels used. Table 2.1 shows estimates of the parameters α and β in the logistic regression of gold deposits against distance from the nearest fault, in the Murchison data, obtained using different pixel grid sizes. Estimates of the slope parameter β are roughly consistent between different grids. The estimate of the intercept parameter α becomes lower (more negative) as the pixels become smaller, so that the predicted presence probabilities also become smaller: this is intuitively reasonable, since a smaller pixel must have a smaller chance of containing a deposit.

The score equations help to explain Table 2.1. If the pixel grid is subdivided into a finer grid, the right-hand sides of (2.5) and (2.6) are unchanged, so the left-hand sides must also be unchanged. Since the number of pixels N has been increased by the subdivision, the predicted probabilities p_j must decrease by the same proportion f , the ratio of pixel areas in the two grids. Using $\log(p/(1-p)) \approx \log p$ for small p , the estimate of α must decrease by approximately $\log f$.

In order to make the results approximately consistent between different pixel sizes, the logistic regression (2.1) could be modified to

$$\log \frac{p}{1-p} = \log A + \alpha + \beta z \quad (2.7)$$

Table 2.1 Fitted logistic regression parameters for Murchison data

Pixel size (km)	α	β
10	-0.260	-0.243
5	-1.321	-0.282
2	-2.947	-0.261
1	-4.303	-0.266
0.5	-5.681	-0.270

Table 2.2 Fitted logistic regression parameters for Murchison data, adjusted for pixel area

Pixel size (km)	α	β
10	-4.844	-0.243
5	-4.532	-0.282
2	-4.332	-0.261
1	-4.302	-0.266
0.5	-4.293	-0.270

where A is the pixel area used. In the language of statistical modelling, the constant $\log A$ plays the role of an *offset* in the model formula. The resulting, adjusted estimates for the parameters α, β from the Murchison data are shown in Table 2.2, and they are indeed approximately consistent across different pixel sizes. They could have been obtained from the results in Table 2.1 by subtracting $\log A$ from the estimates of α .

For reasons explained below, slightly better consistency is achieved by replacing logistic regression (2.1) by *complementary log–log regression*

$$\log(-\log(1 - p)) = \log A + \alpha + \beta z. \quad (2.8)$$

Large Pixels

Large pixel sizes are preferred by some researchers. A common justification is that predictions are desired for large spatial regions, for example, the probability that the entire exploration lease contains at least one deposit. Some researchers also feel that small pixel sizes are inappropriate because they lead to tiny probability values, which may be considered physically unrealistic.

However, large pixels are not needed in order to predict the probability of a deposit in a large spatial region R . Suppose that a logistic regression model has been fitted using a fine grid of pixels. If the region R is decomposed into pixels, the probability $p(R)$ of presence of at least one deposit in R satisfies

$$1 - p(R) = \prod_{j \in R} (1 - p_j), \quad (2.9)$$

where $\prod_{j \in R}$ denotes the product over all pixels in R . The left hand side is the probability that there are no deposits in R . On the right hand side, $(1 - p_j)$ is the probability that there are no deposits in pixel j , and since pixel outcomes are assumed to be independent, these pixel absence probabilities should be multiplied together. Hence, $p(R)$ can be calculated using presence probabilities for a fine pixel grid.

Moreover, the use of large pixels in logistic regression causes difficulties, related to the aggregation of points into geographical areas (Elliott et al. 2000; Waller and Gotway 2004; Wakefield 2007, 2004). The most important of these is the statistical bias due to aggregation ('ecological bias', Wakefield (2004, 2007) or 'aggre-

gation bias', Dean and Balshaw (1997), Alt et al. (2001)). The 'ecological fallacy' (Robinson 1950) is the incorrect belief that a model fitted to aggregated data will apply equally to the original un-aggregated data. The 'modifiable area unit problem' (Openshaw 1984) or 'change-of-support' (Gotway and Young 2002; Banerjee and Gelfand 2002; Cressie 1996) is the problem of reconciling models that were fitted using different pixel sizes or aggregation levels.

Our analysis in Baddeley et al. (2010) shows that aggregation bias is highly dependent on the smoothness of the predictor as a function of spatial location. The distance-to-nearest-fault predictor in the Murchison example, and indeed the distance transform of any spatial feature, is a Lipschitz-continuous function of spatial location, which leads to relatively small aggregation bias. This is illustrated by Table 2.2. However, a predictor which indicates a classification, such as rock type, may have very substantial bias due to aggregation, persisting even at small pixel sizes (Baddeley et al. 2010).

Strictly speaking it can be *impossible* to reconcile two spatial logistic regression models fitted to the same spatial point pattern data using different pixel grids. Two such models are often logically incompatible (Baddeley et al. 2010), because the product rule (2.9) is incompatible with the logistic relation (2.1). It may help to recall that the pixels are artificial. A logistic regression model, using pixels of a particular size, makes an implicit assumption about the spatial random process of points in continuous space. For different pixel sizes, the corresponding assumptions are different, and generally incompatible. There is no random process in continuous space which satisfies a logistic regression model when it is discretised on *every* pixel grid. Two research teams who apply spatial logistic regression to the same data, but using different pixel sizes, may obtain results that cannot be reconciled exactly. This incompatibility can be eliminated by using complementary log–log regression (2.8) instead of logistic regression.

Small Pixels

Mathematical theory suggests that pixels should be as small as possible, in order to reduce the unwanted effects of aggregation (Baddeley et al. 2010). However, if this is taken literally, several practical problems arise. Small pixel size implies a large number of pixels. Software for logistic regression may suffer from numerical overflow. In a fine pixellation, the overwhelming majority of pixels do not contain a data point, so the overwhelming majority of response values y_j are zero. This may cause numerical instability and algorithm failure. The standard algorithm for fitting logistic regression, Iteratively-Reweighted Least Squares (McCullagh and Nelder 1989), relies on second-order Taylor approximation of the log likelihood: the algorithm itself may fail when it encounters a numerically singular matrix, or the associated statistical tools may behave incorrectly due to the *Hauck-Donner effect* (Hauck and Donner 1977).

One valid strategy for avoiding these problems is to take only a random sample of the absence-pixels (the pixels with $y_j = 0$), and to apply logistic regression to the subsampled data, using an additional offset to adjust for the sampling (Baddeley et al. 2015, Sect. 9.10).

A more natural and comprehensive solution is described in the next section.

2.4 Poisson Point Process Models

Pixels are artificial, so it is reasonable to ask whether logistic regression for pixel data has a well-defined meaning in continuous space, without reference to the pixel grid and pixel size. The appropriate meaning is that of the Poisson point process, studied below.

2.4.1 Logistic Regression with Infinitesimal Pixels

Logistic regressions fitted using different pixel sizes may be logically incompatible, except when the pixel size is very small. Accordingly, the only consistent interpretation of logistic regression is obtained by making the pixels *infinitesimal*.

Infinitesimal pixel size is a mathematical rather than a physical concept; it is comparable to the use of infinitesimal increments dx in differential and integral calculus. The practical user will not be required to “construct” infinitesimal pixels; they will exist only in the mathematical theory. Real physical measurements will be expressed as integrals over these infinitesimal pixels.

The presence probability p in an infinitesimal pixel will be infinitesimal. A more tangible quantity is the *intensity* or *rate* λ , loosely defined as the expected number of deposit points per unit area. In a pixel of very small area A , at most one deposit point will be present, so the expected number of points is equal to the probability of presence, and we have $\lambda \approx p/A$.

Logistic regression with infinitesimal pixels can be derived heuristically by letting the pixel size tend to zero. A rigorous argument is laid out in Baddeley et al. (2010), Warton and Shepherd (2010a, b). Assume that, for a small enough pixel size, logistic regression holds in the adjusted form (2.7), and that pixel outcomes are independent. Since p is small, $\log(p/(1-p)) \approx \log p$, so that the logistic regression implies

$$\log p = \log A + \alpha + \beta z$$

or equivalently

$$\log \lambda = \alpha + \beta z.$$

This gives a consistent limit as pixel area tends to zero. In the limit, the intensity $\lambda(u)$ at a spatial location u is a loglinear function of the predictor,

$$\lambda(u) = \exp(\alpha + \beta Z(u)) \tag{2.10}$$

where $Z(u)$ is the predictor value at location u .

Contrary to the claim that logistic regression is a flexible “nonparametric” model, we conclude that logistic regression is tantamount to assuming a loglinear (exponential) relationship between the density of deposits per unit area λ and the predictor variable Z .

2.4.2 Poisson Point Process

Logistic regression, as commonly applied to presence/absence data, implicitly assumes that pixel outcomes are independent of each other. If independence holds for sufficiently small pixel size then, invoking the classical Poisson limit theorem, the random number of deposits falling in any spatial region R must follow a Poisson distribution.

Definition 1 A random variable K taking nonnegative integer values has a Poisson distribution with mean μ if

$$\mathbb{P}\{K = k\} = e^{-\mu} \frac{\mu^k}{k!} \quad (2.11)$$

for any $k = 0, 1, 2, \dots$

Consequently (Warton and Shepherd 2010a, b; Baddeley et al. 2010; Renner et al. 2015)

Theorem 1 *If logistic regression holds in the adjusted form (2.7) for sufficiently small pixels, then the random spatial pattern of deposit points must follow a Poisson point process with intensity of the form (2.10).*

Definition 2 The spatial Poisson point process with intensity function $\lambda(u)$, $u \in \mathbb{R}^2$ is characterised by the following properties:

(PP1) Poisson counts: the number $n(\mathbf{X} \cap B)$ of points falling in any region B has a Poisson distribution;

(PP2) intensity: the number $n(\mathbf{X} \cap B)$ of points falling in a region B has expected value

$$\mu(B) = \mathbb{E}[n(\mathbf{X} \cap B)] = \int_B \lambda(u) du; \quad (2.12)$$

(PP3) independence: if B_1, B_2, \dots are disjoint regions of space then $n(\mathbf{X} \cap B_1)$, $n(\mathbf{X} \cap B_2), \dots$ are independent random variables;

(PP4) conditional property: given that $n(\mathbf{X} \cap B) = n$, the n points are independent and identically distributed, with common probability density

$$f(u) = \frac{\lambda(u)}{I} \quad (2.13)$$

where $I = \int_B \lambda(u) du$.

The intensity function $\lambda(u)$ completely determines the Poisson point process model. It encapsulates both the abundance of points (by Eq. (2.12)) and the spatial distribution of individual point locations (by Eq. (2.13)). Values of intensity have dimension length⁻².

The properties listed above can be used directly to simulate random realisations of the Poisson process. See Daley and Vere-Jones (2003, 2008) for an authoritative treatise on point processes, or (Baddeley et al. 2015, Chaps. 5, 9) for an introduction, and Kutoyants (1998), Møller and Waagepetersen (2004) for further details of statistical theory for point processes.

Theorem 1 establishes a logically consistent, physical meaning in continuous space for the logistic regression model fitted to pixel presence/absence data. Whereas logistic regression models can be somewhat difficult to interpret in practical terms, the infinitesimal-pixel limit of logistic regression is a very simple model, a Poisson point process whose intensity $\lambda(u)$ depends exponentially (log-linearly) on the predictor $Z(u)$ through (2.10). This model is well-studied, and permits highly detailed predictions to be made about various quantities, such as the expected number of points in a target region (using **PP2**), the probability of exactly k points in a target region (using **PP1**), and the probability distribution of distance from a fixed starting location to the nearest random point.

The conclusion of Theorem 1 remains true in the more general case where the pixel outcomes are weakly dependent on each other (Baddeley et al. 2010, Theorem 3).

From a statistical perspective, the Poisson point process is the fundamental model, while logistic regression is a practical technique for fitting this model approximately on a discretised grid. The connection between them is not a surprise: indeed it is strongly suggested by the standard ‘infinitesimal’ description of the Poisson point process (Breiman 1968). It is inconceivable that Tukey (1972) was unaware of this connection.

2.4.3 *Fitting a Poisson Point Process Model*

Fitting Procedures

We emphasise the distinction between a statistical model and the procedure used to fit the model. The statistical model is a description of both the systematic tendencies and the random variability in the observations, and allows us to make predictions. The model must first be fitted to the observed data. The fitting procedure is not uniquely determined by the model (unless we choose to follow a rule such as maximum likelihood) and there may be several possible choices of procedure, each with its own merits.

The Poisson point process, with loglinear intensity (2.10), has been identified as the relevant model for spatial point pattern data in continuous space. We shall now mention several possible fitting procedures for this model.

First we consider maximum likelihood. Suppose that the observed deposit locations are x_1, \dots, x_n in study region W . Then the log likelihood of the Poisson point process with intensity function $\lambda(u)$ is

$$\log L = \sum_{i=1}^n \log \lambda(x_i) + \int_W (1 - \lambda(u)) du. \quad (2.14)$$

This can be derived either from the characteristic properties (PP1)–(PP4) of the Poisson process, or by taking the limit of the logistic regression likelihood (2.4), with appropriate rescaling, as pixel size tends to zero. See Baddeley et al. (2010), Warton and Shepherd (2010a, b), Baddeley et al. (2015, Sect. 9.7).

For the loglinear intensity model (2.10), the score equations are obtained by setting the partial derivatives of (2.14) to zero, giving

$$\int_W \lambda(u) du = n \quad (2.15)$$

$$\int_W Z(u) \lambda(u) du = \sum_{i=1}^n Z(x_i) \quad (2.16)$$

and these are also the infinitesimal-pixel limits of the logistic regression score equations (2.5)–(2.6). The score equations have the same “method-of-moments” interpretation as in the discrete case: namely the left hand side of each equation is the theoretical mean value, under the model, of the statistic that is evaluated for the observed data on the right hand side.

The main practical challenge in fitting the model is the fact that Eqs. (2.14) or (2.15)–(2.16) involve an integral over the study region. Unless this integral can be simplified using calculus, it must be approximated numerically.

An important case where the integral *can* be simplified is where $Z(u)$ takes only the values 0 and 1. This predictor might represent a particular rock type such as the greenstone in the Murchison example. If this is the only predictor, then the integrals in (2.14)–(2.16) can be evaluated exactly, given only the area of the greenstone and non-greenstone regions, because the integrands are constant in each region. Then the model can be fitted exactly. This case is a rare exception.

Pixel Regression

The simplest approximation of an integral is the midpoint rule, using the sum of values of the integrand at a regular grid of sample points. This leads to the logistic regression technique of Sect. 2.3. The observed spatial locations x_1, \dots, x_n of the deposits are discretised into pixel presence-absence indicators y_1, \dots, y_N . The predictor Z is evaluated at the pixel centres c_j to give predictor values $z_j = Z(c_j)$, and logistic regression of y against z is performed.

Procedures of this type are well-established in statistical science. Lewis (1972) and Tukey’s former student Brillinger (Brillinger 1978; Brillinger and Segundo 1979; Brillinger and Preisler 1986) 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 Besag et al. (1982).

This makes it practicable to fit spatial Poisson point process models of general form to point pattern data (Berman and Turner 1992; Clyde and Strauss 1991; Baddeley and Turner 2000, 2005) by enlisting efficient and reliable software already developed for generalized linear models. Approximation of a stochastic process by a generalized linear model is now commonplace in applied statistics (Lindsey 1992, 1995, 1997; Lindsey and Mersch 1992).

Complementary log–log regression is more appropriate than logistic regression in this context. A Poisson random variable K with mean μ has probability $\mathbb{P}\{K = 0\} = e^{-\mu}$ of taking the value zero, by (2.11), and has probability $p = 1 - e^{-\mu}$ of taking a positive value. In a Poisson point process with intensity function $\lambda(u)$, the presence probability of at least one deposit in a given region B is therefore

$$p(B) = 1 - e^{-\mu(B)} = 1 - \exp\left(-\int_B \lambda(u) \, du\right).$$

Inverting this relationship, the expected number of points in B is

$$\int_B \lambda(u) \, du = -\log(1 - p(B))$$

If B is a small pixel of area A , and the intensity has the loglinear form (2.10), then the relationship between presence probability and the predictor variable is

$$\log(-\log(1 - p)) = \log A + \alpha + \beta z,$$

which follows the complementary log–log regression relationship (2.8) rather than the logistic regression (2.1). However, the discrepancy is small in many cases, and the logistic function $\log(p/(1 - p))$ has slightly better numerical and computational properties, because it is the theoretically “canonical” link function (McCullagh and Nelder 1989).

Berman-Turner Device

In numerical analysis, an integral can often be approximated more accurately using a *quadrature rule*, based on a small number of well-chosen sample points, rather than a dense grid of sample points. Berman and Turner (1992) applied this principle to the Poisson point process likelihood (2.14) and developed an efficient fitting procedure based on a relatively small number of sample points.

In the Berman-Turner scheme, the sample points u_1, \dots, u_m consist of the observed deposit locations x_1, \dots, x_n together with a complementary set of “dummy” points u_{n+1}, \dots, u_m . The integral of any function f is approximated using the quadrature rule

$$\int_W f(u) \, du \approx \sum_k w_k f(u_k),$$

where w_1, \dots, w_m are numerical weights chosen appropriately. For example, the weights w_k may be the areas of the tiles of the Dirichlet-Voronoi tessellation (Okabe et al. 1992) of W associated with the quadrature points u_1, \dots, u_m . The Poisson process log likelihood (2.14) is then approximated by

$$\begin{aligned} \log \text{BTL} &= \sum_{i=1} \log \lambda(x_i) + \sum_k w_k (1 - \lambda(u_k)) \\ &= \sum_k (I_k \log \lambda(u_k) + w_k (1 - \lambda(u_k))) \end{aligned} \quad (2.17)$$

where $I_k = 1$ if the quadrature point u_k is a data point, and $I_k = 0$ if it is a dummy point. The approximate log likelihood (2.17) has the same form as the (weighted) log likelihood of a Poisson regression model, and can be fitted reliably using existing statistical software (Berman and Turner 1992). The Berman-Turner technique is the main algorithm for point process modelling in the software package `spatstat` (Baddeley et al. 2015, Chap. 9).

If the predictor variables are smooth functions of spatial location, then the Berman-Turner device is extremely efficient, because of the properties of numerical quadrature (Berman and Turner 1992; Baddeley and Turner 2000). This applies, for example, to the distance function of the geological faults in the Murchison example. The approximation is less accurate when the predictor is discontinuous, such as an indicator of rock type.

Conditional Logistic Regression

An alternative fitting method involves placing the “dummy” sample points at random. This is the equivalent of the procedure, already described for pixel presence/absence data, of randomly selecting a subset of the pixels where no deposit is present.

Suppose the dummy point pattern is randomly generated according to a Poisson point process with known intensity $\delta > 0$. Combine the two point patterns, data \mathbf{x} and dummy \mathbf{d} , into a single pattern $\mathbf{v} = \mathbf{x} \cup \mathbf{d}$; this is a realisation of a random point process with intensity $\kappa(u) = \lambda(u) + \delta$. Given $\mathbf{v} = \{v_1, \dots, v_j\}$, that is, given only the locations of the combined pattern of data and dummy points, let s_1, \dots, s_j be indicators such that $s_j = 1$ if the point v_j is a data point, and $s_j = 0$ if it is a dummy point. The probability $q_j = \mathbb{P}\{s_j = 1\}$ that a given random point v_j is actually a data point, equals

$$q_j = \frac{\lambda(v_j)}{\lambda(v_j) + \delta},$$

the ratio of the intensity of \mathbf{x} to the intensity of \mathbf{v} . Hence

$$\log \frac{q_j}{1 - q_j} = \log \lambda(v_j) - \log \delta = \alpha + \beta Z(v_j) - \log \delta. \quad (2.18)$$

The data/dummy status s_j of each point v_j is independent of other points. It follows that the conditional likelihood of the data/dummy status of the points of \mathbf{v} , given their locations, is the likelihood of logistic regression in the form (2.18). The Poisson point process model with loglinear intensity (2.10) could be fitted by logistic regression of s_j on $z_j = Z(v_j)$ given \mathbf{v} .

This technique relies on the independence properties of the Poisson point process, and is a counterpart of the well-known relationship between logistic regression and loglinear Poisson models in contingency tables (Dobson and Barnett 2008; McCullagh and Nelder 1989).

Several versions of this technique have been used for point pattern data in continuous space (Diggle and Rowlingson 1994; Baddeley et al. 2014). By using random sample points, the technique avoids bias which may occur in numerical quadrature, while potentially increasing variability due to random sampling. The variance contribution due to randomisation can be estimated, and appears to be acceptable in many cases (Baddeley et al. 2014).

Maximum Entropy

The principle of maximum entropy (Dutta 1966) is often used in ecology, for example, to study the influence of habitat variables on the spatial distribution of animals or plants (Dudík et al. 2007; Elith et al. 2011; Phillips et al. 2006). Conceptually this method considers all possible spatial distributions, and finds the spatial distribution which maximises a quantity called entropy, subject to constraints implied by the observed data. The constraints are equivalent to the score equations (2.15)–(2.16) or (2.5)–(2.6). The maximum entropy solution is a probability distribution which is a *loglinear* function of the predictors. It was shown in Renner and Warton (2013) that this solution is equivalent to fitting a loglinear Poisson point process, or equivalent to logistic regression on a fine pixel grid. An analogy could be drawn with the stretching of a string: a string may take on any shape, but if we demand that the string be stretched as tight as possible, it will take up a straight line. Thus, this analysis principle is equivalent to fitting a Poisson point process model with loglinear intensity.

2.4.4 Murchison Example

Here we give a worked example of Poisson point process modelling for the Murchison data of Fig. 2.1. The gold deposit locations are assumed to follow a Poisson process with intensity $\lambda(u)$ assumed to be a loglinear function of distance to the nearest fault,

$$\lambda(u) = \exp(\alpha + \beta d(u)), \quad (2.19)$$

where α, β are parameters and $d(u)$ is the distance from location u to the nearest geological fault. Contours of $d(u)$ are shown in Fig. 2.2. The model (2.19) corresponds

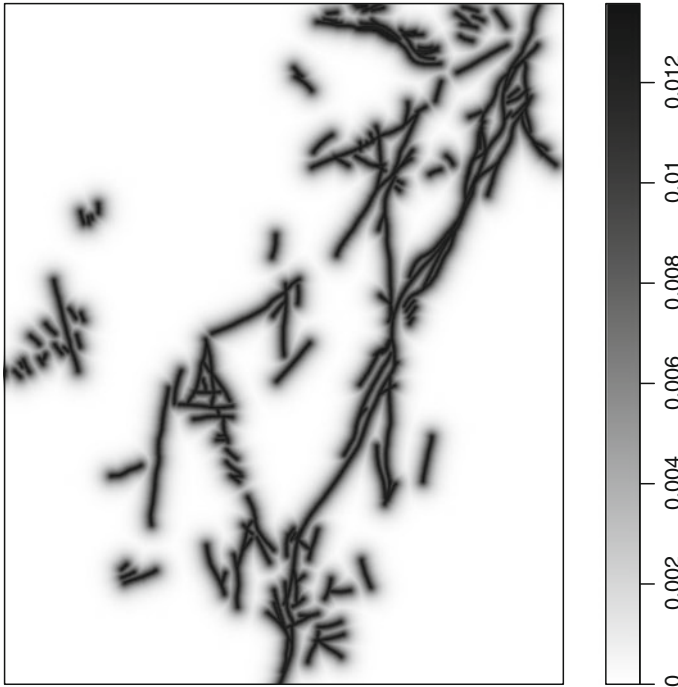


Fig. 2.5 Fitted intensity of gold deposits in the Murchison survey according to the loglinear Poisson point process model. Pixel values are intensities (number of deposits per square kilometre).

to logistic regression of pixel presence/absence indicators against distance to nearest fault.

We used the Berman-Turner device as implemented in the `spatstat` package (Baddeley et al. 2015) in the function `ppm`. The fitted parameters are $\hat{\alpha} = -4.34$ and $\hat{\beta} = -0.26 \text{ km}^{-1}$. These values are quite similar to the estimates in Table 2.2, as expected. The fitted intensity function,

$$\lambda(u) = \exp(-4.34 - 0.26 d(u)), \quad (2.20)$$

is displayed as a greyscale image in Fig. 2.5. Note that the spatial resolution of Fig. 2.5 is finer than the spacing of sample points used to fit the model; indeed $\lambda(u)$ can be evaluated at any location u in continuous space, using (2.20).

The fitted intensity relationship (2.20) can be interpreted directly. The estimated intensity of gold deposits in the immediate vicinity of a geological fault is about $\exp(-4.34) = 0.013$ deposits per square kilometre or 1.3 deposits per 100 km^2 . This intensity decreases by a *factor* of $\exp(-0.26) = 0.77$ for every additional kilometre away from a fault. At a distance of 10 km, the intensity has fallen by a factor of $\exp(10 \times (-0.26)) = 0.074$ to $\exp(-4.34 + 10 \times (-0.26)) = 0.001$ deposits per square kilometre or 0.1 deposits per 100 km^2 . Figure 2.6 shows the effect of the

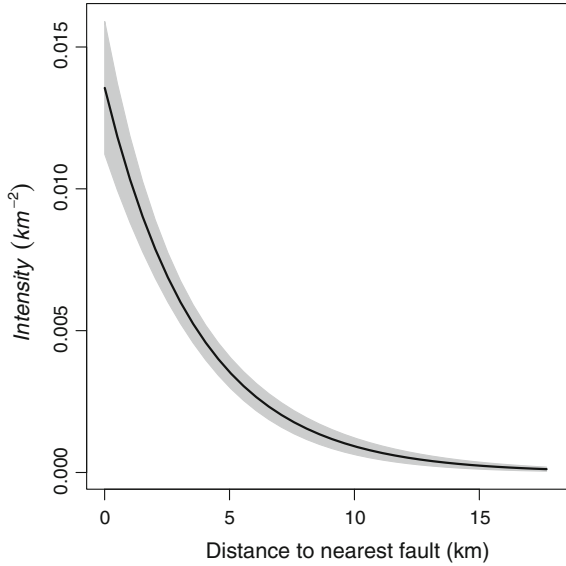


Fig. 2.6 Fitted intensity of Murchison gold deposits as a function of distance to the nearest fault, assuming it is a loglinear function of distance. Solid line: maximum likelihood estimate. Shading: pointwise 95% confidence interval

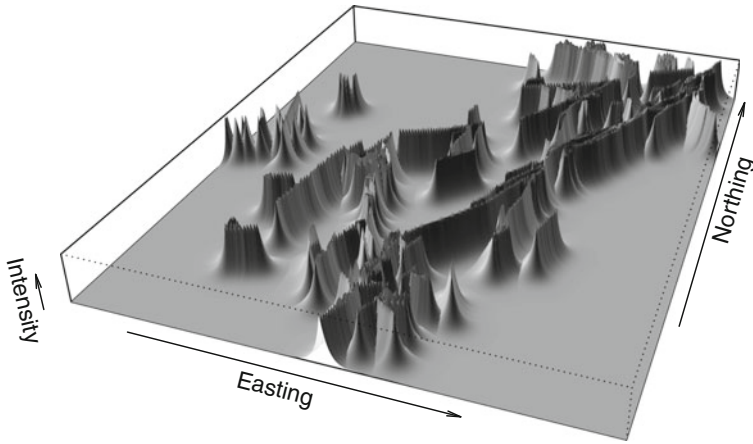


Fig. 2.7 Perspective view of fitted intensity surface of loglinear Poisson point process model of Murchison gold deposits against distance from nearest fault

distance covariate on the intensity function, according to the fitted loglinear Poisson model.

Figure 2.7 shows a perspective view of the fitted intensity function, treated as a surface in three dimensions. Note that, fortuitously, the southern edge of the perspective plot in Fig. 2.7 shows the shape of the fitted intensity curve in Fig. 2.6.

We caution again that this analysis has not fitted a highly flexible model in which the abundance of gold deposits depends, in some unspecified way, on the distance to the nearest fault. Rather, the very specific loglinear relationship (2.19) has been fitted. The flexible part of this analysis is the freedom to choose another predictor variable or variables to replace the distance function $d(u)$. Once the predictors are chosen, the analysis becomes rigidly parametric.

2.4.5 Statistical Inference

The Poisson point process model with loglinear intensity (2.10) belongs to the class of “exponential family” models (McCullagh and Nelder 1989). Statistical inference has been studied in detail for this class (Barndorff-Nielsen 1978) and for the Poisson process in particular (Kutoyants 1998; Rathbun and Cressie 1994).

A full set of standard tools is available for statistical inference. These include standard errors and confidence intervals for the parameter estimates, hypothesis tests (likelihood ratio test, score test), and variable selection and model selection (analysis of deviance, Akaike information criterion). See Baddeley et al. (2015, Chap. 9) for a full implementation.

Table 2.3 shows the estimated standard errors and 95% confidence intervals for the parameters in the loglinear model fitted to the Murchison data. These are asymptotic standard errors based on the Fisher information matrix.

Analysis of variance, or in this case, analysis of deviance (McCullagh and Nelder 1989; Hosmer and Lemeshow 2000; Dobson and Barnett 2008) supports a formal hypothesis test of statistical significance for the dependence on a predictor variable. For example the likelihood ratio test of the null hypothesis $\beta = 0$ against the alternative $\beta \neq 0$ indicates very strong evidence that gold deposit abundance is dependent on the distance to the nearest fault.

Recently-developed tools for model selection in point process models include Sufficient Dimension Reduction (Guan and Wang 2010).

Table 2.3 Standard errors and confidence intervals for parameters in loglinear Poisson model of Murchison data

	Estimate	SE	95% CI
α	-4.30	0.09	[-4.47, -4.13]
β	-0.27	0.02	[-0.31, -0.23]

2.4.6 Diagnostics

A fitted model is not like a fitted shoe. A shoe must approximately match the shape of the wearer’s foot before we call it fitted. On the contrary, statistical software “fits” a model to data on the assumption that the model is true, and does not check that the model describes the data at all.

Diagnostic quantities and diagnostic plots for a fitted model should be used to check the model assumptions. For linear regression and linear models, diagnostics are highly developed in statistical theory and applied statistical practice (Atkinson 1985). For logistic regression in a general context, diagnostics are also available (Landwehr et al. 1984; Dobson and Barnett 2008) and these extend to the “exponential family” class of models, at least in theory.

Diagnostics for the Poisson point process model, corresponding to the well-known diagnostics for logistic regression, were developed in Baddeley et al. (2013a, b). Two of these are shown here for the Murchison data.

The *influence* measure e_i is the effect on the fitted log likelihood of deleting the i th deposit point x_i (Baddeley et al. 2013a). Figure 2.8 shows circles of diameter proportional to e_i centred at the deposit locations x_i . The geological fault pattern is also shown. In this Figure, large circles represent observations which had a large effect on the resulting fitted model. There is one very large circle at middle left of the Figure, and we notice that there are no geological faults near this deposit. That is,

Fig. 2.8 Influence diagnostic for the loglinear Poisson model of gold deposits against distance to nearest fault. Circle diameters are proportional to the influence of each deposit. Grey lines are geological faults

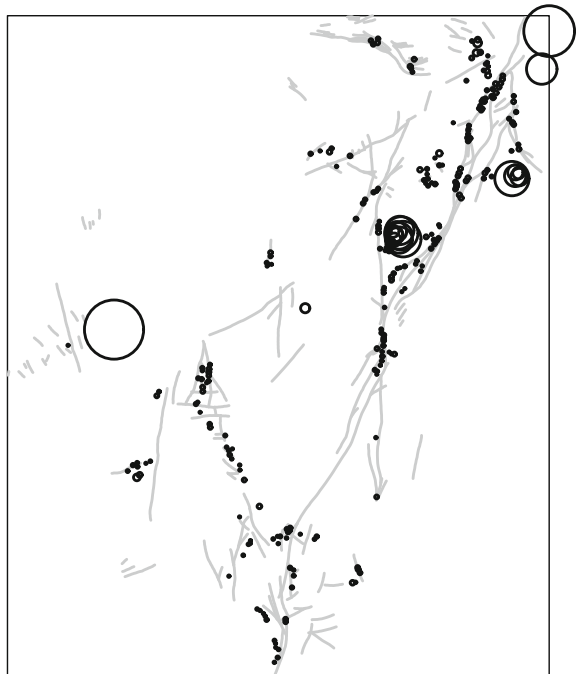
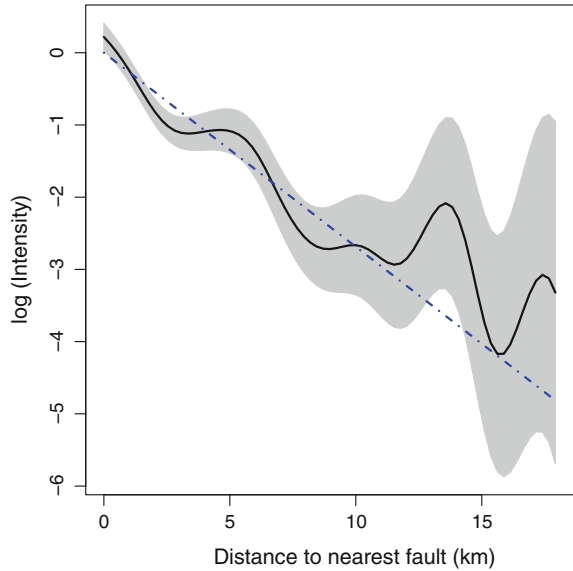


Fig. 2.9 Partial residual plot for loglinear Poisson model of Murchison gold deposits as a function of distance to the nearest fault. Solid line: smoothed partial residual. Shading: pointwise 95% confidence interval. Dot-dash line: fitted model



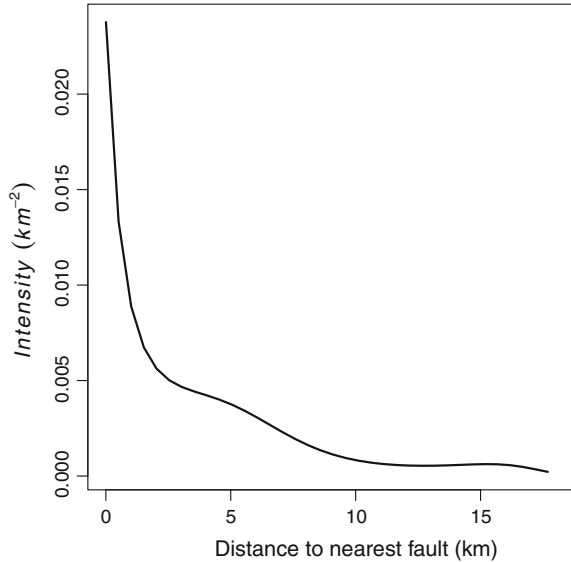
the influence diagnostic identifies this deposit as anomalous, perhaps an “outlier”, with respect to the fitted model in which deposits are most likely to occur close to a geological fault. This is an entirely data-driven diagnostic, and tells us only that this observation is anomalous with respect to the model. It is unable to tell us whether the deposit is truly anomalous in geological terms, or whether the survey perhaps failed to detect an existing geological fault near this location.

Strategies for dealing with anomalous data include outlier detection and removal, and robust model-fitting which is resistant to the effects of outliers. Robust parameter estimation for Poisson point process models was developed in Assunção and Guttorp (1999).

Figure 2.9 shows a *partial residual* plot (Baddeley et al. 2013b) for the Murchison gold deposits against distance to nearest fault. Assuming that the loglinear model (2.19) is approximately true, say $\log \lambda(u) = \alpha + \beta d(u) + H(d(u))$ where the error $H(d)$ is small, this procedure forms an estimate $\hat{H}(d)$ of the error term, adds it to the fitted linear term, and plots $\hat{\alpha} + \hat{\beta}d + \hat{H}(d)$ against values of distance d . If the model is correct, this plot should be a straight line. Departures from the straight line can be interpreted as suggesting the correct form of dependence. Figure 2.9 suggests there is a minor departure from the loglinear model.

An alternative way to explore non-linearity is to fit a polynomial or spline function in place of the linear function on the right hand side of (2.1) or (2.19). In order to avoid over-fitting and numerical instability, the model should be fitted by *penalised* maximum likelihood, in which the log likelihood (2.14) is augmented by a penalty term that discourages extreme values of the parameters which might produce a wildly-oscillating polynomial. Figure 2.10 shows a penalised maximum

Fig. 2.10 Fitted intensity of Murchison gold deposits assuming the log intensity is a fifth-order B-spline function of distance to nearest fault



likelihood fit of a model in which the log intensity is a fifth-order B-spline function of distance to the nearest fault. The model was fitted in the `spatstat` package using code for Generalised Additive Models (GAM) (Hastie and Tibshirani 1990). This fit also suggests minor departure from the loglinear model.

2.4.7 Rationale for Prediction

Up to this point, our commentary on prospectivity analysis applies equally well to the analysis of archaeological finds, plant species distribution, etc., using logistic regression and related tools. However, the key goal of prospectivity analysis is the *prediction of previously-unknown deposits*, and this sets it apart from other applications. This prediction problem deserves more attention from the statistical community, and we shall identify several topics for research.

The rationale for predicting “new” mineral deposits is clearest when we extrapolate from a fully-explored region to an unexplored region. We might extrapolate from a previous, fully-explored mining lease to a newly-granted exploration lease which is geologically analogous. We fit a model to the fully-explored region, obtaining estimates of the model parameters α, β , which we believe can be extrapolated to the unexplored region. Applying the fitted model relationship to the predictor variables for the new region, we obtain explicit predictions for the mineral deposits in the new region. These predictions may include expected numbers of deposits, probability of no deposits, probability distribution of distance to the nearest deposit, and so on. These predictions are valid calculations even if the geological structure in the two

regions was formed at the same epoch, because of the assumption of independence between deposits. Essentially the fully-explored region is used to obtain estimates of the parameters of the “laws” which apply to both regions, and these laws are then applied to the new region.

The statistical reasoning is far more complicated when we wish to predict hitherto-undiscovered mineral deposits from known deposits *in the same region*. It would be futile to assume that the region has been fully explored, since this would imply there are no deposits remaining to be discovered. Instead our statistical model must now recognise two categories of deposits, known and unknown. The methods described above can be re-deployed if we assume that the true spatial pattern of *all* deposits (whether known or unknown) is a Poisson point process with intensity function $\kappa(u)$, say, and that a deposit existing at a location u will be detected with probability $P(u)$, independently of other deposits. Then, by the “thinning” property of the Poisson process, the pattern of *detected* deposits is also a Poisson process, with intensity $\lambda(u) = P(u)\kappa(u)$; the pattern of *undetected* deposits is a Poisson process with intensity $\xi(u) = (1 - P(u))\kappa(u)$; and the detected and undetected deposits are independent of each other. Fitting a Poisson point process model to the observed mineral deposits allows us to estimate $\lambda(u)$ only. If the detection probability $P(u)$ is known, then it becomes feasible to back-calculate $\kappa(u) = \lambda(u)/P(u)$ and

$$\xi(u) = \frac{1 - P(u)}{P(u)} \lambda(u). \quad (2.21)$$

It is then possible to make predictions or conditional simulations of the undetected deposits. The independence property of the Poisson process implies that the prediction or conditional simulation depends only on the fitted model parameters, and does not otherwise depend on the observed deposits. The conditional simulation is a realisation of the Poisson process of the assumed loglinear form with the parameter values fitted from the data: the simulated realisation is independent of the observed deposits, given the fitted model parameters.

This argument is an instance of the prediction approach to survey sampling inference (Royall 1988). The difficulty is that the detection probability $P(u)$ will depend on the detection method, the spatially-varying amount of survey effort, and other factors. If $P(u)$ can be estimated from data, perhaps by comparing the results of successive surveys of the same region, then the form of (2.21) suggests that the appropriate model is a logistic regression for $P(u)$ on explanatory variables. If no information is available about $P(u)$, we could make the simplifying assumption that $P(u) \equiv P$ is constant; then $\xi(u)$ is a constant multiple of $\lambda(u)$, so that at least the relative prospectivity of different locations u can be assessed from a plot of $\lambda(u)$.

Other, non-Poisson point processes can also serve as models of mineral deposits (Baddeley et al. 2015, Chaps. 12 and 13) and support prediction and conditional simulation. In such models, the presence of a point affects the probability of presence of a point at nearby locations. In this case the conditional simulation does depend on the observed deposit locations (Møller and Waagepetersen 2004; Baddeley et al. 2015).

A more realistic model of the detection process would envisage that the discovery of a new deposit will encourage the exploration geologist to survey the surrounding areas more intensively, increasing the detection probability in these surrounding areas. This destroys the independence structure: the pattern of observed deposits is no longer a Poisson point process, and is spatially clustered. Non-Poisson point process models would be needed to describe the spatial pattern of observed deposits, and even if the spatial pattern of all deposits is assumed to be Poisson, the pattern of undiscovered deposits is both non-Poisson and dependent on the observed deposits. A full analysis of the prediction problem would require the deployment of Missing Data principles (Little and Rubin 2002).

In prospectivity analysis it may or may not be desirable to fit any explicit relationship between deposit abundance and predictors such as distance to the nearest fault. Often the objective is simply to select a distance threshold, so as to delimit the area which is considered highly prospective (high predicted intensity) for the mineral. The ROC curve (Sect. 2.7) is more relevant to this exercise. However, if credible models can be fitted, they contain much more valuable predictive information.

2.5 Monotone Regression

The remainder of this article describes three alternative analysis techniques, genuinely different from logistic regression, which do not seem to be widely used in prospectivity analysis. These techniques are genuinely “non-parametric” in the sense that they assume only that the intensity or rate of mineral deposits $\lambda(u)$ is a function of the predictor variable $Z(u)$ at the same location u ,

$$\lambda(u) = \rho(Z(u)) \tag{2.22}$$

where $\rho(z)$ is a function to be estimated. We do not assume that $\rho(z)$ has any particular functional form.

The assumption (2.22) is encountered frequently. In geological applications where the points are the locations of mineral deposits, ρ is an index of the prospectivity (Bonham-Carter 1995) or predicted frequency of deposits as a function of geological and geochemical covariates z . In ecological applications where the points are the locations of individual organisms, ρ is a “resource selection function” (Manly et al. 1993) reflecting preference for particular environmental conditions z .

In *monotone regression*, we assume that $\rho(z)$ is a monotone function of z , either monotone increasing (non-decreasing):

$$z_1 < z_2 \quad \text{implies} \quad \rho(z_1) \leq \rho(z_2)$$

or monotone decreasing (non-increasing):

$$z_1 < z_2 \quad \text{implies} \quad \rho(z_1) \geq \rho(z_2).$$

Sager (1982) considered the log-likelihood of the Poisson point process with intensity (2.22),

$$\log L = \sum_i \log \rho(Z(x_i)) - \int_W \rho(Z(u)) du, \quad (2.23)$$

and showed that the log-likelihood can be maximised over the class of all monotone functions ρ . The optimal function $\hat{\rho}(z)$ is the *nonparametric maximum likelihood* estimate of $\rho(z)$ under the monotonicity constraint, or simply the *monotone regression* estimate.

To simplify the discussion, assume that $\rho(z)$ is monotone decreasing, and that the values of $Z(u)$ are real numbers greater than or equal to zero. Sager (1982) showed that the monotone regression estimate $\hat{\rho}(z)$ is piecewise constant, with jumps occurring only at the observed values $z_i = Z(x_i)$ of the predictor at the deposit point locations. For any z let

$$A(z) = |\{u \in W : Z(u) \leq z\}| \quad (2.24)$$

be the area of the subset of the survey region where the covariate value is less than or equal to z . Also let $N(z) = \sum_i \mathbf{1}\{z_i \leq z\}$ be the number of data points for which the covariate value is less than or equal to z . In the Murchison example, $A(z)$ is the area lying closer than z kilometres from the nearest fault, and $N(z)$ is the number of deposits lying in this region. Then the monotone regression estimate $\hat{\rho}(z)$ is the maximum of simple functions

$$\hat{\rho}(z) = \max_i \rho_i(z) \quad (2.25)$$

where

$$\rho_i(z) = \begin{cases} \frac{N(z_i)}{A(z_i)} & \text{if } z < z_i \\ 0 & \text{if } z \geq z_i. \end{cases} \quad (2.26)$$

The monotone regression estimate $\hat{\rho}(z)$ can be computed rapidly using the Pool Adjacent Violators algorithm (Barlow et al. 1972) or the following Maximum Upper Sets algorithm (Sager 1982):

1. Sort the data values as $z_1 \leq z_2 \leq \dots \leq z_n$.
2. Initialise $K = 0$ and $z_K = 0$. These represent the largest data value whose status has been resolved so far.
3. For each $i > K$ consider the interval $[z_K, z_i]$, evaluate the empirical intensity $\rho_{i,K} = (N(z_i) - N(z_K)) / (A(z_i) - A(z_K))$, and find the value i^* which maximises $\rho_{i,K}$.
4. For all z lying in the interval $[z_K, z_{i^*}]$, set $\rho(z) = \rho_{i^*,K}$.
5. If $i^* = n$, exit. Otherwise, set $K = i^*$ and go to step 3.

Fig. 2.11 Fitted intensity of gold deposits as a function of distance to the nearest fault, using monotone regression

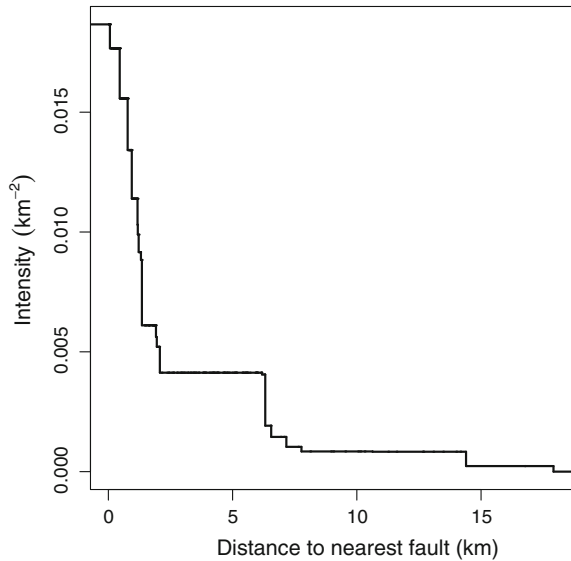


Figure 2.11 shows the monotone regression estimate of the intensity of gold deposits as a function of distance to nearest fault in the Murchison data. The curve has the same overall shape as the exponential curve implied by the loglinear Poisson point process model or logistic regression (Fig. 2.6), except for a prominent plateau between $z = 2$ and $z = 6$ km.

Note also that the monotone regression estimate of intensity at small distances z is higher (Fig. 2.11) than in the loglinear Poisson model (Fig. 2.6). This is expected, since the estimate $\hat{\rho}(z)$ depends primarily on the part of the survey where $Z(u) \leq z$. This is more satisfactory than the behaviour of the loglinear Poisson model for which the fitted curve depends on the entire dataset. If, for example, we were to restrict the study area to the region lying at most 20 km from the nearest fault, the estimates of the parameters α , β in the loglinear Poisson model could change markedly, while the monotone regression in Fig. 2.11 would be unchanged.

Figure 2.12 shows a perspective plot of the predicted intensity implied by the monotone regression. Compared with Fig. 2.7, this shows qualitatively the same effect of a dense concentration close to the geological faults, but with a different profile (again fortuitously displayed at the southern edge of the plot).

Sager (1982) shows that this method extends to multiple predictor variables. The author believes it can also be extended to allow points to have weights determined by the mineral endowment of the deposit, or a similar characteristic.

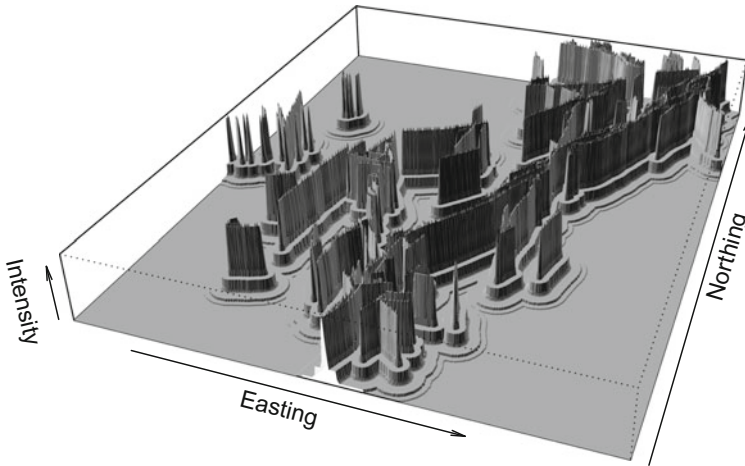


Fig. 2.12 Perspective view of fitted intensity using monotone regression

2.6 Nonparametric Curve Estimation

A second alternative to logistic regression is nonparametric curve estimation, in which we assume that the intensity is a smooth function of the predictor, $\lambda(u) = \rho(Z(u))$, and estimate the function $\rho(z)$ by nonparametric smoothing. This was developed in Baddeley et al. (2012), Guan (2008).

Assume that Eq. (2.22) holds, and that $\rho(z)$ is a continuous function of z , and that $Z(u)$ is at least a continuous function of location u , without further constraints. Nonparametric estimation of ρ is closely connected to estimation of a probability density from biased sample data (Jones 1991; El Barmi and Simonoff 2000) and to the estimation of relative densities (Handcock and Morris 1999). Under the smoothness assumptions, ρ is proportional to the ratio of two probability densities, the numerator being the density of covariate values at the points of the point process, while the denominator is the density of covariate values at random locations in space. Kernel smoothing can be used to estimate the function ρ as a relative density (Baddeley et al. 2012; Guan 2008).

Define the *spatial distribution function* (Lahiri 1999; Lahiri et al. 1999) as the cumulative distribution function of the covariate value $Z(U)$ at a random point U uniformly distributed in W :

$$G(z) = \frac{1}{|W|} \int_W \mathbf{1}\{Z(u) \leq z\} du. \quad (2.27)$$

Here we use the ‘indicator’ notation: $\mathbf{1}\{\dots\}$ equals 1 if the statement ‘...’ is true, and 0 if the statement is false. Equivalently $G(z) = A(z)/A(\infty) = A(z)/|W|$ where $A(z)$, defined in (2.24), is the area of the set of all locations in W where the covariate value

is less than or equal to z . In practice $G(z)$ would often be estimated by evaluating the covariate at a fine grid of pixel locations, and forming the cumulative distribution function

$$G(z) = \frac{\#\{\text{pixels } u : Z(u) \leq z\}}{\#\text{pixels}}. \quad (2.28)$$

Three estimators of ρ proposed in Baddeley et al. (2012) are

$$\hat{\rho}_R(z) = \frac{1}{|W|G'(z)} \sum_i \kappa(Z(x_i) - z) \quad (2.29)$$

$$\hat{\rho}_W(z) = \sum_i \frac{1}{|W|G'(Z(x_i))} \kappa(Z(x_i) - z) \quad (2.30)$$

$$\hat{\rho}_T(z) = \frac{1}{|W|} \sum_i \kappa(G(Z(x_i)) - G(z)) \quad (2.31)$$

where x_1, \dots, x_n are the data points, $Z(x_i)$ are the observed values of the covariate Z at the data points, $|W|$ is the area of the observation window W , and κ is a *one-dimensional* smoothing kernel—smoothing is conducted on the observed values $Z(x_i)$ rather than in the window W . The derivative $G'(z)$ is usually approximated by differentiating a smoothed estimate of G . The estimators (2.29)–(2.31) were developed in Baddeley et al. (2012) by adapting estimators from kernel smoothing (Jones 1991; El Barmi and Simonoff 2000). An estimator similar to (2.29) was proposed in Guan (2008).

Figure 2.13 shows the fitted estimate of intensity for the Murchison gold deposits as a function of distance to the nearest fault. The plot shows the ratio estimator $\hat{\rho}_R(z)$ against z , equation (2.29), together with the pointwise 95% confidence interval for $\rho(z)$ based on asymptotic theory assuming a Poisson process (Baddeley et al. 2012). Tickmarks on the horizontal axis show the observed distance values $z_i = Z(x_i)$ at the deposits.

The overall shape of Fig. 2.13 is consistent with Figs. 2.6 and 2.11. A plateau of intensity is visible between $z = 2.5$ and $z = 5.5$ km, consistent with the plateau seen in Fig. 2.11. The peak of intensity in Fig. 2.13 occurs at about $z = 1$ km, rather than at distance $z = 0$, but this may be an artefact of the smoothing procedure, as it is not seen in the other two estimates (2.30) and (2.31).

Figure 2.14 shows a perspective view of the predicted intensity using nonparametric curve estimation. This is quite similar to the surface obtained by monotone regression, shown in Fig. 2.12.

The nonparametric curve estimate has the attractive property that $\hat{\rho}(z)$ depends only on the survey information from locations where the predictor value is approximately equal to z . In the Murchison example, the estimated intensity $\hat{\rho}(z)$ of gold deposits at a distance z from the nearest fault, is estimated using only the deposits and non-deposit locations which lie approximately z km from the nearest fault. Although smoothing artefacts may be present, this property means that the nonparametric

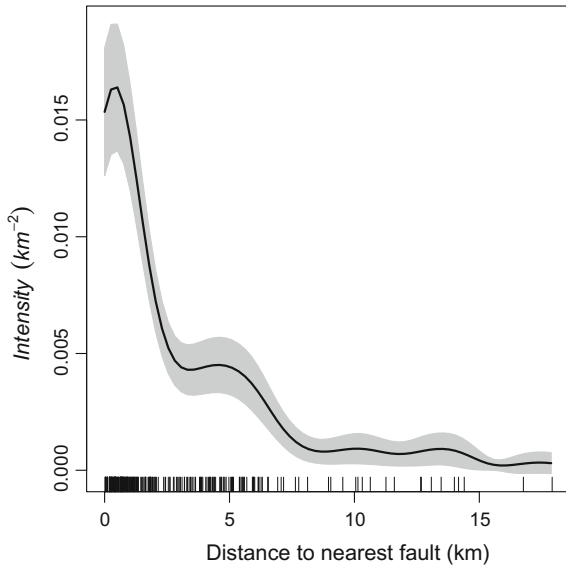


Fig. 2.13 Fitted intensity of gold deposits as a function of distance to the nearest fault, using kernel-based estimator

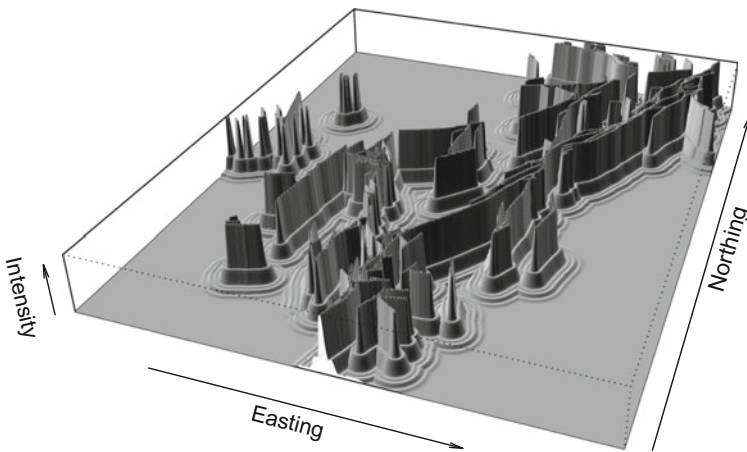


Fig. 2.14 Perspective view of fitted intensity using nonparametric curve estimate of ρ

curve estimate can be treated as an estimate of the true relationship between intensity and predictor.

The estimators (2.29)–(2.31) can be modified to incorporate numerical weights, for example, representing the endowment of each deposit. Then $\rho(z)$ has the interpretation of the expected *total endowment* per unit area, of deposits at a distance z from the nearest fault.

Fig. 2.15 Fitted intensity of Murchison gold deposits as a function of distance to the nearest fault, estimated by three different methods: monotone regression (solid lines), kernel smoothing (dashed lines) and logistic regression (dotted lines)

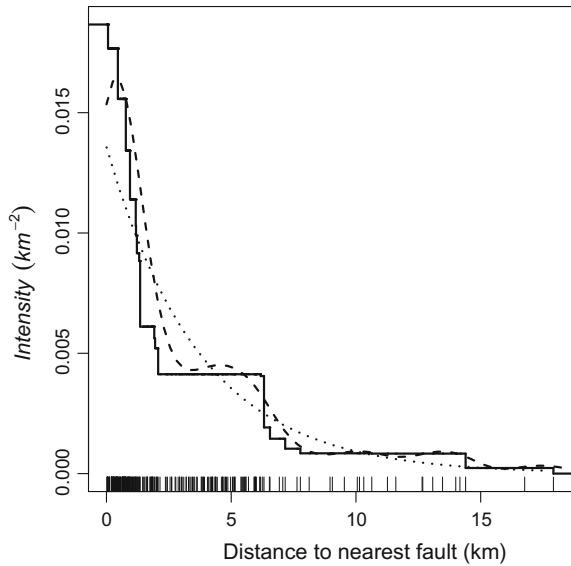


Figure 2.15 shows the three estimates of $\rho(z)$ together. Logistic regression, log-linear Poisson point process modelling, and maximum entropy methods effectively assume that prospectivity is an exponential function $\rho(z) = AB^z = \exp(\alpha + \beta z)$, while monotone regression assumes $\rho(z)$ is a decreasing function of z , and kernel estimation assumes $\rho(z)$ is a smooth function of z without further restriction.

This analysis assumes that the intensity at a location u depends *only* on the covariate value $Z(u)$. To validate the assumption (2.22) we can compare the predicted intensity $\hat{\rho}(Z(u))$ assuming (2.22) with a (spatial) kernel estimate $\hat{\lambda}(u)$ which does not assume (2.22). If the assumption is not true, $\hat{\rho}(z)$ is still meaningful: it is effectively an estimate of the average intensity $\lambda(u)$ over all locations u where $Z(u) = z$.

2.7 ROC Curves

Suppose we agree that the ultimate goal of prospectivity analysis is to decide which parts of an exploration area are most likely to contain a valuable deposit. Then the essential task is to *classify* different parts of the exploration area into areas of high and low prospectivity, rather than necessarily needing to model the degree of prospectivity at every location.

The *Receiver Operating Characteristic (ROC)* curve (Krzanowski and Hand 2009) is a summary of the performance of a classifier. It is often applied to medical diagnostic tests (Nam and D’Agostino 2002) when the test is based on thresholding a quantitative assay. Suppose for example that a medical test returns a “positive” result (predicting a high risk of disease) if the patient’s blood cholesterol level exceeds a

threshold t . For a given choice of threshold t , the “true positive rate” $TP(t)$ is the fraction of patients with the disease who return a correct, positive test result. The “false positive rate” $FP(t)$ is the fraction of patients who do not have the disease but who return an incorrect, positive test result. The ROC curve is a plot of true positive rate $TP(t)$ against false positive rate $FP(t)$ for all thresholds t . A good classifier has a large true positive rate in comparison to the false positive rate, so the ROC curve of a good classifier will lie well above the diagonal line on the graph.

The same technique can be applied to prospectivity analysis (Rakshit et al. 2017), taking the mineral deposits as the “disease”, and using either a spatial predictor $Z(u)$ or a fitted model intensity $\lambda(u)$ to classify pixels into high or low prospectivity classes. Suppose that $Z(u)$ is a real-valued spatial predictor. Calculate the empirical cumulative distribution function of Z at the observed deposit locations,

$$\hat{F}_{\mathbf{x}}(t) = \frac{1}{n(\mathbf{x})} \sum_i \mathbf{1}\{Z(x_i) \leq t\}$$

and the empirical “spatial distribution function” of $Z(u)$ over all locations u in W ,

$$F_W(t) = \frac{1}{|W|} \int_W \mathbf{1}\{Z(u) \leq t\} du.$$

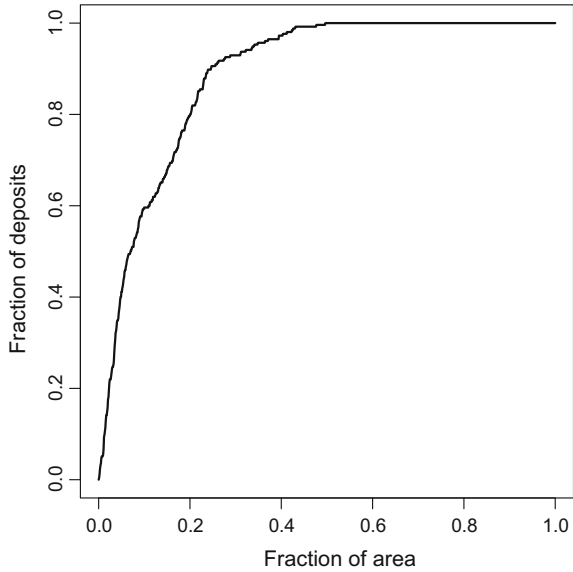
Then the ROC plot is a graph of $1 - \hat{F}_{\mathbf{x}}(t)$ against $1 - F_W(t)$ for all t . Equivalently it is a plot of $R_+(p) = 1 - F_{\mathbf{x}}(F_W^{-1}(1 - p))$ against p . Applied statisticians would recognise this as a form of the classical P–P plot.

The formulae above assume that larger values of Z are more prospective. If smaller values of Z are more prospective, then the appropriate ROC plot is a graph of $F_{\mathbf{x}}(t)$ against $F_W(t)$ for all t , or equivalently a graph of $R_-(p) = F_{\mathbf{x}}(F_W^{-1}(p))$ against p . This is the P–P plot of $F_{\mathbf{x}}$ against F_W .

Figure 2.16 shows the ROC curve for the Murchison gold deposits against distance from nearest fault, assuming smaller distances are more prospective. The horizontal axis shows the fraction of area in the survey region which lies less than t km away from a fault, and the vertical axis shows the fraction of deposits which lie less than t km from a fault. For example, we may read off the plot that 60% of all known deposits lie in a region occupying 10% of the survey area defined by a distance threshold. This has a useful practical interpretation. The threshold itself is not shown on the ROC plot but could be obtained from the spatial cumulative distribution function. The ROC curve depends on the choice of the study region (Jiménez-Valverde 2012).

The interpretation of ROC curves in spatial analysis is controversial. Some writers suggest (Fielding and Bell 1997) that the ROC can be used to evaluate the goodness-of-fit of a species distribution model, or equivalently the goodness-of-fit of a prospectivity analysis. Others disagree (Lobo et al. 2007, p. 146) and argue that the ROC is an indicator of predictive power—the ability to segregate pixels reliably into two classes of high and low prospectivity.

Fig. 2.16 Empirical ROC curve for gold deposits against distance from the nearest fault in the Murchison survey rectangle



The ROC can be based either on a real-valued predictor variable $Z(u)$ or on a fitted model intensity $\lambda(u)$. In the latter case it is tempting to regard the ROC as a summary of the predictive power of the fitted model (Lobo et al. 2007; Austin 2007; Thuiller et al. 2003). However, if the model is logistic regression, or a loglinear Poisson point process, or if the intensity is estimated using monotone regression, then the fitted intensity $\lambda(u)$ is a monotone function of the predictor $Z(u)$. Thresholding $\lambda(u)$ is equivalent to thresholding $Z(u)$, so that the ROC curves derived from any of these models are identical. For example, the ROC cannot be used to compare the predictive power of logistic regression against that of monotone regression. It would be more appropriate to regard the ROC as a summary of the inherent predictive power of the predictor variable $Z(u)$ itself (Rakshit et al. 2017).

The ROC does have a connection with the other techniques described in this article. Suppose that the point process intensity $\lambda(u)$ depends on the predictor $Z(u)$ through a function $\rho(z)$ as in (2.22). Then we show in (Rakshit et al. 2017) that the slope of the ROC curve is closely related to ρ . If large values of the predictor are more prospective, the slope of the ROC curve is

$$\frac{d}{dp}R_+(p) = \frac{d}{dp} [1 - F_x(F_w^{-1}(1 - p))] = \frac{1}{\kappa}\rho(F_w^{-1}(1 - p))$$

while if small values of the predictor are prospective, the slope is

$$\frac{d}{dp}R_-(p) = \frac{d}{dp} [F_x(F_w^{-1}(p))] = \frac{1}{\kappa}\rho(F_w^{-1}(p))$$

where κ is the average intensity over the study region. Analysis using the ROC curve is not fundamentally different from fitting a point process model or pixel presence/absence regression model, but may be a more practically useful presentation of the same information.

2.8 Recursive Partitioning

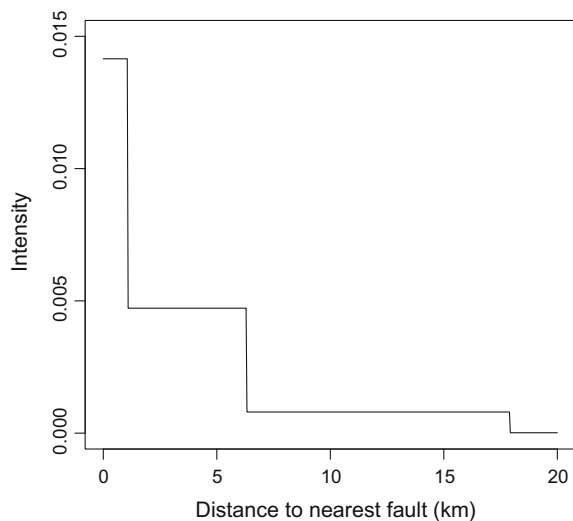
Classification and Regression Tree (CART) (Breiman et al. 1984) or Recursive Partitioning methods offer another alternative approach. Given one or many predictor variables, these methods predict the response by thresholding the predictors. The result is a prediction rule, organised as a logical tree, in which each fork of the tree is a threshold operation on one of the predictors. This kind of rule would appear to be well-suited to the practical needs of prospectivity analysis.

For a single predictor variable, the result of recursive partitioning is a piecewise-constant function $\hat{\rho}(z)$ which is not constrained to be monotone. Figure 2.17 shows the estimated intensity of the Murchison gold deposits as a function of distance to the nearest fault only, using recursive partitioning. Any number of predictor variables can be included in the analysis.

Software and Data

All analyses in this chapter were performed using the `spatstat` library (Baddeley et al. 2015) which is a contributed extension package for the R statistical software system (R Development Core Team 2011). Both R and `spatstat` can be downloaded from <https://cran.r-project.org>. The Murchison data are included in

Fig. 2.17 Intensity of Murchison gold deposits as a function of distance from the nearest fault, estimated by recursive partitioning



spatstat. Software scripts for the analyses in this chapter are available at www.spatstat.org.

Acknowledgements This research would not have been possible without enthusiastic support and collaboration of the Centre for Exploration Targeting (CET) at the University of Western Australia, in particular Professor Eun-Jung Holden. The research was funded by the Australian Research Council under a Discovery Outstanding Researcher Award.

This article includes summaries of the results of joint research with the Perth Spatial Point Processes Group (Ya-Mei Chang, Andrew Hardegen, Thomas Lawrence, Gopalan Nair, Suman Rakshit and Yong Song) and with collaborators Ege Rubak, Rob Foxall and Rolf Turner. Valuable advice was given by Professors Allan Trench and Mike Dentith.

References

- Agterberg F (1974) Automatic contouring of geological maps to detect target areas for mineral exploration. *J Int Assoc Math Geol* 6:373–395
- Agterberg F (2001) Appreciation of contributions by Georges Matheron and John Tukey to a mineral-resources research project. *Nat Resource Res* 10:287–295
- Alt J, King G, Signorino C (2001) Aggregation among binary, count and duration models: estimating the same quantities from different levels of data. *Polit Anal* 9:21–44
- Assunção R, Guttorp P (1999) Robustness for inhomogeneous Poisson point processes. *Ann Inst Stat Math* 51:657–678
- Atkinson A (1985) Plots, transformations and regression. No. 1 in Oxford statistical science series. Oxford University Press, Clarendon
- Austin M (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecol Modell* 200:1–19
- Baddeley A (2017) Local composite likelihood for spatial point processes. *Spat Stat* (in press). <https://doi.org/10.1016/j.spasta.2017.03.001>
- Baddeley A, Turner R (2000) Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Austr New Zealand J Stat* 42(3):283–322
- Baddeley A, Turner R (2005) Spatstat: an R package for analyzing spatial point patterns. *J Stat Softw* 12(6):1–42. www.jstatsoft.org. ISSN 1548-7660
- Baddeley A, Berman M, Fisher N, Hardegen A, Milne R, Schuhmacher D, Turner R (2010) Spatial logistic regression and change-of-support for Poisson point processes. *Electron J Stat* 4:1151–1201. <https://doi.org/10.1214/10-EJS581>
- Baddeley A, Chang Y, Song Y, Turner R (2012) Nonparametric estimation of the dependence of a spatial point process on a spatial covariate. *Stat Interface* 5:221–236
- Baddeley A, Chang Y, Song Y (2013a) Leverage and influence diagnostics for spatial point processes. *Scand J Stat* 40:86–104. <https://doi.org/10.1111/j.1467-9469.2011.00786.x>
- Baddeley A, Chang YM, Song Y, Turner R (2013b) Residual diagnostics for covariate effects in spatial point process models. *J Comput Graph Stat* 22:886–905
- Baddeley A, Coeurjolly JF, Rubak E, Waagepetersen R (2014) Logistic regression for spatial Gibbs point processes. *Biometrika* 101(2):377–392
- Baddeley A, Rubak E, Turner R (2015) Spatial point patterns: methodology and applications with R. Chapman and Hall/CRC, London
- Banerjee S, Gelfand A (2002) Prediction, interpolation and regression for spatially misaligned data. *Sanhkye A* 64:227–245
- Barlow R, Bartholomew D, Bremner J, Brunk H (1972) Statistical inference under order restrictions. Wiley, New York

- Barndorff-Nielsen O (1978) Information and exponential families in statistical theory. Wiley, Chichester, New York, Brisbane, Toronto
- Berkson J (1955) Maximum likelihood and minimum χ^2 estimates of the logistic function. *J Am Stat Assoc* 50:130–162
- Berman M, Turner T (1992) Approximating point process likelihoods with GLIM. *Appl Stat* 41:31–38
- Besag J, Milne R, Zachary S (1982) Point process limits of lattice processes. *J Appl Probab* 19:210–216
- Bonham-Carter G (1995) Geographic information systems for geoscientists: modelling with GIS. No. 13 in computer methods in the geosciences. Pergamon Press, Elsevier, Kidlington, Oxford, UK
- Breiman L (1968) Probability. Addison-Wesley, Reading
- Breiman L, Friedman J, Stone C, Olshen R (1984) Classification and regression trees. Chapman and Hall/CRC
- Brillinger D (1978) Comparative aspects of the study of ordinary time series and of point processes. In: Krishnaiah P (ed) Developments in statistics. Academic Press, New York, London, pp 33–133
- Brillinger D, Preisler H (1986) Two examples of quantal data analysis: a) multivariate point process, b) pure death process in an experimental design. In: Proceedings, XIII international biometric conference. International Biometric Society, Seattle, pp 94–113
- Brillinger D, Segundo J (1979) Empirical examination of the threshold model of neuron firing. *Biol Cybern* 35:213–220
- Brown W, Gedeon T, Baddeley A, Groves D (2002) 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: Bayer U, Burger H, Skala W (eds) IAMG 2002: 8th annual conference of the international association for mathematical geology, vol 1. International Association of Mathematical Geology, pp 257–268
- Chung C, Agterberg F (1980) Regression models for estimating mineral resources from geological map data. *Math Geol* 12:473–488
- Chung C, Fabbri A (1999) Probabilistic prediction models for landslide hazard mapping. *Photogramm Eng Remote Sens* 62(12):1389–1399
- Clyde M, Strauss D (1991) Logistic regression for spatial pair-potential models. In: Possolo A (ed) Spatial statistics and imaging. Lecture notes—monograph series, vol 20, chap. II. Institute of Mathematical Statistics, Hayward, CA, pp 14–30. ISBN 0-940600-27-7
- Cressie N (1996) Change of support and the modifiable area unit problem. *Geogr Syst* 3:159–180
- Daley D, Vere-Jones D (2003) An introduction to the theory of point processes. Volume I: Elementary theory and methods, 2nd edn. Springer, New York
- Daley D, Vere-Jones D (2008) An introduction to the theory of point processes. Volume II: General theory and structure, 2nd edn. Springer, New York
- Dean C, Balshaw R (1997) Efficiency lost by analyzing counts rather than exact times in Poisson and overdispersed Poisson regression. *J Am Stat Assoc* 92:1387–1398
- Diggle P, Rowlingson B (1994) A conditional approach to point process modelling of elevated risk. *J R Stat Soc Ser A* 157(3):433–440
- Dobson A, Barnett A (2008) An introduction to generalized linear models, 3rd edn. CRC Press
- Dudík M, Phillips SJ, Schapire R (2007) Maximum entropy density estimation with generalized regularization and an application to species distribution modeling. *J Mach Learn Res* 8:1217–1260
- Dutta M (1966) On maximum (information-theoretic) entropy estimation. *Sankhya: Indian J Stat Series A* 28:319–328
- El Barmi H, Simonoff J (2000) Transformation based density estimation for weighted distributions. *J Nonparametr Stat* 12:861–878
- Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. *Divers Distrib* 17:43–57

- Elliott P, Wakefield J, Best N, Briggs D (eds) (2000) *Spatial epidemiology: methods and applications*. Oxford University Press, Oxford
- Fringling A, Bell J (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24:38–49
- Foxall R, Baddeley A (2002) Nonparametric measures of association between a spatial point process and a random set, with geological applications. *Appl Stat* 51(2):165–182
- Franklin J (2009) *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge, UK
- Freedman D, Pisani R, Purves R (2007) *Statistics*, 4th edn. Norton, New York
- Gorsevski P, Gessler P, Folz R, Elliott W (2006) Spatial prediction of landslide hazard using logistic regression and ROC analysis. *Trans GIS* 10:395–415
- Gotway C, Young L (2002) Combining incompatible spatial data. *J Am Stat Assoc* 97:632–648
- Groves D, Goldfarb R, Knox-Robinson C, Ojala J, Gardoll S, Yun G, Holyland P (2000) Late-kinematic timing of orogenic gold deposits and significance for computer-based exploration techniques with emphasis on the Yilgarn Block, Western Australia. *Ore Geol Rev* 17:1–38
- Guan Y (2008) On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. *J Am Stat Assoc* 103:1238–1247
- Guan Y, Wang H (2010) Sufficient dimension reduction for spatial point processes directed by Gaussian random fields. *J R Stat Soc Ser B* 72:367–387
- Handcock M, Morris M (1999) *Relative distribution methods in the social sciences*. Springer, New York
- Harrell F (2001) *Regression modeling strategies*. Springer, New York
- Hasenstab R (1983) 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, submitted to the Passaic River Basin Special Studies Branch, Department of the Army, USA. New York District Army Corps of Engineers
- Hastie T, Tibshirani R (1990) *Generalized additive models*. Chapman and Hall/CRC
- Hauck W Jr, Donner A (1977) Wald's test as applied to hypotheses in logit analysis. *J Am Stat Assoc* 72(360, Part 1):851–853
- Hogg R, Craig A (1970) *Introduction to mathematical statistics*, 3rd edn. Macmillan
- Hosmer D, Lemeshow S (2000) *Applied logistic regression*, 2nd edn. Wiley
- Jiménez-Valverde A (2012) Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Glob Ecol Biogeogr* 21:498–507
- Jones M (1991) Kernel density estimation for length-biased data. *Biometrika* 78:511–519
- Knox-Robinson C, Groves D (1997) 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
- Krzanowski W, Hand D (2009) *ROC curves for continuous data*. Chapman and Hall/CRC Press, London/Boca Raton
- Kutoyants Y (1998) *Statistical inference for spatial Poisson processes*. No. 134 in lecture notes in statistics. Springer, New York
- Kvamme K (1983) Computer processing techniques for regional modeling of archaeological site locations. *Adv Comput Archaeol* 1:26–52
- Kvamme K (1995) A view from across the water: the North American experience in archeological GIS. In: Lock G, Stančić Z (eds) *Archaeology and geographical information systems: a European perspective*, pp 1–14. CRC Press
- Kvamme K (2006) There and back again: revisiting archeological locational modeling. In: Mehrer M, KLWescott (eds) *GIS and archaeological site modelling*, pp 3–40. CRC Press
- Lahiri S (1999) Asymptotic distribution of the empirical spatial cumulative distribution function predictor and prediction bands based on a subsampling method. *Probab Theory Rel Fields* 114:55–84

- Lahiri S, Kaiser M, Cressie N, Hsu N (1999) Prediction of spatial cumulative distribution functions using subsampling. *J Am Stat Assoc* 94:86–97
- Landwehr J, Pregibon D, Shoemaker A (1984) Graphical methods for assessing logistic regression models. *J Am Stat Assoc* 79:61–83
- Lewis P (1972) Recent results in the statistical analysis of univariate point processes. In: Lewis P (ed) *Stochastic point processes*. Wiley, New York, pp 1–54
- Lindsey J (1992) *The analysis of stochastic processes using GLIM*. Springer, Berlin
- Lindsey J (1995) *Modelling frequency and count data*. Oxford University Press
- Lindsey J (1997) *Applying generalized linear models*. Springer
- Lindsey J, Mersch G (1992) Fitting and comparing probability distributions with log linear models. *Comput Stat Data Anal* 13:373–384
- Little R, Rubin D (2002) *Statistical analysis with missing data*, 2nd edn. Wiley
- Lloyd C (2011) *Local models for spatial analysis*. CRC Press, Boca Raton, FL
- Loader C (1999) *Local regression and likelihood*. Springer, New York
- Lobo J, Jiménez-Valverde A, Real R (2007) AUC: a misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr* 17(2):145–151
- Manly B, McDonald L, Thomas D (1993) *Resource selection by animals: statistical design and analysis for field studies*. Chapman and Hall, London
- McCullagh P, Nelder J (1989) *Generalized linear models*, 2nd edn. Chapman and Hall
- Møller J, Waagepetersen R (2004) *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton, FL
- Nam BH, D’Agostino R (2002) Discrimination index, the area under the ROC curve. In: Huber-Carol C, Balakrishnan N, Nikulin M, Mesbah M (eds) *Goodness-of-fit tests and model validity*. Birkhäuser, Basel, pp 267–279
- Okabe A, Boots B, Sugihara K (1992) *Spatial tessellations: concepts and applications of Voronoi diagrams*. Wiley series in probability and mathematical statistics. Wiley, Chichester, England, New York
- Openshaw S (1984) *The modifiable area unit problem*. Geo Books, Norwich
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Modell* 190:231–259
- R Development Core Team (2011) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. ISBN 3-900051-07-0
- Rakshit S, Rubak E, Nair G, Baddeley A (2017) Meaning of the ROC curve for a spatial point pattern (manuscript in preparation)
- Rathbun S, Cressie N (1994) Asymptotic properties of estimators of the parameters of spatial inhomogeneous Poisson point processes. *Adv Appl Probab* 26:122–154
- Renner I, Warton D (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics* 69:274–281
- Renner I, Elith J, Baddeley A, Fithian W, Hastie T, Phillips S, Popovic G, Warton D (2015) Point process models for presence-only analysis. *Methods Ecol Evol* 6(4):366–379. <https://doi.org/10.1111/2041-210X.12352>
- Robinson W (1950) Ecological correlations and the behavior of individuals. *Am Sociol Rev* 15:351–357
- Royall R (1988) The prediction approach to sampling theory. In: Krishnaiah P, Rao C (eds) *Handbook of statistics*, vol 6 (Sampling). Elsevier, Amsterdam, pp 399–413
- Sager T (1982) Nonparametric maximum likelihood estimation of spatial patterns. *Ann Stat* 10:1125–1136
- Scholtz S (1981) Location choice models in Sparta. In: III RL, Ottinger J, Scholtz S, Limp W, Watkins B, Jones RD (eds) *Settlement predictions in sparta: a locational analysis and cultural resource assessment on the uplands of Calhoun County, Arkansas*, no. 14 in Arkansas archaeological survey research series. Arkansas Archaeological Survey, Fayetteville, Arkansas, USA, pp 207–222

- Thuiller W, Vayreda J, Pino J, Sabate S, Lavorel S, Gracia C (2003) Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Glob Ecol Biogeogr* 12:313–325
- Tukey J (1972) Discussion of paper by F.P. Agterberg and S.C. Robinson. *Bull Int Stat Inst* 44(1):596; Proceedings, 38th Congress. International Statistical Institute
- Wakefield J (2004) A critique of statistical aspects of ecological studies in spatial epidemiology. *Environ Ecol Stat* 11:31–54
- Wakefield J (2007) Disease mapping and spatial regression with count data. *Biostatistics* 8:158–183
- Waller L, Gotway C (2004) *Applied spatial statistics for public health data*. Wiley
- Warton D, Shepherd L (2010a) Correction note—Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. *Ann Appl Stat* 4:2203–2204
- Warton D, Shepherd L (2010b) Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. *Ann Appl Stat* 4:1383–1402
- Watkins K, Hickman A (1990) Geological evolution and mineralization of the Murchison Province, Western Australia. Bulletin 137, Geological Survey of Western Australia, 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
- Wheatley D, Gillings M (2002) *Spatial technology and archaeology: the archaeological applications of GIS*. Taylor and Francis

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

