

[Click here to view linked References](#)

Natural Resources Research manuscript No.
(will be inserted by the editor)

Optimal Thresholding of Predictors in Mineral Prospectivity Analysis

Adrian Baddeley · Warick Brown · Robin K. Milne ·
Gopalan Nair · Suman Rakshit · Tom Lawrence ·
Aloke Phatak · Shih Ching Fu

Abstract Some methods for analysing mineral prospectivity, especially the Weights of Evidence technique, require the predictor variables to be binary values. When the original evidence data are numerical values, such as geochemical indices, they can be converted to binary values by thresholding. When the evidence layer is a spatial feature such as a geological fault system, it can be converted to a binary predictor by buffering at a suitable cutoff distance. This paper reviews methods for selecting the best threshold or cutoff value, and compares their performance. The review covers techniques which are well-known in prospectivity analysis as well as unfamiliar techniques borrowed from other literature. Methods include maximisation of the estimated contrast, Studentised contrast, χ^2 test statistic, Youden criterion, statistical likelihood, Akman-Raftery criterion, and curvature of the capture-efficiency curve. We identify connections between the different methods, and we highlight a common technical error in their application. Simulation experiments indicate that the Youden criterion has the best performance for selection of the threshold or cutoff value, assuming that a simple binary threshold relationship truly holds. If the relationship between predictor and prospectivity is more complicated, then the likelihood method is the most easily adaptable. The weights-of-evidence contrast performs poorly overall. These conclusions are supported by our analysis of data from the Murchison goldfields, Western Australia. We also propose a bootstrap method for calculating standard errors and confidence intervals for the location of the threshold.

Article Highlights

- Reviews methods for converting a predictor to binary values for use in prospectivity analysis
- Recommends the Youden criterion as the best-performing criterion for selecting the threshold level
- Exposes weaknesses and errors in current practice

Keywords Akman-Raftery criterion · capture-efficiency curve · change-point estimation · likelihood · weights of evidence · Youden index

A. Baddeley School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, GPO Box U1987, Perth WA 6845, Australia E-mail: adrian.baddeley@curtin.edu.au	ORCID: 0000-0001-9499-8382
W. Brown John de Laeter Centre, Curtin University	ORCID: 0000-0002-8257-888X
R. K. Milne Department of Mathematics and Statistics, University of Western Australia	ORCID: 0000-0002-1098-2672
G. Nair Department of Mathematics and Statistics, University of Western Australia	ORCID: 0000-0003-3883-4986
S. Rakshit SAGI-West, School of Molecular and Life Sciences, Curtin University	ORCID: 0000-0003-0052-128X
A. Phatak Centre for Transforming Maintenance through Data Science, Curtin University	ORCID: 0000-0002-0637-7461
T. Lawrence	ORCID: 0000-0003-3648-7634
S. C. Fu Curtin University	ORCID: 0000-0002-9077-6026

INTRODUCTION

In many techniques for mineral prospectivity analysis, especially Weights of Evidence (WofE), the predictor variables are required to be binary (0 or 1) values (Agterberg 1992; Agterberg et al. 1993; Bonham-Carter et al. 1990). Geological survey data of other kinds can be used, provided they are first converted to binary values by thresholding (Goodacre et al. 1993). If the original evidence layer is a geochemical or geophysical variable, such as an isotope abundance ratio, magnetic field anomaly or gravitational field strength, the analyst chooses a threshold above which the value is labelled as anomalous. If the original evidence is a spatial feature, such as a fault system, the analyst will often choose a cutoff distance to define a buffer around the feature. We shall use the term ‘threshold’ in either case.

The crucial step is to choose the best value of the threshold. This can be quite difficult in practice (J. Liu and Cheng 2019, p. 1039). A poor choice of threshold can be costly, leading to poor prediction, wasted exploration effort, and failure to identify important predictor variables.

Several methods for selecting the threshold are commonly used in prospectivity analysis. They include choosing the threshold value which maximises the WofE contrast (Goodacre et al. 1993, pp. 291–292; Bonham-Carter 1995, pp. 321–322), the Studentised contrast (Bonham-Carter 1995, p. 325), the chi-squared test statistic (Bonham-Carter and Agterberg 1990), the Youden criterion (Chen and Wu 2019; Ruopp et al. 2008), or the curvature of the capture-efficiency curve (Porwal et al. 2010). The prospectivity literature contains only limited discussion about the relative performance of these methods (Ford et al. 2016; Yeomans 2018, Chapter 7) and there does not seem to be a consensus about the best method. There are also some concerns that, for example, the Studentised contrast may be unreliable when the number of deposits is small (Goodacre et al. 1993, pp. 291, 295).

Threshold selection is an instance of *change-point estimation*, a classical and widely-studied problem in statistical theory (Basseville and Nikiforov 1993; Chernoff and Rubin 1956; Page 1957; Pons 2018; Zacks 1983). The statistical literature on this problem suggests that the WofE contrast would perform poorly as a criterion for threshold selection, because the contrast is an estimate of effect size rather than of evidence for the threshold. The change-point literature offers some other promising methods for threshold selection, including cumulative sum (CUSUM) methods (Basseville and Nikiforov 1993, Chapter 2), profile likelihood (Galun and Trifonov 1982; Murphy and van der Vaart 2000; Kutoyants 1998, Chapter 5, pp. 183–224) and the method of Akman and Raftery (1986). The existence of so many different competing methods is unusual; statistical theory usually identifies a single method as being “optimal”, but the optimality theory is not applicable in this context, because prospectivity is not a differentiable function of the predictor. Furthermore, in prospectivity analysis the number of known deposits may be quite small, so that asymptotic statistical theory does not apply. Consequently, these alternative methods are not guaranteed to have better practical performance than the currently popular methods listed above.

In this paper we give an overview of all the methods of threshold selection listed above, and evaluate their performance on real and simulated data. We go into detail about each method, identify connections between them, and draw attention to common technical errors. We present graphical diagnostics which may reveal numerical quirks of the data and assist the analyst to reject geologically implausible results. We highlight the central role of the capture-efficiency curve and offer new insights about it. We also show how to calculate standard errors and confidence intervals for the threshold using bootstrap techniques.

Our main findings are as follows. The method of threshold selection based on maximising the WofE contrast is highly unreliable and should be avoided. Threshold selection using the Studentised contrast t has adequate performance. The Youden criterion stands out as having the best performance. These findings assume that a simple binary threshold relationship truly holds (which can be checked by inspecting the shape of the capture-efficiency curve). If the relationship between predictor and prospectivity is more complicated, the maximum profile likelihood method is the most suitable.

Despite their very different origins, we find that all the methods mentioned above are mathematically related, so they should not be regarded as independent sources of evidence; in particular, the Akman-Raftery and chi-squared statistics are mathematically equivalent, and the Youden criterion is equivalent to a CUSUM index.

Typically, the survey data will comprise many predictors (evidence layers). In the WofE approach, each variable is examined separately to determine its merit and suitability as a predictor, and to select the single best threshold value for that variable. This is different from other approaches such as CART, random forest and isolation forest methods (Breiman et al. 1984; Chen and Wu 2019) which use a decision tree of threshold rules that are optimized by considering all the predictors at once. The task of selecting correct thresholds is even more challenging in these multivariate approaches. For simplicity, and because WofE is so popular, we shall broadly follow the WofE approach and consider only one predictor at a time. However, our conclusions are also relevant to other techniques for mineral prospectivity analysis.

77 When many predictor variables are available, the geologist must decide which of the variables have an effect
 78 on prospectivity. Statistical tests of significance are often used. There are connections between significance tests
 79 and the criteria employed to select the threshold (Bonham-Carter 1995; Bonham-Carter and Agterberg 1990). The
 80 Youden criterion, which we found to have the best performance for selecting thresholds, is closely related to the
 81 Kolmogorov-Smirnov test (Conover 1999, pp. 295–301) which is known to have good performance and is calibrated
 82 for any sample size (i.e. for any number of deposits). Consequently, we recommend using the Kolmogorov-Smirnov
 83 test for significance testing.

84 The paper is organised as follows. First we briefly review terms and notation for binary predictors. Then
 85 we define the problem of choosing a threshold value for a numerical (“greyscale”) predictor, define notation, and
 86 introduce two example datasets. The next five sections describe different classes of methods in detail: those based on
 87 the estimated WofE contrast and Studentised contrast; one based on the χ^2 test statistic; ones based on the capture-
 88 efficiency curve; and methods borrowed from the change-point literature. The next section exposes mathematical
 89 connections between the different methods. Then we discuss statistical significance tests. We report on a simulation
 90 experiment and on analysis of data from the Murchison goldfields, Yilgarn craton, Western Australia. A further
 91 section describes bootstrap methods for calculating standard errors and confidence intervals for the threshold. The
 92 final section discusses further aspects of performance, model mis-specification and generalisations. Appendices
 93 provide mathematical details.

94 BINARY PREDICTORS

95 Thresholding converts numerical (“greyscale”) predictor values into binary predictor values, for use in prospectivity
 96 analysis techniques that require binary inputs. This paper is concerned only with the thresholding step, and not
 97 with the prospectivity analysis.

98 However, many threshold selection techniques involve prospectivity analysis: they consider every possible
 99 threshold value in turn, perform a prospectivity analysis using that threshold, and select the threshold value
 100 which gives the best result. In order to describe these techniques, we must first define terms and notation for
 101 prospectivity using a binary predictor.

102 The general scenario is sketched in Figure 1. There is a survey region S in two-dimensional space, within which
 103 we have identified an areal feature or subregion B that is believed to be prospective. The binary predictor takes
 104 the value 1 inside B , and the value 0 outside B . The spatial pattern of known locations of mineral deposits is also
 105 mapped.

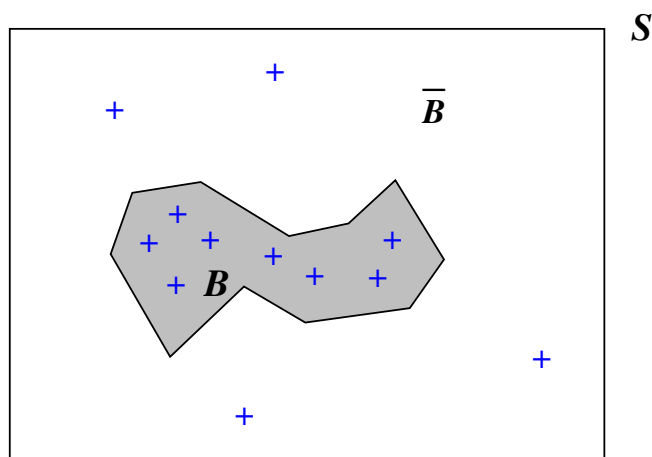


Fig. 1 Sketch of setup for prospectivity analysis once a threshold has been selected. The survey region S (rectangular frame) contains an areal feature B (shaded), believed to be prospective for deposits (blue crosses). The feature B has area a_B and contains n_B deposits. The complementary region \bar{B} (the unshaded part of the survey region outside the prospective feature) has area $a_{\bar{B}}$ and contains $n_{\bar{B}}$ deposits.

For quantitative analysis of prospectivity, the survey region is divided into grid cells of equal area. To simplify the discussion and the equations, we assume throughout this paper that grid cell area is negligible¹ compared to the area of the study region. The appendix **Non-negligible Grid Cell Size** gives the more complicated formulae that apply when grid cell area cannot be ignored.

In each grid cell, the survey records the value of the predictor, and records whether or not there are any (known) mineral deposits in the cell. Overall there are n grid cells that contain deposit locations; amongst them are n_B grid cells containing deposits lying inside feature B , and $n_{\bar{B}}$ grid cells containing deposits lying outside B . The total area of all grid cells, the area of the domain S , is denoted by a ; the area of B is denoted a_B and the area of its complement is $a_{\bar{B}}$.

The data listed above are enough to perform prospectivity analysis using the binary predictor. The most familiar of these methods is WofE (Agterberg 1992; Agterberg and Bonham-Carter 1999; Agterberg et al. 1993; Agterberg and Cheng 2002; Carranza 2004; Bonham-Carter 1995, Chapter 9; Agterberg 2014, Chapter 5). Under our assumption that cell areas are negligible, the estimated weights of evidence are $W_+ = \ln(n_B/a_B) - \ln(n/a)$ for cells inside B , and $W_- = \ln(n_{\bar{B}}/a_{\bar{B}}) - \ln(n/a)$ for cells outside B , where \ln denotes the natural logarithm. The WofE contrast is

$$\hat{C} = W_+ - W_- = \ln \frac{n_B}{a_B} - \ln \frac{n_{\bar{B}}}{a_{\bar{B}}} = \ln \left(\frac{n_B/a_B}{n_{\bar{B}}/a_{\bar{B}}} \right) = \ln \left(\frac{n_B/n_{\bar{B}}}{a_B/a_{\bar{B}}} \right). \quad (1)$$

(For non-negligible cell area, the corresponding formula is given in the appendix **Non-negligible Grid Cell Size** as equation (41).) It is instructive to notice that n_B/a_B is the observed average density of mineral deposit locations (deposits per unit area) inside the prospective region B , while $n_{\bar{B}}/a_{\bar{B}}$ is the observed average density of deposits outside B . The calculated contrast \hat{C} is the logarithm of the ratio of these two observed densities, or equivalently, it is the difference between the logarithms of the two observed densities. A calculated contrast value of $\hat{C} = +1$ means that the observed density of mineral deposit locations inside the prospective region B exceeds the density outside by a multiplicative factor of $e^{+1} = 2.7$. A calculated contrast value of $\hat{C} = 0$ means that the observed densities of deposits inside and outside the prospective region are equal.

For a single binary predictor, WofE is equivalent to logistic regression (Agterberg 1974; Schaeben 2014; Schaeben and Semmler 2016), and is also equivalent to maximum Poisson likelihood estimation and to maximum entropy modelling (Baddeley et al. 2010; Baddeley 2018; Baddeley et al. 2015, Chapter 9). Equivalence means that these techniques use different formulations but yield the same predictions.

WofE has been applied to mineral prospectivity analysis for gold (Agterberg and Bonham-Carter 2005; Boleneus et al. 2001; Cassard et al. 2008; Cheng 2008; J. Harris et al. 2015; Zhang and Zhou 2015), copper (Carranza 2004; D. Harris et al. 2003; Wang et al. 2016), tungsten (Y. Liu et al. 2014; Yeomans 2018), magmatic nickel sulfide (Porwal et al. 2010) and Sn-Cu deposits (Cheng 2007). WofE has also been used to create maps of landslide susceptibility (Cervi et al. 2010; Neuhäuser and Terhorst 2007; Polykretis and Chalkias 2018) and ecological habitat quality (Gorney et al. 2011; Romero-Calcerrada et al. 2010; Romero-Calcerrada and Luque 2006).

Modifications of WofE are described in Agterberg (2011); J. Liu and Cheng (2019). Comparative reviews of the performance of WofE and other methods are given by Ford et al. (2016) and Yeomans (2018, Chapter 7).

THE THRESHOLDING PROBLEM

In this section, we define the thresholding problem, discuss what information is required to select the best threshold, and introduce some notation that will be used throughout the paper. We also describe two datasets — one real and one synthetic — that will be used to motivate and illustrate the different thresholding methods.

Murchison (Western Australia) Gold Survey

The Murchison gold survey data, shown in Figure 2, were extracted from a regional survey (scale 1:500,000) of the Murchison goldfields area by the Geological Survey of Western Australia (GSWA). The figure shows the known locations of gold deposits and occurrences, and the known or inferred locations of geological faults, in a 330×400 kilometre rectangular study region.

These data were first presented and analysed in Knox-Robinson and Groves (1997). Gold deposit and occurrence locations were obtained from a database compiled by the Geological Survey of Western Australia (Geological Survey

¹While we have assumed that grid cell area is negligible in order to simplify and clarify the equations, this assumption is also pragmatically justifiable. If the survey region is 100 km across and grid cells are 100 metre squares, then there are a million grid cells, each with an area fraction of 10^{-6} , and our simplified equations are accurate to 4 decimal places.

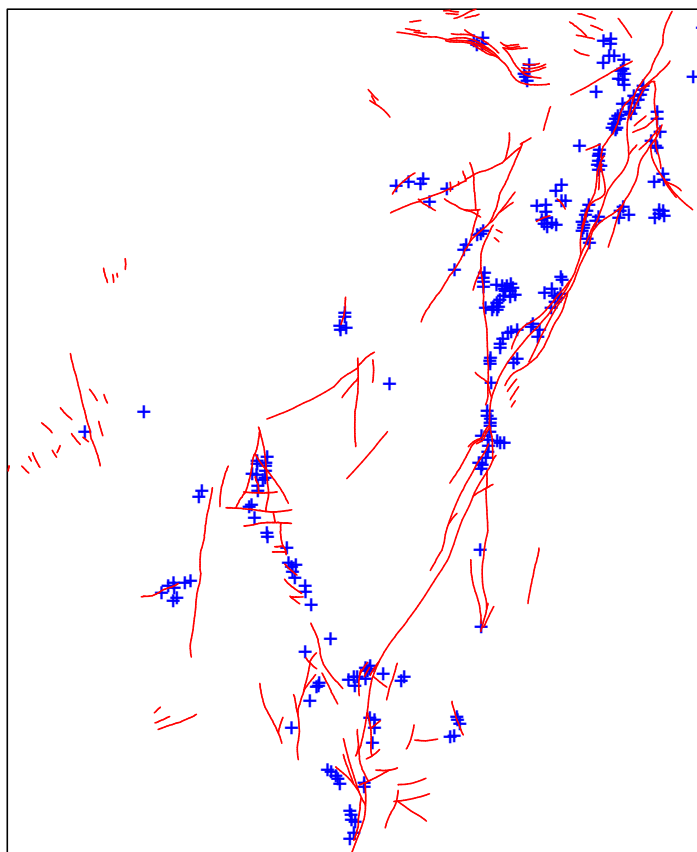


Fig. 2 Murchison gold data. Gold deposits (+) and major faults (—) in a survey region 330 by 400 km.

146 of Western Australia 1994); they include deposits of all sizes and are based on map surveys at a scale of 1:50,000
147 or better (Knox-Robinson and Groves 1997). Fault locations were compiled by Watkins and Hickman (1990). The
148 data were previously discussed in Groves et al. (2000), Foxall and Baddeley (2002), and Baddeley et al. (2015).
149 Data were kindly provided by Dr Carl Knox-Robinson, and permission granted by Dr Tim Griffin, Geological
150 Survey of Western Australia and by Dr Knox-Robinson.

151 Although there exist more up-to-date versions of this survey, we use the original 1994 version for demonstration
152 purposes because this is widely available, allowing comparison between different published analyses of the same
153 data. Users of the R software environment can obtain the data from the open-source contributed package `spatstat`
154 described in Baddeley et al. (2015) and accessible via www.spatstat.org.

155 Figure 2 strongly suggests that proximity to faults is predictive for gold prospectivity. This is expected because
156 crustal to lithospheric-scale faults and shear zones represent a first-order control on the location of orogenic gold
157 deposits at a province scale. These major fluid pathways focus fluid migration along pressure gradients into dilation
158 or damage zones in suitably-oriented, lower-order structures where gold may be deposited (Solomon and Groves
159 1994). Orogenic gold deposits are found world-wide and throughout geological time (Goldfarb and Groves 2015;
160 Groves et al. 2000; Groves and Santosh 2016; Robert et al. 2005).

161 In order to apply mineral prospectivity techniques, we would first have to convert the faults into a binary
162 variable, typically by creating a buffer around them. This process is discussed in the next section.

163 Thresholding a Greyscale Predictor

164 Goodacre et al. (1993) extended the WofE method to predictor variables that are real numbers (“greyscale val-
165 ues”) instead of binary values. Examples of such predictors include isotope ratios, geochemical assay, magnetic or
166 gravitational field strength, and distance from a fault system. The predictor is converted to a binary variable by
167 thresholding. Threshold selection is discussed by Goodacre et al. (1993), Bonham-Carter (1995, Section 9.5, pp.
168 319–325), Cheng (2004), and J. Liu and Cheng (2019).

169 The predictor is a spatial variable Z whose values are real numbers. That is, at any spatial location u in the
 170 survey region S , the value $Z(u)$ of the predictor is a real number that can be extracted or calculated from data.
 171 The task is to find the “best” threshold value z to convert the numerical variable Z to a binary variable, where
 172 the definition of “best” is open to discussion.

173 Figure 3 sketches this scenario. It shows a generic predictor Z which takes numerical values, depicted as a
 174 contour plot; alternatively Z could have been displayed as a greyscale pixel image. Thresholding effectively selects
 175 one of the contour levels, and the region below this contour level is designated as the prospective region B , to be
 176 interpreted as in Figure 1.

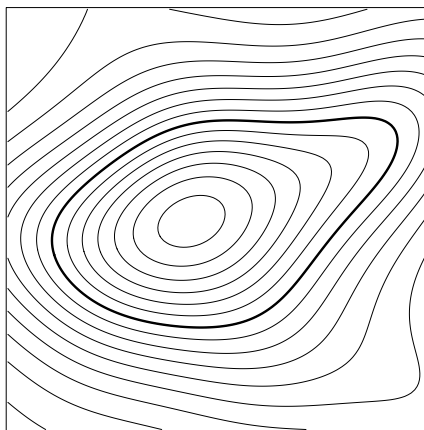


Fig. 3 Thresholding of a numerical predictor such as a geochemical or geophysical variable. Thin lines are contours of the spatial variable. One contour level (thick line) is selected as the optimal threshold.

177 If the original evidence layer is a geochemical variable, then Figure 3 represents contours of this variable.
 178 Thresholding has the effect of classifying each pixel as either “background” or “anomalous”. The “background”
 179 is the spread of geochemical assay values expected to be seen in non-mineralised Earth materials, and the upper
 180 limit of the background range is the “threshold” (Carranza 2009, pp. 9-10). Figure 3 illustrates the selection of
 181 one contour level as the optimal threshold.

182 If the original evidence layer is a spatial feature such as a fault system, then the researcher may choose to
 183 define $Z(u)$ as the shortest distance from the query location u to the feature. Distance values may be computed for
 184 each grid cell in a raster using the distance transform algorithm, or computed by vector geometry for any spatial
 185 location. Thresholding the distance variable at a particular threshold z is equivalent to creating a buffer of distance
 186 z around the original feature using GIS functionality (Cheng 2008, Figure 6). Buffers at different distances are
 187 contours of the distance function $Z(u)$. For the Murchison data in Figure 2, for example, Figure 4 shows a 5 km
 188 buffer around the major faults. Other examples of distance predictors include distance from a moraine, distance
 189 from a relatively steep slope zone, and distance from a relatively thick glacial drift layer (Cheng 2004, Table 1).

190 Once we have chosen a threshold level or threshold distance, the prospectivity analysis reduces to the situation
 191 sketched in Figure 1.

192 For simplicity in the exposition, we shall assume throughout that smaller values of Z are more prospective than
 193 large values. For example, this is true for orogenic gold in the Yilgarn craton (containing the Murchison survey)
 194 when Z is taken to be the distance to the nearest large fault. By adopting a particular threshold value z we are
 195 specifying the region of high prospectivity B as the spatial region consisting of all grid cells or locations u where
 196 the value of Z is at or below the threshold, $Z(u) \leq z$. The region of lower prospectivity \bar{B} consists of all locations
 197 where $Z(u) > z$. In the particular case where $Z(u)$ is the distance to a fault system, the prospective region is a
 198 buffer around the fault system, delimited by the contour at distance z around the faults.

199 Thresholding is a necessary step in some techniques for prospectivity analysis (such as WofE) but it may also
 200 be desirable for pragmatic reasons, because it leads to predictive models which are very simple to apply, and which
 201 drastically reduce the search area. The distance function $Z(u)$ could have been chosen simply because a buffer at
 202 a certain distance from the fault system is easy to delineate for exploration.

203 Mineral prospectivity analyses are often based on a small number of known deposits, although this number
 204 depends on the size of the study region and the inclusion criteria for the study (for example, when considering gold
 205 deposits with a resource of at least one tonne total contained gold). It is important that the technique for selecting
 206 the threshold should still work when the number of deposits is small. We should be able to state the minimum

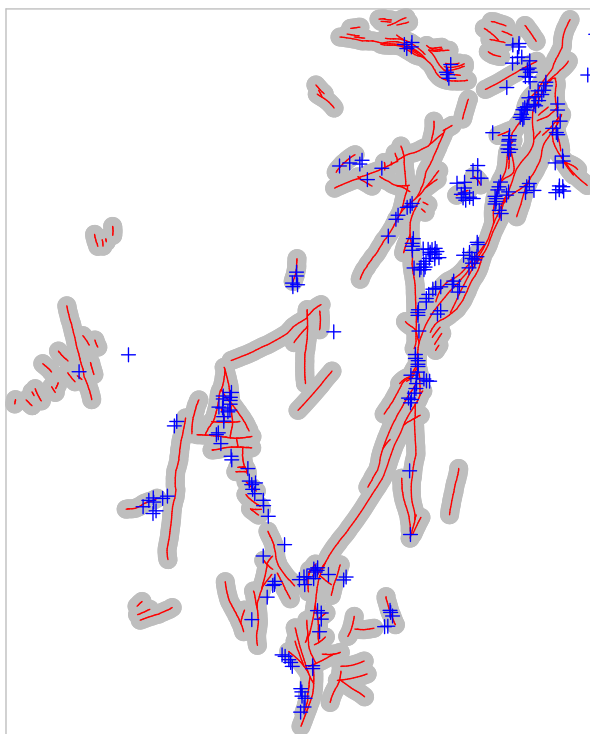


Fig. 4 Murchison gold data of Fig. 2 showing buffer at distance 5 km from the major faults.

207 number of deposits required for validity of the technique. For the thresholding methods described in this paper,
 208 the minimum required number of deposits is discussed in the following five sections (from **THRESHOLD SE-**
 209 **LECTION USING WOFE CONTRAST** to **THRESHOLD SELECTION USING CHANGE-POINT**
 210 **ANALYSIS**) and summarised in Table 10 of the Discussion.

211 Synthetic Example

212 We have also constructed a synthetic dataset that is superficially similar to the Murchison gold data, but has the
 213 virtue that the correct answer is known, so that the performance of threshold selection methods can be evaluated
 214 and compared.

215 The synthetic data, shown in Figure 5, are constructed so that there are strong contrasting densities of deposits
 216 inside and outside a 2 km buffer zone around the fault lines. This provides a basis for evaluating the performance
 217 of threshold selection methods, under controlled conditions. A threshold method which performs optimally in this
 218 example should select a threshold close to 2 km (as we discuss below). Although the synthetic dataset is not very
 219 realistic, any methods which work for complex real-world data must surely be expected to perform well on this
 220 simple dataset.

221 The linear features in the synthetic example are obtained from real data. A survey² of copper deposits in
 222 Queensland, Australia comprised the spatial locations of known copper deposits and of observed lineaments which
 223 are mostly faults. We have discarded the copper deposit points and simplified the lineament pattern by randomly
 224 deleting a fraction of the lineaments. We created a buffer region, consisting of all spatial locations (grid cells)
 225 lying no further than 2 km from the nearest lineament, shown in the left panel of Figure 5. We then simulated a random
 226 spatial pattern of deposits (shown in the right panel of Figure 5), with an average density of 2 deposits per 100 km²
 227 within the buffer, and 1 deposit per 100 km² outside the buffer. For example, in a 100 × 100 metre grid cell (area
 228 0.01 km²) inside the buffer the expected number of deposits inside the grid cell is $\text{number} = \text{density} \times \text{grid cell area} =$
 229 $(2/100) \times 0.01 = 0.0002$. Since the probability of more than one deposit is negligible, the probability that the grid

²The original survey was compiled by Dr. Jonathan Huntington, CSIRO and has been analysed by Berman (1986) and others. Coordinates were kindly provided to us by Dr. Mark Berman and Dr. Andy Green, CSIRO. The survey data are publicly available in the R package `spatstat` (Baddeley et al. 2015; Baddeley and Turner 2005) as the dataset `copper`. The data in Figure 5 are supplied in Online Resource 1.

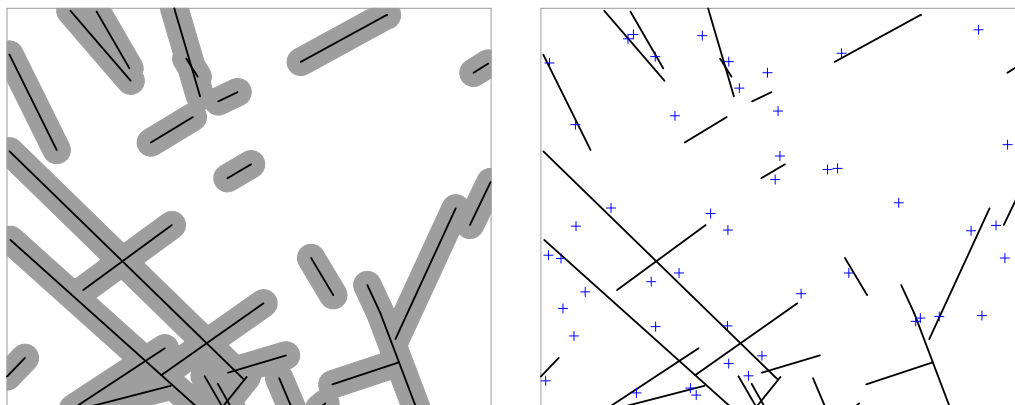


Fig. 5 Synthetic data example in a rectangle approximately 70 by 60 km. *Left:* Lineaments (—) and region of high prospectivity (grey shading) defined as the buffer within 2 km of the nearest lineament. *Right:* Typical simulated outcome. Mineral deposits (+) generated randomly with average density 2 deposits per 100 km² inside the high prospectivity region and 1 deposit per 100 km² outside.

230 cell contains a deposit is also equal to 0.0002. This synthetic model is typical of models that have been applied to
 231 the original dataset.

232 There is a single correct value of the threshold in this synthetic scenario. The true probability of a deposit
 233 depends on whether a grid cell is within or beyond the 2 km threshold. Therefore, a threshold of 2 km must
 234 be chosen in order that any prospectivity method correctly predicts the probability of a deposit. Bonham-Carter
 235 (1995, p. 319) defines the “best” threshold as “[...] the threshold that maximises the spatial association between
 236 the resulting binary map pattern and the point pattern [...]”. According to this definition, the “best” threshold is
 237 also 2 km.

238 The right panel of Figure 5 shows one dataset generated from this model, which we shall use to demonstrate
 239 the calculations and graphics. Since we know the correct threshold distance for this synthetic dataset is 2 km, we
 240 count $n_B = 34$ deposit points inside the buffer and $n_{\bar{B}} = 14$ points outside. The total area of the survey region is
 241 $a = 4102 \text{ km}^2$; the area inside the buffer is $a_B = 1294 \text{ km}^2$, and outside the buffer $a_{\bar{B}} = 4102 - 1294 = 2808 \text{ km}^2$.
 242 The estimated contrast value calculated according to equation (1) is $\hat{C} = \ln \frac{34/1294}{14/2808} = \ln \frac{0.026}{0.005} = 1.66$. In this
 243 synthetic example, the “true” value of the contrast is the logarithm of the ratio of the “true” densities of deposit
 244 points inside and outside B , which is $C_{\text{true}} = \ln(0.02/0.01) = \ln(2) = 0.69$.

245 Essential Data Required for Threshold Selection

246 For any numerical predictor, the minimum information necessary to select the best threshold is contained in the
 247 two graphs shown in Figure 6, using the synthetic data as an illustrative example. The graphs show the cumulative
 248 number of deposits (left panel) and the cumulative area (right panel) as a function of the threshold value on the
 249 horizontal axis. For any candidate threshold value z , let $n(z)$ be the number of grid cells with predictor value
 250 $Z \leq z$ which contain deposits, and $a(z)$ the total area of all grid cells with $Z \leq z$. (Strictly speaking the definition
 251 of $a(z)$ should exclude any grid cells which contain data points, but we have assumed grid cells are small enough to
 252 allow us to neglect this discrepancy.) The left panel of Figure 6 shows the number of deposits $n(z)$ plotted against
 253 the candidate threshold distance z . The right panel shows the area $a(z)$ plotted against z . Other transformations
 254 of these graphs will be used to select the threshold, as described in the next two sections.

255 The minimum data that need to be extracted from the mineral deposits are the predictor values z_1, \dots, z_n at
 256 these deposits, that is, $z_i = Z(x_i)$ for each i , where x_1, \dots, x_n are the deposit locations. Table 1 summarises our
 257 mathematical notation for the main elements of the problem.

258 Overview of Methods

259 In the following sections, we discuss different criteria for selecting the threshold. These criteria are summarised
 260 in Table 2 and the equations defining them are listed. The notation for the Akman-Raftery and Youden criteria

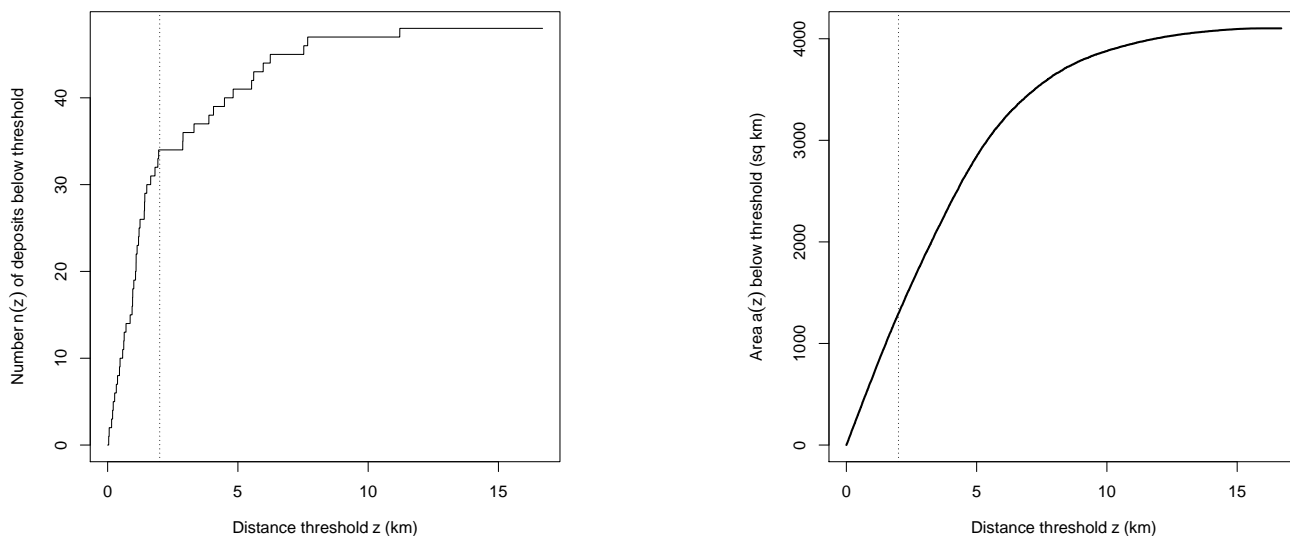


Fig. 6 Essential data for selecting the best threshold in the synthetic data example of Figure 5. *Left:* the number $n(z)$ of deposits lying no further than distance z away from the nearest lineament, plotted against z . *Right:* the area $a(z)$ of the buffer consisting of all locations lying no further than distance z away from the nearest lineament, plotted against z . Dotted vertical lines show the correct value of distance threshold, $z_{\text{true}} = 2$ km.

SYMBOL	MEANING
S	survey region
a	area of survey region
n	total number of deposits
x_1, \dots, x_n	spatial locations of deposits 1, 2, \dots , n
Z	predictor variable (a function of spatial location)
u	any spatial location
$Z(u)$	value of predictor Z at location u
z	any candidate value of threshold for the predictor
$n(z)$	number of grid cells containing deposits where $Z \leq z$
$a(z)$	area of region of survey where $Z \leq z$
$s(z)$	area fraction $s(z) = a(z)/a$
\hat{z}	selected (estimated) threshold value

Table 1 Summary of mathematical notation.

261 is not standard; we have chosen a more mnemonic notation. The performance of these criteria is summarised in
 262 Table 10 in the Discussion section.

SYMBOL	MEANING	EQUATION
$\hat{C}(z)$	WofE contrast	(2)
$t(z)$	Studentised contrast	(4)
$X^2(z)$	χ^2 test statistic	(6)
$Y(z)$	Youden criterion	(8)
$\ln L(z)$	profile log-likelihood	(12)
$\text{AR}(z)$	Akman-Raftery criterion	(13)

Table 2 Notation for threshold selection criteria, and equation numbers for their definitions. Each criterion is expressed as a function of the candidate threshold value z .

263 THRESHOLD SELECTION USING WOFE CONTRAST

264 Threshold selection using the WofE contrast (with or without rescaling) is expounded by Bonham-Carter (1995,
 265 Section 9.5, pp. 319–325) and is widely used.

266 In the original method, the threshold is selected by “progressively altering the threshold value to seek the
 267 greatest contrast” (Goodacre et al. 1993, pp. 291–292), that is, by choosing the threshold value which maximises

268 the WofE contrast (Bonham-Carter 1995, pp. 321–322; Agterberg 2014, p. 150). Recent examples include Cheng
 269 (2008, Section 7.2.3), Romero-Calcerrada and Luque (2006) and Romero-Calcerrada et al. (2010).

As a function of the candidate threshold z , the estimated WofE contrast (1) is given by

$$\hat{C}(z) = \ln \left(\frac{n(z)}{a(z)} \bigg/ \frac{n - n(z)}{a - a(z)} \right) = \ln \frac{n(z)}{a(z)} - \ln \frac{n - n(z)}{a - a(z)}, \quad (2)$$

270 that is, in (1) we replace n_B by $n(z)$ and a_B by $a(z)$, so that $n_{\bar{B}} = n - n_B = n - n(z)$ and $a_{\bar{B}} = a - a_B = a - a(z)$,
 271 giving (2). This formula assumes grid cell area is negligible; the corresponding formula for non-negligible cell area
 272 is given in the appendix entitled **Non-negligible grid cell size** as equation (42).

273 The left panel of Figure 7 shows a plot of the estimated contrast $\hat{C}(z)$ against the candidate threshold distance
 274 z for the synthetic data of Figure 5. The estimated threshold is $\hat{z} = 7.68$, which is a poor estimate of the correct
 275 threshold $z_{\text{true}} = 2$ km. However, very close to the correct threshold value, the graph has a secondary peak, and
 276 this would have been selected if the range of distances z had been suitably restricted. In particular, to avoid
 277 instability associated with small counts, it may be appropriate to restrict z to the range where both the expected
 278 counts $na(z)/a$ and $n(1 - a(z))/a$ are greater than or equal to 5, as discussed below equation (20).

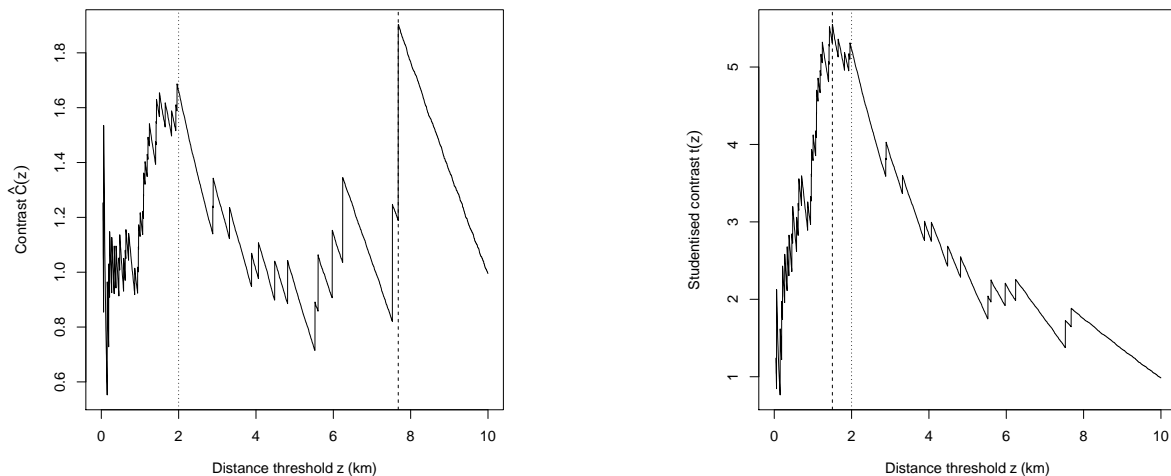


Fig. 7 Weights-of-evidence contrast (*Left*) and Studentised contrast (*Right*) plotted against candidate threshold distance, for the synthetic data of Figure 5. Dotted vertical lines show the correct threshold, $z_{\text{true}} = 2$ km. Dashed vertical lines show estimated thresholds, $\hat{z} = 7.68$ km and $\hat{z} = 1.50$ km, respectively.

279 For efficiency this procedure can be streamlined. Although our algorithm has considered a finely spaced grid of
 280 candidate values z , this is not strictly necessary. The estimated contrast $C(z)$ will be maximised at one of the values
 281 z_1, \dots, z_n which are the predictor values at the deposit locations (for example, the distances from each *deposit*
 282 *location* to the nearest fault), because the cumulative count $n(z)$ is incremented only at these values (assuming
 283 $a(z)$ is a strictly increasing function of z). Accordingly, a more efficient algorithm would involve computing only
 284 $C(z_1), \dots, C(z_n)$ and determining which of these is greatest.

285 We strongly recommend the use of graphical plots like Figure 7 as a tool for evaluating the evidence, identifying
 286 quirks of the data, and detecting nonsensical or geologically implausible results. If prior geological knowledge is
 287 available, then these graphs and numerical procedures can be restricted to the range of threshold values z which
 288 are geologically plausible.

289 THRESHOLD SELECTION USING STUDENTISED CONTRAST

290 The statistical performance of the contrast, as an index for selecting the threshold, becomes erratic at the extreme
 291 ends of the range of thresholds, because of sampling variability. It is natural to adjust for the sampling variability,
 292 as explained below.

Binary Predictor

Returning for a moment to the context of a binary predictor, the estimated standard error of the WofE contrast \hat{C} is

$$\text{se}(\hat{C}) = \sqrt{\frac{1}{n_B} + \frac{1}{n_{\bar{B}}}}. \quad (3)$$

This is the asymptotic (large sample) formula for standard error, derived using the “delta method” or “propagation of error” (Lehmann 1999, Section 2.5, pp. 85–93; Wasserman 2004, Section 9.9, pp. 131–132; Rice 2006, Section 4.6, pp. 149–154) assuming independence of outcomes in each grid cell, and assuming cell areas are negligible. It is reasonably accurate when both the numbers n_B and $n_{\bar{B}}$ are greater than or equal to 5 (implying that the total number n must be at least 10). The formula (3) is valid for any feature B . See the appendix headed **Non-negligible Grid Cell Size** when grid cell areas are large.

The “Studentised contrast” is the ratio of the contrast to its standard error, $t = \hat{C}/\text{se}(\hat{C})$ (Bonham-Carter 1995, p. 325; Cheng 2008, Section 7.2.3; J. Liu and Cheng 2019; Cheng 2007; Porwal et al. 2010, Table 1, p. 187). A value of t greater than 2 can be considered large. Although the term “Studentised” is commonly used when a variable is divided by its standard error, it may be misleading here, because the rescaled variable t does not follow a Student’s t distribution in this context. This is discussed further in the section **SIGNIFICANCE TESTS**.

For the synthetic data in the right panel of Figure 5, the deposit counts are $n_B = 34$ and $n_{\bar{B}} = 14$, so the estimated standard error according to (3) is $\text{se}(\hat{C}) = \sqrt{1/34 + 1/14} = 0.32$. The “Studentised contrast” is $t = \hat{C}/\text{se}(\hat{C}) = 1.66/0.32 = 5.19$.

Numerical Predictor

For a numerical predictor Z as discussed in the section on **THE THRESHOLDING PROBLEM**, we consider the Studentised contrast as a function of the candidate threshold value z ,

$$t(z) = \frac{\hat{C}(z)}{\text{se}(\hat{C}(z))}, \quad (4)$$

where the standard error (3) of the estimated contrast is

$$\text{se}(\hat{C}(z)) = \sqrt{\frac{1}{n(z)} + \frac{1}{n - n(z)}}. \quad (5)$$

Again this assumes grid cell areas are negligible; see the appendix **Non-negligible Grid Cell Size** for the more general formula (44).

The right panel of Figure 7 is a graph of the Studentised contrast $t(z)$ for the synthetic data against the candidate threshold distance z . The estimated threshold obtained by maximising the Studentised contrast is $\hat{z} = 1.50$ km, which is a mediocre estimate of the correct threshold $z_{\text{true}} = 2$ km, but is an improvement on the threshold chosen using the unadjusted contrast.

THRESHOLD SELECTION USING THE χ^2 STATISTIC

Another way to evaluate the strength of evidence for a threshold is to use a statistical significance test. The χ^2 test is described in the section **SIGNIFICANCE TESTS**. Bonham-Carter and Agterberg (1990) used the test statistic associated with the χ^2 test as an alternative criterion for threshold selection. Statistical theory would suggest that this is desirable, because the χ^2 test is valid for smaller samples (fewer deposits) than a test based on the Studentised contrast. This method has been demonstrated on data from the Yilgarn craton by Brown (2002) and Brown et al. (2002).

As a function of the candidate threshold value z , the χ^2 test statistic (equations (18) and (19)) is

$$X^2(z) = n \frac{\left(\frac{n(z)}{n} - \frac{a(z)}{a}\right)^2}{\frac{a(z)}{a} \left(1 - \frac{a(z)}{a}\right)}, \quad (6)$$

while the signed square root (20) is

$$X(z) = \sqrt{n} \frac{\left(\frac{n(z)}{n} - \frac{a(z)}{a}\right)}{\sqrt{\frac{a(z)}{a} \left(1 - \frac{a(z)}{a}\right)}}. \quad (7)$$

322 Note that these formulae do not assume that the grid cell area is negligible.

323 Figure 8 shows the χ^2 test statistic $X^2(z)$ plotted against candidate threshold distance for the synthetic data
 324 example. The estimated threshold obtained by maximising the χ^2 test statistic is $\hat{z} = 1.50$ km, agreeing with the
 325 result from the statistic t . Again, this is a mediocre estimate.

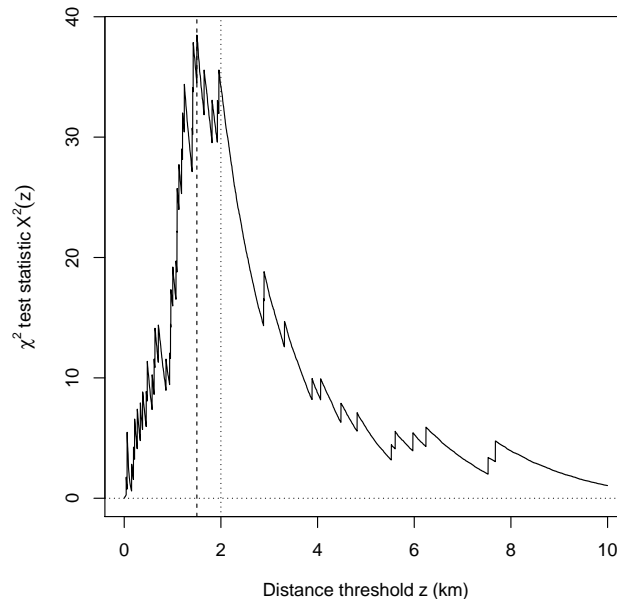


Fig. 8 χ^2 test statistic plotted against candidate threshold distance, for the synthetic data of Figure 5. Dotted vertical lines show the correct threshold, $z_{\text{true}} = 2$ km. Dashed vertical lines show estimated threshold $\hat{z} = 1.50$ km.

326 Values of $X^2(z)$ at each end of the graph are highly unreliable because the χ^2 test is valid only when the
 327 expected count $na(z)/a$ lies between 5 and $n - 5$, as discussed below equation (20).

328 THRESHOLD SELECTION USING THE CAPTURE-EFFICIENCY CURVE

329 This section reviews the role of the capture-efficiency curve, offering new insights and highlighting its central and
 330 unifying role in threshold selection.

331 Principle

332 The methods discussed in this section are based on comparing the distributions of predictor values at the deposit
 333 and non-deposit locations.

334 Two standard statistical ways to perform this comparison are illustrated in Figure 9 for the synthetic data
 335 example. In the left panel, the two curves show the cumulative fraction $F(z) = n(z)/n$ of deposits where the
 336 predictor value is less than or equal to a threshold value z , and the fraction of area $G(z) = a(z)/a$ of all spatial
 337 locations where the predictor is less than or equal to the same threshold z , plotted against the candidate threshold
 338 value z . If deposits were randomly and homogeneously scattered, these two curves would agree, apart from sampling
 339 variation in the counts of deposits. These curves are the *cumulative distribution functions* of the predictor values
 340 for two populations, the deposit and non-deposit locations. That is, $F(z)$ is the cumulative distribution function
 341 of the predictor value at a randomly-selected deposit, while $G(z)$ is the cumulative distribution function of the

342 predictor value at a randomly-selected location in the study region. In the synthetic data example, the predictor
 343 is the distance to the nearest fault, so that $F(z)$ is the cumulative distribution function of the distance from a
 344 randomly-selected deposit to the nearest fault, while $G(z)$ is the cumulative distribution function of the distance
 345 from a randomly-selected *spatial location* to the nearest fault.

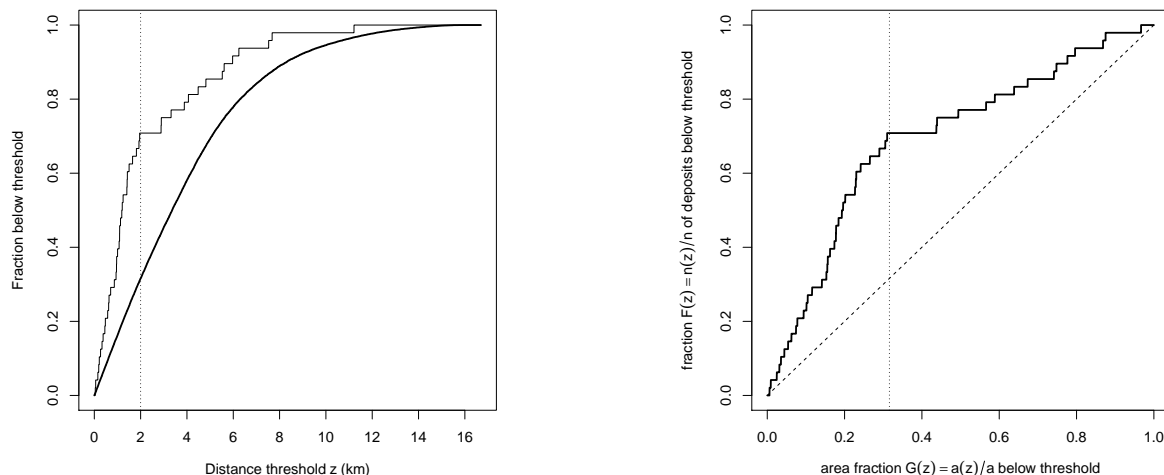


Fig. 9 Distributional comparisons, illustrated using the synthetic data example. *Left*: the fraction $F(z) = n(z)/n$ of deposits at or below threshold (thin solid line) and the fraction of area $G(z) = a(z)/a$ at or below threshold (thick solid line), plotted against candidate threshold value z . Dotted vertical line shows correct threshold $z_{\text{true}} = 2$ km. *Right*: capture-efficiency curve. Thick solid line is the fraction $F(z) = n(z)/n$ of deposits at or below threshold plotted against fraction of area $G(z) = a(z)/a$. Dotted vertical line shows area fraction $G(z_{\text{true}}) = 1296/4102 = 0.32$ corresponding to the correct threshold z_{true} . Diagonal dashed line would correspond to complete agreement between $F(z)$ and $G(z)$, which would occur if deposits were uniformly scattered over the study region.

346 In the right panel of Figure 9, the single curve shows the fraction of deposits $F(z) = n(z)/n$ plotted against the
 347 corresponding fraction of area $G(z) = a(z)/a$ for various candidate values z of the threshold. If deposits were ran-
 348 domly and homogeneously scattered, this curve would agree with the diagonal straight line $y = x$ apart from sam-
 349 pling variation. Such a curve has been reinvented many times and is variously known as a probability-probability
 350 (P–P) plot (Wilk and Gnanadesikan 1968), ROC curve (Krzanowski and Hand 2009; Nam and D’Agostino 2002),
 351 capture-efficiency curve (Porwal et al. 2010) or fitting-rate curve (Fabbri and Chung 2008). We shall use the
 352 term “capture-efficiency curve”, which is more widely recognised in mineral prospectivity analysis.³ The capture-
 353 efficiency curve and the ROC curve are indistinguishable if grid cell area is negligible, but they are slightly different
 354 otherwise.

355 The capture-efficiency curve depicts the proportion of deposits captured by cumulative proportions of the study
 356 area. There are several important reasons for using this curve to select the threshold. The first reason is that the
 357 capture-efficiency curve removes spatial geometry effects. In the left panel of Figure 9, the shapes of the two curves
 358 depend on the spatial arrangement of the fault lines in Figure 5; the area $a(z)$ of the buffer is not a simple function
 359 of the buffer distance z , because it depends on the spatial arrangement of the faults. In the right panel of Figure 9
 360 the effect of geometry is removed; a uniform distribution of deposits would correspond to the diagonal straight
 361 line in the right panel. Moreover, the capture-efficiency curve is unchanged even if the predictor is subjected to
 362 a transformation — for example, if distance-to-nearest-fault is replaced by the squared distance, square root of
 363 distance, or any increasing function of distance.

364 The second reason for using the capture-efficiency curve is that, if a simple threshold exists, the curve will
 365 have an inverted “V” shape, and the position of the elbow corresponds to the threshold. This is discussed in the
 366 next sub-section. The point of significant inflexion, where the slope changes from steep to gentle, can be used as
 367 a threshold for separating high-prospectivity areas from low-prospectivity areas (Porwal et al. 2010).

368 The capture-efficiency curve can also be interpreted as a cumulative distribution function in its own right,
 369 which has useful implications, explained in the appendix **Capture-Efficiency Curve as a CDF**.

³D. Harris and Pan (1999) used a related measure of classification performance which they call *exploration performance*. Exploration is viewed as the task of selecting the optimum threshold for allocating grid cells to the mineralized class, and it is assumed that all grid cells classified by the system as mineralized are retained as ground for further exploration. Classification performance is measured by the percentage of total grid cells that must be retained to ensure that different percentages of mineralized grid cells are retained when various cutoff probabilities for mineralization are applied to neural network output.

Relation to Threshold

The capture-efficiency curve gives the clearest indication of existence of a threshold (Goodacre et al. 1993; Porwal et al. 2010) for the following reason.

Suppose that the predictor variable Z does indeed have a threshold-like effect. That is, there is a true threshold value z_{true} so that, in the region where $Z \leq z_{\text{true}}$, deposits are homogeneous with expected density μ_1 deposits per unit area, while in the complementary region where $Z > z_{\text{true}}$, deposits are homogeneous with expected density μ_0 per unit area. If this true or correct threshold z_{true} were chosen, the data would conform to a WofE model with true contrast $C_{\text{true}} = \ln(\mu_1/\mu_0)$.

If this simple threshold model is true, then the capture-efficiency curve is expected to be linear on each side of the true threshold. That is, ignoring sampling variability, the capture-efficiency curve should consist of a straight line between the origin $(0, 0)$ and the point P , and another straight line between P and the top corner $(1, 1)$, where P has coordinates $(a(z_{\text{true}})/a, n(z_{\text{true}})/n)$.

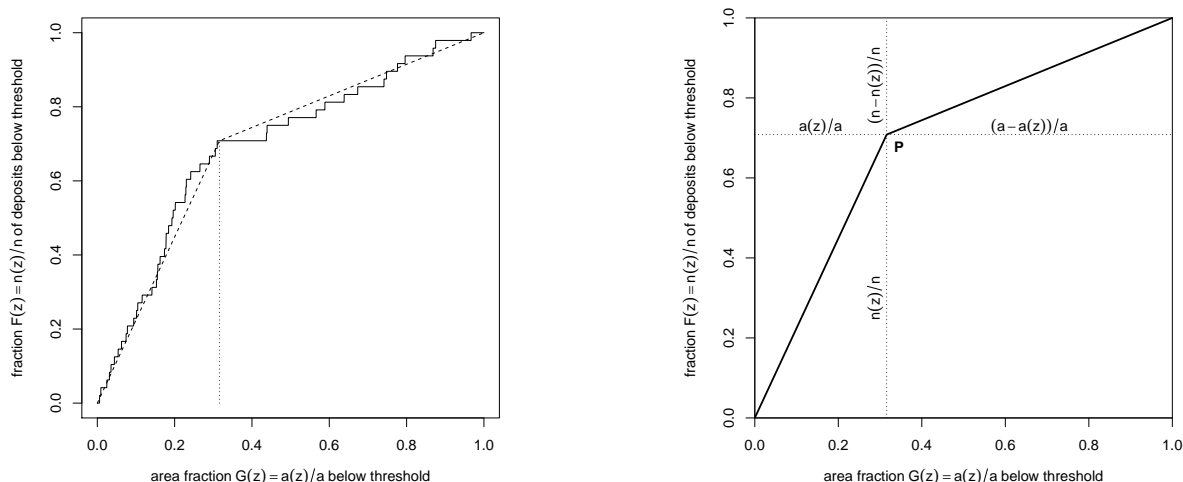


Fig. 10 Use of the capture-efficiency curve. *Left*: capture-efficiency curve (solid line) for synthetic data of Figure 5 and straight line approximation (dashed line) on either side of the point of maximum change in slope (dotted vertical line). *Right*: diagram for calculation (see text).

The synthetic data example conforms to this simple model, and its capture-efficiency curve (in the right panel of Figure 9) is approximately linear on each side of the true threshold value $z_{\text{true}} = 2$ km. The left panel of Figure 10 shows the same capture-efficiency curve augmented by the two lines described above, using knowledge of the true threshold $z_{\text{true}} = 2$ km.

There is also a connection with estimation of the WofE contrast: the slopes of the initial and final segments of the capture-efficiency curve (or in general, the ROC curve) are the average densities of deposits on either side of the threshold. The logarithm of the ratio of these slopes is equal to $\widehat{C}(z_m)$, the estimate of the WofE contrast at the elbow point z_m . We shall explain this in detail, referring to the right panel of Figure 10. For any point (x, y) on the capture-efficiency curve, consider the two straight lines joining $(0, 0)$ to (x, y) and joining (x, y) to $(1, 1)$ as illustrated in Figure 10. These lines have slope y/x and $(1 - y)/(1 - x)$ respectively. The logarithm of the ratio of the two slopes is $\ln(y/x) - \ln((1 - y)/(1 - x))$. Since the point (x, y) lies on the capture-efficiency curve, its coordinates are $x = a(z)/a$ and $y = n(z)/n$ for the corresponding candidate threshold z , so the logarithm of the ratio of the two slopes is equal to (2), the estimated contrast $\widehat{C}(z)$.

It follows that a technique which selected the threshold z by maximising the ratio of the slopes of the two lines, as described above, would be equivalent to maximising the WofE contrast.

Youden Criterion

If deposits were uniformly scattered over the study region, then the fraction $F(z) = n(z)/n$ of deposits below a certain threshold z would be approximately equal to the fraction $G(z) = a(z)/a$ of study area below this threshold.

Discrepancies between the “observed” fraction $F(z)$ and the “expected” fraction $G(z)$ constitute evidence against uniform scattering.

If the simple threshold model is true, as described in the previous section, then the estimated threshold \hat{z} can be obtained as the position of greatest discrepancy between the observed fraction $F(z)$ and “expected” fraction $G(z)$. This location can be identified in several ways. In the left panel of Figure 9, the threshold \hat{z} is the location where the two curves have the greatest vertical separation, that is, the location z where $F(z) - G(z)$ is greatest. In the right panel of Figure 9, the threshold \hat{z} corresponds to the location where the capture-efficiency curve has the greatest vertical deviation from the diagonal line; the x coordinate of this location is the area fraction $G(\hat{z}) = a(\hat{z})/a$ and from this we can determine the estimated threshold value \hat{z} . A third, graphical technique for identifying the estimated threshold is described below (see Figure 12).

The vertical deviation at a candidate threshold value z is

$$Y(z) = F(z) - G(z) = \frac{n(z)}{n} - \frac{a(z)}{a}. \quad (8)$$

The maximum positive vertical deviation is

$$Y = \max_z Y(z). \quad (9)$$

Under the assumption that grid cell areas are negligible, Y is equivalent to the *Youden index* (Chen and Wu 2019; Ruopp et al. 2008; Youden 1950) of classifier performance, and we call (8) the *Youden criterion*.⁴ The general formula for $Y(z)$, not assuming that grid cell area is negligible, is given in the appendix **Non-negligible Grid Cell Size** as equation (45).

The Youden index was originally defined (Youden 1950) as a measure of optimal performance of a medical diagnostic classifier. As explained by Ruopp et al. (2008), suppose that a patient’s disease status were reported as positive (disease present) if the result of a biomedical assay returned a value lower than a specified cutoff, and otherwise were reported as negative (disease absent). Sensitivity (**Se**) and specificity (**Sp**) are the probabilities of correctly identifying diseased and non-diseased individuals, respectively. If equal importance is given to sensitivity and specificity, then the predictive power or discriminating ability of the technique can be measured by $\text{Se} - (1 - \text{Sp})$ and the cutoff should be chosen to maximise this quantity. The maximised value is the Youden (1950) index.

In prospectivity analysis, the analogous procedure (Chen and Wu 2019) is to label a grid cell as “prospective” when the value of the predictor variable Z is below a threshold z . The sensitivity **Se** is the probability of correctly identifying grid cells that contain deposits, that is, the probability that a grid cell which actually contains a deposit will be correctly labelled as prospective by this rule. Under our assumption that grid cell areas are negligible, the sensitivity is equal to $n(z)/n = F(z)$, the fraction of deposit locations where the predictor is less than or equal to z . The specificity **Sp** is the probability of correctly identifying grid cells that do not contain a deposit, and is equal to the fraction $1 - G(z) = [a - a(z)]/a$ of study area where the predictor value exceeds z . The discriminating power of the classifier, $\text{Se} - (1 - \text{Sp})$, is equal to $F(z) - G(z)$.

Figure 11 shows the calculation of the Youden criterion for the synthetic data example. The left panel shows the two cumulative distribution functions $F(z) = n(z)/n$ and $G(z) = a(z)/a$ plotted against candidate threshold distance z . The right panel shows the Youden criterion $Y(z) = F(z) - G(z)$ plotted against z . In this synthetic data example, the estimated threshold obtained by maximising the Youden criterion is $z = 1.96$ km, an excellent estimate of the true value $z_{\text{true}} = 2$ km.

In statistical terminology, the graph of the Youden criterion in the right panel of Figure 11 is a “control chart” (Shewhart 1983) of the CUSUM type (Page 1954) and these are known to perform well for estimation of change-point locations (Basseville and Nikiforov 1993, Chapter 2; Barnard 1959; Hinkley 1971).

An alternative graphical technique for identifying the estimated threshold by the Youden method is sketched in Figure 12. Parallel lines with slope equal to 1 are drawn above the capture-efficiency curve, and displaced downward until a line touches the curve. The tangent point determines the selected threshold: from the tangent point we read off the horizontal coordinate, which is equal to the optimal area fraction, and this corresponds to the optimal threshold for the predictor.

In the original definition of the Youden index as a measure of diagnostic performance, equal importance was given to sensitivity and specificity. In the context of mineral prospectivity, this corresponds to assigning a monetary figure to the cost of thoroughly exploring each 1% of survey area that is equal to the value of each 1% of existing mineral deposits. If a different cost-benefit ratio α had been used, so that the value of 1% of mineral deposits was equal to the cost of testing $\alpha\%$ of the survey area, this would correspond to using parallel lines of slope α instead of slope 1 in Figure 12. A valuable insight from Figure 12 is that we would have selected the same threshold even if we had used parallel lines with some other slopes, because of the inverted “V” shape of the capture-efficiency

⁴The Youden index is traditionally denoted by the letter J , but we have called it Y for a more mnemonic notation.

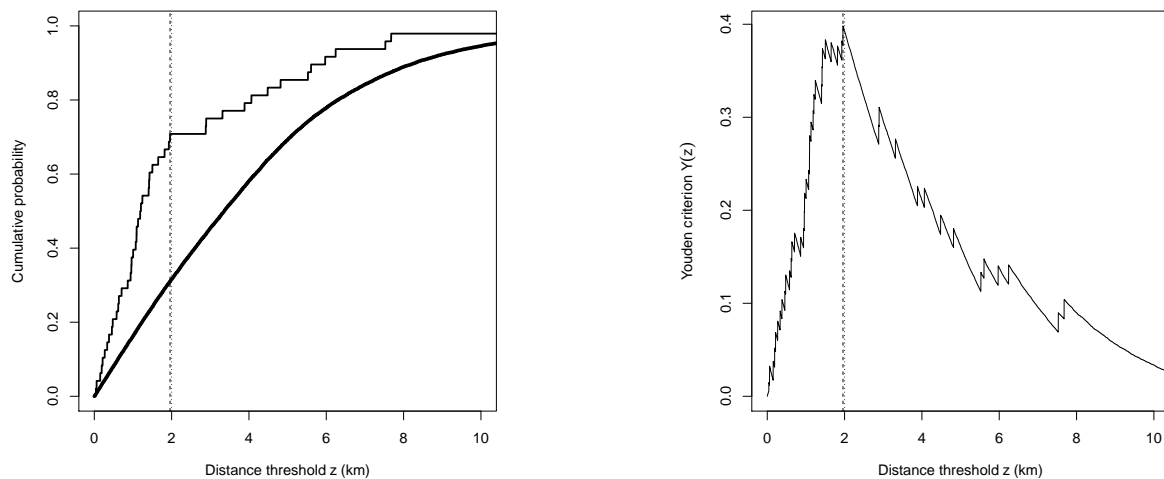


Fig. 11 Youden criterion for the synthetic data of Figure 5. *Left*: cumulative fractions $n(z)/n$ (thin curve) and $a(z)/a$ (bold curve) plotted against candidate threshold distance z . *Right*: Youden criterion $Y(z) = n(z)/n - a(z)/a$ plotted against z . Vertical line shows the correct threshold, $z_{\text{true}} = 2$ km and the estimated threshold, $\hat{z} = 1.96$ km, which are visually indistinguishable.

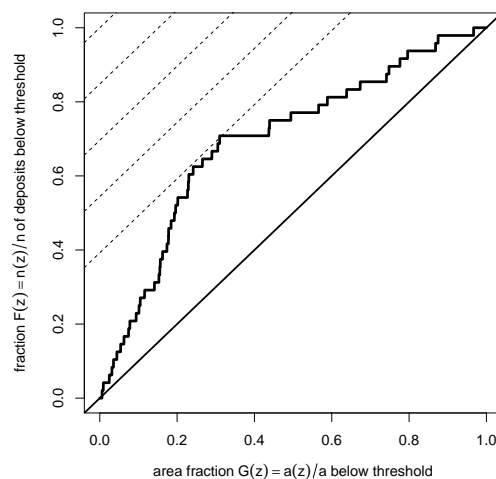


Fig. 12 Alternative graphical technique for the Youden method. The capture efficiency curve for the synthetic data is drawn. A line with slope 1 (dashed) is moved down the graph until it intersects the capture efficiency curve. The tangent point corresponds to the threshold that maximises the Youden criterion.

449 curve. That is, when the simple threshold model is true, the cost-benefit ratio is not very influential in choosing
 450 the threshold.

451 Slope of Curve

452 We have seen that, if the simple threshold model is true, the capture-efficiency curve is expected to be linear on
 453 each side of the threshold, and the threshold position can be discerned by any of the procedures described above.
 454 It would be more realistic to assume that the threshold model is not exactly true, but that there is nevertheless a
 455 large difference between the densities of deposit points occurring at low and high values of the predictor variable.
 456 In that case, one can exploit the fact that, at any position along the capture-efficiency curve, say the point
 457 $(a(z)/a, n(z)/n)$, the *slope* of the curve is proportional to the probability of a deposit when the predictor value is
 458 equal to z (Baddeley et al. 2012; Goodacre et al. 1993; Porwal et al. 2010). It would then be appropriate to select
 459 the threshold by finding “the point of significant inflexion where the slope changes from steep to gentle” (Porwal
 460 et al. 2010, p. 193). However, the procedure to be used for finding the inflection point or for measuring changes of
 461 slope was not described in detail in the paper by Porwal et al. (2010).

462 A possible weakness of this method is that estimates of derivatives can be highly unstable, and some form of
 463 smoothing is required. Goodacre et al. (1993) used spline smoothing to estimate the local slope of the capture-
 464 efficiency curve. The closely related technique of Baddeley et al. (2012) used kernel smoothing.

465 THRESHOLD SELECTION USING CHANGE-POINT ANALYSIS

466 In statistical theory, an abrupt change in the characteristics of a signal is called a *change-point* (Basseville and
 467 Nikiforov 1993; Page 1957; Zacks 1983). Detection of abrupt change, and estimation of the location of the change-
 468 point, are challenging problems because they do not conform to the standard assumptions of differentiability that
 469 justify most statistical techniques.

470 From the viewpoint of this literature, it would be expected that a procedure which estimates the change-
 471 point location by maximising the estimated contrast \hat{C} would not perform well. The estimated contrast $\hat{C}(z)$ is
 472 an estimate of effect size, that is, an estimate of the *magnitude* of the change at z , rather than a measure of
 473 the *evidence* for a change at location z . Adjusting for the variability of $\hat{C}(z)$ could improve performance, so the
 474 Studentised contrast $t(z)$ would be expected to perform better. As remarked above, the Youden criterion $Y(z)$
 475 would be expected to perform well, because it can be regarded as a control chart of CUSUM type, and these are
 476 known to perform well for estimation of change-point locations (Basseville and Nikiforov 1993, Chapter 2; Barnard
 477 1959; Hinkley 1971). For testing statistical significance of a change-point, it is important that the Youden criterion
 478 is associated with the Kolmogorov-Smirnov test (see the subsection on the **Kolmogorov-Smirnov test** in the
 479 section **SIGNIFICANCE TESTS** below).

480 Profile Likelihood

481 A common way to estimate the threshold z_{true} would be to maximise the *profile likelihood*.

482 Likelihood

483 Likelihood (Fisher 1922) is a fundamental tool in statistical analysis. For a parametric statistical model, the
 484 likelihood is a function $L(\theta)$ of the model parameters θ ; it is defined as the probability of observing the data that
 485 were actually observed, assuming the model is true with the specified parameter values. The likelihood function
 486 serves as a comparative measure of the support given by the data to different choices of the parameter values.

487 One important use of the likelihood function is to estimate the model parameters from the data. In the *method*
 488 *of maximum likelihood*, after observing the data we select the parameter value which maximises the likelihood,
 489 that is, the parameter value which enjoys the greatest support given by the data.

490 Expositions of likelihood can be found in Hogg and Craig (1970, Section 8.2), Rice (2006, Section 8.5), Lindsey
 491 (1996, Chapter 3), Severini (2000, Chapter 3), Kalbfleisch (1985, Chapters 9–10) and Murphy and van der Vaart
 492 (2000).

493 In prospectivity analysis, the techniques of logistic regression and WofE are both applications of the method
 494 of maximum likelihood (Baddeley et al. 2010; Schaeben 2014). Furthermore, the hypothesis tests reviewed in the
 495 section on **SIGNIFICANCE TESTS** are all obtained from the likelihood.

496 Maximising the likelihood is equivalent to maximising the *logarithm* of the likelihood, and it is usual to work
 497 with the log-likelihood for convenience. For further simplification, any terms in the log-likelihood equation that do
 498 not depend on the parameters can be removed because they do not affect the maximisation.

For the simple threshold model, the observed data are the two curves depicted in Figure 6, and the log-likelihood⁵ of the candidate threshold z and densities μ_0, μ_1 is

$$\ln L(z, \mu_0, \mu_1) = n(z) \ln \mu_1 + (n - n(z)) \ln \mu_0 - a(z)\mu_1 - (a - a(z))\mu_0. \quad (10)$$

499 The maximum likelihood estimate of the three parameters (z, μ_0, μ_1) would be obtained by finding the values of
 500 (z, μ_0, μ_1) which yield the maximum value of (10).

⁵For a detailed derivation of this likelihood, see the appendix **Likelihood Function for Threshold Model**. For a different, more intuitively accessible explanation, see Baddeley et al. (2015, pp. 132–135, 342–343).

Profile Likelihood for the Threshold Model

The *profile likelihood* method is a general technique for dealing with multiple parameters when one parameter is more “important” than the others. The profile likelihood is a function of the “important” parameter only, and is defined as the maximum value of likelihood over all possible values of the “less important” parameters.

For the threshold model, the “important” parameter is the threshold value z . To apply profile likelihood to the threshold model, for each candidate threshold value z we maximise the likelihood over all possible values of the densities μ_0, μ_1 . Using calculus the maximum is achieved when $\mu_1 = n(z)/a(z)$ and $\mu_0 = (n - n(z))/(a - a(z))$; these are just the average densities of deposits per unit area inside and outside the buffer, as we might expect. Substituting these into (10) we obtain the profile log-likelihood for a candidate threshold z ,

$$\ln L(z) = n(z) \ln \frac{n(z)}{a(z)} + (n - n(z)) \ln \frac{n - n(z)}{a - a(z)} - n. \quad (11)$$

Using (2) this is equivalent to

$$\ln L(z) = n \ln \frac{n(z)}{a(z)} - (n - n(z))\widehat{C}(z) - n. \quad (12)$$

Maximising the profile log-likelihood over candidate threshold values z is equivalent to maximising the likelihood of the three parameters (z, μ_0, μ_1) , enabling us to estimate the threshold by the method of maximum likelihood.

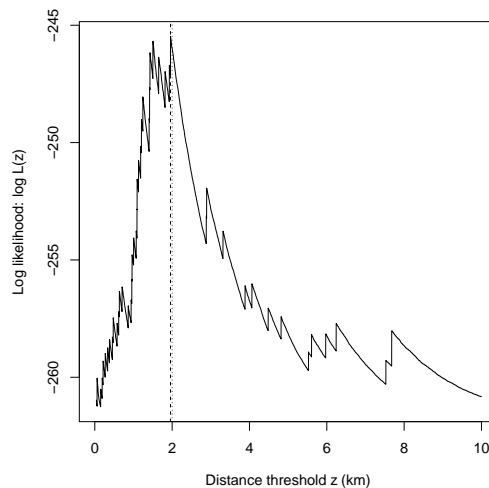


Fig. 13 Log-likelihood plotted against candidate threshold distance, for the synthetic data of Figure 5. Dotted vertical line shows correct threshold, $z_{\text{true}} = 2$ km. Dashed vertical line shows estimated threshold, $\widehat{z} = 1.96$ km, virtually indistinguishable from the correct value.

Figure 13 shows the application of likelihood methods to the threshold selection problem for the synthetic data. The profile log-likelihood $\ln L(z)$ from (12) is plotted against candidate threshold distance z . The estimated threshold, $\widehat{z} = 1.96$ km, obtained by maximising the profile log-likelihood, delivers an excellent estimate of the correct value, $z = 2$ km. In particular, this method yields a better estimate than is obtained by maximising the Studentised contrast t .

Although likelihood methods are theoretically optimal (i.e. they have the best possible statistical performance) in many contexts, they may not be optimal for the threshold model. The standard theory justifying maximum likelihood assumes that the likelihood is a differentiable function of the parameters. This is not true for the threshold model because the log-likelihood (10) or (12) is not differentiable as a function of the candidate threshold z . This has the important consequence that maximum likelihood is not necessarily the optimal method for estimating the threshold.

Another important consequence is that standard errors and confidence intervals cannot be calculated in the usual way. In the section **BOOTSTRAP STANDARD ERRORS AND CONFIDENCE INTERVALS** below, however, we propose a bootstrap method for calculating standard errors and confidence intervals for the threshold location.

Other Statistical Approaches to Change-Point Estimation

Since the method of maximum likelihood is not necessarily optimal for estimating a change-point, there is extensive literature aimed at finding better techniques (Basseville and Nikiforov 1993; Bhattacharya and Brockwell 1976; Bhattacharya and Johnson 1968; Chernoff and Zacks 1964; Darkhovsky 1976; Deshayes 1984; Galun and Trifonov 1982; Gardner 1969; Hájek and Rényi 1955; Hinkley 1970, 1971; Kander and Zacks 1966; Pratt 1959; Sen and Srivastava 1975; Smith 1975; Zacks 1983). For surveys, see Basseville and Nikiforov (1993); Pons (2018).

Thresholding is equivalent to estimating a *change-point in probability density*. Recall that the capture-efficiency curve can be regarded as a cumulative distribution function. The slope of the capture-efficiency curve can therefore be regarded as a probability density function. If the capture-efficiency curve follows the ideal, inverted “V” shape, then its slope jumps abruptly from a high value to a low value, i.e. there is a change-point in the probability density.

Techniques for estimating a change-point in a probability density go back at least as far as Chernoff and Rubin (1956) and Rubin (1961), and there is now extensive literature⁶ on this problem (Kutoyants 1998, Chapter 5, pp. 183–224). Other literature, with possible relevance to prospectivity, includes models of a change-point plus gradual trend (Loader 1992), a change-point in discretised data (West and Ogden 1997), two change-points (Chernoyarov, Kutoyants, and Top 2018), change-points in waiting time (Commenges and Seal 1985), and smoothing-based estimates (Leonard 1978).

Akman–Raftery Method

Akman and Raftery (1986) and Raftery and Akman (1986) proposed another method for estimating a change-point in the probability density of a variable. For mathematical convenience, they assumed the variable ranges between 0 and 1, the change-point is a value s lying strictly between 0 and 1, and that the probability density of the variable is constant on either side of the change-point.

To apply the Akman-Raftery method to the thresholding problem, we must first transform the threshold value z to the change-point coordinate s by setting $s = s(z) = a(z)/a = G(z)$, the fraction of area of the survey where the predictor value does not exceed the threshold z . Equivalently, s is the horizontal coordinate in the capture-efficiency curve. Then the Akman-Raftery criterion is defined as⁷

$$\text{AR}(z) = \sqrt{s(z)(1-s(z))} \left(\frac{n(z)}{s(z)} - \frac{n-n(z)}{1-s(z)} \right) = \sqrt{a(z)(a-a(z))} \left(\frac{n(z)}{a(z)} - \frac{n-n(z)}{a-a(z)} \right). \quad (13)$$

The term in parentheses on the right hand side of (13) is the *difference* between the estimated densities of deposits in the regions above and below the putative threshold z . The standard error of this difference, under the null hypothesis that a threshold is absent, is equal to $1/\sqrt{ns(z)(1-s(z))}$. That is, $\text{AR}(z)$ is $1/\sqrt{n}$ times the standardised difference between estimated intensities above and below the putative threshold. This is a “standardised” rather than “Studentised” quantity because the standard error is based on the null hypothesis that a threshold is absent.

The change-point is estimated by maximising $|\text{AR}(z)|$. However, when $s = s(z)$ is close to 0 or 1, the sampling fluctuations in $\text{AR}(z)$ become large. Accordingly Akman and Raftery (1986) stipulate that the maximum should be taken only for area fractions s in the range $[s_0, s_1]$ with $0 < s_0 < s_1 < 1$. For example, restricting to the range $[0.1, 0.9]$ would mean that the candidate threshold z must lie between the 10th and 90th percentiles of the values of the predictor variable over the study region.

This constraint is impractical in many applications to mineral prospectivity. If the predictor is distance to a map feature like a fault line or lithological boundary, it is not acceptable to exclude the smallest percentiles, as these correspond to the region where the effect of the feature is likely to be greatest. However, we shall study the effect of this constraint.

In our discussion, the Akman-Raftery *criterion* is the quantity $\text{AR}(z)$ defined in (13) which could be evaluated for any value of z , while the Akman-Raftery *method* (abbreviated ARM) is the procedure which maximises $\text{AR}(z)$ subject to the constraint on z .

For the synthetic data, Figure 14 shows the Akman-Raftery criterion $\text{AR}(z)$ plotted against the candidate threshold distance z in the left panel, and plotted against area fraction s in the right panel. Oblique shading covers the predictor values which lie below the 10th percentile or above the 90th percentile of predictor values in the study region. In this example, the maximum value of $\text{AR}(z)$ over *all possible* thresholds z is achieved at distance

⁶For experts in statistics, we note that the central limit theorem does not apply to this problem: the asymptotic law of the log-likelihood is a compound Poisson process, not a Gaussian process (Pflug 1983; Kutoyants 1998, eq. (5.3), p. 184).

⁷Akman and Raftery (1986) denoted their criterion by the letter Y , but we use AR for a more mnemonic notation.

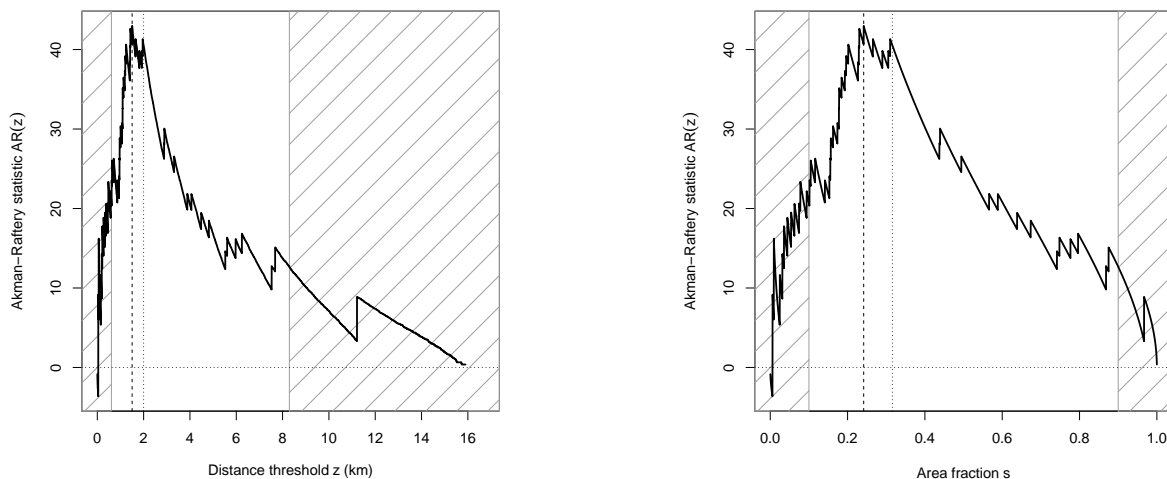


Fig. 14 Akman-Raftery criterion plotted against candidate threshold distance z (*Left*) and plotted against area fraction s (*Right*) for the synthetic data of Figure 5. Dotted vertical line shows correct threshold, $z_{\text{true}} = 2$ km (left panel) and corresponding correct area fraction $s_{\text{true}} = 0.316$ (right panel). Dashed vertical line shows estimated threshold, $\hat{z} = 1.50$ km (left panel) and corresponding area fraction $\hat{s} = 0.242$ (right panel). Oblique shading covers predictor values below the 10th percentile or above the 90th percentile.

565 threshold $z = 1.50$ km corresponding to area fraction $s = 0.242$; this happens to lie between the 10th
 566 percentiles, so that if we had applied the Akman-Raftery constraint that s should lie between $[s_0, s_1] = [0.1, 0.9]$,
 567 the outcome would have been identical. The Akman-Raftery estimated threshold $\hat{z} = 1.50$ km is equal to the
 568 estimates obtained by maximising the χ^2 statistic or the Studentised contrast, which we described as a mediocre
 569 estimate of the true value 2 km.

570 The theory developed by Akman and Raftery (1986) and Raftery and Akman (1986) has other important
 571 advantages, explained in the section **SIGNIFICANCE TESTS** below.

572 MATHEMATICAL CONNECTIONS BETWEEN THE CRITERIA

The threshold selection criteria described above are qualitatively similar to one another; they are all measures of the discrepancy between the observed average densities of deposits above and below the threshold value. Indeed we have been able to prove algebraically that some of them are directly related. The most striking relationship, which we prove in the appendix **Connections between the Threshold Criteria**, is that

$$\text{AR}(z) = \sqrt{n} X(z). \quad (14)$$

573 That is, the Akman-Raftery criterion $\text{AR}(z)$ defined in (13) is *proportional* to the signed χ^2 test statistic $X(z)$
 574 defined in (7). This means that maximising the Akman-Raftery criterion $\text{AR}(z)$ over all possible thresholds,
 575 i.e., without constraints, would be equivalent to maximising the χ^2 test statistic $X^2(z)$. However, the Akman-
 576 Raftery *method* constrains the range of candidate threshold values, so that the Akman-Raftery method is not
 577 quite equivalent to maximising $X^2(z)$ over all possible thresholds.

The second relationship can be seen directly by comparing (7) and (8):

$$X(z) = \frac{Y(z)}{\text{se}_0(Y(z))}, \quad (15)$$

where

$$\text{se}_0(Y(z)) = \sqrt{s(z)(1 - s(z))/n} \quad (16)$$

578 is the standard deviation of $Y(z)$ under the null hypothesis that the density of deposits is uniform. That is, the
 579 signed χ^2 statistic $X(z)$ is a *standardised* version of the Youden criterion $Y(z)$. See the appendix **Connections**
 580 **between the Threshold Selection Criteria**, subsection *Relation between χ^2 statistic and Youden criterion*.
 581 This is a “standardised” rather than “Studentised” version because the standard error is calculated under the null
 582 model. The relationship (15) does not imply redundancy between $X(z)$ and $Y(z)$, but provides some insight.

Thirdly, there are qualitative similarities. The WofE contrast (2) is analogous to the Youden criterion (8) except for the use of the logarithm in (2). The statistic t in (4) is the *Studentised* version of the WofE contrast (2) while the signed χ^2 statistic $X(z)$ is the *standardised* version of the Youden criterion (8).

The overall conclusion is that the different methods for threshold selection described here are not independent methodologies, despite their very different origins. In practice, these different methods cannot be relied upon to provide independent assessments of prospectivity.

SIGNIFICANCE TESTS

Often it is desired to conduct a statistical test of significance of a binary predictor, or a test for the *existence* of a threshold effect in a numerical predictor variable Z .

Significance Test for a Binary Predictor

First consider the scenario sketched in Figure 1 where a binary predictor is already given, and we wish to test whether the predictor has a statistically significant effect.

The predictor has no effect if the WofE contrast is zero. To test whether the WofE contrast is statistically significantly different from zero, Goodacre et al. (1993, p. 291) used the Studentised contrast

$$t = \frac{\hat{C}}{\text{se}(\hat{C})}, \quad (17)$$

and determined statistical significance by referring the value of t to a standard Normal distribution. For example, at the usual 5% level of significance, the contrast is declared to be significant if $|t| > 1.96$, or equivalently if $|\hat{C}| > 1.96 \text{ se}(\hat{C})$. See also Cheng (2007, p. 321).

The null hypothesis is that the predictor has no effect, so that the density of deposits is uniform over the survey area; the densities inside and outside the buffer are equal, so that the true contrast is zero. The alternative hypothesis is the two-sided alternative that the true contrast is not equal to zero.

The test based on the statistic in equation (17) treats large positive and large negative values of t as being statistically significant. Negative values of t (or \hat{C}) suggest that the feature B is antithetical to deposits, and that the complementary feature \bar{B} is prospective. Attaching significance to negative values of contrast may be appropriate in some applications; in their study of seismicity, Goodacre et al. (1993, Table 2, p. 295) reported that some rock types had statistically significant negative values of contrast.

However, in prospectivity analysis, features are usually selected for investigation because our domain knowledge suggests they are likely to be prospective. It is then appropriate to perform a “one-sided” or “one-tailed” test in which only large positive values of t are considered significant; the 5% critical value 1.96 is changed to 1.64, that is, the contrast is declared significant if $t > 1.64$ or equivalently $\hat{C} > 1.64 \text{ se}(\hat{C})$. This is the familiar distinction between “two-tailed” and “one-tailed” tests (Conover 1999, pp. 78–79). The choice of test depends on the purpose, as discussed above. For brevity, we use two-tailed tests unless otherwise stated.

It may be confusing that the test rule uses the critical points of the Normal distribution, while the test statistic is called “Studentised” and is denoted by the symbol t , suggesting that we should use the critical points of the Student’s t distribution. This would not be correct because the test statistic t in (17) does not have a Student’s t -distribution.⁸

Goodacre et al. (1993, pp. 291, 295) point out that a test based on t is valid when the total number of deposits n is large, but would be inaccurate if either of the numbers n_B or $n_{\bar{B}}$ is “small”. For that eventuality they propose an alternative test based on the binomial distribution (Goodacre et al. 1993, pp. 295–296). The total number of deposits n is treated as fixed, and statistical significance is assessed by referring the observed count n_B to the binomial distribution with n trials and success probability $s = a_B/a$. This procedure is slightly more complicated, as it involves looking up a table or using software to compute the p -value. The binomial distribution can be approximated by the Normal distribution when ns and $n(1-s)$ are both greater than 10 (Carlton and Devore 2014, p. 181), so we believe that the more refined test of Goodacre et al. (1993) is needed only when n_B or $n_{\bar{B}}$ is less than 10.

⁸The technical requirements for the Student’s t distribution are not satisfied: in particular the numerator and denominator of (17) are not independent variables.

A closely-related test which may be more familiar is the χ^2 (chi-squared) test of goodness-of-fit, based on the test statistic

$$X^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \frac{(n_B - e_B)^2}{e_B} + \frac{(n_{\bar{B}} - e_{\bar{B}})^2}{e_{\bar{B}}}, \quad (18)$$

where $e_B = (n/a)a_B = n(a_B/a)$ and $e_{\bar{B}} = (n/a)a_{\bar{B}} = n(a_{\bar{B}}/a)$ are the expected counts in B and \bar{B} , respectively, when the deposits are randomly distributed with constant density. The test is performed by referring X^2 to the χ^2 distribution with 1 degree of freedom (Hogg and Craig 1970, Section 10.2, pp. 308–316; Kendall and Stuart 1973, Chapter 30; Pearson 1900). A χ^2 variable with one degree of freedom is equivalent to the square of a standard Normal variable, so this particular χ^2 goodness-of-fit test is equivalent to computing the signed square root $X = \pm\sqrt{X^2}$ and referring X to the standard Normal distribution. In order to take the square root of X^2 , we simplify equation (18) as explained in the appendix **Connections between the Threshold Selection Criteria**, yielding

$$X^2 = n \frac{\left(\frac{n_B}{n} - \frac{a_B}{a}\right)^2}{\frac{a_B}{a} \left(1 - \frac{a_B}{a}\right)}, \quad (19)$$

so that

$$X = \sqrt{n} \frac{\frac{n_B}{n} - \frac{a_B}{a}}{\sqrt{\frac{a_B}{a} \left(1 - \frac{a_B}{a}\right)}}. \quad (20)$$

Note that X is a standardised version of the count n_B . The contrast is declared statistically significant (i.e. different from zero) at the 5% significance level if $|X| > 1.96$. The one-tailed test would declare significance if $X > 1.64$.

The χ^2 test is also based on large-sample asymptotic theory and suffers from inaccuracy when the sample size (number of deposits) is small. The established rule of thumb is that both e_B and $e_{\bar{B}}$ should be at least 5 (Rice 2006, p. 487; Hogg and Craig 1970, p. 309) implying that n should be at least 10.

Statistical theory shows that the test with the best performance in large samples is the Likelihood Ratio Test, based on the statistic

$$\text{LRTS} = 2 \sum \text{observed} \times \ln \left(\frac{\text{observed}}{\text{expected}} \right) = 2 \left(n_B \ln \frac{n_B}{a_B} + n_{\bar{B}} \ln \frac{n_{\bar{B}}}{a_{\bar{B}}} - n \ln \frac{n}{a} \right), \quad (21)$$

which is referred to the χ^2 distribution with 1 degree of freedom, so that the estimated contrast \hat{C} is declared significant at the 5% level if $\text{LRTS} > 1.96^2 = 3.84$ (Kendall and Stuart 1973, eq. (30.6), p. 438). The Likelihood Ratio Test also depends on asymptotics but performs well, even in small samples (small numbers of deposits).

For the WofE contrast method for estimating thresholds, the Studentised contrast statistic (17) is $t = 1.66/0.32 = 5.19$, which exceeds the 5% critical value 1.96 for the Normal distribution for the two-sided test. For the χ^2 test, the signed square root statistic (20) is

$$X = \sqrt{48} \frac{34/48 - 1294/4102}{\sqrt{(1294/4102)(1 - 1294/4102)}} = 5.86,$$

which exceeds 1.96. The Likelihood Ratio Test statistic (21) is

$$\text{LRTS} = 2(34 \ln(34/1294) + 14 \ln(14/2808) - 48 \ln(48/4102)) = 31.12,$$

which far exceeds the 5% critical value 3.84 for the χ^2 distribution with 1 degree of freedom. Such agreement between the outcomes of different hypothesis tests is expected in large samples (large numbers of deposits).

All the significance tests described above perform well, and give similar outcomes, for the synthetic data set. More exotic tests, which have even better small-sample performance, are also available (Read and Cressie 1988).

In summary, the customary test based on the Studentised contrast (17), referred to the standard Normal distribution, is valid provided the total number of deposits n is at least 30 and the expected numbers $e_B, e_{\bar{B}}$ both exceed 5. Otherwise, one should use the χ^2 test based on (20), or the binomial test described by Goodacre et al. (1993, pp. 295–296), or the Likelihood Ratio Test using (21), or the tests described by Read and Cressie (1988). Our overall recommendation is to use the Likelihood Ratio Test.

The χ^2 test is widely used in exploration geoscience (Filzmosera et al. 2005; Garrett 1989; Ghannadpour and Hezarkhani 2016) with caveats (Ballantyne and Cornish 1979).

644 The Problem of “Peeking”

645 An important caveat about statistical significance tests is that the test conditions must be fixed in advance; it is
 646 not valid to change the test procedure in the light of the observed data. The statistical significance tests described
 647 in the previous subsection, for testing whether a spatial feature B is prospective, are only valid when the feature
 648 B is known and fixed in advance. It would not be valid to perform the same test calculation several times with
 649 different choices of B , to select the most favorable test result, and then to judge statistical significance using the
 650 rules described above.

651 In particular, if we construct a buffer around a fault system, select the optimal buffer distance by maximising
 652 the Studentised contrast t , and find that $t(\hat{z})$ for this optimised buffer exceeds 1.96, then this does *not* imply
 653 statistical significance at the 5% level. The test based on the statistic t assumes that the buffer distance was fixed
 654 in advance. Selecting the “optimal” buffer distance using the observed deposit location data would be a form of
 655 cheating, equivalent to selecting the best result among several test outcomes. This is more likely to achieve a
 656 spuriously significant outcome, that is, the probability of erroneously declaring “significance” is much greater than
 657 0.05, because of the multiple testing (“look elsewhere”) effect (Hochberg and Tamhane 1987; Hsu 1996; Shaffer
 658 1995). See Rice (2006, pp. 60, 321–324) and the XKCD cartoon on jellybean statistics (Anonymous 2011).

659 Colleagues often ask us why this procedure should be condemned as an instance of “multiple” testing, when
 660 they have only performed a single hypothesis test calculation, using the optimised threshold value. Consider that,
 661 for each candidate threshold z , the value of the statistic $t(z)$ could have been translated into a p -value (whether we
 662 compute it or not) associated with the hypothesis test for threshold z ; larger values of $t(z)$ correspond to smaller
 663 (i.e. more significant) p -values; consequently, by choosing the largest value of $t(z)$ we have equivalently chosen the
 664 smallest p -value (most significant outcome) out of all the possible significance tests available.

665 This is tantamount to performing many different statistical tests, each associated with a different value of z ,
 666 but *applying the standard of statistical significance that is appropriate for a single test*. This is likely to produce
 667 spurious significant outcomes, since we will have multiple opportunities to reach the standard of significance for a
 668 single test.

669 When the threshold \hat{z} is estimated from data, the correct critical value for a 5% significance test based on $t(\hat{z})$
 670 is much larger than 1.96, and is more complicated to compute, even for large samples (large numbers of deposits).

671 Kolmogorov-Smirnov Test

672 The problem identified above is that we require a test for the *existence* of a threshold effect, where the threshold
 673 level z is not known in advance. The null hypothesis is that there is no threshold effect, prospectivity is uniform
 674 across the survey region, and the deposits are homogeneously randomly scattered. The alternative hypothesis is
 675 the simple threshold model, in which prospectivity is higher below the threshold than above the threshold, where
 676 the true threshold value z_{true} is unknown and must be estimated from data. Such a test could be deployed to select
 677 the predictor variables to be used in the analysis, from amongst a list of candidate predictors.

678 Effectively this is a test of significance of the disagreement between two curves, that is, a test of agreement or
 679 disagreement between cumulative distribution functions.

The most famous test of this kind is the *Kolmogorov-Smirnov (K-S) test* (Conover 1999, pp. 295–301), which
 uses a rescaled version of the maximum discrepancy between the curves $F(z)$ and $G(z)$. In our context, we use the
 one-sided K-S test statistic, KS_+ , which is closely related to the Youden criterion Y defined in (9):

$$KS_+ = \sqrt{n} Y_+, \quad (22)$$

where n is the number of deposits, Y is the Youden criterion, and Y_+ is its positive value, $Y_+ = Y$ if $Y \geq 0$, and
 $Y_+ = 0$ if $Y < 0$. The K-S test declares the discrepancy to be significant if $KS_+ > c_n$, where c_n is the critical
 value of the test, which depends on the total number of deposits n . Reinterpreting Table 4.2 of Stephens (1986, p.
 105) we get, for the 5% level of significance,

$$c_n \approx 1.224 / (1 + 0.12 / \sqrt{n} + 0.11 / n). \quad (23)$$

680 The critical value c_n is a slowly-increasing function of the number of deposits n , for example $c_5 = 1.14$, $c_{10} = 1.17$,
 681 $c_{100} = 1.21$ and $c_\infty = 1.224$. We emphasise that the K-S test does not require large numbers of deposits; the test
 682 can be performed for small n , using the appropriate critical value c_n .

683 For the synthetic data example in Figure 5, we have $Y = 0.40$ and $n = 48$ so that $KS_+ = \sqrt{48} \times 0.40 = 2.77$
 684 and the 5% critical point is $c_{48} = 1.20$, so the test declares the discrepancy to be significant at the 5% level.

685 Other tests related to the capture-efficiency curve include the Anderson-Darling and Cramér-Von Mises tests,
 686 and the tests of Berman (1986). The Berman tests were motivated by problems in mineral exploration; Berman’s
 687 Z_2 test is based on the area under the capture-efficiency curve. These other tests will be investigated further in a
 688 subsequent article.

689 Akman-Raftery (Constrained Chi-Squared) Test

690 The method of Akman and Raftery (1986) and Raftery and Akman (1986) also accounts for the effect of searching
 691 for the optimal threshold, enabling us to perform a statistical significance test for existence of a change-point.

Akman and Raftery (1986) show that, if \hat{z} is the estimated change-point location obtained by the Akman-Raftery method with area fractions s constrained to lie between s_0 and s_1 , then a test of the hypothesis of no change (i.e. equal densities on each side of the change-point) can be performed using the test statistic

$$d = (1/\sqrt{n}) |\text{AR}(\hat{z})|, \quad (24)$$

where n is the number of deposits. The statistic d is the standardised Akman-Raftery criterion maximised subject to the constraint on area fractions. The approximate p -value is (Akman and Raftery 1986, p. 1586)

$$p \approx 2\sqrt{\frac{2}{\pi}} \exp(-d^2/2)(\alpha d - \alpha/d + 1/d), \quad (25)$$

692 where $\alpha = (1/2) \ln(s_1(1-s_0)/(s_0(1-s_1)))$. For the synthetic data example with the constraints $s_0 = 0.1$, $s_1 = 0.9$,
 693 we get $d = 6.2$ and $p = 9.5 \times 10^{-8}$ indicating strong evidence of a threshold.

694 Since we have established in the appendix **Connections between the Threshold Selection Criteria** that
 695 the χ^2 test statistic satisfies $X(z) = \text{AR}(z)/\sqrt{n}$, the results of Akman and Raftery give us a test, based on the χ^2
 696 test statistic, for the presence of a threshold. One simply replaces d by $X(\hat{z})$ in (25). That is, the signed square
 697 root of the χ^2 test statistic, at the selected threshold, has the p -value given by (25) with d replaced by $X(\hat{z})$.

698 Akman and Raftery (1986, p. 1587) also mention that their p -value formula (25) could be used to construct
 699 a confidence interval for the true threshold value, by “inversion” of the formula. To do this, we would consider
 700 $d(z) = (1/\sqrt{n}) |\text{AR}(z)|$ as a function of z , evaluate $p = p(z)$ for each $d = d(z)$ according to (25), plot $p(z)$ against
 701 z , and find the range of z values where the p -values exceed 0.05. However, (25) is accurate only for large n and
 702 large d , so this confidence interval is unlikely to be valid when the number of deposits is small.

703 SIMULATION EXPERIMENTS

704 Simulation experiments based on synthetic data were conducted in order to objectively evaluate and compare the
 705 performance of the different techniques. This cannot be done with real data, because the true threshold is not
 706 known. The simulation experiments involve a simple, ideal scenario where there is a single best threshold in the
 707 sense of Bonham-Carter (1995, p. 319). Performance of each technique is measured by its ability to identify this
 708 correct threshold.

709 Experiment 1

710 In this experiment we generated 10,000 simulated outcomes of the synthetic model described in subsection **Syn-**
 711 **thetic Example** of the section on **THE THRESHOLDING PROBLEM**. The prospective region is the buffer
 712 zone shown in the left panel of Figure 5. Deposit locations were generated randomly with an average density of
 713 0.02 and 0.01 points per km^2 inside and outside the buffer, respectively. The right panel of Figure 5 shows a typical
 714 simulated outcome. The correct threshold distance is $z_{\text{true}} = 2$ km. The prospective region B within distance 2
 715 km has area $a_B = 1294 \text{ km}^2$, and its complement has area $a_{\bar{B}} = 2809 \text{ km}^2$. The expected number of deposits is
 716 $0.02 \times 1294 = 25.9$ inside B and $0.01 \times 2809 = 28.1$ outside B , making an expected total of 54.0 deposits.

717 For each simulated outcome, the threshold distance z was estimated by maximising the estimated contrast, the
 718 Studentised contrast, the log-likelihood, the Akman-Raftery criterion with constraints, the Youden criterion and
 719 the χ^2 test statistic. We denote these methods by C, t, LL, ARM, Y, and X2, respectively.

720 For the Akman-Raftery method (ARM), following their recommendation, we set $s_0 = 0.1$, $s_1 = 0.9$ so that
 721 solutions were constrained to lie between the first and ninth deciles of the predictor (i.e. of the predictor values
 722 at all locations in the study region), which are 0.609 km and 8.290 km respectively. This constraint prevents the

723 algorithm from choosing extreme values of the threshold, and confers a somewhat unfair or unrealistic advantage
 724 on the Akman-Raftery method in our comparison experiment.

725 For each estimation method, the spread of 10,000 threshold values \hat{z} obtained by applying the method to
 726 the simulated patterns could be portrayed as a histogram. To simplify the results, we have used kernel density
 727 estimation, effectively smoothing the histogram. The left column of Figure 15 shows probability densities of the
 728 threshold values obtained from each estimation method. Probability densities were computed by Gaussian kernel
 729 density estimation (Silverman 1986; Wand and Jones 1995) with bandwidth selected by Silverman's rule (Silverman
 730 1986, p. 48, eq. (3.31)), both implemented in the function `density.default` in the R system (R Development Core
 731 Team 2018).

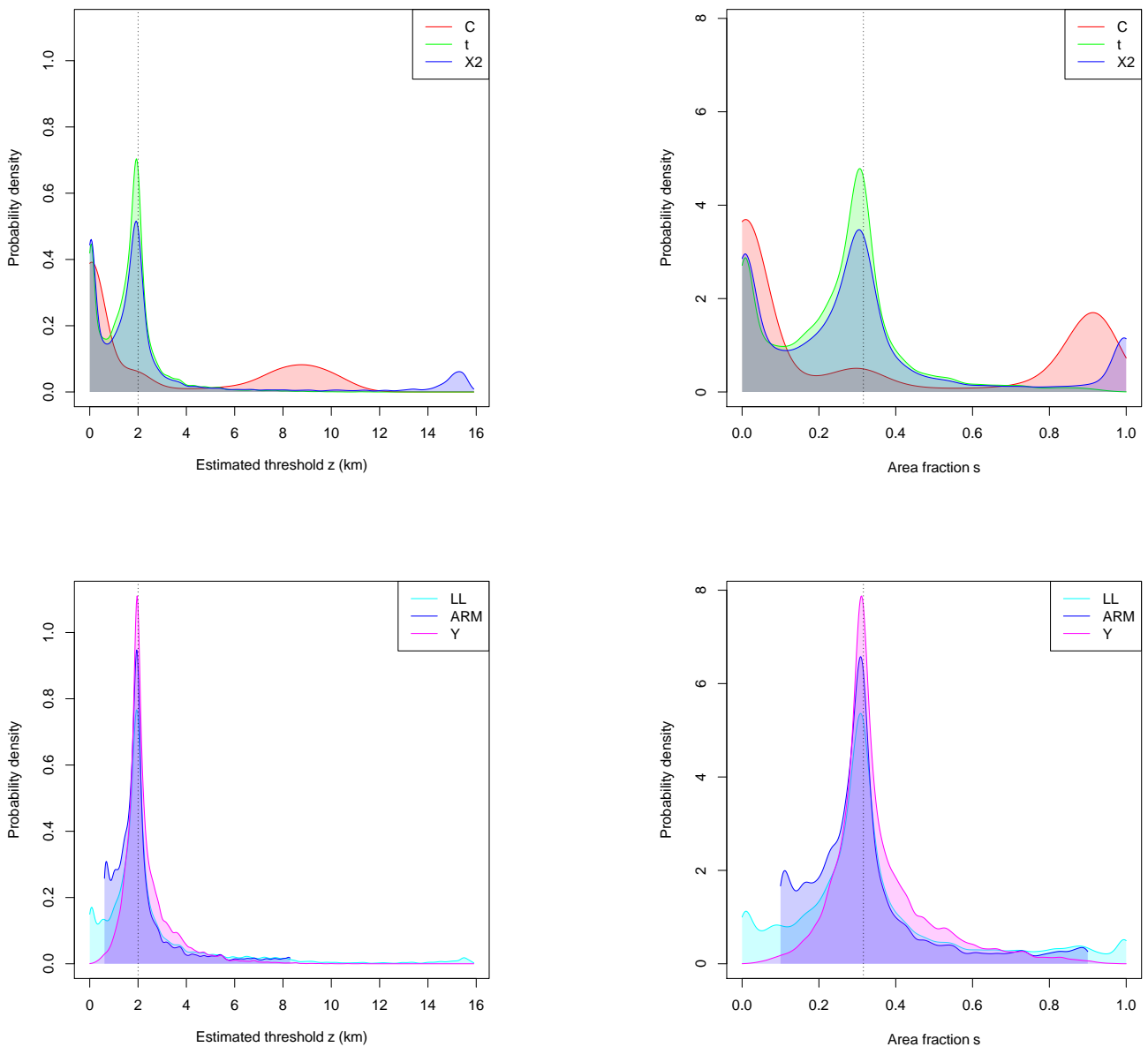


Fig. 15 Performance of methods for selecting the threshold in Experiment 1. Kernel density estimates of the estimated threshold (*Left column*) and the corresponding buffer area fraction (*Right column*). In the *top row*, thresholds were selected by methods C, t, and X2, which maximise the estimated contrast, Studentised contrast, and the χ^2 test statistic, respectively. In the *bottom row*, thresholds were selected using methods LL, ARM, and Y, which maximise the log-likelihood, the Akman-Raftery criterion with constraints, and the Youden criterion, respectively. Dotted vertical lines indicate the correct threshold, $z_{\text{true}} = 2$ km (left column), and corresponding area fraction, 0.32 (right column).

To avoid misinterpretation, we emphasise that each curve in the left column of Figure 15 is a smoothed histogram of the threshold values selected by a given method applied to the 10,000 simulations. The horizontal axis represents the threshold value that was selected; the vertical axis represents the relative frequency with which this value was selected. While all curves have a peak near the correct threshold value, indicating that the selected thresholds are often close to the correct one, many of the curves also have a steep incline at the left, indicating that they have a substantial chance of erroneously selecting a threshold close to zero.

In the top left panel of Figure 15, the curve for method C does not have a peak around 2 km, indicating that this method is unlikely to produce a threshold value \hat{z} close to the correct value $z_{\text{true}} = 2$. Instead the curve has a steep incline at the left, and a substantial bump around 9 km. This method very often produces a value close to 0 km, and frequently produces values around 9 km, which are both unacceptable. Areas under the density curve are probabilities; the area under the density bump around 9 km is 0.31, implying that *in about 30% of all outcomes*, method C would have selected a threshold between 6 and 12 km.

The selected threshold values \hat{z} were also converted to fractions of survey area, $\hat{a} = a(\hat{z})/a$. That is, \hat{a} is the fraction of survey area deemed to be prospective by the threshold selection procedure. As discussed previously, this transformation eliminates the effect of spatial structure, such as the spatial pattern of faults. Kernel estimates of probability densities of the area fractions \hat{a} are shown in the right column of Figure 15.

Table 3 shows objective performance measures for each method of threshold estimation. If the estimation error is $e = \hat{z} - z_{\text{true}}$, then “bias” is the mean value of e , “median deviation” is the median of e , “SD” (standard deviation) is the standard deviation of e , “MAD” (mean absolute deviation) is the mean of the absolute value of e , “RMSE” is the root-mean-square of e (the square root of the mean of e^2), and the last two rows indicate the fraction of outcomes where the absolute value of e exceeded 0.5 and 1, respectively. In statistical literature, overall performance is often evaluated using RMSE or MAD, while the other indices provide insight into the nature of the errors. Table 4 shows the corresponding performance measures on the scale of the area fractions.

PERFORMANCE MEASURES	METHODS					
	C	t	X2	ARM	LL	Y
BIAS	0.94	-0.39	0.74	0.13	0.58	0.34
MEDIAN DEVIATION	-1.82	-0.30	-0.27	-0.11	-0.06	0.03
SD	3.89	1.27	4.00	1.37	2.62	1.06
MAD	3.18	0.93	2.13	0.82	1.36	0.65
RMSE	4.01	1.33	4.07	2.68	2.68	1.11
$\mathbb{P}(\text{deviation} > 0.5)$	0.93	0.55	0.64	0.48	0.53	0.38
$\mathbb{P}(\text{deviation} > 1)$	0.90	0.38	0.50	0.26	0.36	0.19

Table 3 Performance of threshold estimation methods in Experiment 1, based on deviation between the estimated and true threshold distances. Boldface numerals indicate the best performance in each row.

PERFORMANCE MEASURES	METHODS					
	C	t	X2	ARM	LL	Y
BIAS	0.008	-0.070	-0.012	0.004	0.031	0.041
MEDIAN DEVIATION	-0.288	-0.044	-0.039	-0.017	-0.008	0.003
SD	0.399	0.171	0.286	0.164	0.226	0.129
MAD	0.361	0.136	0.205	0.108	0.151	0.086
RMSE	0.399	0.185	0.286	0.164	0.229	0.136
$\mathbb{P}(\text{deviation} > 0.05)$	0.940	0.630	0.700	0.571	0.610	0.486
$\mathbb{P}(\text{deviation} > 0.1)$	0.913	0.477	0.577	0.383	0.451	0.281

Table 4 Performance of threshold estimation methods in Experiment 1, based on deviation between the estimated and true area fractions of the prospective region. Boldface text indicates best performance in each row.

The most striking result is that method C, based on maximising the WofE contrast, performs very poorly. The methods t and X2, based respectively on the Studentised contrast and the χ^2 test statistic, perform much better than C, and frequently produce good estimates of the threshold. However, the steep inclines at the left of each panel in Figure 15 and the large values in the last two rows of Tables 3 and 4 indicate that t and X2 are also liable to produce a value close to zero km. In a single outcome (which a geologist would work with), these methods could well have chosen an unacceptably low value.

Three methods which perform quite well are the profile log-likelihood LL, the Akman-Raftery (constrained chi-squared) method ARM, and the Youden method Y. These methods are very likely to produce a threshold value close to the correct value.

The Akman-Raftery method ARM includes the restriction that the threshold should lie between the 10th and 90th percentiles of the predictor. The statistical performance of ARM is expected to be better than for X2, and this is borne out by the results in Tables 3 and 4 and Figure 15. The constraint has a substantial effect; ARM and X2 give different results in 37% of simulated realisations. The probability density plots in Figure 15 show the results for X2 in the top row, and for ARM in the bottom row; note the substantial bumps of probability mass near the endpoints for X2, and the truncation of the curves for ARM.

This simulation experiment is just one instance of an infinite universe of possible spatial patterns of faults. The results in Table 3 and in the left column of Figure 15 depend very specifically on the spatial pattern of faults in Figure 5. Fortunately, this effect is eliminated if we measure performance on the scale of the area fractions, so that the simulation results in Table 4 and in the right column of Figure 15 can be extrapolated to predict performance in a wide range of real situations.

Experiment 2

Since performance is expected to depend on sample size (i.e. number of deposits), our second experiment is a modification of the first experiment, in which the expected sample size is reduced by a factor of 4 (i.e. each of the density values is reduced by a factor of 4). There is now an expected total number of $54/4 = 13.5$ deposits; different simulated datasets will have different numbers of deposits, with an average of 13.5 deposits. The true contrast C_{true} and true threshold z_{true} are the same as in Experiment 1.

Figure 16 and Tables 5–6 show the corresponding results. The performance of all methods is degraded by the reduction in the amount of data, as we would expect. The threshold selection methods C and t, based respectively on the estimated contrast and the Studentised contrast, both perform substantially worse than they did in Experiment 1, with a much greater propensity to choose small threshold values. The Youden method Y has the best performance, closely followed by ARM, then the profile likelihood method, and then X2. We also found that ARM and X2 gave different results in 45% of simulations.

PERFORMANCE MEASURES	METHODS					
	C	t	X2	ARM	LL	Y
BIAS	0.04	-0.43	1.38	0.65	1.23	0.79
MEDIAN DEVIATION	-0.92	-0.75	-0.39	-0.13	-0.06	0.21
SD	2.28	1.58	4.38	2.10	3.56	1.76
MAD	1.92	1.30	2.89	1.50	2.26	1.28
RMSE	2.28	1.64	4.60	2.19	3.77	1.93
$\mathbb{P}(\text{deviation} > 0.5)$	0.88	0.75	0.82	0.71	0.75	0.63
$\mathbb{P}(\text{deviation} > 1)$	0.79	0.58	0.68	0.50	0.58	0.42

Table 5 Performance of threshold estimation methods in Experiment 2 (expected total number of deposits reduced from 54 to 13.5), based on deviation between the estimated and true threshold distances. Boldface text indicates best performance in each row.

PERFORMANCE MEASURES	METHODS					
	C	t	X2	ARM	LL	Y
BIAS	-0.031	-0.082	0.044	0.058	0.073	0.088
MEDIAN DEVIATION	-0.140	-0.113	-0.057	-0.018	-0.008	0.029
SD	0.293	0.211	0.342	0.241	0.292	0.203
MAD	0.264	0.188	0.278	0.187	0.232	0.162
RMSE	0.295	0.226	0.345	0.248	0.302	0.220
$\mathbb{P}(\text{deviation} > 0.05)$	0.910	0.810	0.861	0.784	0.811	0.721
$\mathbb{P}(\text{deviation} > 0.1)$	0.843	0.681	0.766	0.631	0.680	0.538

Table 6 Performance of threshold estimation methods in Experiment 2, based on deviation between the estimated and true area fractions of the prospective region.

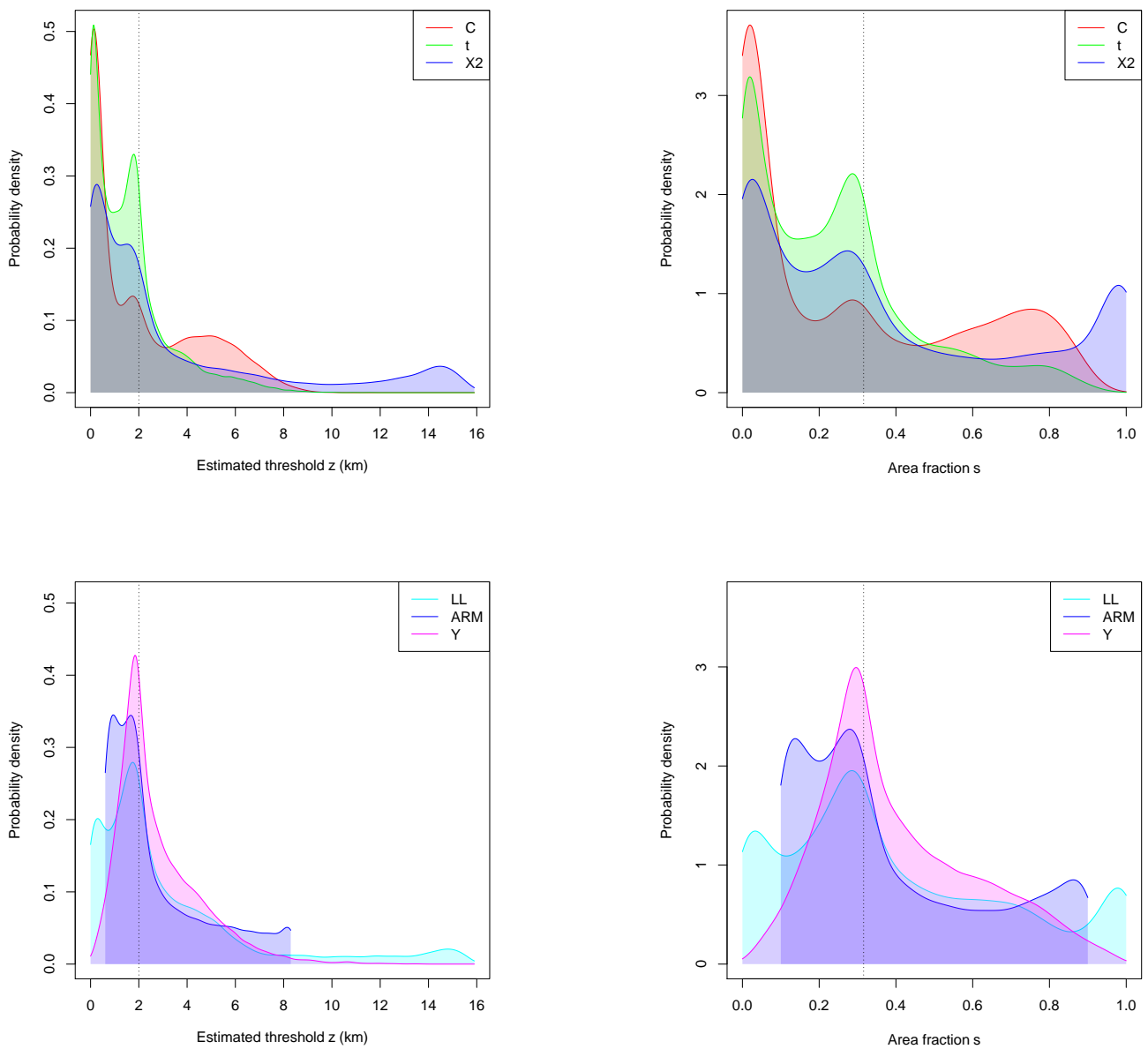


Fig. 16 Experiment 2 counterpart of Figure 15, that is, with the expected total number of deposits reduced from 54 to 13.5.

ANALYSIS OF MURCHISON GOLD DATA

Analysis Based on Full Dataset

Our analysis uses the distance-to-nearest-fault as a predictor. Figure 17 shows the cumulative number of gold deposits plotted against distance from the nearest fault (left panel) and plotted against buffer area (right panel). These are the distribution comparison plots analogous to those in Figure 9, but their behaviour is quite different from those in the synthetic example. The plots in Figure 17 show evidence of three or possibly four segments having distinct slopes, decreasing with increasing distance. Some reasons for such behaviour are discussed in the **DISCUSSION** section, under **What Happens if the Simple Threshold Model Does Not Hold?**.

Note that the left panel of Figure 17 has been truncated to a maximum distance of 15 km, which captures 95% of the deposits but only 40% of the survey area. This explains why the dashed cumulative curve in the left panel does not reach 100%.

The first outcome of the analysis is that buffer distances larger than 10 km are largely irrelevant for the purpose of selecting a threshold. This is consistent with the findings of Groves et al. (2000, Fig. 8a, p. 22) that over 95% of large deposits (with $> 1.5t$ Au production) in the Kalgoorlie Terrane are located within 5 km of regional-scale faults, and 95% of small deposits ($\leq 1.5t$ Au) within 10km of regional scale faults. Bierlein et al. (2006, Fig. 10b, p. 119) showed that gold endowment falls to a very low level beyond 5–10 km from major faults in the Yilgarn Craton. Bierlein et al. (2008, Fig. 11, p. 74) found that 90% of gold production occurs within 10 km from first, second and third order faults in the north to central Sierra Nevada Foothills (USA) Metamorphic Belt. Brown (2002) found that a statistically significant positive association between deposits and proximity to faults persisted up to 4 km distance in the Kalgoorlie Terrane. However, the geometry of the faults and shear zones in all of the abovementioned provinces shows that major faults are separated by 20 km at most, so that a distance-to-nearest-fault greater than 10 km is achieved only at locations which are remote from the fault system.

It is not clear whether thresholding is appropriate for the Murchison data. Witt et al. (2013, Appendix, Table A1.58) found a gradual decrease in prospectivity with distance from faults, for the Murchison Domain, using different data which are more complete than those shown in Figure 2. Analysis may need to consider the size or endowment of each deposit; whereas very large gold deposits are likely to have been formed in response to large crustal disturbances, very small deposits may behave like “noise” in the analysis (Knox-Robinson and Groves 1997). For example, the spatial resolution of the predictor data may not be fine enough to support useful prediction of very small deposits. Furthermore, spatial patterns of known gold deposits are influenced by the spatially-varying intensity of survey effort. For the purposes of this paper, we set aside such considerations, and simply demonstrate application of the thresholding method, noting when the method provides warnings about its own applicability.

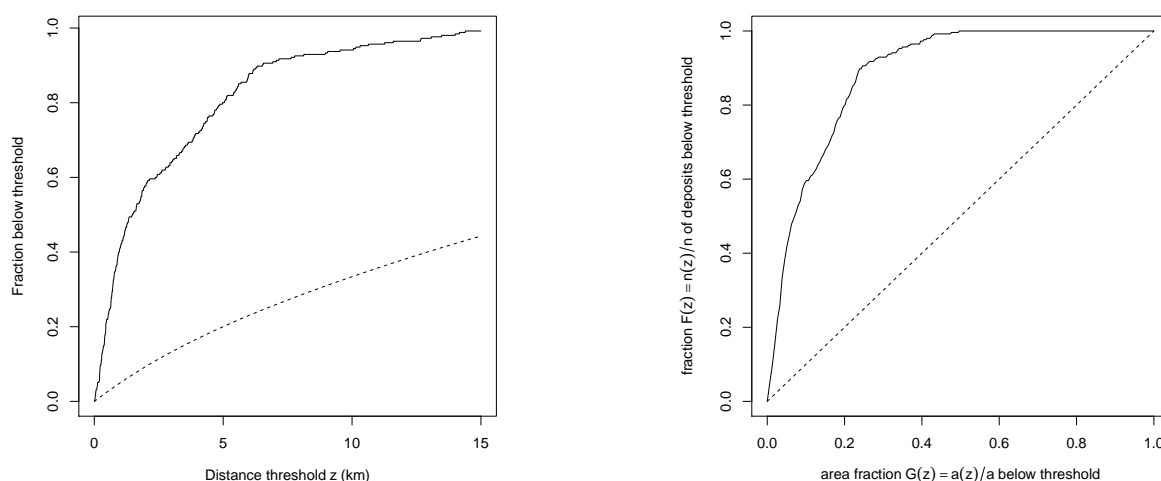


Fig. 17 Counterpart of Figure 9 for the Murchison gold survey data. *Left:* Cumulative fraction $n(z)/n$ of gold deposits (solid lines) and fraction $a(z)/a$ of survey area (dashed lines) lying within a given distance z of the nearest major fault, plotted against candidate threshold distance z . *Right:* Capture-efficiency curve; fraction of gold deposits against fraction of survey area, for different threshold distances.

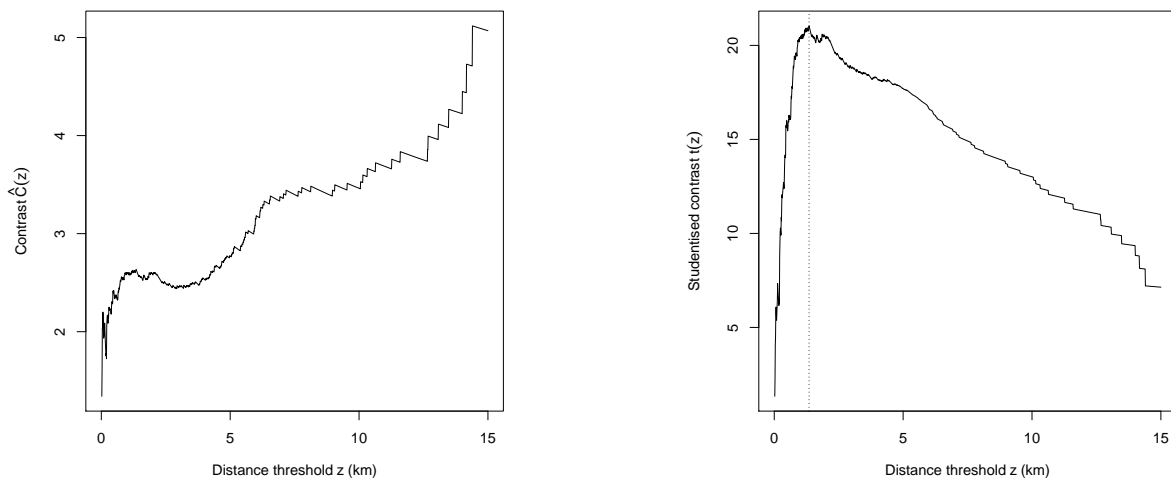


Fig. 18 Weights-of-evidence contrast (*Left*) and Studentised contrast (*Right*) plotted against candidate threshold distance z for the Murchison gold data. Vertical line in right panel shows estimated value of threshold, $\hat{z} = 1.35$ km.

818 In Figure 18 the WofE contrast (left panel) and the Studentised contrast (right panel) are plotted against the
 819 candidate threshold distance z for the Murchison data. The figure shows erratic behaviour for the WofE contrast,
 820 and reasonably stable behaviour with a clear maximum for the Studentised contrast. Again this indicates that the
 821 Studentised contrast is to be preferred over the WofE contrast. The estimated threshold using the Studentised
 822 contrast is $\hat{z} = 1.35$ km.

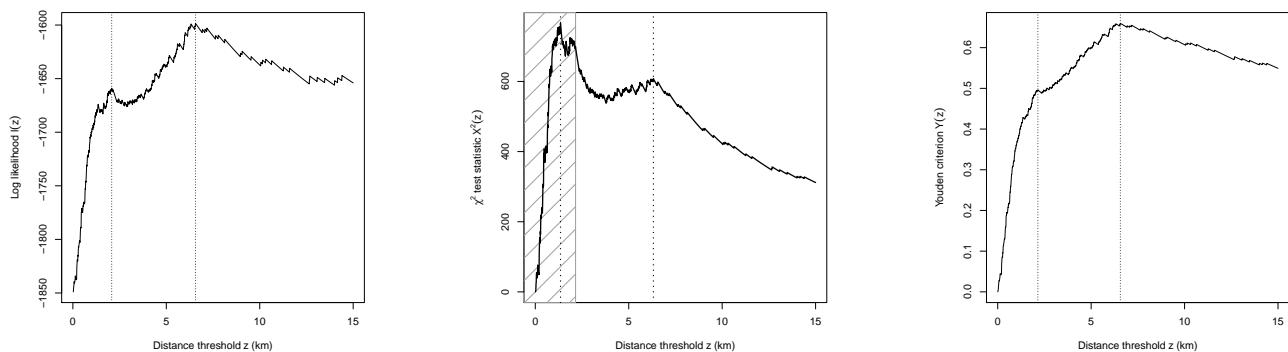


Fig. 19 Profile log-likelihood (*Left*), χ^2 criterion (*Middle*) and Youden criterion (*Right*) plotted against candidate threshold distance z for the Murchison gold data. Vertical lines show the estimated threshold and the position of a secondary peak. Shading (*Middle*) shows the region excluded from consideration in the Akman-Raftery method.

823 Figure 19 shows the profile log-likelihood (12), χ^2 statistic (6) and Youden criterion (8) plotted against the
 824 candidate threshold distance z , for the Murchison gold survey data. These graphs show evidence of two, or possibly
 825 three, local maxima for each criterion. The profile log-likelihood has a maximum at $\hat{z} = 6.56$ km and a secondary
 826 peak at $z = 2.08$ km. The χ^2 statistic has a maximum at $\hat{z} = 1.35$ km and a secondary peak at $z = 6.32$ km. For
 827 the Youden method the estimated threshold is $\hat{z} = 6.56$ km and there is a secondary maximum of curvature
 828 at about $z = 2.15$ km. Again, some reasons for such behaviour are considered in the **DISCUSSION** section, under
 829 **What Happens if the Simple Threshold Model Does Not Hold?**.

830 In the middle panel of Figure 19 the shading covers the region below the 10th percentile of the distance
 831 predictor; the 90th percentile lies beyond 15 km and is not shown. The main peak of the χ^2 statistic at $\hat{z} = 1.35$
 832 km lies below the 10th percentile of distance values (2.16 km). The Akman-Raftery method ARM does not select
 833 this main peak; instead it selects the boundary value $z = 2.16$ km which maximises the statistic subject to the
 834 constraint.

835 The upper limit (90th percentile, 68 km) in the Akman-Raftery method is not objectionable, since a distance
 836 of more than 20 km is not geologically plausible (in terms of the understanding of gold mineralization as previously
 837 discussed).

838 Table 7 summarises the results for the Murchison gold data, with \hat{z} indicating the selected distance threshold,
 839 and $s(\hat{z})$ the corresponding fraction of survey area.

METHOD	PRIMARY		SECONDARY	
	\hat{z}	$s(\hat{z})$	\hat{z}	$s(\hat{z})$
t	1.35	0.066		
X2	1.35	0.066	6.32	0.239
ARM	2.10	0.100	6.32	0.239
LL	6.56	0.246	2.08	0.097
Y	6.56	0.246	2.15	0.098

Table 7 Summary of estimates \hat{z} of distance threshold (in km), and corresponding fraction $s(\hat{z})$ of the survey area, using each method applied to the Murchison gold survey data. PRIMARY indicates the estimate obtained by maximising the criterion. SECONDARY is the estimate associated with a smaller peak, if present.

840 Analysis Based on Restricted Dataset

841 An important caveat about WofE, and most other mineral prospectivity techniques, is that the results depend
 842 on the survey region: different results can be obtained if the data are restricted to a smaller subset of the survey
 843 region.

844 The left panel of Figure 17 highlights the fact that there are parts of the survey region where the distance-to-
 845 nearest-fault is far greater than it is for any deposit. It might be reasonable to restrict attention to a subset of the
 846 survey region where geological considerations predict that there is a good chance of encountering a gold deposit.

847 One such calculation which is easy to perform is to restrict attention to a buffer around the fault system, of
 848 say 20 kilometres. The preceding formulae are altered only by restricting the distance values z to $0 \leq z \leq 20$,
 849 and replacing the area a of the full survey region by the area $a(20)$ of the buffer under consideration. Fractions of
 850 survey area $s(z)$ are adjusted by the factor $a/a(20) = 1.89$.

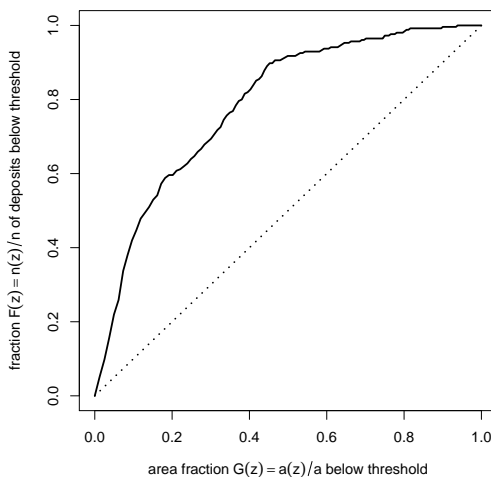


Fig. 20 Capture-efficiency curve for the Murchison gold data within a 20 km buffer of the fault system.

851 Figure 20 shows the capture-efficiency curve for the Murchison data in the 20 km buffer. Figure 21 shows the
 852 WofE contrast $\hat{C}(z)$ and the Studentised contrast $t(z)$ plotted against the candidate threshold distance value z .
 853 The WofE contrast value is maximised at unrealistically large distances and is not practicable. The Studentised
 854 contrast has a clear maximum at about $\hat{z} = 1.35$ km.

855 Figure 22 shows the profile log-likelihood and the Youden criterion plotted against z . These both select $\hat{z} = 6.32$
 856 km as the optimal threshold, with a secondary peak at $\hat{z} \approx 1.5$ km.

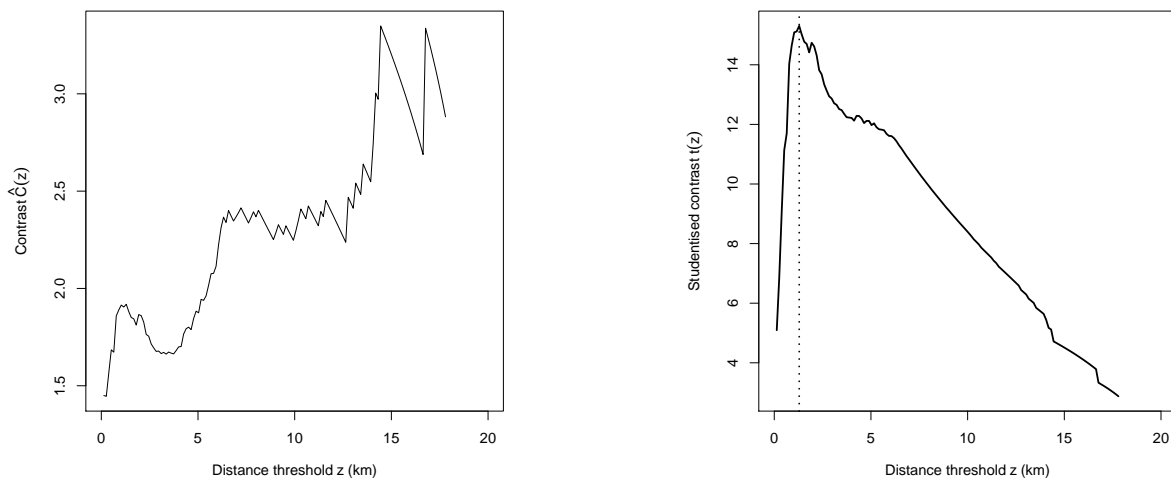


Fig. 21 Weights-of-evidence contrast (*Left*) and Studentised contrast (*Right*) plotted against candidate threshold distance z , for the Murchison gold data within a 20 km buffer of the fault system. Vertical line in the right panel shows estimated threshold $\hat{z} = 1.35$ km.

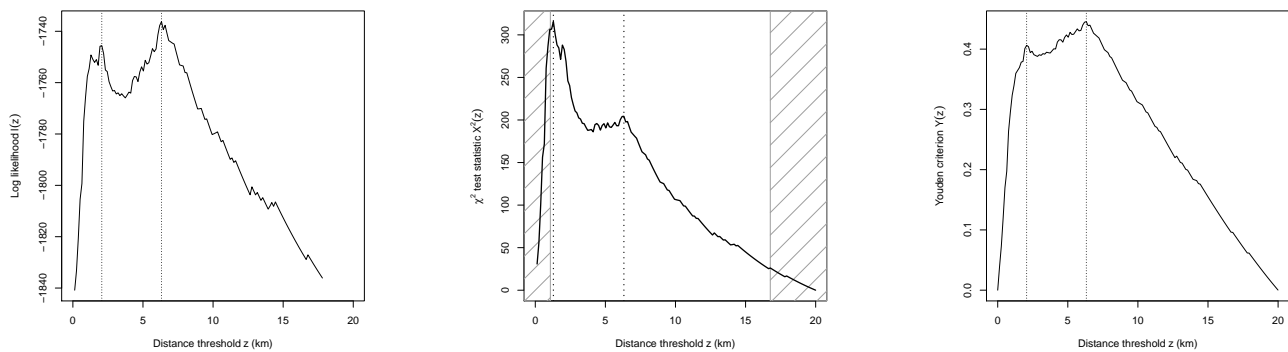


Fig. 22 Profile log-likelihood (*Left*), χ^2 statistic (*Middle*) and Youden criterion (*Right*) plotted against candidate threshold distance z , for the Murchison gold data within a 20 km buffer of the fault system. Vertical lines show the selected threshold and a secondary peak. Shading (*Middle*) shows the region excluded from consideration in the Akman-Raftery method.

METHOD	PRIMARY		SECONDARY	
	\hat{z}	$s(\hat{z})$	\hat{z}	$s(\hat{z})$
t	1.29	0.119		
X2	1.29	0.119	6.32	0.452
ARM	1.29	0.119	6.32	0.452
LL	6.32	0.452	2.06	0.182
Y	6.32	0.452	2.06	0.182

Table 8 Summary of estimates \hat{z} of distance threshold (in km), and corresponding fraction $s(\hat{z})$ of survey area, using each method applied to the Murchison gold survey data within a 20 km buffer of the fault system.

857 In this subset of the data, the 10th percentile of distance is 1.07 km, and the χ^2 and Akman-Raftery methods
858 are in agreement.

859 Estimated distance thresholds around 1 km seem reasonable for regional-scale faults, that is, faults with a strike
860 length of between 10 and 100 km. These distances fall within the range (0.8 to 2.4 km) obtained using a variety
861 of methods for crustal- and regional-scale faults in a study of orogenic gold deposits in the Kalgoorlie Terrane by
862 Brown (2002, pp. 248–249 and Table 6.4, p. 252).

863 Distances of 6 km are quite high for a regional-scale fault system, but would be consistent with the existence of
864 (a) crustal-scale shear zones (large, complex fault systems consisting of multiple slip surfaces with ductile as well
865 as brittle deformation); or (b) small unmapped local-scale secondary- and tertiary-scale faults associated with the
866 regional faults which control the location of gold deposits; or (c) existence of regional-scale faults which were not
867 detected in the data provided. Both (b) and (c) are plausible in the Yilgarn Craton which includes the Murchison
868 goldfields.

BOOTSTRAP STANDARD ERRORS AND CONFIDENCE INTERVALS

The uncertainty (lack of precision) of the estimated threshold is important in practice. Unfortunately, the familiar methods for determining standard errors and confidence intervals cannot be applied to the estimated thresholds calculated in this paper, because those methods are based on the Normal distribution, which is not appropriate here, as explained in the final paragraph of the subsection **Profile Likelihood for the Threshold Model** in the section **THRESHOLD SELECTION USING CHANGE-POINT ANALYSIS**.

Bootstrap methods (Efron and Tibshirani 1993) can be used to compute standard errors and confidence intervals for any target quantity. They are particularly useful when the estimator of the target quantity is not Normally distributed. In essence, the variability of estimates of the target quantity is captured by repeating the estimation procedure with many randomised versions of the original data.

In the “*nonparametric*” bootstrap, the original data are randomly resampled many times; the threshold estimation procedure is applied to these randomised data sets; and the spread of the resulting threshold estimates is used to calculate the standard error or confidence interval. In the “*parametric*” bootstrap, the simple threshold model is assumed to be true; simulated data are generated from the simple threshold model; the threshold estimation procedure is applied to these simulated data sets, and the resulting threshold estimates are used to calculate the standard error or confidence interval.

Nonparametric bootstrap

1. Record the predictor values $z_i = Z(x_i)$ at the deposit locations x_1, \dots, x_n .
2. Repeat the following procedure K times:
 - (a) Generate a random dataset $z_1^\dagger, \dots, z_n^\dagger$ of the same size as the original data, by “resampling” from the original data, that is by selecting a random sample of size n with replacement from amongst the values z_1, \dots, z_n .
 - (b) Apply the threshold estimator to the random dataset, yielding an estimate of the threshold, \hat{z} .
3. Collect the K threshold estimates $\hat{z}^{(1)}, \dots, \hat{z}^{(K)}$ computed by the previous step.
4. For the bootstrap standard error, compute the empirical standard deviation of the \hat{z} values,

$$\text{SE} = \frac{1}{K-1} \sum_{k=1}^K (\hat{z}^{(k)} - \bar{\hat{z}})^2$$

where $\bar{\hat{z}} = \frac{1}{K} \sum_{k=1}^K \hat{z}^{(k)}$ is the average estimated threshold. For a bootstrap 95% confidence interval, compute quantiles of the \hat{z} values as described below.

The number of repetitions K depends on computer resources, but should typically be at least 10000.

Since the distribution of estimates $\hat{z}^{(k)}$ is expected to be asymmetrical, we recommend computing an asymmetric confidence interval. The upper endpoint of this interval is computed as the 95th percentile of the bootstrap values $\hat{z}^{(k)}$ greater than the data estimate \hat{z} , and the lower endpoint is the 5th percentile of the values $\hat{z}^{(k)}$ smaller than the data estimate \hat{z} , yielding a 95% confidence interval.

Parametric bootstrap

1. Estimate the threshold z from the data.
2. Compute the corresponding area fraction $s = a(z)/a$ below the threshold and record the number $m = n(z)$ of deposits below the threshold.
3. Repeat the following procedure K times:
 - (a) Generate m random numbers distributed uniformly between 0 and s , and $n - m$ random numbers distributed uniformly between s and 1, yielding n random numbers $s_1^\dagger, \dots, s_n^\dagger$.
 - (b) Back-transform the area fractions s_i^\dagger to predictor values z_i^\dagger using the graph of $G(z)$ (for example the thin solid curve in the left panel of Figure 9). This yields a random dataset of predictor values $z_1^\dagger, \dots, z_n^\dagger$ at the deposits.
 - (c) Apply the threshold estimator to the random dataset $z_1^\dagger, \dots, z_n^\dagger$ yielding an estimate of the threshold, \hat{z} .
4. For the bootstrap standard error, compute the empirical standard deviation of the \hat{z} values. For a bootstrap CI, compute quantiles of the \hat{z} values, as described above for the nonparametric bootstrap.

The parametric bootstrap is typically more accurate, provided the assumed model (in this case the simple threshold model) is true. The nonparametric bootstrap is more robust against violation of the model assumptions. Furthermore, the standard errors will provide a good summary of variability only if the distribution of the \hat{z}

917 values is reasonably symmetric. A pragmatic approach is to inspect the capture-efficiency curve to assess whether
 918 it follows the inverted “V” shape expected for the simple threshold model, shown in Figure 10. If so, then the
 919 parametric bootstrap can be used; if not, then the nonparametric bootstrap should be used. After calculation, any
 920 glaring discrepancy between the parametric and nonparametric bootstrap results would suggest that the simple
 921 threshold model does not hold.

922 We applied the two bootstrap techniques to our example datasets, using 10,000 simulations for each calculation,
 923 and using the Youden criterion to select the threshold. Table 9 shows the estimates, bootstrap standard errors and
 924 confidence intervals.

DATASET	THRESHOLD (km)				AREA FRACTION (%)			
	Estimate \hat{z}	Bootstrap			Estimate $s(\hat{z})$	Bootstrap		
		Method	SE	CI		Method	SE	CI
synthetic	1.96	non	0.26	1.27–2.37	31.6	non	3.9	20.7–36.8
		par	0.23	1.58–2.70		par	3.2	25.6–41.3
Murchison	6.56	non	0.41	6.00–7.76	24.6	non	1.0	23.0–27.8
		par	0.07	6.41–6.77		par	0.2	24.2–25.2
Murchison (< 20 km)	6.32	non	1.15	2.06–6.61	45.2	non	7.0	18.2–46.8
		par	1.12	5.99–6.44		par	0.6	43.5–45.8

Table 9 Bootstrap standard errors and bootstrap 95% confidence intervals (*non* = nonparametric, *par* = parametric) calculated for the example datasets using the Youden method to estimate the threshold.

925 For the synthetic data in the right panel of Figure 5, there is close agreement between the results from the
 926 nonparametric and parametric bootstrap methods. This is expected because the synthetic data were created
 927 according to the simple threshold model.

928 For the Murchison gold survey data, the parametric bootstrap gives a smaller standard error and narrower
 929 confidence interval than the nonparametric bootstrap. Since the capture-efficiency curve suggests non-conformity
 930 to the simple threshold model, the parametric results may be discarded as over-optimistic. Similar comments apply
 931 to the restricted Murchison data.

932 DISCUSSION

933 What Happens if the Simple Threshold Model Does Not Hold?

934 The methods of threshold selection described in this paper all assume a “simple threshold model”, in which
 935 density of gold deposits is constant inside a certain region, and constant outside the region, as discussed in the
 936 section on **THRESHOLD SELECTION USING THE CAPTURE-EFFICIENCY CURVE**, subsection
 937 **Relation to Threshold**. Clearly, the Murchison data do not conform to this simple model (Baddeley 2018) and
 938 it is important to understand how the different techniques will respond.

939 Figure 23 sketches three different kinds of relationship between the predictor and prospectivity.⁹ The left column
 940 shows the simple threshold model; the middle column is an alternative “plateau” model in which prospectivity is
 941 highest when the predictor has a middling value; and the right column is a model in which prospectivity declines
 942 gradually as the predictor value increases. The top row depicts the relationship between the predictor value (on the
 943 horizontal axis) and the prospectivity (on the vertical axis). The bottom row is the corresponding capture-efficiency
 944 curve relating the cumulative fraction of deposits to the cumulative fraction of survey area.

945 The “plateau” model involves two threshold positions, and the prospectivity is highest when the predictor
 946 value falls between these thresholds. This might apply when the predictor is a mechanical property such as the
 947 rheological strength of a host rock. Rocks with intermediate strength may allow some brittle fracturing, facilitating
 948 focused fluid flow, which is favourable for the formation of orogenic gold deposits. However, very weak rocks with
 949 extensive fracturing may allow fluids to disperse without focused fluid flow, while very strong rocks may not
 950 fracture at all.

951 For predictors which represent angles, there cannot be a single threshold, for mathematical reasons. Two cuts
 952 are needed to divide a circle into pieces, so the simplest non-trivial model has two thresholds or breakpoints. The
 953 strike angles of faults and shear zones are examples of predictor variables where at least two thresholds may apply.
 954 Faults and shear zones within a particular range of orientations may be favorable for gold deposit formation (for

⁹For a given value z of the predictor, the prospectivity $\rho(z)$ is defined as the expected density of deposits per unit area amongst those grid cells with predictor value equal to z .

955 example because faults formed, or preferentially reactivated, in different epochs may be distinguished by their
 956 different orientations). Faults and shear zones that are favourably oriented to the regional-scale stress field may
 957 be reactivated to produce dilational zones, resulting in fluid influx and gold deposition (Groves et al. 2000). At
 958 a district scale (ca. 10 km), gold deposits are often located near *jogs* (deviations of $\sim 10^\circ$ to 20°) in the overall
 959 trend of crustal-scale faults (Groves et al. 2000).

960 Change-point problems involving several change-points can easily be handled using maximum likelihood, but the
 961 other methods (Youden, Akman-Raftery and Studentised contrast) do not easily adapt to this setting. Techniques
 962 for applying maximum likelihood with multiple change-points include nonparametric maximum likelihood (Sager
 963 1982) and CART (Breiman et al. 1984).

964 Multiple thresholds of a single predictor variable are discussed by J. Liu and Cheng (2019) in the context
 965 of hydrothermal Au deposits. Polykretis and Chalkias (2018, p. 258) deal with categorical predictors with many
 966 possible levels (e.g. rock type variable with many possible types) and propose merging the levels by grouping them
 967 according to the sign of the estimated contrast, before applying WofE.

968 Gradual Decline in Prospectivity

969 The right-hand column of Figure 23 sketches a gradual decline in prospectivity with increasing value of the
 970 predictor.

971 One could argue that, in this situation, the data should not be subjected to thresholding, and WofE should
 972 not be applied; instead, a more general version of logistic regression should be applied. Here we simply consider
 973 what would happen if thresholding were inappropriately applied to such data.

974 Assume that the density of deposits per unit area is a decreasing function $\rho(z)$ of the predictor value z . That
 975 is, in grid cells with predictor value z , the density of deposits is $\rho(z)$ deposits per unit area, where $\rho(z)$ decreases
 976 as z increases. With some algebra (given in the appendix **Derivation of Optimal Threshold for Gradual**
 977 **Decline in Prospectivity**) it can be shown that the optimal threshold for the Youden method is the threshold
 978 \hat{z} at which $\rho(\hat{z})$ is equal to the average density of deposits over the study region.

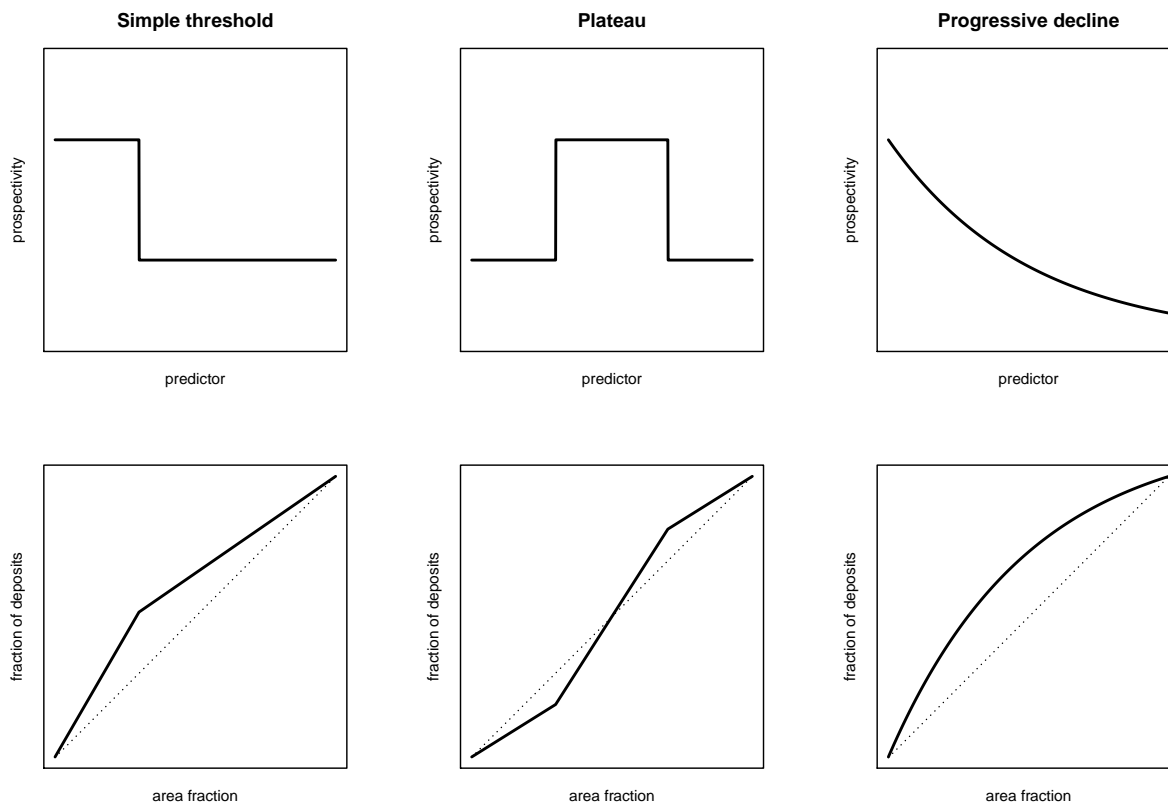


Fig. 23 Three different types of relationship between the predictor and prospectivity. *Top row:* prospectivity as a function of predictor value. *Bottom row:* capture-efficiency curve. *Left column:* simple threshold model. *Middle column:* plateau model. *Right column:* gradual decline in prospectivity.

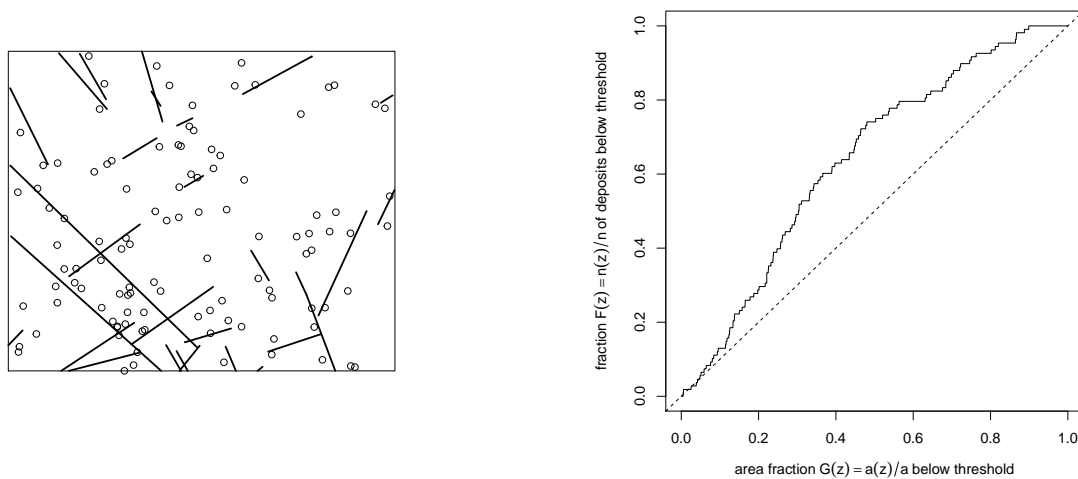


Fig. 24 Synthetic example exhibiting gradual decline in prospectivity with distance. *Left*: faults and simulated deposit locations. *Right*: capture-efficiency curve.

979 Figure 24 shows a synthetic example of gradually-declining prospectivity. It is based on the same spatial pattern
 980 of faults as the synthetic example in Figure 5, and the predictor variable Z is once again taken to be the distance
 981 to the fault pattern. However in this new example, the density of deposits declines exponentially with distance
 982 z , in the form $\rho(z) = be^{-cz}$, where b is the constant density that would give an average of 200 deposits throughout
 983 the study region, and $c = 1/5$. This is equivalent to a logistic regression model (Baddeley 2018). The left panel of
 984 Figure 24 shows the fault pattern and the simulated deposit locations. The right panel shows the capture-efficiency
 985 curve.

986 The left panel of Figure 25 shows the WofE contrast $\hat{C}(z)$ plotted against z . The threshold which maximises the
 987 contrast is 7.95 km. The contrast is generally increasing with the threshold value, which does not seem plausible,
 988 and the obtained threshold does not seem appropriate.

989 The right panel of Figure 25 shows the Studentised contrast $t(z)$ plotted against z . The threshold which
 990 maximises the Studentised contrast is 3.21 km.

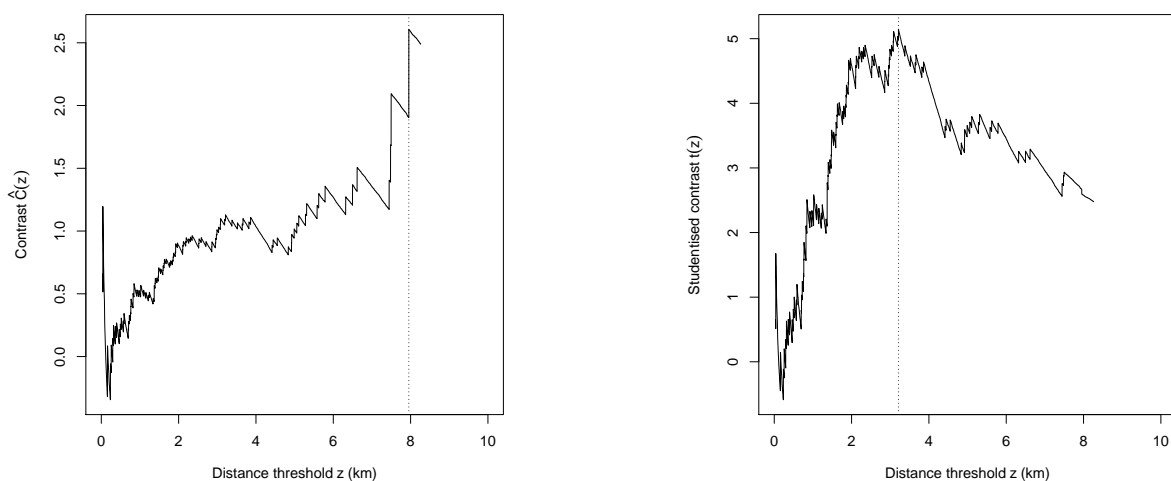


Fig. 25 Weights-of-evidence contrast (*Left*) and Studentised contrast (*Right*) for the synthetic example in Figure 24.

991 Figure 26 shows the profile log-likelihood (left panel), the Akman-Raftery criterion (middle panel) and the
 992 Youden criterion (right panel), plotted against z . The selected threshold is 3.21 km in all three cases.

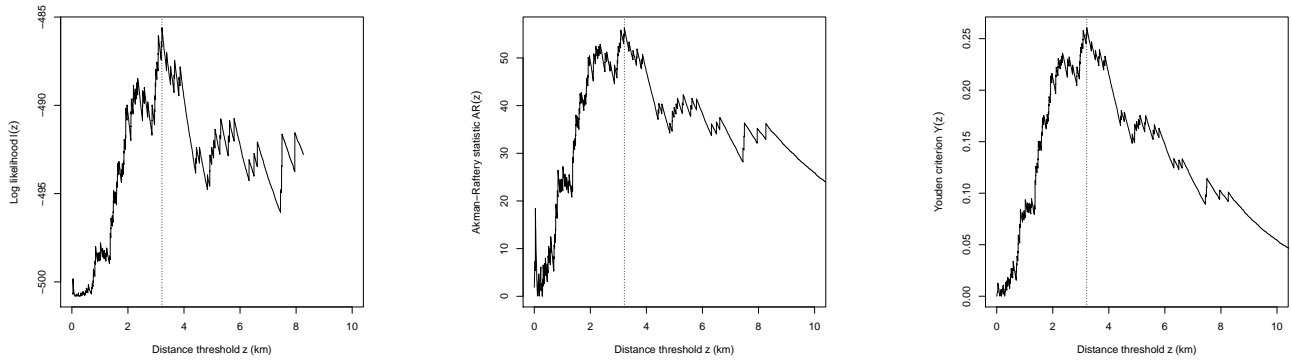


Fig. 26 Profile log-likelihood (*Left*), Akman-Raftery criterion (*Middle*) and Youden criterion (*Right*) for the synthetic example in Figure 24.

993 The simulated pattern has 108 points, giving an average density of $\hat{\mu} = 108/4102 = 0.0263$ deposits per square
 994 kilometre. Solving the equation $\rho(z) = \hat{\mu}$ gives a predicted optimal threshold of $z = 3.08$ km, reasonably close to
 995 the value obtained by the Youden method, the profile log-likelihood and the Akman-Raftery method.

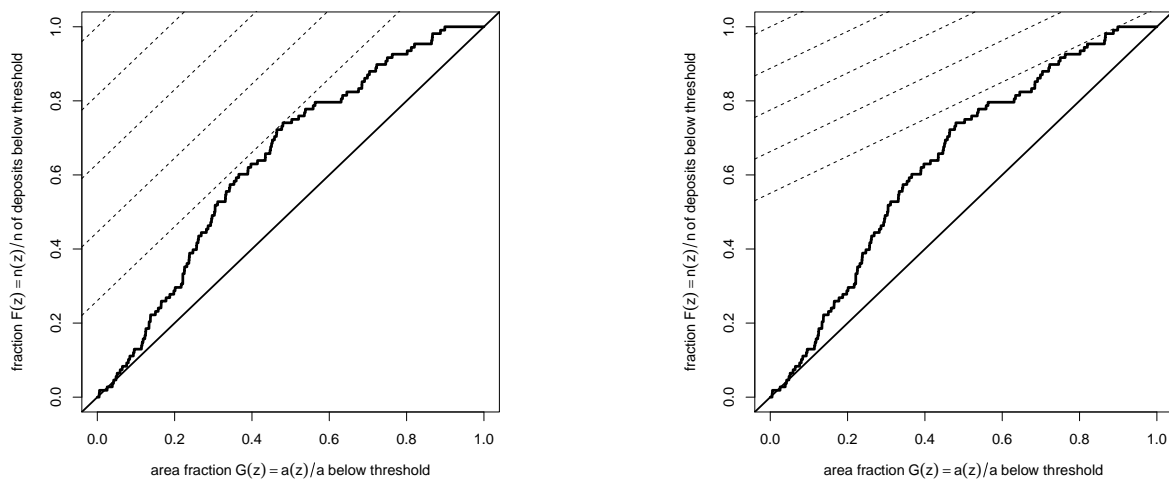


Fig. 27 Effect of cost-benefit ratio on selection of threshold for the synthetic example in Figure 24. Dashed lines have slope equal to cost-benefit ratio. *Left*: cost-benefit ratio equals 1. Selected threshold is 3.06 km. *Right* cost-benefit ratio equals 0.5. Selected threshold is 9.32 km.

996 When prospectivity declines gradually as in this example, the cost-benefit ratio (exploration cost to exploitation
 997 benefit) becomes more important. The left panel of Figure 27 shows the application of the graphical procedure
 998 for maximising the Youden criterion using parallel lines of slope 1. The right panel shows the same procedure
 999 using parallel lines of slope 1/2, (i.e. assuming that the value of 1% of mineral deposits is equal to the cost of
 1000 exploring 2% of survey area), which selects a different threshold of $z = 8.26$ km. This is quite different from the
 1001 behaviour observed in Figure 12. In this case the cost-benefit ratio does have a substantial influence on the choice
 1002 of threshold.

1003 Overall Conclusion

1004 Table 10 lists the threshold selection procedures studied in this paper, and summarises their performance. Ticks
 1005 and crosses represent good and bad performance, respectively. Selection Performance is performance of the threshold
 1006 selection procedure in estimating the threshold value. Reliability at Extremes is the reliability (inverse variance) of
 1007 the criterion near the ends of the range of possible values of the threshold. Significance Test indicates whether there
 1008 is a valid test of statistical significance for the presence of a threshold, directly linked to the criterion (validity

1009 means that the test does not violate the rule against cheating discussed in the subsection **The Problem of**
 1010 **“Peeking”** in the section on **SIGNIFICANCE TESTS**. Min number of deposits gives a rule of thumb estimate
 1011 of the minimum total number of deposits required for validity of the method. Robustness to failure of assumptions
 1012 indicates the ability of the technique to give acceptable results when the simple threshold model does not hold.

1013 The rule of thumb for the minimum total number of deposits assumes that the number of deposits inside the
 1014 prospective region is at least twice the number outside the prospective region. Then the minimum total number of
 1015 deposits is equal to three times the minimum number outside the prospective region, and the latter is determined
 1016 by statistical rules for the various test procedures.

METHOD	Selection Performance	Reliability at extremes	Significance test	Min number of deposits	Robustness to failure of assumptions	Adaptability to other models
C	XX	X		30	X	X
t	✓	X		30	✓	X
X2	✓	X		15	✓	✓
ARM	✓✓	X	A-R	15	✓	✓
LL	✓✓	✓		15	✓✓	✓✓✓
Y	✓✓✓	✓✓✓	K-S	5	✓✓	✓

Table 10 Summary of methods for threshold selection reviewed in this paper, and their performance. Method abbreviations: C (WofE contrast), t (Studentised contrast), X2 (χ^2 statistic), ARM (Akman-Raftery criterion with constraints), LL (profile log-likelihood), and Y (Youden criterion). For column abbreviations, see text.

1017 In our investigations, the method which selects the threshold value by maximising the WofE contrast \hat{C} had poor
 1018 performance, both on real data and in simulation experiments. The Studentised contrast t and the Akman-Raftery
 1019 criterion performed reasonably well, although they have erratic behaviour when the candidate threshold is close to
 1020 the smallest possible, or largest possible, threshold. The profile log-likelihood and Youden methods performed very
 1021 well, and the Youden method had the best overall performance. The Youden, Akman-Raftery and χ^2 methods are
 1022 related (see the section **MATHEMATICAL CONNECTIONS BETWEEN THE CRITERIA**) in such
 1023 a way that the Youden method will almost always have the best performance of the three. We recommend the
 1024 Youden method because of its performance, its simplicity and because it can be used immediately to conduct a
 1025 (Kolmogorov-Smirnov) test for existence of a threshold. Where the capture-efficiency curve indicates there may
 1026 be more than one threshold, the profile log-likelihood is the most suitable.

1027 WofE approaches have recently made the step from 2D to 3D modelling (Li et al. 2016; Payne et al. 2015;
 1028 Wang et al. 2015; Xiao et al. 2015; Yang et al. 2017). The methods proposed in this paper do not depend on the
 1029 spatial dimension, and could be applied to these 3D techniques.

1030 Caveats

1031 Our presentation has ignored numerous real-world complications, for lack of space.

1032 The handling of missing data is important in real applications and is a major research field in statistical science
 1033 (Alison 2002; Dempster et al. 1977; Harrell 2001; Hosmer and Lemeshow 2000; Little and Rubin 2002; Schafer
 1034 1997; Vach 1994; van Buuren 2012). It is not possible to do justice to this issue in the space available, but all
 1035 the methods discussed here can be modified to handle missing data in the same way that is traditionally done for
 1036 WofE.

1037 Sample bias, due to preferential sampling or non-uniform survey effort, affects the data analysis because it
 1038 affects the probability of detecting a deposit. The first two subsections above have discussed “misspecification”,
 1039 that is, the case where the real data do not obey the assumed model of a threshold relationship. One possible
 1040 cause for this is a sampling bias effect. If the sampling bias is known (for example if the amount of survey effort
 1041 in each region is known), then it can easily be accounted for in logistic regression (Baddeley 2018, p. 50, Baddeley
 1042 et al. 2015, p. 302, 304). Further discussion is postponed to another paper.

1043 For the two datasets considered in the paper, sample density bias is not a great concern. The synthetic dataset
 1044 has been constructed artificially to follow an ideal threshold relationship. The Murchison dataset is regional in
 1045 scale, and the source of the data is Geological Survey mapping at 1:250,000 scale, so that only major features are
 1046 shown. Faults and shear zones were mapped using airborne magnetic survey data so the entire study area is likely
 1047 to have been covered with equal detail.

1048 Our study does not apply to *unordered* categorical predictors, such as rock types which are not organised in
 1049 a sensible sequence. This is a different, and more challenging, problem because of the “peeking” fallacy (multiple

testing effect) explained in the subsection **The Problem of “Peeking”** in the section on **SIGNIFICANCE TESTS**.

Other complications include possible dependence between grid cells due to spatial clustering of deposits, the different endowments of different deposits, and the physical size of the deposits.

Acknowledgements

We thank Kassel Hingee for his insightful contributions to the initial research work which led to this article. In initial phases of this research, Adrian Baddeley was funded by the Australian Research Council under a Discovery Outstanding Researcher Award, and hosted by Professor Eun-Jung Holden of the Centre for Exploration Targeting (CET) at the University of Western Australia. Alope Phatak was partially supported by the Australian Government through the Australian Research Council’s Industrial Transformation Training Centres scheme (project IC180100030). We warmly thank the reviewers for their comments, which have greatly helped us to improve the paper.

Electronic Supplementary Material

The online version of this article contains supplementary material which is available to authorised users.

REFERENCES

- Agterberg, F. (1974). Automatic contouring of geological maps to detect target areas for mineral exploration. *Journal of the International Association for Mathematical Geology*, 6, 373–395.
- Agterberg, F. (1992). Combining indicator patterns in weights of evidence modeling for resource evaluation. *Nonrenewable Resources*, 1, 39–50.
- Agterberg, F. (2011). A modified weights-of-evidence method for regional mineral resource evaluation. *Natural Resources Research*, 20(2), 95–101.
- Agterberg, F. (2014). *Geomathematics: Theoretical Foundations, Applications, and Future Developments* (No. 18). Cham: Springer.
- Agterberg, F., & Bonham-Carter, G. (1999). Logistic regression and weights of evidence modeling in mineral exploration. In K. Dagdalen (Ed.), *Proceedings, 28th International Symposium on Computer Applications in the Mineral Industries – APCOM 99* (pp. 483–590). Golden, Colorado: Colorado School of Mines. (ISBN 0-918062-12-8)
- Agterberg, F., & Bonham-Carter, G. (2005). Measuring the performance of mineral-potential maps. *Natural Resources Research*, 14(1), 1–17.
- Agterberg, F., Bonham-Carter, G., Cheng, Q., & Wright, D. (1993). Weights of evidence modeling and weighted logistic regression for mineral potential mapping. In J. Davis & U. Herzfeld (Eds.), *Computers in Geology — 25 Years of Progress* (pp. 13–32). New York: Oxford University Press.
- Agterberg, F., & Cheng, Q. (2002). Conditional independence test for weights-of-evidence modeling. *Natural Resources Research*, 11, 249–255.
- Akman, V., & Raftery, A. (1986). Asymptotic inference for a change-point Poisson process. *Annals of Statistics*, 14(4), 1583–1590.
- Alison, P. (2002). *Missing Data*. Thousand Oaks, CA: Sage.
- Anonymous. (2011). “Significant”. Retrieved from <https://xkcd.com/882> (Web comic published on the XKCD website on 06 April 2011)
- Baddeley, A. (2018). A statistical commentary on mineral prospectivity analysis. In B. D. Sagar, Q. Cheng, & F. Agterberg (Eds.), *Handbook of Mathematical Geosciences: Fifty Years of IAMG* (pp. 25–65). International Association for Mathematical Geosciences.
- Baddeley, A., Berman, M., Fisher, N., Hardegen, A., Milne, R., Schuhmacher, D., Shah, R., & Turner, R. (2010). Spatial logistic regression and change-of-support for Poisson point processes. *Electronic Journal of Statistics*, 4, 1151–1201. doi: 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. *Statistics and its Interface*, 5, 221–236.
- Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. London: Chapman and Hall/CRC.
- Baddeley, A., & Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6), 1–42. (URL: www.jstatsoft.org, ISSN: 1548-7660)

- 1101 Ballantyne, C., & Cornish, R. (1979). Use of the chi-square test for the analysis of orientation data. *Journal of*
1102 *Sedimentary Research*, 49(3), 773–776.
- 1103 Barnard, G. (1959). Control charts and stochastic processes. *Journal of the Royal Statistical Society, Series B*,
1104 21, 239–271.
- 1105 Basseville, M., & Nikiforov, I. (1993). *Detection of Abrupt Changes: Theory and Applications*. Englewood Cliffs,
1106 NJ: Prentice-Hall.
- 1107 Berman, M. (1986). Testing for spatial association between a point process and another stochastic process. *Applied*
1108 *Statistics*, 35, 54–62.
- 1109 Bhattacharya, G., & Brockwell, P. (1976). The minimum of an additive process with applications to signal
1110 estimation and storage theory. *Zeitschrift fuer Wahrscheinlichkeitstheorie und verwandte Gebiete*, 37, 51–
1111 75.
- 1112 Bhattacharya, G., & Johnson, R. (1968). Nonparametric tests for shift at unknown time point. *Annals of*
1113 *Mathematical Statistics*, 39, 1731–1743.
- 1114 Bierlein, F., Murphy, F., Weinberg, R., & Lees, T. (2006). Distribution of orogenic gold deposits in relation to
1115 fault zones and gravity gradients: targeting tools applied to the Eastern Goldfields, Yilgarn Craton, Western
1116 Australia. *Mineralium Deposita*, 41, 107–126.
- 1117 Bierlein, F., Northover, H., Groves, D., Goldfarb, R., & Marsh, E. (2008). Controls on mineralisation in the
1118 Sierra Foothills gold province, central California, USA: a GIS-based reconnaissance prospectivity analysis.
1119 *Australian Journal of Earth Sciences*, 55, 61–78.
- 1120 Boleneus, D., Raines, G., Causey, J., Bookstrom, A., Frost, T., & Hyndman, P. (2001). *Assessment method for*
1121 *epithermal gold deposits in northeast Washington State using weights-of-evidence GIS modeling* (Open-File
1122 Report Nos. 2001–501). US Geological Survey.
- 1123 Bonham-Carter, G. (1995). *Geographic Information Systems for Geoscientists: Modelling with GIS* (No. 13).
1124 Kidlington, Oxford, UK: Pergamon Press/ Elsevier.
- 1125 Bonham-Carter, G., & Agterberg, F. (1990). Application of a microcomputer-based geographic information
1126 system to mineral-potential mapping. In J. T. Hanley & D. F. Merriam (Eds.), *Microcomputer Applications*
1127 *in Geology 2* (pp. 49–74). Amsterdam: Pergamon.
- 1128 Bonham-Carter, G., Agterberg, F., & Wright, D. (1990). Weights of evidence modelling: a new approach to
1129 mapping mineral potential. In F. Agterberg & G. Bonham-Carter (Eds.), *Statistical Applications in the*
1130 *Earth Sciences* (pp. 171–183). Ottawa: Geological Survey of Canada. (Proceedings of the Colloquium on
1131 Statistical Applications in the Earth Sciences hosted by the Geological Survey of Canada in Ottawa on 14–18
1132 November, 1988)
- 1133 Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). *Classification and Regression Trees*. Chapman and
1134 Hall/CRC.
- 1135 Brown, W. (2002). *Artificial Neural Networks: A New Method for Mineral-Prospectivity Mapping* (PhD thesis).
1136 University of Western Australia.
- 1137 Brown, W., Gedeon, T., Baddeley, A., & Groves, D. (2002). Bivariate J-function and other graphical statistical
1138 methods help select the best predictor variables as inputs for a neural network method of mineral prospec-
1139 tivity mapping. In U. Bayer, H. Burger, & W. Skala (Eds.), *IAMG 2002: 8th Annual Conference of the*
1140 *International Association for Mathematical Geology* (Vol. 1, pp. 257–268).
- 1141 Carlton, M., & Devore, J. (2014). *Probability with Applications in Engineering, Science, and Technology*. New
1142 York: Springer. doi: 10.1007/978-1-4939-0395-5
- 1143 Carranza, E. (2004). Weights of evidence modeling of mineral potential: a case study using small number of
1144 prospects, Abra, Philippines. *Natural Resources Research*, 13(3), 173–187.
- 1145 Carranza, E. (2009). Data-driven modeling of mineral prospectivity. In M. Hale (Ed.), *Handbook of Exploration*
1146 *and Environmental Geochemistry 11: Geochemical Anomaly and Mineral Prospectivity Mapping in GIS* (pp.
1147 249–310). Elsevier.
- 1148 Cassard, D., Billa, M., Lambert, A., Picot, J., Husson, Y., & Lassere, J. (2008). Gold predictivity mapping in
1149 French Guiana using an expert-guided data-driven approach based on a regional-scale GIS. *Ore Geology*
1150 *Reviews*, 34(3), 471–500.
- 1151 Cervi, F., Berti, M., Borgatti, L., Ronchetti, F., Manenti, F., & Corsini, A. (2010). Comparing predictive capability
1152 of statistical and deterministic methods for landslide susceptibility mapping: a case study in the northern
1153 Apennines (Reggio Emilia Province, Italy). *Landslides*, 7(4), 433–444.
- 1154 Chen, Y., & Wu, W. (2019, January). Isolation forest as an alternative data-driven mineral prospectivity mapping
1155 method with a higher data-processing efficiency. *Natural Resources Research*, 28(1), 31–46.
- 1156 Cheng, Q. (2004). Application of weights of evidence method for assessment of flowing wells in the greater Toronto
1157 area, Canada. *Natural Resources Research*, 13, 77–86.

- 1158 Cheng, Q. (2007). Mapping singularities with stream sediment geochemical data for prediction of undiscovered
1159 mineral deposits in Gejiu, Yunnan Province, China. *Ore Geology Reviews*, *32*, 314–324.
- 1160 Cheng, Q. (2008). Non-linear theory and power-law models for information integration and mineral resources
1161 quantitative assessments. *Mathematical Geosciences*, *40*(5), 503–532.
- 1162 Chernoff, H., & Rubin, H. (1956). The estimation of the location of a discontinuity in density. In *Proceedings*,
1163 *Third Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 19–37).
- 1164 Chernoff, H., & Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to
1165 change in time. *Annals of Mathematical Statistics*, *35*, 999–1018.
- 1166 Chernoyarov, O., Kutoyants, Y., & Top, A. (2018). On multiple change-point estimation for Poisson process. *Com-*
1167 *munications in Statistics – Theory and Methods*, *47*(5), 1215–1233. doi: 10.1080/03610926.2017.1317810
- 1168 Commenges, D., & Seal, J. (1985). The analysis of neuronal discharge sequences: change-point estimation and
1169 comparison of variances. *Statistics in Medicine*, *4*, 91–104.
- 1170 Conover, W. (1999). *Practical Nonparametric Statistics* (Third ed.). New York: John Wiley and Sons.
- 1171 Darkhovsky, B. (1976). A nonparametric method for the a posteriori detection of the “disorder” time of a sequence
1172 of independent random variables. *Theory of Probability and its Applications*, *21*, 178–183.
- 1173 Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the E–M
1174 algorithm. *Journal of the Royal Statistical Society B*, *39*, 1–22.
- 1175 Deshayes, J. (1984). Ruptures de modèles pour les processus de Poisson. *Annales Scientifiques Univ Clermont-*
1176 *Ferrand II*, *78*, 1–7.
- 1177 Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap* (Vol. 57). London: Chapman and Hall.
- 1178 Fabbri, A., & Chung, C.-J. (2008). On blind tests and spatial prediction models. *Natural Resources Research*,
1179 *17*(2), 107–118.
- 1180 Filzmosera, P., Garrett, R., & Reimann, C. (2005). Multivariate outlier detection in exploration geochemistry.
1181 *Computers and Geosciences*, *31*, 579–587.
- 1182 Fisher, R. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the*
1183 *Royal Society, Series A*, *222*(594–604), 309–368.
- 1184 Ford, A., Miller, J., & Mol, A. (2016). A comparative analysis of Weights of Evidence, evidential belief functions,
1185 and fuzzy logic for mineral potential mapping using incomplete data at the scale of investigation. *Natural*
1186 *Resources Research*, *25*, 19–33.
- 1187 Foxall, R., & Baddeley, A. (2002). Nonparametric measures of association between a spatial point process and a
1188 random set, with geological applications. *Applied Statistics*, *51*(2), 165–182.
- 1189 Galun, S., & Trifonov, A. (1982). Detection and estimation of the time when the Poisson flow intensity changes.
1190 *Automation and Remote Control*, *43*(6), 782–790.
- 1191 Gardner, L. (1969). On detecting changes in the mean of normal variables. *Annals of Mathematical Statistics*,
1192 *40*, 116–126.
- 1193 Garrett, R. (1989). The chi-square plot: a tool for multivariate outlier recognition. *Journal of Geochemical*
1194 *Exploration*, *32*(1), 319–341.
- 1195 Geological Survey of Western Australia. (1994). *MINEDEX database*.
1196 (<https://dmp.wa.gov.au/Mines-and-mineral-deposits-1502.aspx>)
- 1197 Ghannadpour, S., & Hezarkhani, A. (2016). Exploration geochemistry data-application for anomaly separation
1198 based on discriminant function analysis in the Parkam porphyry system. *Geosciences Journal*, *20*(6), 837–
1199 850.
- 1200 Goldfarb, R., & Groves, D. (2015). Orogenic gold: Common or evolving fluid and metal sources through time.
1201 *Lithos*, *233*, 2–26.
- 1202 Goodacre, A., Bonham-Carter, G., Agterberg, F., & Wright, D. (1993). A statistical analysis of the spatial
1203 association of seismicity with drainage patterns and magnetic anomalies in western Quebec. *Tectonophysics*,
1204 *217*, 285–305.
- 1205 Gorney, R., Ferris, D., Ward, A., & Williams, L. (2011). Assessing channel-forming characteristics of an impacted
1206 headwater stream in Ohio, USA. *Ecological Engineering*, *37*(3), 418–430.
- 1207 Groves, D., Goldfarb, R., Knox-Robinson, C., Ojala, J., Gardoll, S., Yun, G., & Holyland, P. (2000). Late-kinematic
1208 timing of orogenic gold deposits and significance for computer-based exploration techniques with emphasis
1209 on the Yilgarn Block, Western Australia. *Ore Geology Reviews*, *17*, 1–38.
- 1210 Groves, D., & Santosh, M. (2016). The giant Jiaodong gold province: The key to a unified model for orogenic gold
1211 deposits? *Geoscience Frontiers*, *7*, 409–417.
- 1212 Hájek, J., & Rényi, A. (1955). Generalization of an inequality of Kolmogorov. *Acta Math. Acad. Sci. Hungar.*, *6*,
1213 281–283.
- 1214 Harrell, F. (2001). *Regression Modeling Strategies*. New York: Springer.

- 1215 Harris, D., & Pan, G. (1999). Mineral favourability mapping: a comparison of artificial neural networks, logistic
1216 regression, and discriminant analysis. *Natural Resources Research*, *8*, 93–109.
- 1217 Harris, D., Zurcher, L., Stanley, M., Marlow, J., & Pan, G. (2003). Comparative analysis of favorability mappings
1218 by weights of evidence, probabilistic neural networks, discriminant analysis, and logistic regression. *Natural*
1219 *Resources Research*, *12*(4), 241–255.
- 1220 Harris, J., Grunsky, E., Behnia, P., & Corrigan, D. (2015). Data- and knowledge-driven mineral prospectivity
1221 maps for Canada’s North. *Ore Geology Reviews*, *71*, 788–803.
- 1222 Hinkley, D. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, *57*, 1–17.
- 1223 Hinkley, D. (1971). Inference about the change-point from the cumulative sum test. *Biometrika*, *58*, 509–523.
- 1224 Hochberg, Y., & Tamhane, A. (1987). *Multiple Comparison Procedures*. New York: John Wiley and Sons.
- 1225 Hogg, R., & Craig, A. (1970). *Introduction to Mathematical Statistics* (Third ed.). Macmillan.
- 1226 Hosmer, D., & Lemeshow, S. (2000). *Applied Logistic Regression* (Second ed.). John Wiley and Sons.
- 1227 Hsu, J. (1996). *Multiple Comparisons: Theory and Methods*. Chapman and Hall.
- 1228 Kalbfleisch, J. (1985). *Probability and Statistical Inference. Volume 2: Statistical Inference* (Second ed.). New
1229 York: Springer.
- 1230 Kander, Z., & Zacks, S. (1966). Test procedures for possible changes in parameters of statistical distributions
1231 occurring at unknown time points. *Annals of Mathematical Statistics*, *37*, 1196–1210.
- 1232 Kendall, M. G., & Stuart, A. (1973). *The Advanced Theory of Statistics* (Third ed., Vol. 2). London: Charles
1233 Griffin and Company Ltd.
- 1234 Knox-Robinson, C., & Groves, D. (1997). Gold prospectivity mapping using a geographic information system
1235 (GIS), with examples from the Yilgarn Block of Western Australia. *Chronique de la Recherche Minière*,
1236 *529*, 127–138.
- 1237 Krzanowski, W., & Hand, D. (2009). *ROC Curves for Continuous Data*. London/Boca Raton: Chapman and
1238 Hall/CRC Press.
- 1239 Kutoyants, Y. (1998). *Statistical Inference for Spatial Poisson Processes* (No. 134). New York: Springer.
- 1240 Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. New York: Springer-Verlag.
- 1241 Leonard, T. (1978). Density estimation, stochastic processes and prior information (with discussion). *Journal of*
1242 *the Royal Statistical Society, Series B*, *40*, 113–146.
- 1243 Li, N., Bagas, L., Li, X., Xiao, K., Li, Y., Ying, L., & Song, X. (2016). An improved buffer analysis technique for
1244 model-based 3D mineral potential mapping and its application. *Ore Geology Reviews*, *76*, 94–107.
- 1245 Lindsey, J. (1996). *Parametric Statistical Inference*. Oxford: Clarendon Press.
- 1246 Little, R., & Rubin, D. (2002). *Statistical Analysis with Missing Data* (Second ed.). John Wiley and Sons.
- 1247 Liu, J., & Cheng, Q. (2019). A modified Weights-of-Evidence method for mineral potential prediction based on
1248 structural equation modeling. *Natural Resources Research*, *28*, 1037–1053.
- 1249 Liu, Y., Cheng, Q., Xia, Q., & Wang, X. (2014). Mineral potential mapping for tungsten polymetallic deposits in
1250 the Nanling metallogenic belt, South China. *Journal of Earth Science*, *25*, 689–700.
- 1251 Loader, C. (1992). A log-linear model for a Poisson process changepoint. *Annals of Statistics*, *20*, 1391–1411.
- 1252 Murphy, S., & van der Vaart, A. (2000). On profile likelihood. *Journal of the American Statistical Association*,
1253 *95*(450), 449–465.
- 1254 Nam, B.-H., & D’Agostino, R. (2002). Discrimination index, the area under the ROC curve. In C. Huber-Carol,
1255 N. Balakrishnan, M. Nikulin, & M. Mesbah (Eds.), *Goodness-of-Fit Tests and Model Validity* (pp. 267–279).
1256 Basel: Birkhäuser.
- 1257 Neuhäuser, B., & Terhorst, B. (2007). Landslide susceptibility assessment using “weights-of-evidence” applied to
1258 study area at the Jurassic escarpment (SW-Germany). *Geomorphology*, *86*(1–2), 12–24.
- 1259 Page, E. (1954). Continuous inspection schemes. *Biometrika*, *41*, 100–115.
- 1260 Page, E. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, *44*,
1261 248–252.
- 1262 Payne, C., Cunningham, F., Peters, K., Nielsen, S., Puccioni, E., Wildman, C., & Partington, G. (2015). From
1263 2D to 3D: Prospectivity modelling in the Taupo Volcanic Zone, New Zealand. *Ore Geology Reviews*, *71*,
1264 558–577.
- 1265 Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a
1266 correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling.
1267 *Philosophical Magazine*, *50*(302), 157–175. (Series 5)
- 1268 Pflug, G. (1983). The limiting log-likelihood process for discontinuous density families. *Zeitschrift fuer Wah-*
1269 *scheinlichkeitstheorie und verwandte Gebiete*, *64*, 15–35.
- 1270 Polykretis, C., & Chalkias, C. (2018). Comparison and evaluation of landslide susceptibility maps obtained from
1271 weights of evidence, logistic regression, and artificial neural network models. *Natural Resources Research*,

- 1272 93, 249–274.
- 1273 Pons, O. (2018). *Estimations and Tests in Change-Point Models*. Singapore: World Scientific.
- 1274 Porwal, A., Gonzalez-Alvarez, I., Markwitz, V., McCuaig, T., & Mamuse, A. (2010). Weights-of-evidence and
1275 logistic regression modeling of magmatic nickel sulfide propectivity in the Yilgarn Craton, Western Australia.
1276 *Ore Geology Reviews*, 38(3), 184–196.
- 1277 Pratt, J. (1959). On a general concept of “In Probability”. *Annals of Mathematical Statistics*, 30, 549–558.
- 1278 R Development Core Team. (2018). R: A language and environment for statistical computing [Computer software
1279 manual]. Vienna, Austria. Retrieved from <http://www.R-project.org/> (ISBN 3-900051-07-0)
- 1280 Raftery, A., & Akman, V. (1986). Bayesian analysis of a Poisson process with a change-point. *Biometrika*, 73(1),
1281 85–89.
- 1282 Read, T., & Cressie, N. (1988). *Goodness-of-Fit Statistics for Multivariate Data*. New York: Springer-Verlag.
- 1283 Rice, J. (2006). *Mathematical Statistics and Data Analysis* (Third ed.). New York: Duxbury.
- 1284 Robert, F., Poulson, K., Cassidy, K., & Hodgson, C. (2005). Gold metallogeny of the Superior and Yilgarn Cratons.
1285 In J. Hedenquist, J. Thompson, R. Goldfarb, & J. Richards (Eds.), *Economic Geology One Hundredth*
1286 *Anniversary Volume* (pp. 1001–1033). Littleton, Colorado, USA: Society of Economic Geologists. (ISBN
1287 978-1-887483-01-8)
- 1288 Romero-Calcerrada, R., Barrio-Parra, F., Millington, J., & Novillo, C. (2010). Spatial modeling of socioeconomic
1289 data to understand patterns of human-caused wildfire ignition risk in the SW of Madrid (central Spain).
1290 *Ecological Modelling*, 221(1), 34–45.
- 1291 Romero-Calcerrada, R., & Luque, S. (2006). Habitat quality assessment using Weights-of-Evidence based GIS
1292 modelling: The case of *Picoides tridactylus* as species indicator of the biodiversity value of the Finnish forest.
1293 *Ecological Modelling*, 196(1–2), 62–76.
- 1294 Rubin, H. (1961). The estimation of discontinuities in multivariate densities, and related problems in stochastic
1295 process. In *Proceedings, Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp.
1296 563–574). University of California Press.
- 1297 Ruopp, M., Perkins, N., Whitcomb, B., & Schisterman, E. (2008, June). Youden index and optimal cut-point
1298 estimated from observations affected by a lower limit of detection. *Biometrical journal*, 50(3), 419–430.
- 1299 Sager, T. (1982). Nonparametric maximum likelihood estimation of spatial patterns. *Annals of Statistics*, 10,
1300 1125–1136.
- 1301 Schaeben, H. (2014). Targeting: logistic regression, special cases and extensions. *ISPRS International Journal of*
1302 *Geo-Information*, 3, 1387–1411.
- 1303 Schaeben, H., & Semmler, G. (2016). The quest for conditional independence in prospectivity modeling: weights-
1304 of-evidence, boost weights-of-evidence, and logistic regression. *Frontiers of Earth Science*, 10(3), 389–408.
- 1305 Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.
- 1306 Sen, A., & Srivastava, M. (1975). On tests for detecting change in mean. *Annals of Statistics*, 3, 98–108.
- 1307 Severini, T. (2000). *Likelihood Methods in Statistics*. Oxford: Oxford University Press.
- 1308 Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, 46, 561 - 584.
- 1309 Shewhart, W. (1983). *Economic Control of Quality of Manufactured Product*. Princeton, NJ: Van Nostrand
1310 Reinhold.
- 1311 Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- 1312 Smith, A. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables.
1313 *Biometrika*, 62, 407–416.
- 1314 Solomon, M., & Groves, D. (1994). *The Geology and Origin of Australia’s Mineral Deposits*. New York: Oxford
1315 University Press.
- 1316 Stephens, M. (1986). Tests based on EDF statistics. In R. D’Agostino & M. Stephens (Eds.), *Goodness-of-Fit*
1317 *Techniques* (Vol. 68, pp. 97–193). New York: Marcel Dekker.
- 1318 Vach, W. (1994). *Logistic Regression with Missing Values in the Covariates*. Berlin: Springer-Verlag.
- 1319 van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Boca Raton: Chapman and Hall.
- 1320 Wand, M., & Jones, M. (1995). *Kernel Smoothing*. Chapman and Hall.
- 1321 Wang, G., Du, W., & Carranza, E. (2016). Remote sensing and GIS prospectivity mapping for magmatic-
1322 hydrothermal base- and precious-metal deposits in the Honghai district, China. *Journal of African Earth*
1323 *Sciences*, 128, 97–115.
- 1324 Wang, G., Li, R., Carranza, E., Zhang, S., Yan, C., Zhu, Y., Qu, J., Hong, D., Song, Y., Han, J., Ma, Z., Zhang,
1325 H., & Yang, F. (2015). 3D geological modeling for prediction of subsurface Mo targets in the Luanchuan
1326 district, China. *Ore Geology Reviews*, 71, 592–610.
- 1327 Wasserman, L. (2004). *All of Statistics: a Concise Course in Statistical Inference*. New York: Springer.
- 1328 Watkins, K., & Hickman, A. (1990). *Geological evolution and mineralization of the Murchison Province, Western*

- 1329 *Australia* (Bulletin No. 137). Geological Survey of Western Australia. (Published by Department of Mines,
1330 Western Australia, 1990. Available online from Department of Industry and Resources, State Government
1331 of Western Australia, www.doir.wa.gov.au)
- 1332 West, W., & Ogden, T. (1997). Continuous-time estimation of a changepoint in a Poisson process. *Journal of*
1333 *Statistical Computation and Simulation*, 56(4), 293–302.
- 1334 Wilk, M., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55,
1335 1–17.
- 1336 Witt, W., Ford, A., Hanrahan, B., & Mamuse, A. (2013). *Regional-scale targeting for gold in the Yilgarn Cra-*
1337 *ton: Part 1 of the Yilgarn Gold Exploration Targeting Atlas* (Report No. 125). Perth, Western Australia:
1338 Geological Survey of Western Australia.
- 1339 Xiao, K., Li, N., Porwal, A., Holden, E., Bagas, L., & Lu, Y. (2015). GIS-based 3D prospectivity mapping: A case
1340 study of Jiama copper-polymetallic deposit in Tibet, China. *Ore Geology Reviews*, 71, 611–632.
- 1341 Yang, F., Wang, G., Santosh, M., Li, R., Tang, L., Cao, H., Guo, N., & Liu, C. (2017). Delineation of potential
1342 exploration targets based on 3D geological modeling: A case study from the Laoangou Pb-Zn-Ag polymetallic
1343 ore deposit, China. *Ore Geology Reviews*, 89, 228–252.
- 1344 Yeomans, C. (2018). *Enhancing the Geological Understanding of Southwest England using Machine Learning*
1345 *Algorithms* (Unpublished doctoral dissertation). Camborne School of Mines.
- 1346 Youden, W. (1950). Index for rating diagnostic tests. *Cancer*, 3, 32–35.
- 1347 Zacks, S. (1983). Survey of classical and Bayesian approaches to the change-point problem: Fixed sample and
1348 sequential procedures of testing and estimation. In M. Rizvi, J. Rustagi, & D. Siegmund (Eds.), *Recent*
1349 *Advances in Statistics: Papers in Honour of Herman Chernoff on His Sixtieth Birthday* (pp. 245–269). New
1350 York/London: Academic Press.
- 1351 Zhang, N., & Zhou, K. (2015). Mineral prospectivity mapping with weights of evidence and fuzzy logic methods.
1352 *Journal of Intelligent and Fuzzy Systems*, 29(6), 2639–2651.

APPENDICES

Capture-Efficiency Curve as a CDF

In the section on **THRESHOLD SELECTION USING THE CAPTURE-EFFICIENCY CURVE**, subsection **Principle**, we mentioned that the capture-efficiency curve can be regarded as a cumulative distribution function (cdf) in its own right. Here we clarify that comment.

A very subtle interpretation of the capture-efficiency curve used by statisticians is the “transformation to uniformity” or “probability integral transformation” (Kendall and Stuart 1973, p. 459 ff.; Hogg and Craig 1970, pp. 349–350). Suppose that the original predictor Z is replaced by a new predictor V , defined at each spatial location u by $V(u) = G(Z(u)) = a(Z(u))/a$, which is the area fraction of spatial locations where the predictor value does not exceed the value z . In words, at a given spatial location u , the value of $V(u)$ is the fraction of area of the survey region where the original predictor Z does not exceed the value $Z(u)$ which it takes at u . If Z is distance-to-nearest-fault, then $V(u)$ is the area fraction occupied by the buffer at distance equal to the distance from u to the faults, that is, the buffer whose boundary passes through the point u . Then the capture-efficiency curve is the cumulative distribution function of the transformed predictor V at the deposit points, while the diagonal line is the cumulative distribution function of V over all spatial locations in the survey region.

Likelihood Function for Threshold Model

This appendix provides elementary explanations for the appearance of the binomial probability distribution discussed in the section on **SIGNIFICANCE TESTS**, subsection **Significance Test for a Binary Predictor**, and for the form of the likelihood function based on this distribution, and for the form of the likelihood discussed in the section on **THRESHOLD SELECTION USING CHANGE-POINT ANALYSIS**, sub-subsection *Profile Likelihood for the Threshold Model*, which results from the assumption that the grid cells are very small in area relative to the survey region.

To simplify discussion, we assume that the geometry of the survey region S and of the prospective map feature B are fixed and known in advance, while the mineral deposit locations are discovered during the survey. In the notation of the section on **BINARY PREDICTORS**, the survey region has area a and the feature B has area a_B .

The survey process begins by dividing the survey region into N grid cells of equal area, then determining whether each grid cell contains or does not contain a deposit. For our purposes the results of the survey are the count n_B of grid cells inside B which contain a deposit, and the count $n_{\bar{B}}$ of grid cells outside B which contain a deposit. The total number of grid cells containing deposits is $n = n_B + n_{\bar{B}}$.

Uniform Prospectivity Model Using Binomial Probabilities

We first consider the simple model in which prospectivity is uniform over the entire survey region. Each grid cell has the same probability p of containing a deposit. The outcomes in different grid cells are assumed to be statistically independent. There are N grid cells altogether. Therefore the number n of grid cells containing a deposit follows a binomial distribution on N trials with success probability p ; the probability that exactly n grid cells contain deposits is

$$\binom{N}{n} p^n (1-p)^{N-n}. \quad (26)$$

The expected total number of grid cells which contain deposits is Np .

The same principle applies to the grid cells inside the feature B ; the probability that exactly n_B grid cells *inside* B contain deposits is

$$\binom{N_B}{n_B} p^{n_B} (1-p)^{N_B-n_B}, \quad (27)$$

where $N_B = (a_B/a)N$ is the number of grid cells that constitute the feature B . The expected number of grid cells *inside* B that contain deposits is $pN_B = p(a_B/a)N$.

Again this applies to the grid cells outside B ; the probability of obtaining $n_{\bar{B}}$ grid cells outside B which contain deposits is

$$\binom{N_{\bar{B}}}{n_{\bar{B}}} p^{n_{\bar{B}}} (1-p)^{N_{\bar{B}}-n_{\bar{B}}}, \quad (28)$$

1387 where $N_{\bar{B}} = (a_{\bar{B}}/a)N$ is the number of grid cells in \bar{B} . The expected number of grid cells *outside* B that contain
 1388 deposits is $pN_{\bar{B}} = (a_{\bar{B}}/a)pN$.

Combining (27) and (28), the probability of obtaining n_B cells with deposits inside B and $n_{\bar{B}}$ cells with deposits
 outside B is

$$\binom{N_B}{n_B} p^{n_B} (1-p)^{N_B-n_B} \binom{N_{\bar{B}}}{n_{\bar{B}}} p^{n_{\bar{B}}} (1-p)^{N_{\bar{B}}-n_{\bar{B}}}$$

which simplifies to

$$\binom{N_B}{n_B} \binom{N_{\bar{B}}}{n_{\bar{B}}} p^n (1-p)^{N-n}. \quad (29)$$

1389 This combining is justified because individual grid cell outcomes, and hence outcomes in disjoint parts of the
 1390 survey region, are independent.

The *likelihood function* of the model is the probability (29) of observing the data $(n_B, n_{\bar{B}})$, treated as a function
 of the parameter p :

$$L(p) = \binom{N_B}{n_B} \binom{N_{\bar{B}}}{n_{\bar{B}}} p^n (1-p)^{N-n}.$$

Constant factors that do not depend on p can be omitted, so the likelihood would often be reported as

$$L(p) = p^n (1-p)^{N-n}. \quad (30)$$

1391 *Simple Threshold Model Using Binomial Probabilities*

In the simple threshold model we assume that prospectivity is higher inside the feature B . Each grid cell inside
 B has probability p_B of containing a deposit, while each grid cell outside B has a different probability $p_{\bar{B}}$
 of containing a deposit, where $p_{\bar{B}} < p_B$. We simply replace p by p_B in equation (27) to find that the probability of
 obtaining exactly n_B cells with deposits inside B is

$$\binom{N_B}{n_B} p_B^{n_B} (1-p_B)^{N_B-n_B}. \quad (31)$$

Replacing p by $p_{\bar{B}}$ in equation (28), the probability of obtaining $n_{\bar{B}}$ grid cells outside B which contain deposits is

$$\binom{N_{\bar{B}}}{n_{\bar{B}}} p_{\bar{B}}^{n_{\bar{B}}} (1-p_{\bar{B}})^{N_{\bar{B}}-n_{\bar{B}}}. \quad (32)$$

Combining (31) and (32), the probability of obtaining n_B cells containing deposits inside B and $n_{\bar{B}}$ cells containing
 deposits outside B is

$$\binom{N_B}{n_B} p_B^{n_B} (1-p_B)^{N_B-n_B} \binom{N_{\bar{B}}}{n_{\bar{B}}} p_{\bar{B}}^{n_{\bar{B}}} (1-p_{\bar{B}})^{N_{\bar{B}}-n_{\bar{B}}}.$$

The likelihood function for the simple threshold model is (again omitting constant factors)

$$L(p_B, p_{\bar{B}}) = p_B^{n_B} (1-p_B)^{N_B-n_B} p_{\bar{B}}^{n_{\bar{B}}} (1-p_{\bar{B}})^{N_{\bar{B}}-n_{\bar{B}}}. \quad (33)$$

1392 The likelihood now has two arguments p_B and $p_{\bar{B}}$ representing the probabilities of a deposit for grid cells inside
 1393 and outside B , respectively. Notice that if the two probabilities were equal, $p_B = p_{\bar{B}} = p$, then (33) would collapse
 1394 to (30).

1395 *Rescaling in Terms of Density*

1396 Since the grid cells are artificial (and in particular their size is an arbitrary choice) it is useful to rescale the
 1397 equations so that they depend as little as possible on the grid geometry. This can be done by using the average
 1398 *density* of deposits, μ , defined as the expected number per unit area.

In the uniform prospectivity model, the expected total number of grid cells that contain deposits is Np . The
 average density is therefore $\mu = Np/a$. Noting that the area of one grid cell is $\epsilon = a/N$, we see that the average
 density is equal to $\mu = p/\epsilon$. Equivalently, $p = \mu\epsilon$, that is, the probability of a deposit in any given cell is equal to
 the average density of deposits times the area of the grid cell. Replacing p by $\mu\epsilon$ in equation (30), and removing
 constant factors, we get the likelihood

$$L(\mu) = \mu^n (1 - \mu\epsilon)^{N-n}. \quad (34)$$

In the simple threshold model, we have two different densities inside and outside the feature B . Inside B ,
 the expected number of grid cells containing deposits is $N_B p_B$ and the area is $a_B = N_B \epsilon$ so the density is
 $\mu_B = N_B p_B / a_B = p_B / \epsilon$. Outside B , the density is $\mu_{\bar{B}} = p_{\bar{B}} N_{\bar{B}} / a_{\bar{B}} = p_{\bar{B}} / \epsilon$. Substituting into (33) we get the
 likelihood

$$L(\mu_B, \mu_{\bar{B}}) = \mu_B^{n_B} (1 - \epsilon\mu_B)^{N_B-n_B} \mu_{\bar{B}}^{n_{\bar{B}}} (1 - \epsilon\mu_{\bar{B}})^{N_{\bar{B}}-n_{\bar{B}}}. \quad (35)$$

1399 *Small Grid Cells*

Finally, we suppose that the grid cells are very small. Then N is very large, $\epsilon = a/N$ is very small. For the uniform prospectivity model, in the likelihood (30) the term $(1 - \mu\epsilon)^{N-n}$ converges to the exponential $\exp(-\mu N\epsilon) = \exp(-\mu a)$ as N becomes large, giving the likelihood

$$L(\mu) = \mu^n \exp(-\mu a).$$

For the simple threshold model, the likelihood (33) similarly converges to

$$L(\mu_B, \mu_{\bar{B}}) = \mu_B^{n_B} \mu_{\bar{B}}^{n_{\bar{B}}} \exp(-\mu_B a_B - \mu_{\bar{B}} a_{\bar{B}}).$$

Taking logarithms, the log-likelihood for the null model of uniform prospectivity is

$$\ln L(\mu) = n \ln \mu - \mu a, \tag{36}$$

and for the simple threshold model

$$\ln L(\mu_B, \mu_{\bar{B}}) = n_B \ln \mu_B + n_{\bar{B}} \ln \mu_{\bar{B}} - \mu_B a_B - \mu_{\bar{B}} a_{\bar{B}}. \tag{37}$$

1400 If the feature B is the region determined by thresholding a spatial predictor function $Z(u)$ to have values less
 1401 than or equal to a threshold z , then $a_B = a(z)$ is the area of this region, $n_B = n(z)$ is the number of deposits with
 1402 predictor values less than or equal to z , and we have $a_{\bar{B}} = a - a_B$ and $n_{\bar{B}} = n - n_B = n - n(z)$, so that equation
 1403 (37) is translated into equation (10) of the paper.

1404 **Connections Between the Threshold Selection Criteria**

1405 In this appendix, we provide proofs of claims made in the sections **SIGNIFICANCE TESTS** and **MATHE-**
 1406 **MATICAL CONNECTIONS BETWEEN THE CRITERIA** regarding relationships between various of
 1407 the threshold selection criteria.

1408 *Forms of the χ^2 Statistic*

First we prove the claim in the section on **SIGNIFICANCE TESTS** that the general form of the χ^2 statistic (18) reduces to the special form (19) in this case. For any feature B , define $e_B = (n/a)a_B = (a_B/a)n$ and $e_{\bar{B}} = (n/a)a_{\bar{B}} = (a_{\bar{B}}/a)n$, the expected counts in B and \bar{B} , respectively, when the deposits are randomly distributed with constant density. Then

$$n_{\bar{B}} - e_{\bar{B}} = (n - n_B) - (n - e_B) = -(n_B - e_B)$$

1409 so that

$$\begin{aligned} X^2 &= \frac{(n_B - e_B)^2}{e_B} + \frac{(n_{\bar{B}} - e_{\bar{B}})^2}{e_{\bar{B}}} = \frac{(n_B - e_B)^2}{e_B} + \frac{(n_B - e_B)^2}{e_{\bar{B}}} = (n_B - e_B)^2 \left(\frac{1}{e_B} + \frac{1}{e_{\bar{B}}} \right) \\ &= (n_B - e_B)^2 \frac{e_{\bar{B}} + e_B}{e_B e_{\bar{B}}} = \frac{n}{e_B e_{\bar{B}}} (n_B - e_B)^2, \end{aligned}$$

1410 which is equivalent to (19).

1411 *Connection Between Akman-Raftery and χ^2 Criteria*

1412 Next we prove the connection between $AR(z)$ and $X(z)$ stated in equation (14), in the section on **MATHE-**
 1413 **MATICAL CONNECTIONS BETWEEN THE CRITERIA**, and again in the section on **SIGNIFICANCE**
 1414 **TESTS**, subsection **Akman-Raftery (Constrained χ^2) Test**.

1415 For candidate threshold value z , let $n(z)$ be the number of deposits below the threshold, $a(z)$ the area of study
 1416 region below the threshold, n the total number of deposits, a the total area of study region. Define $s(z) = a(z)/a$,
 1417 the area fraction.

1418 Then from the definition (13) of the Akman-Raftery criterion,

$$\begin{aligned}
 \text{AR}(z) &= \sqrt{s(z)(1-s(z))} \left(\frac{n(z)}{s(z)} - \frac{n-n(z)}{1-s(z)} \right) \\
 &= \sqrt{s(z)(1-s(z))} \frac{n(z)(1-s(z)) - s(z)(n-n(z))}{s(z)(1-s(z))} \\
 &= \sqrt{s(z)(1-s(z))} \frac{n(z) - ns(z)}{s(z)(1-s(z))} \\
 &= n \frac{\frac{n(z)}{n} - s(z)}{\sqrt{s(z)(1-s(z))}} \\
 &= \sqrt{n} X(z).
 \end{aligned}$$

1419 This proves (14).

1420 *Relation Between χ^2 Statistic and Youden Criterion*

1421 Here, we prove the claim, made in the section **MATHEMATICAL CONNECTIONS BETWEEN THE**
 1422 **CRITERIA**, that the χ^2 statistic is the standardised version of the Youden criterion.

1423 Under the null hypothesis that the density of deposits is uniform, if we treat the total number of deposits as
 1424 fixed, then for any threshold z the count $n(z)$ follows a binomial distribution with n trials and success probability
 1425 $p = a(z)/a = s(z)$. This distribution has variance $np(1-p) = ns(z)(1-s(z))$. Accordingly $Y(z) = (n(z)/n) - s(z)$
 1426 has variance $s(z)(1-s(z))/n$ so that the standard error of $Y(z)$, under the null hypothesis of a uniform density
 1427 of deposits, is $\text{se}_0(Y(z)) = \sqrt{s(z)(1-s(z))/n}$. Finally, inspecting (7) yields (15).

1428 **Derivation of Optimal Threshold for Gradual Decline in Prospectivity**

1429 This appendix provides the proof of the claim in the subsection **Gradual Decline in Prospectivity** of the
 1430 **DISCUSSION** that, if the prospectivity (i.e. the density of deposit points) is a decreasing function of the
 1431 predictor value, then the optimal threshold for the Youden criterion is the threshold at which that function equals
 1432 the average density of deposits over the survey region.

1433 We define the prospectivity as the spatially-varying density (intensity) of deposit points considered as a function
 1434 $\lambda(u)$ of spatial location u . In any given grid cell, the expected number of deposits is equal to $\lambda(u)\epsilon$ where u is the
 1435 location of the cell centre and $\epsilon > 0$ is the cell area. See Baddeley (2018) for an explanation.

Assume first that the density of deposits depends only on the predictor Z . That is, we assume that $\lambda(u)$
 depends on $Z(u)$ through the relation $\lambda(u) = \rho(Z(u))$, where ρ is a nonnegative function. Several examples of the
 relationship between predictor and prospectivity are shown in the top row of Figure 23; these are graphs of $\rho(z)$
 against z . Assume further, as we did in subsection **Gradual Decline in Prospectivity**, that $\rho(z)$ is a decreasing
 function of z , such as those in the left and right columns of Figure 23. Then, $G(z)$ is the spatial cumulative
 distribution function of the predictor over the study region S , that is,

$$G(z) = \frac{1}{a} \int_S \mathbf{1}\{Z(u) \leq z\} du,$$

where $\mathbf{1}\{Z(u) \leq z\}$ is the indicator function, equal to 1 if $Z(u) \leq z$ and equal to 0 otherwise. The expected total
 number of deposits $\mathbb{E}[n]$ is equal to the integral of the intensity function,

$$\mathbb{E}[n] = \int_S \lambda(u) du = \int_S \rho(Z(u)) du$$

This integral over the spatial domain S can be transformed into a one-dimensional integral

$$\mathbb{E}[n] = a \int_{-\infty}^{\infty} \rho(z) dG(z).$$

The cumulative distribution function of the values of Z at the deposit points, averaged over all random outcomes,
 is

$$\tilde{F}(z) = \frac{a \int_{-\infty}^z \rho(v) dG(v)}{a \int_{-\infty}^{\infty} \rho(v) dG(v)} = \frac{1}{\mu} \int_{-\infty}^z \rho(v) dG(v), \quad (38)$$

1436 where $\mu = \mathbb{E}[n]/a$ is the average density of deposits over the whole domain, and v is the dummy variable of
1437 integration.

The expected capture-efficiency curve is the graph of $\tilde{F}(z)$ against $G(z)$ for all z ; equivalently it is the graph of the function $s \mapsto \tilde{F}(G^{-1}(s))$ where G^{-1} is the inverse function of G . Assuming differentiability, the slope of the expected capture-efficiency curve is

$$\frac{d}{ds} \tilde{F}(G^{-1}(s)) = \frac{\tilde{F}'(G^{-1}(s))}{G'(G^{-1}(s))}.$$

1438 But from (38) we have $\tilde{F}'(z) = \rho(z)G'(z)/\mu$, so that the slope of the expected capture-efficiency curve at a given
1439 area fraction s is equal to $\rho(z)/\mu$, where $z = G^{-1}(s)$.

1440 The Youden method selects the point on the capture-efficiency curve with slope equal to 1. Ignoring sampling
1441 variability, that is, if we replace the observed capture-efficiency curve by the expected capture-efficiency curve, the
1442 Youden method selects the point with slope $\rho(z)/\mu = 1$, i.e., it selects the threshold value z for which $\rho(z) = \mu$.

1443 Non-negligible Grid Cell Size

1444 Equations in the paper assume, for simplicity and clarity, that the area of a grid cell is negligible. To be precise,
1445 if ϵ denotes the area of one grid cell, then we assume that $n\epsilon$, the total area of all cells containing deposits, can
1446 be treated as zero. This appendix lists the modifications to these equations that are necessary when $n\epsilon$ is not
1447 negligible.

1448 In equation (1), the revised formulae are

$$W_+ = \ln \frac{n_B}{a_B - \epsilon n_B} - \ln \frac{n}{a} \quad (39)$$

$$W_- = \ln \frac{n_{\bar{B}}}{a_{\bar{B}} - \epsilon n_{\bar{B}}} - \ln \frac{n}{a} \quad (40)$$

$$\hat{C} = \ln \frac{n_B}{a_B - \epsilon n_B} - \ln \frac{n_{\bar{B}}}{a_{\bar{B}} - \epsilon n_{\bar{B}}}. \quad (41)$$

Equation (2) becomes

$$\hat{C}(z) = \ln \left(\frac{n(z)}{a(z) - \epsilon n(z)} \bigg/ \frac{n - n(z)}{a - a(z) - \epsilon(n - n(z))} \right). \quad (42)$$

Equation (3) becomes

$$\text{se}(\hat{C}) = \sqrt{\frac{1}{n_B} + \frac{1}{n_{\bar{B}}} + \frac{1}{a_B/\epsilon - n_B} + \frac{1}{a_{\bar{B}}/\epsilon - n_{\bar{B}}}} \quad (43)$$

and equation (5) becomes

$$\text{se}(\hat{C}(z)) = \sqrt{\frac{1}{n(z)} + \frac{1}{n - n(z)} + \frac{1}{a(z)/\epsilon - n(z)} + \frac{1}{(a - a(z))/\epsilon - (n - n(z))}}. \quad (44)$$

Equation (8) for the Youden criterion becomes

$$Y(z) = \frac{n(z)}{n} - \frac{a(z) - \epsilon n(z)}{a - \epsilon n}. \quad (45)$$

1449 There are no changes in other equations, except that equation (12) does not hold. In the right panel of Figure 10,
1450 the ROC curve should be used instead of the capture-efficiency curve.

Online Resource 1

for the article

Optimal thresholding of predictors in mineral prospectivity analysis

Baddeley *et al.* (2020)

This is a demonstration of our computer code for optimal thresholding of greyscale predictors that was used in the original article. The code and original data are provided in auxiliary files.

This document was produced from the original script file `OnlineResource1.Rnw` by applying the R command `Sweave`.

The original script file contains a mixture of R code and \LaTeX document source. The `Sweave` command executes the R code and inserts the results into the relevant place in the document.

To repeat the analysis, the reader can simply re-type all the R commands shown here. Alternatively, use R CMD `Stangle OnlineResource1.Rnw` to extract the R commands from this document into the file `OnlineResource1.R`. Then `source("OnlineResource1.R")` will run all the commands.

1 Load the code

This code depends on the `spatstat` package:

```
> library(spatstat)
> library(spatstat.utils)
```

Load the special code:

```
> source("usercode.R")
```

Create a folder to store the generated graphics files (this also checks that we have the necessary file permissions):

```
> if(!dir.exists("./pix-auto")) dir.create("./pix-auto")
```

2 Load data

2.1 Synthetic data

First, here are the synthetic data described in the paper:

```
> load("synthetic.rda")
> X
```

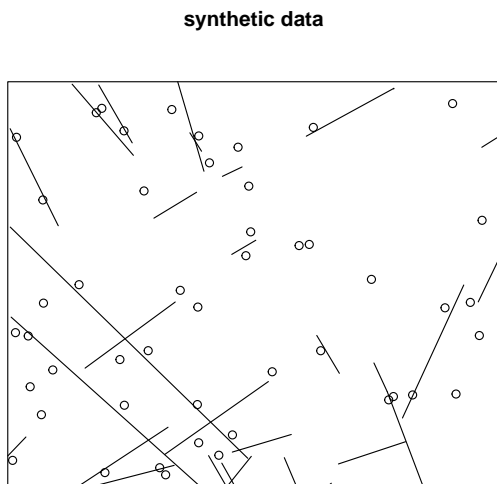
```
Planar point pattern: 48 points
window: rectangle = [-0.335, 70.11] x [100, 158.233] units
```

```
> L
```

```
planar line segment pattern: 30 line segments
window: rectangle = [-0.335, 70.11] x [100, 158.233] km
```

Here X is the (synthetic) spatial point pattern of deposits, and L is the spatial pattern of fault lines.

```
> plot(L, main="synthetic data")
> plot(X, add=TRUE)
```



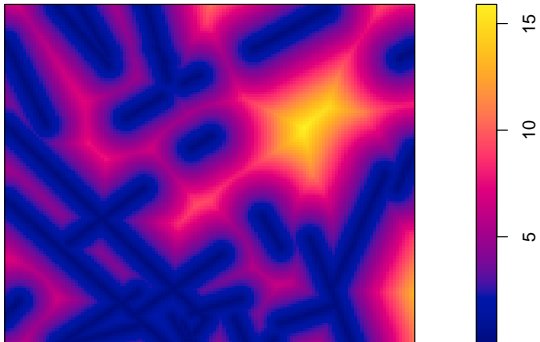
For the covariate Z we will use the distance-to-nearest-fault. This can be computed in the `spatstat` package by either of the commands `distmap` or `distfun` with slightly different behaviour (documented in the package).

```
> Z <- distmap(L)
> Z
```

```
real-valued pixel image
128 x 128 pixel array (ny, nx)
enclosing rectangle: [-0.335, 70.11] x [100, 158.23] km
```

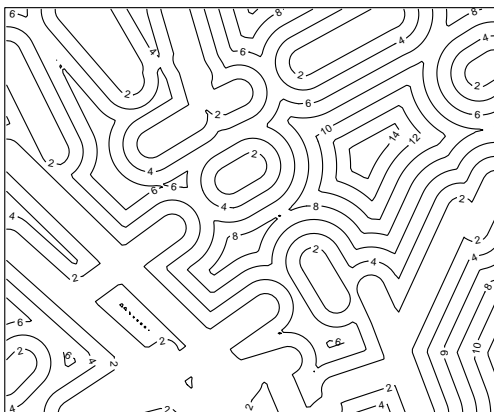
```
> plot(Z)
```

Z



```
> contour(Z)
```

Z



2.2 Murchison data

The Murchison gold data are included in the `spatstat` package.

```
> murchison
```

List of spatial objects

gold:

```

Planar point pattern: 255 points
window: rectangle = [352782.9, 682589.6] x [6699742, 7101484] metres

faults:
planar line segment pattern: 3252 line segments
window: rectangle = [352782.9, 682589.6] x [6699742, 7101484] metres

greenstone:
window: polygonal boundary
enclosing rectangle: [352782.9, 681699.6] x [6706467, 7100804] metres

```

In this case the dataset `murchison` is a list, containing the different spatial objects (gold deposit pattern, fault pattern, and greenstone feature). The list elements can be extracted individually using `$`:

```
> summary(murchison$gold)
```

```

Planar point pattern: 255 points
Average intensity 1.924569e-09 points per square metre

Coordinates are given to 1 decimal place
i.e. rounded to the nearest multiple of 0.1 metres

Window: rectangle = [352782.9, 682589.6] x [6699742, 7101484] metres
              (329800 x 401700 metres)
Window area = 1.32497e+11 square metres
Unit of length: 1 metre

```

Alternatively, they can be accessed using the `with` operator, which is often neater:

```
> with(murchison, summary(gold))
```

```

Planar point pattern: 255 points
Average intensity 1.924569e-09 points per square metre

Coordinates are given to 1 decimal place
i.e. rounded to the nearest multiple of 0.1 metres

Window: rectangle = [352782.9, 682589.6] x [6699742, 7101484] metres
              (329800 x 401700 metres)
Window area = 1.32497e+11 square metres
Unit of length: 1 metre

```

Let's rescale all the Murchison data from metres to kilometres:

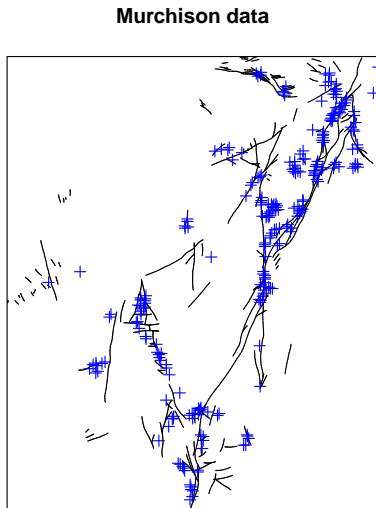
```
> mur <- solapply(murchison, rescale, s=1000, unitname="km")
```

and add the distance-to-nearest-fault covariate:

```
> mur$dfault <- distfun(mur$faults)
```

The new list `mur` contains the rescaled Murchison data and the derived covariate.

```
> plot(mur$faults, main="Murchison data")  
> plot(mur$gold, add=TRUE, pch=3, col="blue")
```



3 Threshold selection

The command `selectThresh` defined in `usercode.R` selects the threshold automatically. The user can specify the method.

For the synthetic data:

```
> selectThresh(X, Z, method="t")
```

```
      Z  
1.424575
```

```
> selectThresh(X, Z, method="Y")
```

```
      Z  
1.966457
```

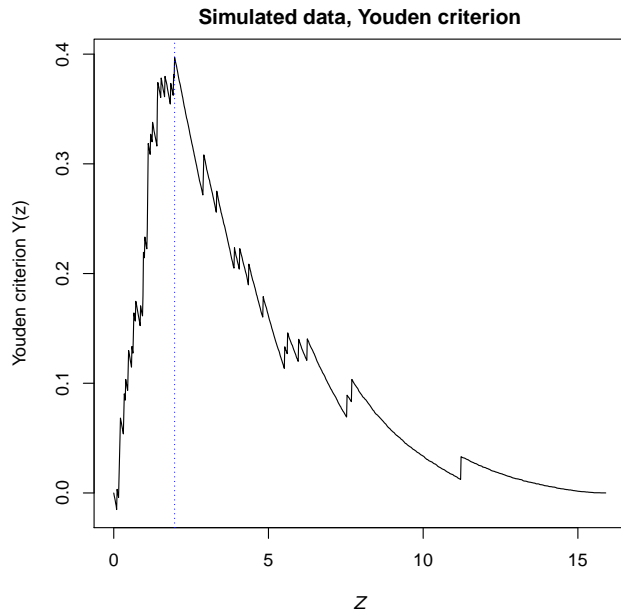
The first result indicates that the threshold obtained by maximising the Studentised contrast t was 1.425 km. The second result indicates that the threshold obtained by maximising the Youden criterion was 1.966 km.

The result of `selectThresh` can also be plotted to show the criterion as a function of the threshold value:

```

> b <- selectThresh(X, Z, method="Y")
> plot(b, main = "Simulated data, Youden criterion",
+       ylab = "Youden criterion Y(z)")

```

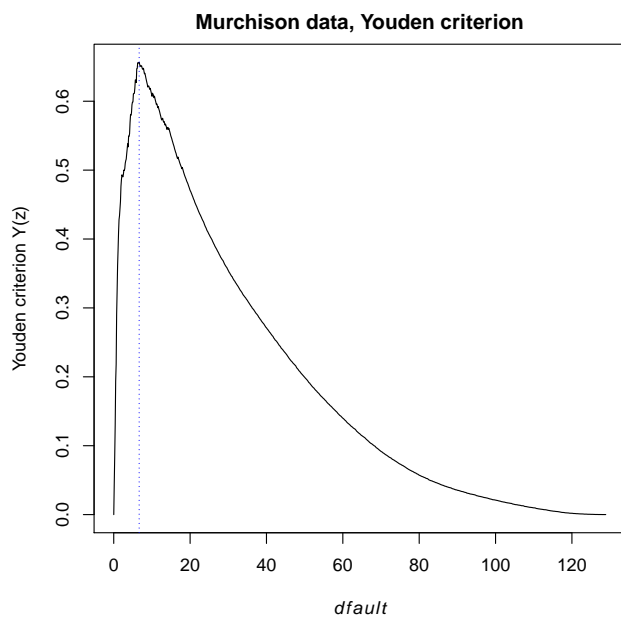


For the Murchison data:

```

> b <- with(mur, selectThresh(gold, dfault, method="Y"))
> b
      dfault
6.655075
> plot(b, main="Murchison data, Youden criterion",
+       ylab = "Youden criterion Y(z)")

```



4 Bootstrap confidence intervals

The functions `bootSE` and `bootCI` defined in `usercode.R` compute (respectively) bootstrap standard errors and bootstrap confidence intervals for the threshold, using the Youden criterion. The usage of the two functions is very similar, so we only show the computation for confidence intervals.

4.1 Synthetic data

For the synthetic data:

```
> bootCI(X, Z)

      lo      hi  conf
z 1.2038740 2.2558028 0.949
s 0.1956055 0.3521729 0.949
```

The first row shows the nonparametric bootstrap confidence interval for the threshold distance z in kilometres (true value 2 km). The second row shows the corresponding confidence interval for the fraction of area s .

The parametric bootstrap confidence interval is:

```
> bootCI(X, Z, parametric=TRUE)

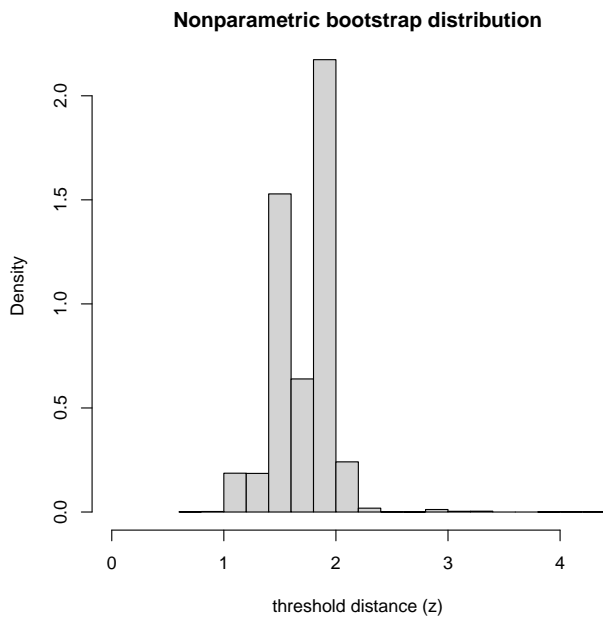
      lo      hi  conf
z 1.592587 2.6448494 0.95
s 0.256665 0.4057007 0.95
```

In the section **Bootstrap standard errors and confidence intervals** of the main paper, we suggested that the parametric bootstrap is typically more accurate, provided the assumed model (for the simulated data, the simple threshold model) is true. Indeed, we see this to be the case here as the nonparametric bootstrap distribution is bimodal, but the parametric distribution is unimodal and asymmetric, as was also suggested in the main paper.

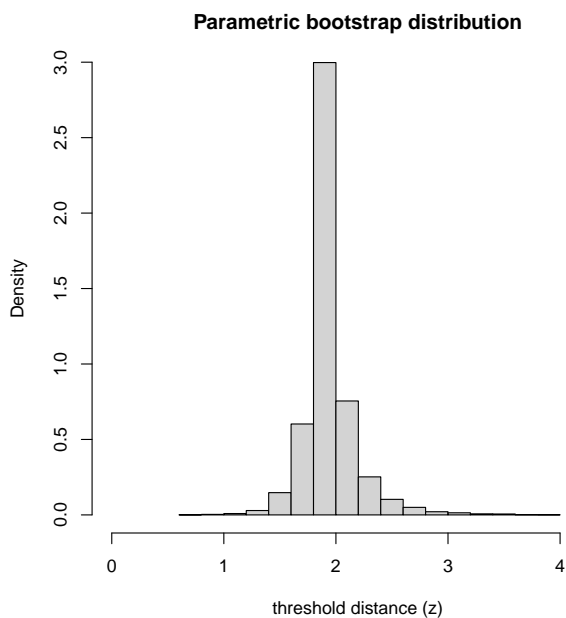
```
> bootDistN <- bootSIM(X, Z, nsim = 10000, parametric = FALSE)
> bootDistP <- bootSIM(X, Z, nsim = 10000, parametric = TRUE)

> maxz <- max(c(max(bootDistN$z), max(bootDistP$z)))

> with(bootDistN,
+     hist(z, main = "Nonparametric bootstrap distribution",
+         prob = TRUE,
+         xlab = "threshold distance (z)",
+         xlim = c(0, maxz)))
```

```
> with(bootDistP,
+     hist(z, main = "Parametric bootstrap distribution",
+         prob = TRUE,
+         xlab = "threshold distance (z)",
+         xlim = c(0, maxz)))
```



4.2 Murchison gold data

For the Murchison gold data:

```
> with(mur, bootCI(gold, dfault))
```

```

          lo          hi conf
z 5.9882296 7.7575571 0.95
s 0.2300171 0.2768555 0.95

```

```
> with(mur, bootCI(gold, dfault, parametric=TRUE))
```

```

          lo          hi conf
z 6.1756566 6.5136096 0.95
s 0.2352295 0.2457275 0.95

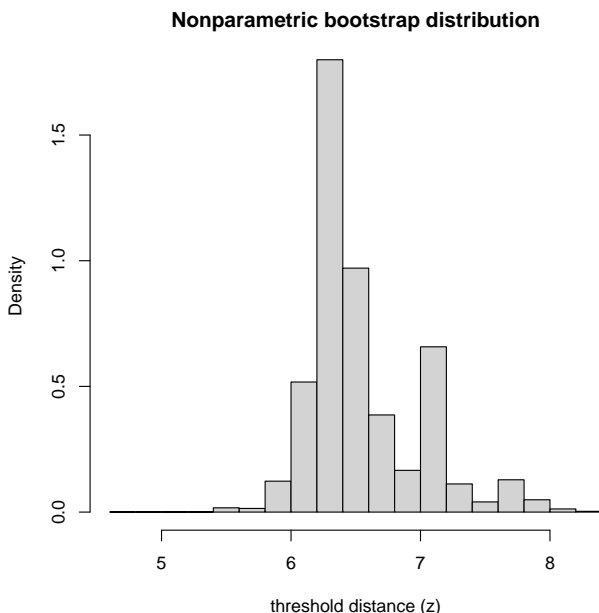
```

The parametric bootstrap confidence interval (obtained with the option `parametric=TRUE`) is not reliable for the Murchison data because they do not seem to conform to the simple threshold model. As pointed out in the main paper, the nonparametric bootstrap is more robust against violation of the model assumptions, and we see below that that the nonparametric bootstrap distribution is both unimodal and asymmetric.

```

> bootDistN <- with(mur,
+                   bootSIM(gold, dfault, nsim = 10000,
+                           parametric = FALSE))
> with(bootDistN,
+      hist(z, main = "Nonparametric bootstrap distribution",
+          prob = TRUE,
+          xlab = "threshold distance (z)"))

```



Online Resource 2

Online supplement to
Optimal thresholding of predictors in mineral prospectivity analysis
by Baddeley *et al.* (2020)

The file `OnlineResource1.pdf` demonstrates how to calculate the threshold of the distance-to-nearest-fault for the Murchison data and for synthetic data. In addition, the code calculates bootstrap confidence intervals for the threshold based on the Youden criterion.

In this document, we show the results of carrying out the same analysis but with additional realizations of the synthetic data. Using a fixed pattern of fault lines and fixed threshold distance, additional spatial point patterns of deposits have been generated using the same simple threshold model used to generate the synthetic data in Figure 5 of the main paper. The results below confirm that when the simple threshold model applies, the parametric bootstrap yields more accurate and sensible confidence intervals for the threshold distance. In addition, the following discussion outlines the practical value of examining different realizations of synthetic data.

Geologists have only a single pattern of deposit points to estimate a threshold and classify an area into prospective and non-prospective zones. Although geological features like faults, rock unit boundaries and areas of favourable host rock determine the sites of gold deposition, the threshold estimate is very dependent on the point pattern of the known deposits. It is important to recognize that this structural framework of rock units, and geometry of faults and shear zones controls the potential location of gold deposits and therefore the prospectivity. But our knowledge of the deposits and occurrences which provide evidence for this prospectivity is incomplete and the precise location of deposits at the scale of minor faults (a few kilometres or less) that they are associated with is random. There are two further aspects to the randomness of the deposit pattern:

1. As additional exploration is carried out, new mineral occurrences and deposits are often discovered so with time additional points are added to the pattern.
2. Generally, only gold deposits located within 300 m of the surface are discoverable by drilling and geochemical exploration, yet orogenic gold deposits were originally formed at depths from 5 to at least 15 km. There is no reason to consider that the present-day surface zone is more or less prospective, in terms of deposit numbers or gold content, than any other slice through the upper crust. A complex network of faults and shear zones together with differential uplift and erosion determined which depths at the time of gold mineralization are exposed at the surface today. Consequently, the pattern of deposits we have discovered or can potentially discover is the result of random processes within areas where the prospectivity is determined by a larger scale framework of geodynamic events and processes and structural features.

Given that the pattern of deposits we observe is only one of many possible patterns, it makes sense to examine the effect of alternative sets of random locations of deposits within the favourable zones on our estimates of threshold distance. It should be noted though, that although the location of the deposit points change in the simulations, the numbers and densities of deposits in the prospective zones that define the optimal threshold remain the same as those described for the synthetic dataset in the main text.

Original synthetic data

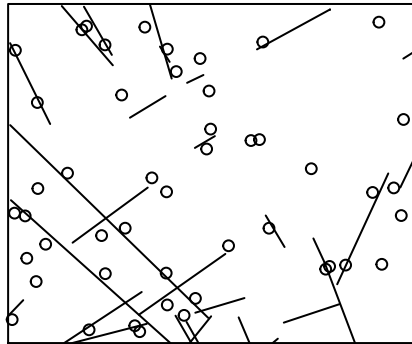


Figure 1: Faults shown as solid lines, deposits as open circles.

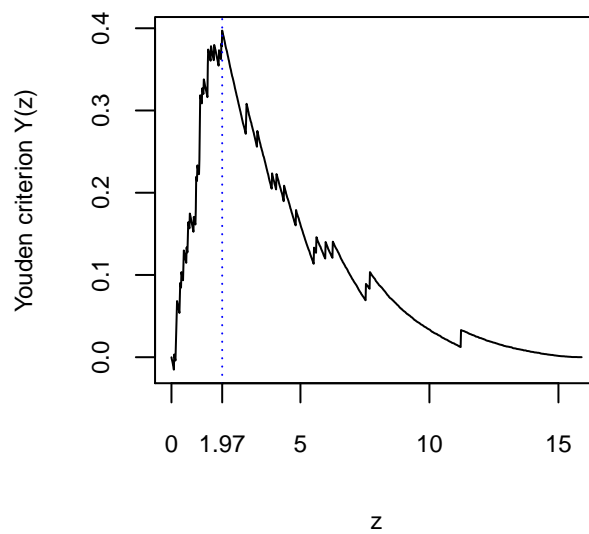


Figure 2: Youden criterion, showing the estimate of the threshold

For the nonparametric and parametric bootstrap, 95% confidence intervals are shown followed by the corresponding bootstrap distributions.

Table 1: 95% nonparametric bootstrap confidence interval for the threshold estimate (z) and fraction of area (s).

	lo	hi	conf
z	1.24	2.93	0.95
s	0.20	0.44	0.95

Table 2: 95% parametric bootstrap confidence interval for the threshold estimate (z) and fraction of area (s).

	lo	hi	conf
z	1.59	2.50	0.95
s	0.26	0.39	0.95

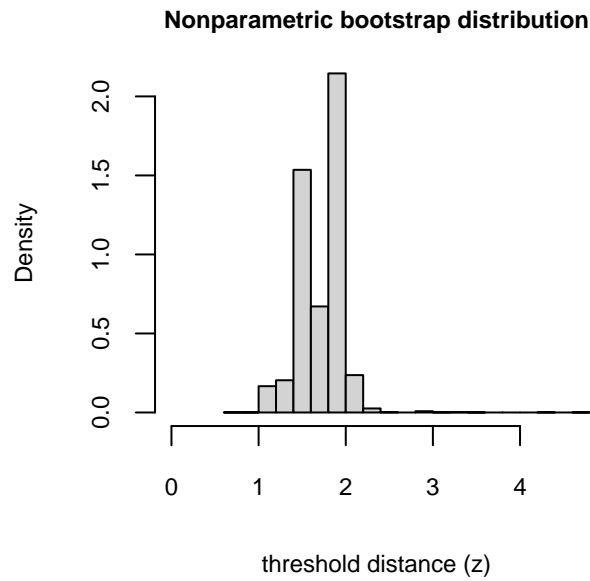


Figure 3: Nonparametric bootstrap distribution of the threshold distance selected by the Youden criterion.

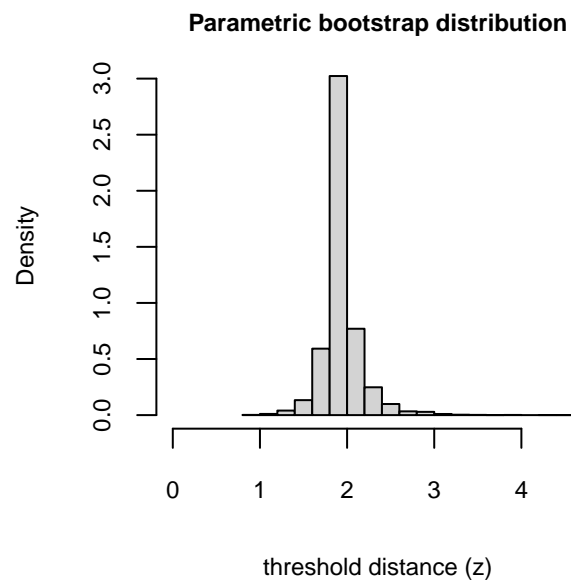


Figure 4: Parametric bootstrap distribution of the threshold distance selected by the Youden criterion.

Realization 2

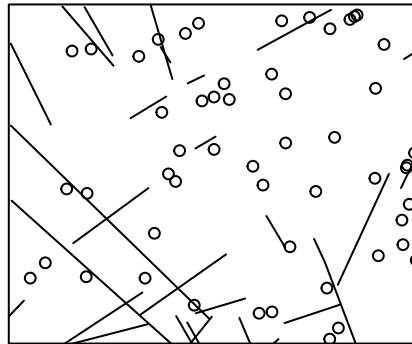


Figure 5: Faults shown as solid lines, deposits as open circles.

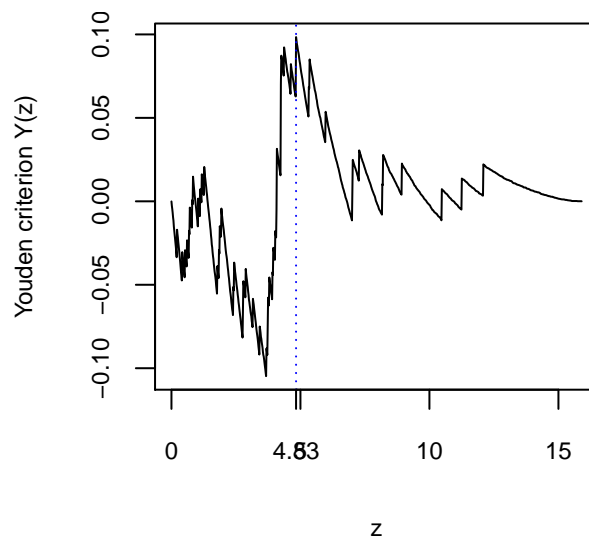


Figure 6: Youden criterion, showing the estimate of the threshold

For the nonparametric and parametric bootstrap, 95% confidence intervals are shown followed by the corresponding bootstrap distributions.

Table 3: 95% nonparametric bootstrap confidence interval for the threshold estimate (z) and fraction of area (s).

	lo	hi	conf
z	0.89	10.92	0.95
s	0.15	0.96	0.95

Table 4: 95% parametric bootstrap confidence interval for the threshold estimate (z) and fraction of area (s).

	lo	hi	conf
z	1.14	7.75	0.95
s	0.19	0.88	0.95

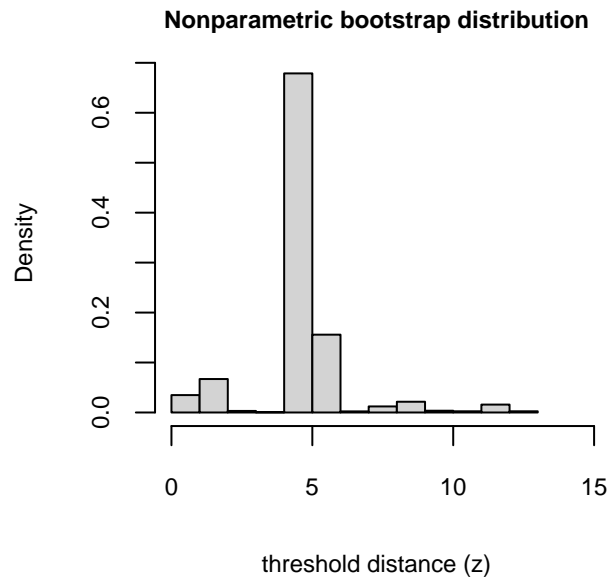


Figure 7: Nonparametric bootstrap distribution of the threshold distance selected by the Youden criterion.

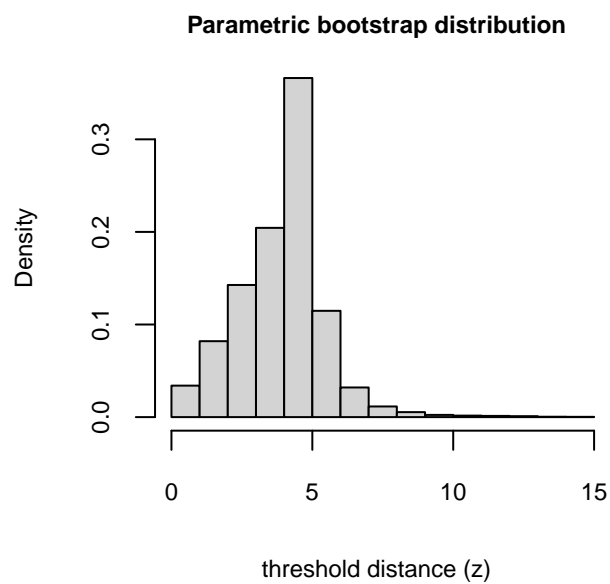


Figure 8: Parametric bootstrap distribution of the threshold distance selected by the Youden criterion.

Realization 3

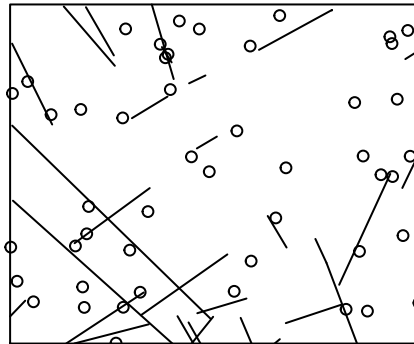


Figure 9: Faults shown as solid lines, deposits as open circles.

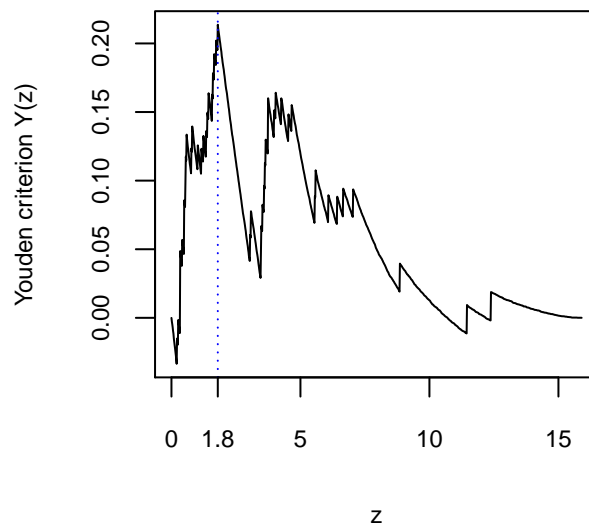


Figure 10: Youden criterion, showing the estimate of the threshold

For the nonparametric and parametric bootstrap, 95% confidence intervals are shown followed by the corresponding bootstrap distributions.

Table 5: 95% nonparametric bootstrap confidence interval for the threshold estimate (z) and fraction of area (s).

	lo	hi	conf
z	0.6	5.18	0.95
s	0.1	0.69	0.95

Table 6: 95% parametric bootstrap confidence interval for the threshold estimate (z) and fraction of area (s).

	lo	hi	conf
z	1.06	3.86	0.95
s	0.17	0.56	0.95

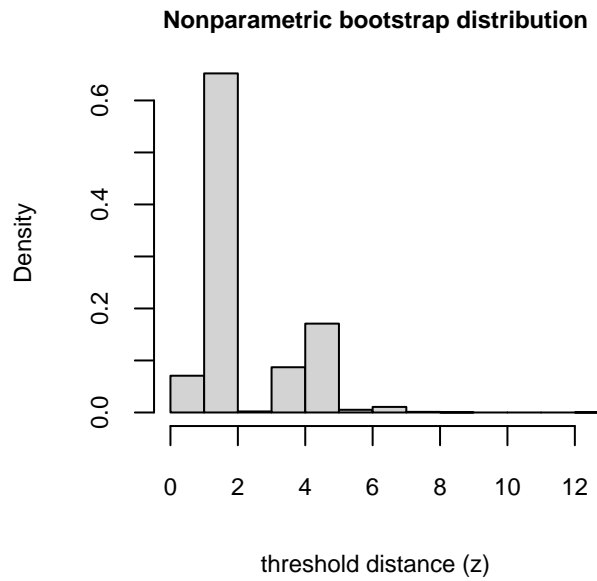


Figure 11: Nonparametric bootstrap distribution of the threshold distance selected by the Youden criterion.

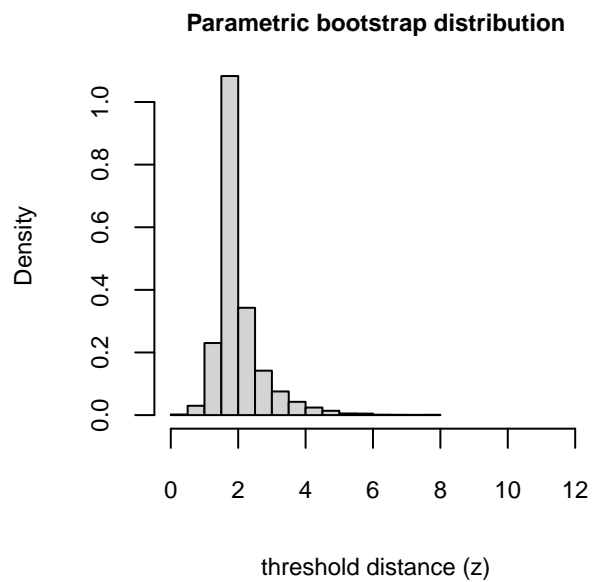


Figure 12: Parametric bootstrap distribution of the threshold distance selected by the Youden criterion.

Realization 4

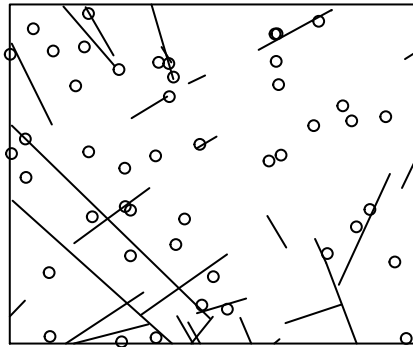


Figure 13: Faults shown as solid lines, deposits as open circles.

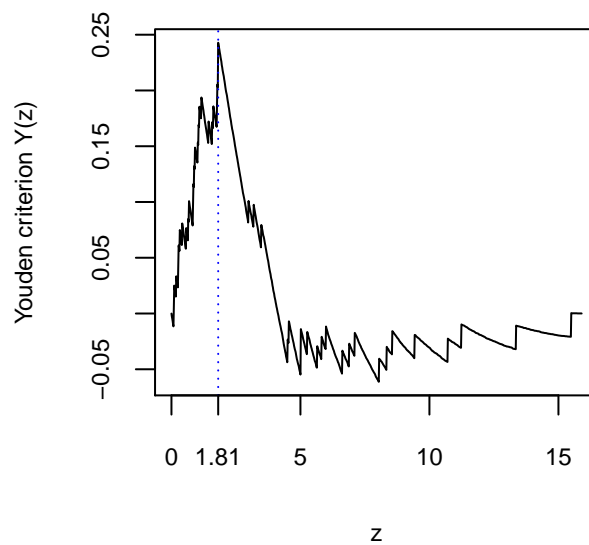


Figure 14: Youden criterion, showing the estimate of the threshold

For the nonparametric and parametric bootstrap, 95% confidence intervals are shown followed by the corresponding bootstrap distributions.

Table 7: 95% nonparametric bootstrap confidence interval for the threshold estimate (z) and fraction of area (s).

	lo	hi	conf
z	0.91	1.92	0.95
s	0.15	0.30	0.94

Table 8: 95% parametric bootstrap confidence interval for the threshold estimate (z) and fraction of area (s).

	lo	hi	conf
z	1.09	3.57	0.95
s	0.18	0.52	0.95

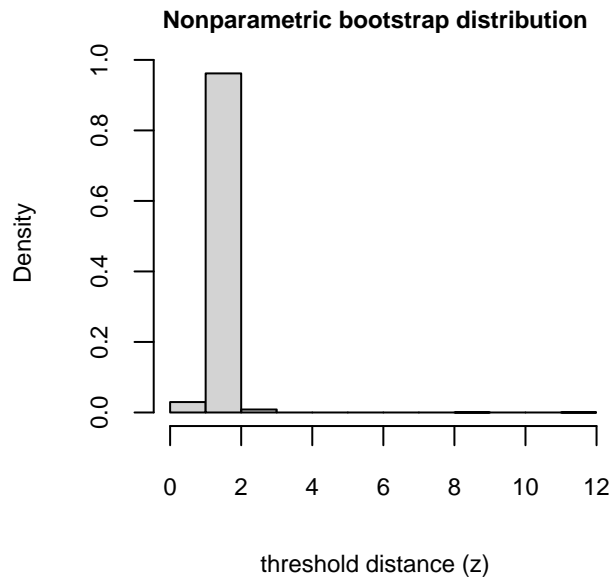


Figure 15: Nonparametric bootstrap distribution of the threshold distance selected by the Youden criterion.

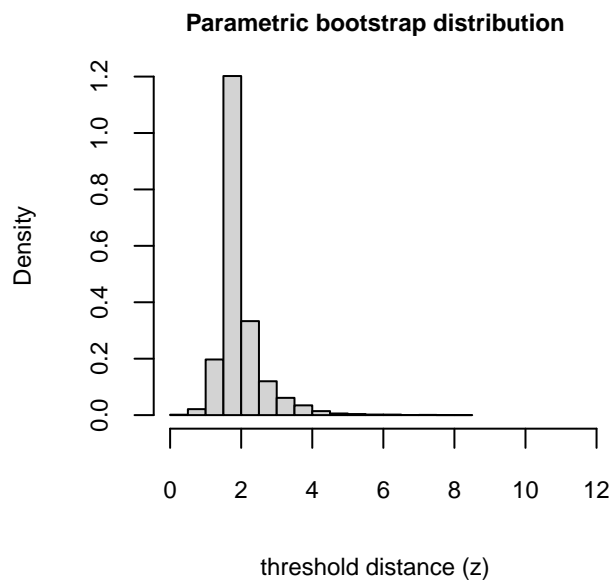


Figure 16: Parametric bootstrap distribution of the threshold distance selected by the Youden criterion.