Objective selection of suitable unit cell size in data-driven modeling of mineral prospectivity

Emmanuel John M. Carranza *

Department of Earth Systems Analysis, International Institute for Geo-Information Science and Earth Observation (ITC), Hengelosestraat 99, PO Box 6, 7500AA Enschede, The Netherlands

A R T I C L E   I N F O

Article history:
Received 29 July 2008
Received in revised form 2 February 2009
Accepted 5 February 2009

Keywords:
Point pattern analysis
Spatial contrast
Fractal analysis
Weights-of-evidence
GIS

A B S T R A C T

In GIS-based data-driven modeling of mineral prospectivity, a suitably fine unit cell size is used for spatial representation of known occurrences of mineral deposits of the type sought \((D)\) in a study area \((T)\). However, until now, the unit cell size is chosen subjectively. In this paper, a methodology is proposed for objective selection of the most suitable unit cell size for data-driven modeling of mineral prospectivity using a raster-based GIS. A set of choices of suitable unit cell sizes is first derived via point pattern analysis of a set of known occurrences of mineral deposits of the type sought. Then, (a) the lower limit of a set of choices of suitable unit cell sizes is considered and defined according to the map scales from which spatial data for mineral prospectivity mapping were derived, and (b) the upper limit of the same set of choices of suitable unit cell sizes is considered (and revised as necessary) based on knowledge of spatial extents of mineral deposits of the type sought or via analysis of reflexive nearest neighbour points. Finally, it is shown that fractal analysis of spatial contrast between unit cells containing \(D\) and unit cells not containing \(D\) in \(T\) provides for objective selection of the most suitable unit cell size. In a case study application of the weight-of-evidence method to mineral prospectivity mapping, using the most suitable unit cell size, found via the proposed methodology, results in spatial evidence weights and weight uncertainties that are nearly identical to those derived by using the finest (i.e., lower limit) unit cell size. In contrast to using the most suitable unit cell size, using coarser unit cell sizes result in higher positive weights, lower negative weights and higher weight uncertainties of spatial evidence of mineral prospectivity. The proposed methodology for objective selection of the most suitable unit cell size in data-driven modeling of mineral prospectivity using a raster-based GIS is robust and can easily be implemented.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

The term mineral prospectivity implies the chance or likelihood that mineral deposits of the type sought can be found in a region or district under investigation. It is similar to the terms mineral potential and mineral favourability, both of which imply the chance or likelihood that mineral deposits of the type sought are contained in a region or district under investigation. The terms mineral prospectivity, mineral potential, and mineral favourability are thus synonymous and can be used interchangeably. For consistency, the term mineral prospectivity is used in this paper.

Modeling of mineral prospectivity is a regional- or district-scale mapping activity, whether field based or GIS based, which aims to delineate prospective areas for further exploration at the next higher scales of mapping. Notwithstanding the scale of mapping, mineral prospectivity is related to the degrees of presence and degrees of importance of individual pieces of spatial evidence of occurrence of mineral deposits of the type sought. That is, in a region or district under investigation, if there are more important pieces of spatial evidence in an area than in another area, then the former is considered to have higher prospectivity than the latter.

In GIS-based data-driven modeling of mineral prospectivity, known occurrences of a particular type of mineral deposits \((D)\) in a region or district \((T)\) are each considered to occupy a small unit cell for the purpose of calculating the weights of individual pieces of spatial evidence of mineral prospectivity. Many studies of GIS-based data-driven modeling of mineral prospectivity make use of small equal-sized (usually square) unit cells or pixels. A unit cell size, denoted hereafter as \((\bullet)\), is chosen and used to discretize \(T\) into a regular grid of unit cells such that \(T\) has \(N(T)\) total number of unit cells and has \(N(D)\) number of unit cells containing \(D\). The \((\bullet)\) that is chosen, however, is roughly related to the lateral extents of every \(D\) but it is considered to be relatively...
information theory. Mineral prospectivity is one type of mineral prospectivity consistent with the general principles of mineral prospectivity. By doing so makes data-driven modeling of spatial objects (Bonham-Carter, 1994, p. 29).

Sampling density in data-driven modeling of mineral prospectivity the choice of a suitable (●) is subjective. Hengl (2006) recently discussed general guidelines in selecting a suitable grid resolution in cartographic and statistical analysis of spatial data. The guideline Hengl (2006) discussed regarding the application of point pattern analysis (Boots and Getis, 1988; Rowlingson and Diggle, 1993) has some relevance to data-driven modeling of mineral prospectivity using a raster-based GIS. In this paper, point pattern analysis is applied to a set of points representing D in order to derive a range of choices of suitable unit cell sizes. Then, it is proposed and demonstrated here (with real data sets) that analysis of spatial contrast between cells containing D and cells not containing D provides for objective selection of the most suitable (●) in data-driven modeling of mineral prospectivity using a raster-based GIS. Finally, to demonstrate the usefulness of the proposed methodology, a case study is presented to compare weights of spatial evidence of mineral prospectivity based on the most suitable (●), a finer (●) and coarser unit cell sizes.

2. Objective selection of suitable (●)

2.1. Concept

In a region or district under investigation, each D represents a sample of mineralized land. Thus, it is essential to consider and determine a suitable sampling density in data-driven modeling of mineral prospectivity. By doing so makes data-driven modeling of mineral prospectivity consistent with the general principles of the statistical theory, sampling theory and spatial (or geographical) information theory. Mineral prospectivity is one type of spatial information, which must be derived systematically via (statistical) analysis of various layers of spatial evidence (i.e., data) of mineral prospectivity at known deposit-type locations and at non-deposit locations. Prospective areas are not only definition-limited but also sampling-limited spatial objects (Bonham-Carter, 1994, p. 29).

Sampling density in data-driven modeling of mineral prospectivity is defined by the (●) that is used for spatial representation of D in T and for discretizing individual layers of spatial evidence. A unit cell is equivalent to a grid cell in sampling. It is also equivalent to a pixel in a raster map. A (●) is described by its length or width, if it is a rectangle, or by its diameter or radius, if it is a circle. A (●) determines the scale of a raster map. The finer the unit cell, the larger the scale of a raster map; the coarser the unit cell, the smaller the scale of a raster map. However, map scale does not symbolize and, thus, should not be confused with the size of T. A (●) defines the spatial (grid or pixel) resolution of models of geographical data. Fine unit cells represent high spatial resolutions, whereas coarse unit cells represent low spatial resolutions. A (●) also defines the spatial accuracy of derived pieces of spatial information portrayed in raster maps or images. Fine unit cells represent high sampling density and imply superior accuracy of spatial information, whereas coarse unit cells represent low sampling density and imply inferior accuracy of spatial information. The preceding comparisons of unit cell sizes and, thus, sampling densities, in the context of mineral prospectivity mapping, are pertinent only to a particular area, because the spatial pattern of D varies from one area to another. Thus, the choice of the most suitable (●) for data-driven modeling of mineral prospectivity must be based on the spatial pattern of D in T.

2.2. Deriving a set of choices of suitable (●)

Because each D is usually symbolized as a point in regional- to district-scale maps, algorithms of point pattern analysis for measures of dispersion (Boots and Getis, 1988; Rowlingson and Diggle, 1993), which is independent of the size of T, can be used to determine distances from every D and the corresponding probabilities associated with these distances that there is one neighbour D situated next to another D. This analysis is chiefly used to determine if a set of points assumes a random, clustered or regular pattern. Examples of application of this analysis are made for individual sets of points representing mineral deposits in three areas and to geothermal prospects in one area (Fig. 1). The steep slopes of the curves in Fig. 2a and b indicate that the spatial pattern of epithermal Au deposits in the Aroroy district (Philippines) and the spatial pattern of epithermal Au deposits in the Cabo de Gata area (Spain) tend to be regular rather than random (see Fig. 1a and b). The gentle slopes of the curves in Fig. 2c and d indicate that the spatial pattern of geothermal prospects in West Java (Indonesia) and the spatial pattern of alkalic porphyry Cu–Au deposits in British Columbia (Canada) tend to be clustered rather than random (see Fig. 1c and d).

While point pattern analysis for measures of dispersion provides for interpretation of the spatial pattern of a set of occurrences of a particular type of mineral deposits, which further provides the possibility to define plausible geological controls on mineralization (e.g., Carranza, 2008), the analysis also indicates distances in which there is zero probability of one neighbour D situated next to another D (Fig. 2). Any of these distances, except zero, can be used as a (●) because each of them satisfies the assumption that each D occupies only one unit cell. Thus, the results shown in Fig. 2 suggest the following. For data-driven modeling of prospectivity for epithermal Au deposits in the Aroroy district (Philippines), the choices of suitable unit cell sizes are 0 < (●) ≤ 560 m (Fig. 2a). For data-driven modeling of prospectivity for epithermal Au deposits in the Cabo de Gata area (Spain), the choices of suitable unit cell sizes are 0 < (●) ≤ 160 m (Fig. 2b). For data-driven modeling of geothermal prospectivity in West Java (Indonesia), the choices of suitable unit cell sizes are 0 < (●) ≤ 750 m (Fig. 2c). For data-driven modeling of prospectivity for alkalic porphyry Cu–Au deposits in British Columbia (Canada), the choices of suitable unit cell sizes 0 < (●) ≤ 330 m (Fig. 2d).

It is emphasized here that the upper limit of a set of choices of suitable unit cell sizes derived from point pattern analysis for measures of dispersion can be extended if, based on geological knowledge, one considers it to be an unreasonable spatial representation of a mineral deposit of the type sought. For example, in mineral resource assessment for porphyry Cu deposits, Singer et al. (2005) and Singer and Menzie (2008) apply consistently an operational rule to combine occurrences of porphyry Cu within 2 km of each into one deposit. Thus, based on this operational rule for combining porphyry Cu occurrences in to a deposit, one may consider an upper limit 330 m for spatial representation of alkalic porphyry Cu–Au deposits to be unreasonable. For example, in British Columbia some of the alkalic porphyry Cu–Au deposits (e.g., Axe (Adit Zone), Axe (South Zone) and Axe (West Zone)) described in the MINFILE mineral inventory database (BCGS, 2007), probably represent one large alkalic porphyry Cu–Au deposit and they can be contained in a unit cell...
size larger than 330 m. Thus, for the analysis of the most suitable \( \cdot \) for data-driven modeling of prospectivity for alkalic porphyry Cu–Au deposits, rather than for assessment of resources for this type of deposit, in British Columbia (Canada), the choices of suitable unit cell sizes is extended to 2000 m.

If there is lack of geological information to justify using a \( \cdot \) that results in some unit cells containing more than one \( D \), one may perform an analysis of reflexive (or reciprocal) nearest neighbour (RNN) points to test if a set of points assumes a clustered or a regular pattern (Boots and Getis, 1988). Two points in a point pattern are considered 1st-order RNN if they are each other's nearest neighbour, whereas 2nd-order RNNs are points that are each other's 2nd-nearest neighbours, and so on. RNNs are always pairs of points, so that the number of RNNs is always an even number. The observed number of RNNs in a point pattern under examination is compared with the expected number of RNNs in a pattern of points with complete spatial randomness (CSR). The points in CSR are simulated for the same area and the same number of points. The expected number of \( j \)th-order RNNs is estimated according to the probability that a point in a point pattern of CSR is the \( j \)th-nearest neighbour of its own \( j \)th-nearest neighbour (see Cox (1981) for details). If the observed number of \( j \)th-order RNNs is higher than the expected number of \( j \)th-order RNNs in CSR expectations, then those RNNs in a point pattern under examination assume a regular pattern. If some \( D \)s in a point pattern of \( D \) are clustered, then using a \( \cdot \) that is coarser than the unit cell sizes indicated by the distance–probability relation (Fig. 1) is justifiable in prospectivity mapping for alkalic porphyry Cu–Au deposits in British Columbia (Canada) and for geothermal resources in West Java (Indonesia). The results shown in Table 1 also suggest that using a \( \cdot \) that is coarser than the unit cell sizes indicated by the distance–probability relation (Fig. 1) is not justifiable in prospectivity mapping for epithermal Au deposits in the Cabo de Gata area (Spain) and in the Aroroy district (Philippines). Nonetheless, one must determine if results of a RNN analysis are more-or-less consistent with knowledge of lateral extents of some (if not all) \( D \) in a study area before proceeding with the analysis of the most suitable \( \cdot \) described here.

2.3. Determining lower limit of a set of choices of suitable \( \cdot \)

Once a set of choices of suitable unit cell sizes has been derived (and the upper limit has been adapted as necessary), one must determine a lower limit of this set of choices of suitable unit cell size.
sizes because of the following arguments. A rough estimate (i.e., in the absence of spatial evidence) of prospectivity for $D$ in $T$ is the ratio $[N(D)]/[N(T)−N(D)]$, where the denominator is the number of unit cells not containing $D$. This ratio, which represents the spatial contrast between unit cells containing $D$ and unit cells not containing $D$, must empirically be a very small value. That is because mineralization is a relatively rare geological phenomenon such that areas occupied by $D$ invariably represent a very small proportion of $T$, meaning that the $\frac{1}{N(T)}$ must be suitably fine with respect to the size of a study area. If it is infinitesimally fine, the spatial contrast between unit cells containing $D$ and unit cells not containing $D$ converges to $1/N(T)$, which can be approximately

---

**Table 1**

Numbers of 1st- and 2nd-order reflexive nearest neighbours (RNNs) in point patterns of mineral and/or geothermal occurrences in four different areas.

<table>
<thead>
<tr>
<th>Point objects (area)</th>
<th>RNNs</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Order</td>
<td>Observed no.</td>
<td>Expected no. in CSR*</td>
</tr>
<tr>
<td>Epithermal Au deposits (Aroroy, Philippines)</td>
<td>1st</td>
<td>8</td>
<td>8.08</td>
</tr>
<tr>
<td></td>
<td>2nd</td>
<td>4</td>
<td>4.28</td>
</tr>
<tr>
<td>Epithermal Au deposits (Cabo de Gata, Spain)</td>
<td>1st</td>
<td>22</td>
<td>25.48</td>
</tr>
<tr>
<td></td>
<td>2nd</td>
<td>12</td>
<td>13.49</td>
</tr>
<tr>
<td>Geothermal occurrences (West Java, Indonesia)</td>
<td>1st</td>
<td>34</td>
<td>32.94</td>
</tr>
<tr>
<td></td>
<td>2nd</td>
<td>16</td>
<td>17.44</td>
</tr>
<tr>
<td>Alkalic porphyry Cu–Au deposits (British Columbia, Canada)</td>
<td>1st</td>
<td>228</td>
<td>221.25</td>
</tr>
<tr>
<td></td>
<td>2nd</td>
<td>106</td>
<td>117.16</td>
</tr>
</tbody>
</table>

CSR* = complete spatial randomness.
equal to zero, such that using the expression \([N(D)] - [N(T) - N(D)]\) becomes less meaningful. However, it is impractical to spatially represent \(D\) with infinitesimally small unit cell sizes because the lateral extents of mineral deposits are not infinitely small. In addition, although a set of \(D\) is portrayed as a set points in regional- to district-scale maps, it does not mean that mineral deposits are ‘dimensionless’ objects and, therefore, there should be a lower limit for a set of choices of suitable \((\bullet)\).

Based on the typical size for legible point symbols in a map (i.e., 0.25 mm or 0.00025 m) and a map scale \((MS)\) factor, the finest legible spatial (grid or pixel) resolution, \((\bullet)_0\), can be estimated as (cf. Vink, 1975; Valenzuela and Baumgardner, 1990; Hengl, 2006):

\[
(\bullet)_0 = MS \times 0.00025.
\]  

(1)

Accordingly, the lower limit of a set of suitable choices of \((\bullet)\) can be estimated based on the scales of the map \((s)\) from which locations of every \(D\) and spatial evidences of mineral prospectivity were derived. For the case study of data-driven modeling of prospectivity for epithermal Au deposits in the Aroroy district (see further below), the spatial data sets used were derived mostly from a 1:50,000 scale geological map; thus, \((\bullet)_0 = 12.5\ m\) (or, say, 10 m). In data-driven modeling of prospectivity for epithermal Au deposits in the Cabo de Gata area, Carranza et al. (2008a) used spatial data sets derived mostly from 1:50,000 scale geological maps; thus, \((\bullet)_0 = 12.5\ m\) (or, say, 10 m). In data-driven modeling of geothermal prospectivity in West Java, Carranza et al. (2008c) used spatial data sets derived mostly from 1:100,000 scale geological maps; thus, \((\bullet)_0 = 25\ m\). In data-driven modeling of prospectivity for alkalic porphyry Cu–Au deposits in British Columbia, Carranza et al. (2008b) used spatial evidences derived from geological maps and airborne geophysical images with a median scale of about 1:500,000; thus, \((\bullet)_0 = 125\ m\). Each of these estimates of lower limit unit cells represents the largest possible practical scale for prospectivity mapping for the type of earth-resource in each of the areas under examination (Fig. 1). Using finer unit cells would imply that the spatial data sets used have higher spatial accuracy than the maps from which they were derived.

### 2.4. Analysis of spatial contrast between unit cells containing \(D\) and unit cells not containing \(D\)

In this analysis and the succeeding analyses of the most suitable \((\bullet)\) for prospectivity mapping for the type of earth-resource in each of the areas under examination (Fig. 1), the following ‘final’ sets of choices of suitable \((\bullet)\) were used:

1. \(10 \leq (\bullet) \leq 560\ m\), epithermal Au deposits in Aroroy district;
2. \(10 \leq (\bullet) \leq 160\ m\), epithermal Au deposits in Cabo de Gata area;
3. \(25 \leq (\bullet) \leq 750\ m\), geothermal prospects in West Java; and
4. \(125 \leq (\bullet) \leq 2000\ m\), alkalic porphyry Cu–Au deposits in British Columbia.

Any \((\bullet)\) based on each set of choices of suitable unit cell sizes results in a very small value of spatial contrast between unit cells containing \(D\) and unit cells not containing \(D\). The spatial contrast between unit cells containing \(D\) and unit cells not containing \(D\) increases non-linearly as the \((\bullet)\) increases linearly (Fig. 3). Because \(N(D)\) is constant when each \(D\) is contained in only one unit cell (except when \((\bullet)_0\) is \(\geq 330\ m\) (Fig. 2d)), the non-linear increase in the spatial contrast between unit cells containing \(D\) and unit cells not containing \(D\) is due to the non-linear decrease in value of \([N(T) - N(D)]\) as the \((\bullet)\) increases linearly. The graphs in Fig. 3 suggest that mineral prospectivity can be underestimated by using a relatively coarse \((\bullet)\) (e.g., the largest zero-probability distance derived from the point pattern analyses, Fig. 2 except Fig. 2d).

Two procedures can be performed in order to determine how the ratio \([N(D)] : [N(T) - N(D)]\) varies as \((\bullet)\) varies. Firstly, we can determine the rate (expressed in %) of increase in the ratio \([N(D)] : [N(T) - N(D)]\) based on each \((\bullet)\) to the next coarser \((\bullet)_{i+1}\) (where 1 denotes the \((\bullet)\) interval used to create the graphs in Fig. 3). Because this procedure involves analysis of spatial information

![Fig. 3. Graphs of ratio \([N(D)] : [N(T) - N(D)]\) as a function of unit cell size, \((\bullet)\), for spatial representation of locations of (a) epithermal Au deposits in Aroroy (Philippines), (b) epithermal Au deposits in Cabo de Gata (Spain), (c) geothermal occurrences in West Java (Indonesia) and (d) alkalic porphyry Cu–Au deposits in British Columbia (Canada).](image-url)
from high-to-low spatial resolutions, let us call it ‘down-scaling’ analysis. Alternatively, we can determine the rate (expressed in %) of decrease in the ratio \([N(D)] : [N(T)−N(D)]\) based on a \((\bullet)\), to the next finer \((\bullet)\), where 1 denotes the \((\bullet)\) interval used to create the graphs in Fig. 3. Because this procedure involves analysis of spatial information low-to-high spatial resolutions, let us call it ‘up-scaling’ analysis. Based on the down-scaling analysis, the graphs in Fig. 4 indicate that linear (i.e., equal interval) increase from fine to intermediate unit cells results in non-linearly decreasing rates of increase in the spatial contrast between unit cells containing D and unit cells not containing D, whereas linear (i.e., equal interval) increase from intermediate to coarse unit cells result in linearly decreasing rates of increase in the spatial contrast between unit cells containing D and unit cells not containing D. Based on the up-scaling analysis, the graphs in Fig. 5 indicate that linear (i.e., equal interval) decrease from coarse to intermediate unit cells result in linearly increasing rates of decrease in the spatial contrast between spatial information content of unit cells containing D and unit cells not containing D, whereas linear (i.e., equal interval) decrease from intermediate to fine unit cells result in non-linearly increasing rates of decrease in the spatial contrast between unit cells containing D and unit cells not containing D. Because \(N(D)\) is constant when each \(D\) is contained in only one unit cell (except when \((\bullet)\) is ≥330 m (Fig. 2d)), the rates of increase and decrease in the spatial contrast between unit cells containing \(D\) and unit cells not containing \(D\) are, respectively, due to the rates of decrease and increase in the value of \([N(T)−N(D)]\) as a function of \((\bullet)\).

The results shown in Figs. 4 and 5 are consistent with the knowledge that, in raster-based GIS (cf. Stein et al., 2001; Hengl, 2006; Nykänen and Raines, 2006), the overall spatial information content of a raster map is artificially deflated and inflated as the spatial resolution (i.e., pixel size) becomes coarser and finer, respectively. In data-driven modeling of mineral prospectivity, a binary map of \(D\) (i.e., a map of unit cells containing \(D\) and unit cells not containing \(D\)) is the vital source of spatial information. The results shown in Figs. 4 and 5 imply, therefore, that as \((\bullet)\) increases from fine to coarse the overall spatial information content of unit cells not containing \(D\) (i.e., \([N(T)−N(D)]\)) is artificially deflated while the overall spatial information content of unit cells containing \(D\) is artificially inflated. Conversely, the results shown in Figs. 4 and 5 imply that as \((\bullet)\) decreases from coarse to fine the overall spatial information content of unit cells not containing \(D\) is artificially inflated while the spatial information content of unit cell containing \(D\) is artificially deflated. So, there is a need to find a threshold \((\bullet)\) that provides a ‘balance’ between spatial information content of unit cells containing \(D\) and spatial information content of unit cells not containing \(D\).

Because the graphs in Figs. 4 and 5 depict monotonically diminishing rates of increase in the ratios of \([N(D)] : [N(T)−N(D)]\) as the \((\bullet)\) increases, it is difficult to find a \((\bullet)\) about the transition from non-linear to linear decreasing (or increasing) rates of increase (or decrease) in the ratio \([N(D)] : [N(T)−N(D)]\) plausibly provides a ‘balance’ between spatial information content of unit cells containing \(D\) and spatial information content of unit cells not containing \(D\).

2.5. Finding the most suitable \((\bullet)\)

If the individual data sets shown in Fig. 3 are plotted on log–log graphs (Fig. 6), then the positive relation between the ratio \([N(D)] : [N(T)−N(D)]\) and \((\bullet)\) can be described by power-law functions each with an exponent (or a slope) of approximately 2. The exponent of \((\bullet)\) in the relation in Fig. 6d is slight less than 2, which is probably due of the fact that \(N(D)\) is not constant because some unit cells contained more than one \(D\) when \((\bullet)\) is ≥330 m (Fig. 2d). However, the plots in Fig. 6 imply that \(N(T)\) is approximately proportional to \(1/(\bullet)^2\) and the almost constant exponent of \((\bullet)\) probably refers to the Euclidean dimension of \((\bullet)\).
Likewise, if the individual data shown in Figs. 4 and 5 are plotted on log–log graphs (Figs. 7 and 8, respectively), then the negative relation between the rate of change in the ratio \( \frac{N(D)}{N(T)} \) and equal-interval change in \( \frac{N(T) - N(D)}{C_0 \cdot N(D)} \) can be described by power-law functions with exponents (or slopes) varying between 0 and \(-2\). Inspection of the log–log plots in either Fig. 7 or 8
indicate that the curves are not actually straight but they can be fitted, by least squares method, with (at least) two straight lines, indicating that the rates of change in the ratio \( \frac{N(D)}{N(T) - N(D)} \) as function of equal-interval change in \((C_15/C_138)\) can be described by at least two power-law relations.

On the one hand, the graphs in Fig. 7 can be described by the following power-law function (cf. Mandelbrot, 1985):

\[
\Delta \frac{N(D)}{N(D)} = C(\Delta \phi)^{p-2},
\]

(2)
where \( \Delta(N(D)/N(D')) \) is the rate of increase in the ratio \( N(D)/N(D') \) (i.e., \( [N(D)]: [N(T)−N(D')] \)). \( \Delta(\bullet) \) represents equal-interval increase in \( \bullet \), \( C \) represents the constant of proportionality between \( \Delta(N(D)/N(D')) \) and \( \Delta(\bullet) \), and \( D_\Delta \) represents the scaling component or fractal dimension of the distribution of \( \Delta(N(D)/N(D')) \) with respect to \( \Delta(\bullet) \). Thus, the log–log plots in Fig. 7 imply the following. For the set of choices of unit cells suitable for spatial representation of epithermal Au deposits in the Aroyoroy district (Philippines), the rates of increase in the ratio \( N(D)/N(D') \) with respect to fine unit cells and to coarse unit cells have \( D_\Delta = 0.47 \) and 0.94, respectively. For the set of choices of unit cells suitable for spatial representation of epithermal Au deposits in the Cabo de Gata area (Spain), the rates of increase in the ratio \( N(D)/N(D') \) with respect to fine unit cells and to coarse unit cells have \( D_\Delta = 0.26 \) and 0.82, respectively. For the set of choices of unit cells suitable for spatial representation of geothermal prospects in West Java (Indonesia), the rates of increase in the ratio \( N(D)/N(D') \) with respect to fine unit cells and to coarse unit cells have \( D_\Delta = 0.38 \) and 0.89, respectively. For the set of choices of unit cells suitable for spatial representation of alkalic porphyry Cu–Au deposits in British Columbia (Canada), the rates of increase in the ratio \( N(D)/N(D') \) with respect to fine unit cells and to coarse unit cells have \( D_\Delta = 0.16 \) and 0.71, respectively.

On the other hand, the graphs in Fig. 8 can be described by the following power-law function (cf. Mandelbrot, 1983; Feder, 1988):

\[
\frac{N(D)}{N(D')} = C (\nabla(\bullet))^{-D_v},
\]

where \( \Delta(N(D)/N(D')) \) is the rate of decrease in the ratio \( N(D)/N(D') \) (i.e., \( [N(D)]: [N(T)−N(D')] \)). \( \nabla(\bullet) \) represent equal-interval decrease in \( \bullet \), \( C \) represents the constant of proportionality between \( \Delta(N(D)/N(D')) \) and \( \nabla(\bullet) \), and \( D_\nablax \) represents the scaling component or fractal dimension of the distribution of \( \Delta(N(D)/N(D')) \) with respect to \( \nabla(\bullet) \). Thus, the log–log plots in Fig. 8 imply the following. For the set of choices of unit cells suitable for spatial representation of epithermal Au deposits in the Aroyoroy district (Philippines), the rates of increase in the ratio \( N(D)/N(D') \) with respect to fine unit cells and to coarse unit cells have \( D_\nablax = 0.86 \) and 0.96, respectively. For the set of choices of unit cells suitable for spatial representation of epithermal Au deposits in the Cabo de Gata area (Spain), fine unit cells have \( D_\nablax = 0.82 \) and 0.94, respectively. For the set of choices of unit cells suitable for spatial representation of geothermal prospects in West Java (Indonesia), fine unit cells have \( D_\nablax = 0.84 \) and 0.97, respectively. For the set of choices of unit cells suitable for spatial representation of alkalic porphyry Cu–Au deposits in British Columbia (Canada), fine unit cells have \( D_\nablax = 0.81 \) and 1.03.

Thus, whereas the graphs in Figs. 4 and 5 are not essentially different, the graphs in Figs. 7 and 8 are apparently different. It is also apparent that the two values of either \( D_\Delta \) or \( D_\nablax \) describing the relationship between the rate of change in the ratio \( [N(D)]: [N(T)−N(D')] \) and equal-interval change in \( \bullet \) allow classification of a set of choices of suitable unit cell sizes derived from the distance–probability relation (Fig. 2) into a subclass of fine unit cell sizes and a subclass of coarse unit cell sizes. Fine unit cell sizes are associated with a lower value of either \( D_\Delta \) or \( D_\nablax \), whereas coarse unit cell sizes are associated with a higher value of either \( D_\Delta \) or \( D_\nablax \). Therefore, the \( \bullet \) representing the break-of-slope of two straight lines fitted into the log–log graphs in either Fig. 7 or 8 is either the coarsest fine \( \bullet \) or the finest coarse \( \bullet \). This implies that \( \bullet \) representing the break-of-slope of two straight lines fitted into the log–log graphs in either Fig. 7 or 8 represents a threshold \( \bullet \), or the most suitable \( \bullet \), that possibly provides a ‘balance’ between spatial information content of unit cells containing D and spatial information content of unit cells not containing D. Thus, the results shown in either Fig. 7 or 8 imply that (a) the most suitable \( \bullet \) for data-driven modeling of prospectivity for epithermal Au deposits in the Aroyoroy district (Philippines) is 100 m, (b) the most suitable \( \bullet \) for data-driven modeling of prospectivity for epithermal Au deposits in the Cabo de Gata area (Spain) is 60 m, (c) the most suitable \( \bullet \) for data-driven modeling of geothermal prospectivity in West Java (Indonesia) is 200 m, and (d) the most suitable \( \bullet \) for data-driven modeling of prospectivity for alkalic porphyry Cu–Au deposits in British Columbia (Canada) is 625 m.

3. Discussion

The relevance of the analysis of the most suitable \( \bullet \) for data-driven modeling of prospectivity, as illustrated in either Fig. 7 or 8, can be explained as follows. At district to regional scales, the spatial patterns of mineral deposits of certain types have been demonstrated to satisfy certain power-law functions indicating that they can be described by either a box-count fractal dimension or a radial-density fractal dimension (Carlson, 1991; Cheng and Agterberg, 1995; Cheng et al., 1996, Wei and Pengda, 2002; Ford and Blenkinsop, 2008; Raines, 2008; Carranza, 2008). Similarly, at deposit scales (i.e., at millimeter scales), the distributions of thicknesses of mineralized veins have been described by certain power-law functions (Sanderson et al., 1994; Clark et al., 1995; McCaffrey and Johnston, 1996; Roberts et al., 1998; Monecke et al., 2001). Because the spatial patterns and the distributions of particular properties of certain types of mineral deposits can be characterized by certain power-law functions, it follows that the rates of change in the ratio \( [N(D)]: [N(T)−N(D')] \) as function of equal-interval change in \( \bullet \) can be characterized by certain power-law functions. Accordingly, the results in either Fig. 7 or 8 show that the distribution of the rates of change in the ratio \( [N(D)]: [N(T)−N(D')] \) as function of equal-interval change in \( \bullet \) has at least two fractal dimensions, which allow distinction of a threshold \( \bullet \) in a set of choices of suitable unit cell sizes derived from the distance–probability relation (Fig. 2). Therefore, the results imply that fractal analysis of the distribution of the rates of change in the ratio \( [N(D)]: [N(T)−N(D')] \) as function of equal-interval change in \( \bullet \) is essential in the selection of the most suitable \( \bullet \) for data-driven modeling of mineral prospectivity.

The analysis illustrated in either Fig. 7 or 8 is comparable to the fractal analysis proposed by Cheng et al. (1994) to separate geochemical anomalies from background. However, it is not clear why the power-law relation depicted in Eq. (2) is similar to the power-law relation describing the radial-density fractal dimension of a spatial pattern of set of point objects (Mandelbrot, 1983) and why the power–law relation depicted in Eq. (3) is similar to the power–law relation describing the box-count fractal dimension of spatial pattern of a set of (point) objects (Mandelbrot, 1985; Feder, 1988). These are outstanding questions that need further research. In addition, the results depicted in Figs. 7 and 8 raise the following questions for further research, not only for geoscientists involved with spatial data analysis for geo-information extraction, but especially for geo-information scientists. Firstly, “Is the distribution of information content of a map fractal”? Secondly, “In down-scaling (or up-scaling) a map to a lower (or higher) resolution, is there a threshold grid (or pixel) size at which the predictive capabilities or information content of a map is diminished”?

In contrast to and notwithstanding of the results of the analyses (illustrated in either Fig. 7 or 8) for objective selection of the most suitable \( \bullet \) for data-driven modeling of mineral prospectivity, the following previous works of GIS-based data-driven modeling of mineral prospectivity each used a \( \bullet \) based on subjective judgment and/or consideration of the
distance–probability relation in the spatial pattern of $D$ (Fig. 2). In data-driven modeling of prospectivity for epithermal Au deposits in the Cabo de Gata area (Spain), Carranza et al. (2008a) selected and used a ($\bullet$) of 100 m, which is within the range of distances in which there is zero probability of one neighbour epithermal Au deposit location situated next to another epithermal Au deposit location (Fig. 2b) and which is coarser than the most suitable ($\bullet$) of 60 m suggested by the results shown in either Fig. 7b or 8b. In data-driven modeling of geothermal prospectivity in West Java (Indonesia), Carranza et al. (2008c) chose and used a ($\bullet$) of 500, which is within the range of distances in which there is zero probability of one neighbour geothermal prospect situated next to another geothermal prospect (Fig. 2c) and which is coarser than the most suitable ($\bullet$) of 200 m suggested by the results shown in either Fig. 7c or 8c. In data-driven modeling of prospectivity for alkalic porphyry Cu–Au deposits in British Columbia, Carranza et al. (2008b) used a ($\bullet$) of 1 km, which is outside the range of distances in which there is zero probability of one neighbour alkalic porphyry Cu–Au deposit location situated next to another alkalic porphyry Cu–Au deposit location (Fig. 2d) and which is coarser than the most suitable ($\bullet$) of 625 m suggested by the results displayed in either Fig. 7d or 8d. In these case studies, the common reason for arbitrarily choosing and using a ($\bullet$) that is coarser than the most suitable ($\bullet$) is simplicity of area calculations (i.e., as multiples of number of unit cells or pixels) in the application of a raster-based or pixel-based GIS.

Unit cell sizes that are much finer than the most suitable ($\bullet$) are considered impractical for spatial representation of $D$ because they are likely to result in artificial deflation of spatial information content from unit cells containing $D$ and artificial inflation of spatial information content from unit cells not containing $D$. Unit cell sizes that are much coarser than the most suitable ($\bullet$) are also considered impractical for spatial representation of $D$ because they are likely to result in artificial inflation of spatial information content from unit cells containing $D$ and artificial deflation of spatial information content from unit cells not containing $D$. Thus, in the case studies mentioned in the preceding paragraph, it is likely that estimates of prospectivity for the type of earth-resource examined have been overestimated. In the following section, a case study in the Aroroy district is provided to compare weights of spatial evidence estimated by using different unit cell sizes including the most suitable ($\bullet$) implied by the results shown in either Fig. 7a or 8a.

4. Case application

4.1. Study area: Aroroy district (Philippines)

The Aroroy district, in the northwestern portion of Masbate Island in the Philippines (Fig. 9), is one of the several districts in the Philippines endowed with epithermal Au deposits (Mitchell and Balce, 1990; Mitchell and Leach, 1991). In this district, measuring about 130 km$^2$, the oldest rocks belong to the Eocene-Oligocene Mandaon Formation (Fig. 9), consisting mostly andesitic-dacitic agglomerates (Baybayan and Matos 1986; JICA-MMAJ 1986). The Mandaon Formation is intruded by the Miocene Aroroy Diorite, which varies in composition from quartz diorite to hornblende diorite. Unconformably overlying the Mandaon Formation and the Aroroy Diorite are feldspathic wackes belonging to the Early Miocene Sambulawan Formation. Andesitic lithic tuffs (Late Miocene to Early Pliocene Lanang Formation) disconformably overlap the Mandaon and Sambulawan Formations. The series of dacitic-andesitic volcano-sedimentary rocks (i.e., the Mandaon, Sambulawan and Lanang Formations) and the Aroroy Diorite are intruded by the Pliocene Nabongsoran Andesite, which consists of porphyritic stocks, plugs and dikes. The Nabongsoran Andesite porphyry intrusions are probably responsible for hydrothermal alteration and epithermal Au

![Fig. 9. Aroroy district (Philippines): simplified geological map (adapted from Baybayan and Matos (1986) and JICA-MMAJ (1986)). Map coordinates are in meters (UTM projection, zone 51). Map limits of Aroroy district serve as scale for Masbate Island inset, which in turn serves as scale for Philippine Islands inset.](image-url)
mineralization in the intruded rocks (Mitchell and Balce, 1990; Mitchell and Leach, 1991). Gold, in at least 13 mineral deposit occurrences in the district (Fig. 5), is associated with sulphide (mainly pyritic) minerals in wide-sheeted and manganese-bearing quartz veins or silicified breccias in generally northwest-trending faults/fractures that cut the volcano-sedimentary rocks.

Carranza (2008) used geological and geochemical data sets in the Aroroy district to describe various GIS-aided and/or GIS-based techniques for geochemical anomaly and mineral prospectivity mapping. Among the layers of spatial evidence that he used for mapping epithermal Au prospectivity in the Aroroy district are distances to intersections of north-northwest (NNW) and north-west (NW) trending faults/fractures (representing structural controls and proxy for heat source controls; hereafter denoted as PROXIMITY) and integrated principal component (PC3 × PC4) scores of stream sediment geochemical data (representing geochemical anomaly evidence; hereafter denoted as ANOMALY). In the following section, a comparative analysis of weights of classes of PROXIMITY and weights of classes of ANOMALY, using different unit cell sizes including the most suitable (●), is given in order to show the usefulness of the proposed methodology for objective selection of the most suitable (●).

4.2. Analysis of weights of spatial evidence mineral prospectivity

Several methods exist for calculation of weights of spatial evidence of mineral prospectivity and these methods can be classified into two types of techniques: bivariate and multivariate. On the one hand, bivariate techniques involve pairwise analysis of spatial association between a map of D and a map of spatial evidence discretized into a number of classes. On the other hand, multivariate techniques involve simultaneous analysis of spatial associations between a map of D and maps of spatial evidence each of which is discretized into a number of classes. Here, a popular bivariate technique – weights-of-evidence (WoE) method – is applied to show the usefulness of the proposed methodology for objective selection of the most suitable (●).

The WoE method, which is based on a Bayesian conditional probability framework (Good, 1950), is applicable for data-driven mapping of mineral prospectivity in areas where a number of occurrences of mineral deposits of the type sought are known (Bonham-Carter et al., 1989; Agterberg et al., 1990). WoE modeling of mineral prospectivity is a 3-stage process: (1) estimation of prior probability of mineral deposit occurrence; (2) estimation of weights of spatial evidence with respect to known mineral deposit occurrences; and (3) updating of the prior probability by the weights of spatial evidence in order to estimate posterior probability of mineral deposit occurrence. Here, the focus is to show changes in the weights and uncertainties of weights of spatial evidence as (●) changes.

Suppose that T (in this case Aroroy district, Fig. 1a or 8) is divided into a number of N(T) equal-area unit cells, of which there are N(D) number of unit cells each containing just one D (in this case epithermal Au deposits). Suppose further that in T there are N(Ci) and N(Ći) numbers of unit cells where spatial evidence Ci is present and absent, respectively. A positive (W+) weight and a negative (W-) weight are calculated for N(Ci) and N(Ći), respectively, given presence of D (i.e., N(D)) and absence of D (i.e., N(D) or [N(T)–N(D)]). The W+ and W- can be estimated, respectively, as

$$W^+ = \log_e \frac{N(C_i \cap D) - N(D)}{N(C_i \cap D) - N(D)};$$

and

$$W^- = \log_e \frac{N(Ć_i \cap D) - N(D)}{N(Ć_i \cap D) - N(D)}.$$
Fig. 11. Aroroy district (Philippines): results of using different unit cell sizes ($10 \times 10$, $100 \times 100$, $150 \times 150$ and $200 \times 200$ m$^2$) to derive, with respect to epithermal Au deposits, values of $W^-$ for (a) cumulative ascending distances to intersections of NNW- and NW-trending faults and (b) cumulative descending values of integrated PC3 and PC4 scores of stream sediment geochemical data.

Fig. 12. Aroroy district (Philippines): results of using different unit cell sizes ($10 \times 10$, $100 \times 100$, $150 \times 150$ and $200 \times 200$ m$^2$) to derive, with respect to epithermal Au deposits, values of $s(W^-)$ for (a) cumulative ascending distances to intersections of NNW- and NW-trending faults and (b) cumulative descending values of integrated PC3 and PC4 scores of stream sediment geochemical data.
As measures of uncertainties of weights of spatial evidence, the standard deviation \( s(W^+\|D) \) of \( W^+\) and the standard deviation \( s(W^-\|D) \) of \( W^-\) can be estimated, respectively, as follows:

\[
 s(W^+) = \sqrt{\frac{1}{\mu(D) N(C_i \cap D)} + \frac{1}{\mu(D) N(C_i \cap D)}},
\]

\[
 s(W^-) = \sqrt{\frac{1}{\mu(D) N(C_i \cap D)} + \frac{1}{\mu(D) N(C_i \cap D)}}.
\]

It is clear in Eqs. (4)–(7) that, since \( N(D) \) is constant when each \( D \) is contained in only one unit cell, the weights and their uncertainties depend on \( N(D) \). Because \( N(D) \) decreases as the \( \bullet \) increases, we can expect from Eqs. (4) and (5) that the standard deviations are lower than those derived based on a \((\circ)\) of either 50 or 100 m. In addition, Figs. 10 and 12 both show that based on a \((\circ)\) of either 50 or 100 m the calculated values of \( W^-\) are almost identical and the calculated standard deviations of the weights are also almost identical. Figs. 11 and 13 show, respectively, that based on a \((\circ)\) of either 150 or 200 m the calculated values of \( W^-\) are smaller than those derived based on a \((\bullet)\) of either 10 or 100 m. In addition, Figs. 10 and 12 both show that the standard deviations of the weights derived based on a \((\circ)\) of either 50 or 100 m are almost identical and the calculated standard deviations of the weights are also almost identical. The almost identical weights and almost identical standard deviations of the weights derived by using a \((\circ)\) of 10 and 100 m indicate that (a) the latter \((\bullet)\) is ‘suitably small’ for WofE modeling of epithermal Au prospectivity in the Aroroy district and (b) the proposed method for objective selection of the most suitable \((\circ)\) is satisfactory.

5. Discussion

It is considered here that the analysis of the weights and uncertainties of the weights of spatial evidence is sufficient to demonstrate the usefulness of the proposed methodology for objective selection of the most suitable \((\circ)\) in data-driven modeling of mineral prospectivity via the WofE method using a raster-based GIS. The reason for this is that, in WofE, if weights of spatial evidence are over estimated and are associated with high uncertainty, then the following undesirable results are likely to be obtained (Agterberg, 1992; Carranza, 2004): (a) over-estimation of posterior probability of mineral deposit occurrence; and (b)
violation of the assumption of conditional independence among layers of spatial evidence with respect to the mineral deposit occurrences. The results in Figs. 10–13 suggest, therefore, that among the four unit cell sizes tested, undesirable results are likely to be obtained by using a (●) of either 150 or 200 m, while satisfactory results are likely to be obtained by using a (○) of either 10 or 100 m. However, a (●) of 10 m, compared to a (○) of 100 m, is an impractical spatial representation of the sizes (i.e., lengths or widths) of areas mineralized with auriferous epithermal veins/brecias in the Aroroy district (Mitchell and Leach, 1991, p. 281). Therefore, the lateral extents of mineral deposits under examination must still be considered in the final choice of the (○) to be used in data-driven modeling of mineral prospectivity, although in the Aroroy district complete data on lateral extents of individual epithermal Au deposits are unavailable to the author’s disposal. Nevertheless, it can be postulated that, because certain properties of mineral deposits at deposit scales (e.g., vein thickness) can be described by certain power-law functions (Sanderson et al., 1994; Clark et al., 1995; McCaffrey and Johnston, 1996; Roberts et al., 1998; Monecke et al., 2001), it is likely fractal analysis will be useful in objective selection of the most suitable (●) based on data of lateral extents of mineral deposits of the type sought. This proposition remains to be investigated and demonstrated.

6. Summary and conclusions

A unit cell size chosen for spatial representation of mineral deposit occurrences of the type sought and for discretization of individual layers of evidential data sets can be crucial aspect in data-driven modeling of mineral prospectivity using a raster-based GIS. Until now, the selection of a suitable unit cell size in most, if not all, studies of GIS-based data-driven modeling of mineral prospectivity has been subjective. By applying point pattern analysis to a set of points representing known occurrences of mineral deposits of the type sought, a range of distances in which there is zero probability of one deposit occurrence situated next to another deposit occurrence can be obtained as a set of choices of suitable unit cell sizes, each of which satisfies the assumption in prospectivity mapping that each mineral deposit occurrence is contained in only one unit cell. The upper limit of the choices of suitable unit cell sizes indicated by the distance–probability relation in a point pattern of known occurrences of mineral deposits of the type sought can be extended in consideration of either knowledge of lateral extents of mineral deposit of the type sought or a rule-of-thumb distance for combining mineral occurrences into one mineral deposit. Alternatively, an analysis of reflexive nearest neighbours in a point pattern of known occurrences of mineral deposits of the type sought can be useful in deciding whether to extend the upper limit of a set of choices of suitable unit cell sizes indicated by the distance–probability relation. Whether the upper limit of the choices of suitable unit cell sizes indicated by the distance–probability relation in point pattern of known occurrences of mineral deposits of the type sought is extended or not, a lower limit of the choices of suitable unit cell sizes must also be defined. This lower limit unit cell size can be estimated as the product of the finest legible size of point objects in a map and the map scale factor. Finally, fractal analysis of the rates of increase or decrease in the spatial contrast between unit cells containing mineral deposit occurrences and unit cells without mineral deposit occurrences as a function of equal-interval change in unit cell size is favourable in objective selection of the most suitable unit cell size. This and the implications of fractal analysis of the most suitable unit cell size need further investigation however.

The most suitable unit cell sizes derived via the proposed methodology are finer than the unit cell sizes chosen subjectively for data-driven modeling of prospectivity for earth resources of the type sought in three different areas (Cabo de Gata (Spain), West Java (Indonesia), British Columbia (Canada). In the Aroroy district (Philippines) case study, the most suitable unit cell size derived via the proposed methodology results in weights and uncertainties of weights of spatial evidence of mineral prospectivity that are closely identical to those derived by using the finest unit cell size and they are lower than those derived by using coarser unit cell sizes. The methodology presented here for objective selection of the most suitable unit cell size for data-driven modeling of mineral prospectivity using a raster-based GIS is reproducible and robust regardless of the number of known occurrences of mineral deposits of the type sought and the size of the study region or district.

Acknowledgement

The author would like to thank Frits Agterberg and an anonymous reviewer for their constructive comments on earlier drafts of the manuscript.

References


