Landscape Ecology **00:** 1–4, 2002.

A mathematical comment on the formulae for the aggregation index and the shape index

Jan Bogaert[1,2,*], Ranga B. Myneni[1] and Yuri Knyazikhin[1]
*[1]Department of Geography, Climate and Vegetation Research Group, Boston University, Boston, MA 02215-1401, USA; [2]Department of Biology, Research Group of Plant and Vegetation Ecology, University of Antwerp, Universiteitsplein 1, B-2610 Wilrijk, Belgium; *Author for correspondence (e-mail: jan.bogaert@ua.ac.be)*

Received 9 February 2001; accepted in revised form 19 December 2001

Key words: Aggregation index, Index redundancy, Landscape metric, Perimeter, Pixel edge, Pixel geometry, Shape index, Spatial pattern

Abstract

In a recent paper [Landscape Ecol. 15: 591–601 (2000)] He et al. described an aggregation index AI_i to measure pixel aggregation within a single class i. We show that the commonly used shape index SI_i is related to the proposed aggregation metric as $SI_i = \Phi(A_i) + AI_i(1 - \Phi(A_i))$, with $\Phi(A_i)$ dependent on class area A_i only. Moreover, it is shown that the normalized shape index, SI_i^*, equals $(1 - AI_i)$. We conclude that AI_i does not provide any information not provided by SI_i, or SI_i^*.

Introduction

Recently, an aggregation index (*AI*), designed to quantify landscape spatial patterns for raster data, was presented and tested (He et al. 2000). It was initially applied by Miller et al. (1997) as a 'cluster/interspersion index' in a study relating landscape pattern metrics to land use and to changes in biotic communities. The index responds to demands for quantification of aggregation levels within a single class. Calculation of AI_i for class i is based on the pixel edges $e_{i,i}$ shared with itself, i.e.,

$$AI_i = \frac{e_{i,i}}{\max\_e_{i,i}}, \tag{1}$$

with $\max\_e_{i,i}$ the largest number of possible edges shared for class i. The value of $e_{i,i}$ is also known as the 'contact perimeter' (Bribiesca 1997). The procedure for calculating $\max\_e_{i,i}$ is based on the largest square integer (n^2) smaller than the area A_i of class i. Note that A_i is an integer value, exceeding zero, i.e. $A_i \in \mathbb{Z}_+^0$. The shape index *SI* is another class specific index, calculated as:

$$SI_i = \frac{P(A_i)}{4\sqrt{A_i}}, \tag{2}$$

with SI_i the value for the i-th class, and $P(A_i)$ the total perimeter. SI_i is based on the area-to-perimeter relationship $P = 4\sqrt{A}$ for a square shape and measures deviation of the perimeter from that of a perfect square. Three justifications were given by He et al. (2000) for proposing *AI* as a substitute for *SI*. First, *SI* assumes that a square shape is the most aggregated shape, although due to pixel geometry, a square cannot always be constructed using A_i pixels, as was already mentioned by Milne (1991). Second, *SI* has no upper bound, which makes its interpretation difficult. Third, *SI* is based on perimeter calculation, and hence uses the edges $e_{i,j}$ between different classes i and j. He et al. (2000) argue that *AI*, although similar to *SI*, quantifies aggregation more precisely. Here, we present formulae for use of SI_i with raster data, and to normalize SI_i. Consequently, we provide mathematical arguments to illustrate a direct relationship between AI_i and SI_i.

Relationship between the aggregation index and the shape index

The minimum perimeter $P_{min}(A_i)$ for aggregates of A_i square pixels is calculated as (Milne 1991; Bogaert et al. 2000):

$$P_{\min}(A_i) = 4\text{int}\sqrt{A_i} + c, \tag{3}$$

int(x) is a function that truncates x at its decimal point, i.e. $|x - \text{int}(x)| < 1$, and,

$$c = \begin{cases} 0 & \text{if} \quad \text{int}\sqrt{A_i} = \sqrt{A_i}; \\ 2 & \text{if} \quad (\text{int}\sqrt{A_i})^2 < A_i \leqslant (\text{int}\sqrt{A_i})(1 + \text{int}\sqrt{A_i}); \\ 4 & \text{if} \quad A_i > (\text{int}\sqrt{A_i})(1 + \text{int}\sqrt{A_i}). \end{cases}$$

Note that $\text{int}\sqrt{A_i} = n$, and for $A_i \gg 1$, $4\text{int}\sqrt{A_i} + c \simeq 4\sqrt{A_i}$, as in Equation (2). This is illustrated in Figure 1A, and is mathematically written as:

$$\lim_{A_i \to \infty} \frac{4\text{int}\sqrt{A_i} + c}{4\sqrt{A_i}} = 1. \tag{4}$$

As a consequence of pixel geometry, Equation (2) can be applied if $\text{int}\sqrt{A_i} = \sqrt{A_i}$, hence if $A_i = n^2$, or, by approximation, for $A_i \gg 1$. For vector data, Equation (2) is replaced by

$$SI_i = \frac{P(A_i)}{2\sqrt{\pi A_i}}, \tag{5}$$

based on the relation $P = 2\sqrt{\pi A}$ for disks (McGarigal and Marks 1995). The application of Equation (2) is hence mortgaged by (i) erroneous outcomes for values of A_i not equal to n^2 or not large enough to use the approximation for $A_i \gg 1$, and by (ii) the existence of an alternate formula, i.e. Equation (5), for vector data structures, where the circle shape can always be used to find the smallest perimeter of a given area. Note that both formulae (Equations 2 and 5) were initially designed for shape assessment of single patches (McGarigal and Marks 1995). They can be just as well applied to multiple patches of a single class, because these formulae are composed only of area and perimeter data. The index outcomes should be therefore interpreted with caution, because multiple patches characterized by a perfect isodiametric shape will generate an overall outcome exceeding that of a single isodiametric patch.

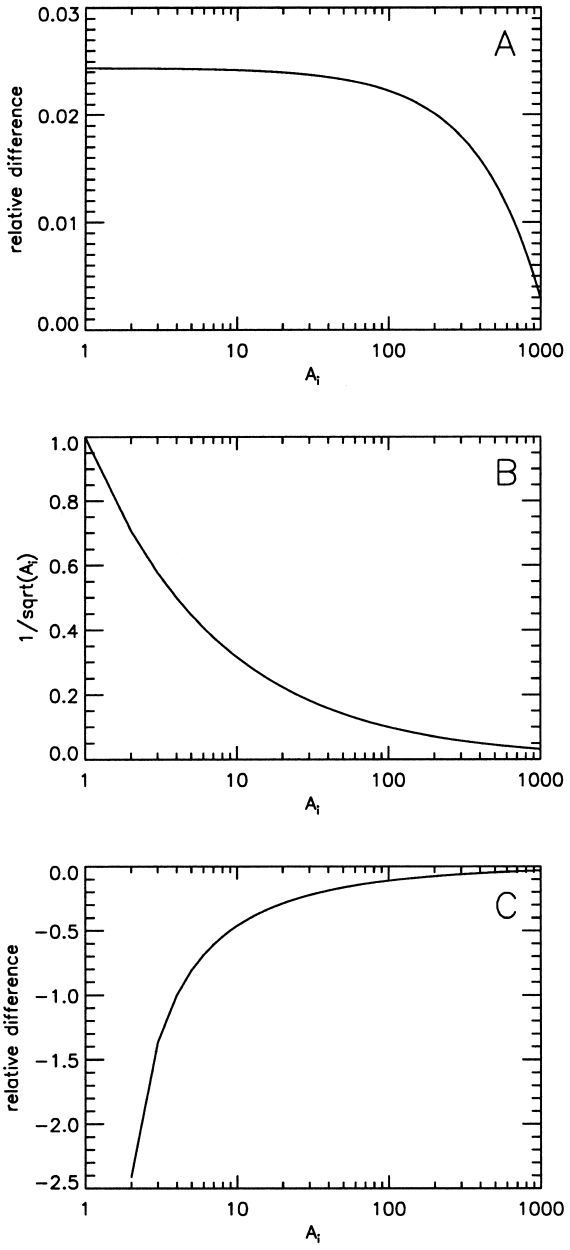

Figure 1. Approximations for $A_i \gg 1$, illustrated for $1 \leqslant A_i \leqslant 1{,}000$, leading to a more simple calculation of Equations (3) and (9). (A) Polygon fit ($\chi^2 = 0.139$) showing the decreasing trend in the relative difference between $(4\text{int}\sqrt{A_i} + c)$ and $4\sqrt{A_i}$, calculated as $(4\text{int}\sqrt{A_i} + c - 4\sqrt{A_i})/(4\text{int}\sqrt{A_i} + c)$. (B) Decreasing trend of $1/\sqrt{A_i}$ as a function of A_i. (C) Decreasing trend of the relative difference between $(\sqrt{A_i} - 1)$ and $\sqrt{A_i}$, calculated as $(\sqrt{A_i} - 1 - \sqrt{A_i})/(\sqrt{A_i} - 1) = (-1)/(\sqrt{A_i} - 1)$.

The general formula for perimeter calculation of pixel aggregates is given by (Bribiesca 1997; Bogaert

et al. 2000):

$$P(A_i) = 4A_i - 2e_{i,i}. \quad (6)$$

Consequently, for $e_{i,i} = \max\_e_{i,i}$, $P(A_i) = P_{\min}(A_i)$. Note that for $e_{i,i} = 0$, $P(A_i) = P_{max}(A_i) = 4A_i$, with $P_{max}(A_i)$ the maximum perimeter (Bribiesca 1997; Bogaert et al. 2000). $P_{max}(A_i)$ is the theoretical upper limit of $P(A_i)$, and requires that no neighboring pixels belong to the same class i. However, this will exceptionally occur in classified remote sensing data or in any thematic data layer that describes a natural phenomena (Johnsson 1995). Using Equation (3) and Equation (6), Equation (2) can be substituted by

$$SI_i = \frac{P(A_i)}{P_{\min}(A_i)} = \frac{4A_i - 2e_{i,i}}{4A_i - 2\max\_e_{i,i}} = \frac{4A_i - 2e_{i,i}}{4\text{int}\sqrt{A_i} + c}, \quad (7)$$

for use with raster data. Note that for Equation (7), SI_i has an upper and lower limit, i.e.

$$1 \leqslant SI_i \leqslant \frac{P_{\max}(A_i)}{P_{\min}(A_i)}, \quad (8)$$

and hence SI_i can be normalized as SI_i^*, i.e.

$$SI_i^* = \frac{SI_i - 1}{P_{\max}(A_i)(P_{\min}(A_i))^{-1} - 1} = \frac{P(A_i) - P_{\min}(A_i)}{P_{\max}(A_i) - P_{\min}(A_i)}. \quad (9)$$

The upper limit of SI_i reflects the effect of pixel geometry on the data, and hence accounts for resolution when comparing data sets. It should be noted that the use of pixel number in defining a reference value is also applicable to AI_i in Equation (1), where $\max\_e_{i,i}$ is dependent on A_i, and that it is the only objective parameter for standardizing a metric for pixel configuration (Milne 1991; Bogaert et al. 2000). For $A_i \gg 1$, it can be accepted that $(\sqrt{A_i})^{-1} \simeq 0$ and $\sqrt{A_i} - 1 \simeq \sqrt{A_i}$. The validity of these approximations is illustrated in Figures 1B and 1C, and is mathematically given by:

$$\lim_{A_i \to \infty} \frac{1}{\sqrt{A_i}} = 0, \quad (10)$$

$$\lim_{A_i \to \infty} \frac{\sqrt{A_i} - 1}{\sqrt{A_i}} = 1 - \frac{1}{\sqrt{A_i}} = 1. \quad (11)$$

Thus, SI_i^* can be approximated for large values of A_i as (cf. Equations (4), (11) and (12)),

$$SI_i^* \simeq \frac{P(A_i)}{4A_i} \simeq \frac{SI_i}{\sqrt{Ai}}, \quad (12)$$

or, mathematically,

$$\lim_{A_i \to \infty} \frac{SI_i^*}{SI_i}\sqrt{A_i} = 1. \quad (13)$$

The use of SI_i^* instead of SI_i hence addresses the two limitations of Equation (2); SI_i^* generates the correct outcome for all values of A_i by accounting for pixel geometry, and is easy to interpret because of its range [0,1]. The relation between SI_i and AI_i is found by combining Equations (1) and (7):

$$SI_i = \frac{4A_i - 2AI_i\max\_e_{i,i}}{4\text{int}\sqrt{A_i} + c}, \quad (14)$$

which can be simplified into

$$SI_i = \Phi(A_i) + AI_i(1 - \Phi(A_i)), \quad (15)$$

with $\Phi(A_i)$ dependent on A_i only, i.e.

$$\Phi(A_i) = \frac{4A_i}{4\text{int}\sqrt{A_i} + c} = \frac{P_{\max}(A_i)}{P_{\min}(A_i)}, \quad (16)$$

with $\Phi(A_i) > 1$ if $A_i > 1$, and $\Phi(A_i)$ equal to the upper bound of SI_i. Note that Equation (15) does not apply for $A_i = 1$ which generates $P_{max}(1) = P_{min}(1) = 4$ and $\Phi(1) = 1$, and because for a class composed of a single pixel, $e_{i,i} = \max\_e_{i,i} = 0$, and Equation (1) cannot be used. To obtain the relation between SI_i^* and AI_i, Equation (1) is rewritten using $e_{i,i} = (4A_i - P(A_i))/2$ and $\max\_e_{i,i} = (4A_i - P_{min}(A_i))/2$ as

$$AI_i = \frac{4A_i - P(A_i)}{4A_i - P_{\min}(A_i)} = \frac{P_{\max}(A_i) - P(A_i)}{P_{\max}(A_i) - P_{\min}(A_i)}, \quad (17)$$

or,

$$P(A_i) = P_{max}(A_i) - AI_i(P_{\max}(A_i) - P_{\min}(A_i)). \quad (18)$$

$P(A_i)$ in Equation (9) can be substituted by the above which results,

$$SI_i^* = 1 - AI_i. \quad (19)$$

These obvious relations between AI_i, SI_i^*, and SI_i (cf. Equations (15) and (19)) indicate that these indices are not independent, and that they duplicate information, i.e. exhibit index redundancy, which is preferably avoided (O'Neill et al. 1988).

Note that we assume

$$e_i = e_{i,i} + e_{i,j}, \quad (20)$$

with e_i the total number of edges for class i, and that $e_{i,j} = P(A_i)$. If pixels of class i touch the grid or landscape edge, they are included in the count of $e_{i,j}$ (Pearson et al. 1999). The edges between pixels of class i are tallied once (He et al. 2000), such that

$$4A_i = 2e_{i,i} + e_{i,j}. \quad (21)$$

Thus, we note that (i) Equation (7) adapts SI_i for raster data, (ii) SI_i^* defined in Equation (9) is standardized by its maximum and minimum, and thus ranges from 0 to 1, and (iii) Equations (15) and (19) directly relate AI_i to SI_i and SI_i^*, respectively. Therefore we conclude that AI_i does not provide any information not provided by SI_i, or SI_i^*.

Summary

The aggregation metric AI_i, proposed as a substitute for the shape index SI_i, was supported by three justifications (He et al. 2000): it accounts for pixel geometry, it has an upper bound, and it measures $e_{i,i}$, the total edges shared by class i with itself. AI_i was suggested to measure aggregation more precisely than SI_i. Using perimeter equations, we define SI_i for use with raster data, together with its upper and lower limit, which leads to normalization of SI_i as SI_i^*. Using the ratio $\Phi(A_i)$ of the theoretical maximum and minimum perimeter, we show that SI_i equals $\Phi(A_i) + AI_i(1 - \Phi(A_i))$. Further, we find that $SI_i^* = 1 - AI_i$. To avoid index redundancy, only one of the metrics AI_i, SI_i, and SI_i^* should be used when quantification of aggregation levels within a single class is required.

Acknowledgements

The authors acknowledge D. Mladenoff for bringing the Miller et al. (1997) article to our attention. The authors thank the referees for their comments and critiques. J. Bogaert is a Postdoctoral Fellow of the Fund for Scientific Research – Flanders (Belgium) (F.W.O. – Vlaanderen).

References

Bogaert J., Rousseau R., Van Hecke P. and Impens I. 2000. Alternative area-perimeter ratios for measurement of 2-D shape compactness of habitats. Applied Mathematics and Computation 111: 71–85.

Bribiesca E. 1997. Measuring 2-D shape compactness using the contact perimeter. Computers and Mathematics with Applications 33: 1–9.

He H.S., DeZonia B.E. and Mladenoff D.J. 2000. An aggregation index (*AI*) to quantify spatial patterns of landscapes. Landscape Ecology 15: 591–601.

Johnsson K. 1995. Fragmentation index as a region based GIS operator. International Journal of Geographical Information Systems 9: 211–220.

McGarigal K. and Marks B.J. 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Rep. PNW–GTR–351. US Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR, USA, 122 p.

Miller J.N., Brooks R.P. and Croonquist M.J. 1997. Effects of landscape patterns on biotic communities. Landscape Ecology 12: 137–153.

Milne B.T. 1991. Lessons from applying fractal models to landscape patterns. In: Turner M.G. and Gardner R.H. (eds), Quantitative Methods in Landscape Ecology – The Analysis and Interpretation of Landscape Heterogeneity. Springer-Verlag, New York, NY, USA, pp. 199–235.

O'Neill R.V., Krummel J.R., Gardner R.H., Sugihara G., Jackson B., DeAngelis D.L. et al. 1988. Indices of landscape pattern. Landscape Ecology 1: 153–162.

Pearson S.M., Turner M.G. and Urban D.L. 1999. Effective exercises in teaching landscape ecology. In: Klopatek J.M. and Gardner R.H. (eds), Landscape Ecological Analysis – Issues and Applications. Springer-Verlag, New York, NY, USA, pp. 335–368.