Landsc Ecol
https://doi.org/10.1007/s10980-021-01393-6

PERSPECTIVE

A simplified perspective on the index of spatial autocorrelation

Thomas J. DeWitt

Received: 30 July 2021 / Accepted: 26 December 2021

Abstract
Context A recent article in *Landscape Ecology* presented a method to true Moran's I to its conceptual ideals and existing intuition regarding correlations. It's scope included multiple methods, exploration of designed and empirical datasets, sensitivity analyses, and extensive mathematical treatment. The editor, reviewers, and lead author feared that the article due to complexity would not be accessible to empirical landscape ecologists.
Objectives This perspective aims to highlight critical problems with the traditional autocorrelation metric and make standout material from the larger analysis accessible by paring to essentials and presenting a simple recipe to calculate an improved metric that also serves as a statistic.
Methods Desirable traits for an autocorrelation metric were reviewed followed by distillation of best practices discerned in the larger project to attain those traits. A minimal method to obtain the superior metric was formulated.
Results Moran's I met only 2 of 14 desirable qualities for indexing autocorrelation. An improved metric was found to be achievable in 7 steps. The new metric, now a statistic, realized 14 of 14 desirable traits. The new statistic fit existing intuition for regular correlation and facilitated comparisons across disparate contexts.
Conclusions Spatial autocorrelation is a common focus in landscape ecology. The new statistic enabled intuitive interpretation and meaningful comparison within and among studies. It provided for meta-analysis and meta-research, such as co-use with other spatial pattern statistics. These improvements should foster sustained use and impact of the new autocorrelation statistic I_r.

Keywords Spatial statistics · Moran's I · Geospatial analysis · Landscape pattern · Spatial regression · Matrix correlation · Index of clumping

Introduction

In a recent *Landscape Ecology* article, DeWitt et al. (2021) enumerated 14 qualities logically and functionally desirable for a metric of spatial autocorrelation (Table 1). The traditional measure, Moran's I, demonstrated only two of the 14 desirable qualities. Yet I has a long history in landscape analysis and the

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10980-021-01393-6.

T. J. DeWitt (✉)
Department of Ecology & Conservation Biology, Texas A&M University, College Station 77843-2258, TX, USA
e-mail: tdewitt@tamu.edu

Table 1 Qualities expected, assumed, or desirable for autocorrelation metrics

Expected or desired property of autocorrelation metric	I .	I_r
Responsive to gradient data dispersion	+	+
Responsive to non-gradient data clumping	+	+
Distinguishes clumping from gradient effects	−	**~1**
Strictly bounded between -1 and 1	−	+
Value for perfect interspersion:		
− 1 or commensurate with expectations for regular correlation	−	+
Value for perfect gradient:		
1 or commensurate with expectations for regular correlation	−	+
Value for gradient-free clustering:		
1 or commensurate with expectations for regular correlation	−	~
Central tendency for random patterns:		
Mean is 0	−	+
Median is 0	−	+
Distribution of random pattern metrics is symmetric (oppositive)	−	+
Metric has stand-alone interpretability (contextualized by n)	−	+
Metric can be compared among studies (contextualized by n)	−	+
Metric itself is a standard effect size	−	+
Metric itself is a test statistic	−	+
Derived null distributions retain original type I error rate	n/a	+

Key: +, −, and ∼ indicate the metric meets, fails, or approximately (or weakly) fits the criterion

[1] Hotspot and gradient structures can be separated by using I_r and multiple r together as I_r^2 / (I_r^2 + r^2) plotted against I_r.

ideals it is widely thought to represent are valuable existing capital if the metric can be transformed to have most or all of the desired qualities. In the article, two approaches were conceived and tested to reconcile I with its existing image and ideals. One method fit traditional I to the theoretical distribution of regular correlations. This was found to retain the 2 qualities that Moran's I meets while also satisfying the remaining 12. The main study extensively developed both methods, for example to include sensitivity analyses, and tested them with empirical and designed datasets. This *Perspective* presents an overview of tenets and best practices revealed by the broader article.

Moran's index of spatial autocorrelation is often thought of and taught as scaling between − 1 and 1 where − 1 indicates perfect interspersion (overdispersion), 0 indicates no spatial patterning, and 1 represents perfect data clumping or gradient structure. Although the metric is responsive to spatial dispersion, its values are not readily interpretable. The metric is not in fact bound between − 1 and 1 and unpatterned data do not average 0. Its null and empirical distributions are not oppositive (symmetrical and centered; Fig. 1a). Nor even is the magnitude of I a guide to the degree of spatial pattern (extensively documented in DeWitt et al. 2021). In one of the surveyed empirical studies, data from 229 sites for avian clutch size yielded a significant I of +0.005 (corrected for mean bias; Marrot et al. 2015). I for incubation duration in the same study was 2.6-fold greater in magnitude but was nonsignificant. In contrast with those counter-intuitive results, rectified metrics for these two variables were +0.15 and − 0.02, which can readily be recognized on the scale of correlations as weak positive and negligible autocorrelation.

Although I has often been thought of and interpreted as a correlation, perhaps due to parallel nomenclature—'index of autocorrelation', 'coefficient of correlation'—and the widely cited notion that it was developed 'from' r which implies homology. I and r are both cross products (∼co-variance), but homology ends there. Correlation is

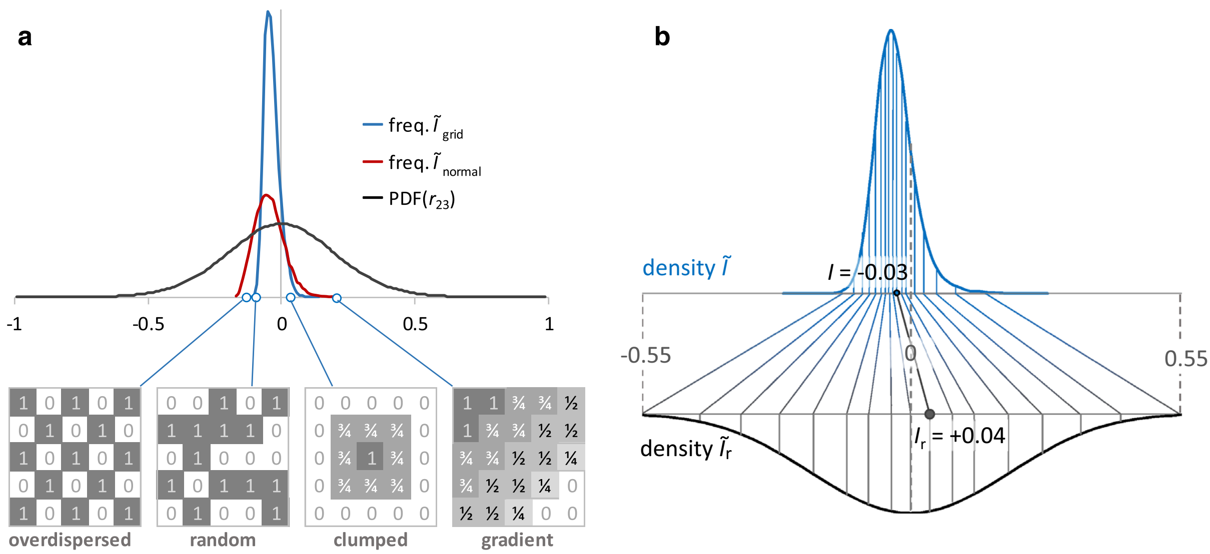

Fig. 1 Null distributions of Moran's I and r. **a** For 5×5 sampling grids (blue), random normal sample locations (red), and r for random normal data pairs (grey). I for selected spatial patterns as illustrated are given for context. **b** Projection of cumulative densities from a distribution of $\tilde{I}$ to the r-normalized distribution of I_r. Position of an observed I of -0.03 and its rectified value of $+0.04$ are shown to illustrate disparity in metric values. The true auto*correlation* in the example is $+0.04$

the cross product of two standard normal variates—each is singularly dimensioned, normally distributed, and scaled to unit variance. I is the crossproduct of a standard normal variate and a variably-dimensioned, variably-distributed χ^2-related hollow matrix of inverted distance elements scaled to unit sum. Exploration using the supplementary material revealed that sampling from 25 uniform (0−1) random coordinates resulted in weights matrices of typically 7 to 9 dimensions and volume (summed eigenvalues) from 0.08 to 0.125. Sampling from a uniform 5×5 grid resulted consistently in 6 dimensions of total volume 0.067. Dimensionality also increased with sample number. Thus, the basis for I is dimensionally, volumetrically, and distributionally variable among studies with alternative spatial conformations (e.g. uniform transects, random coordinates) and sample sizes. This variable geometry is responsible for I's inconsistent properties and lack of a characteristic null distribution (e.g. Fig. 1a). The null distribution is therefore best discerned empirically through data randomizations. The location of an actual I or related term in its empirically resolved null distribution indicates a cumulative probability that can be transformed to any metric with a defined distribution, such as t, χ^2, or r.

The procedure to calculate a correlation-equivalent I, termed I_r, is given in Table 2. Briefly, one calculates a crossproduct between the centered data and the proximity (unscaled weights) matrix for the actual data and for k randomizations. From the position (as cumulative frequency) of the actual crossproduct in the distribution of randomization results, a t statistic is derived and converted to the equivalent r, which is deemed I_r. I_r is then interpreted as a regular (Pearson or Spearman) correlation. It will be useful to those familiar with formulation of I using summation notation to consult Supplementary File S1 ("Equivalence of summation and linear algebra I"). This brief note describes the equivalence of the traditional and linear algebraic formulation of I (after Chen 2013). These formulations are demonstrated by example and shown to yield the same result in Supplemental File S2 ("Calculation of I and I_r"), page 1. Page 2 of that file demonstrates the step-by-step procedure to calculate I_r as given in Table 2. I_r may also be calculated using the R software package Irescale (Fuentes et al. 2020).

Rectifying I to its correlation equivalent normalizes it to existing perceptive capital regarding the ideals expected or desired for autocorrelation as well as that widely familiar regarding regular correlation. The rectified metric I_r has the ideal properties of r but also responds to dispersion as does I. The average (and

Table 2 Calculation of I_r

1	Calculate the triple crossproduct, C, between elements of the centered data vector and the proximity (unscaled weights) matrix
2	Randomize the data or location matrix k times
3	Calculate $\tilde{C}$ for each randomization (a tilde in superposition denotes randomization outcomes)
4	Find the proportion of $\tilde{C}$ that is less than or equal to actual C
5	Convert that proportion to a t-statistic using the inverse t distribution
	Caveat—where the proportion of $\tilde{C} \leq C$ is 0 or 1, then t may be underestimated
	Solution—increase k to the next higher order of magnitude
6	Convert t to $r = t/\sqrt{(n-2+t^2)}$ and deem this I_r
7	Interpret I_r as a regular correlation
8	Option: Assess non-dispersion due to hotspot clumping v. gradient structure

These steps are demonstrated in Supplementary File S2 ("Calculation of I and Ir")

median) I_r in the absence of pattern is 0, it's scale is oppositive, it's distribution is unskewed, and it's values are truly bound at − 1 and 1 (Fig. 1b). Thus I_r has standalone interpretability in a manner already intuitive to researchers. Since I_r follows a well characterized (modified t) distribution (Eq. 5 in DeWitt et al. 2021) it is also its own test statistic. When squared it yields the familiar effect strength R^2. The fit of I_r to r means one may correctly speak of I_r as an auto*correlation*. I without rectification is a mere 'index'—one that is not interpretable outside singular contexts and cannot be compared among studies.

If spatial autocorrelation is taken to be the 1st law of geography—measurements taken near each other tend to be similar (Tobler 1970)—it is important that it be represented by a consistently meaningful and intuitive statistic. Due to the qualities reviewed above (Table 1), I_r can be compared among variables and studies even for those with different spatial conformations or contiguity/proximity definitions, or alternative data types. I_r can be compared to and used with other statistics related to autocorrelation, such as spatial regression (multiple r). DeWitt et al. (2021) used the ratio of $I_r^2/(I_r^2 +$ multiple $r^2)$ plotted with I_r to partition non-dispersion due to clumping and gradient effects (Fig. 9 in DeWitt et al. 2021), although one could also detect such differences using correlograms (Dale and Fortin 2014) or autocorrelation analysis with residuals from spatial regression (Chen 2016). This new perspective and method of representing autocorrelation may help sustain or increase research on the topic of autocorrelation as well as promote reviews and synthetic work such as meta-research.

Authors' contributions Not applicable.

Funding Funded by a Triads for Transformation (T3) grant from Texas A&M University.

Data availability Example data and software to replicate this study are provided in Supplementary file S1 and is also archived at the Open Access to Knowledge (OAK) digital repository at https://oaktrust.library.tamu.edu/handle/1969.1/194479.

Code availability Not applicable.

Declarations

Conflict of interest The author declares no competing interests.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Chen Y (2013) New approaches for calculating Moran's index of spatial autocorrelation. PLoS ONE 8:e68336

Chen Y (2016) Spatial autocorrelation approaches to testing residuals from least squares regression. PLoS ONE 11:e0146865

Getis A, Ord JK (2010) The analysis of spatial association by use of distance statistics. Perspectives on spatial data analysis. Springer, pp 127–145

DeWitt TJ, Fuentes JI, Ioerger TR, Bishop MP (2021) Rectifying I: three point and continuous fit of the spatial autocorrelation metric, Moran's *I*, to ideal form. Landscape Ecol 36:2897–2918.

Dale MR, Fortin M-J (2014) Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge, UK

Fuentes JI, DeWitt TJ, Ioerger TR, Bishop MB (2020) Irescale: Calculate and rectify Moran's I. R package v. 2.3.0. https://cran.r-project.org/package=Irescale. Accessed 10 Dec 2020

Marrot P, Garant D, Charmantier A (2015) Spatial autocorrelation in fitness affects the estimation of natural selection in the wild. Methods Ecol Evol 6:1474–1483

Moran PAP (1950) Notes on continuous stochastic phenomena. Biometrika 37:17–23

Tobler W (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46(Supplement):234–240