



# Local indicators for categorical data: impacts of scaling decisions

JED A. LONG

Spatial Pattern Analysis and Research (SPAR) Laboratory, Department of Geography, University of Victoria, Victoria, British Columbia, Canada V8W 3R4  
(e-mail: jlong@uvic.ca)

TRISALYN A. NELSON

Spatial Pattern Analysis and Research (SPAR) Laboratory, Department of Geography, University of Victoria, Victoria, British Columbia, Canada V8W 3R4  
(e-mail: trisalyn@uvic.ca)

MICHAEL A. WULDER

Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, Victoria, British Columbia, Canada V8Z 1M5 (e-mail: mwulder@pfc.nrcan.gc.ca)

*When the geographic distribution of landscape pattern varies, global indices fail to capture the spatial nonstationarity within the dataset. Methods that measure landscape pattern at a spatially local scale are advantageous, as an index is computed at each point in the dataset. The geographic distribution of local indices is used to discover spatial trends. Local indicators for categorical data (LICD) can be used to statistically quantify local spatial patterns in binary geographic datasets. LICD, like other spatially local methods, are impacted by decisions relating to the spatial scale of the data, such as the data grain ( $p$ ), and analysis parameters such as the size of the local neighbourhood ( $m$ ). The goal of this article is to demonstrate how the choice of the  $m$  and  $p$  parameters impacts LICD analysis. We also briefly discuss the impacts spatial extent can have on analysis; specifically the local composition measure. An example using 2006 forest cover data for a region in British Columbia, Canada where mountain pine beetle mitigation and salvage harvesting has occurred is used to show the impacts of changing  $m$  and  $p$ . Selection of local window size ( $m = 3, 5, 7$ )*

*Les indicateurs locaux pour l'analyse de données catégorielles : les incidences des choix d'échelle*

*Lorsqu'il existe une variation dans la distribution géographique de la structure paysagère, les indices globaux n'arrivent pas à détecter la nonstationnarité spatiale dans les données. Les méthodes qui visent à évaluer la structure paysagère à une échelle locale offrent des avantages en produisant un indice pour chacun des points de la base de données. L'intérêt d'utiliser une distribution géographique des indices locaux consiste à dégager les tendances spatiales. Les indicateurs locaux pour données catégorielles (ILDC) se prêtent à la quantification statistique des schémas spatiaux d'échelle locale présents dans les bases de données géographiques de type binaire. Les ILDC, comme les autres méthodes spatiales de niveau local, sont affectés par les choix relatifs à l'échelle spatiale des données, tels que la granularité des données ( $p$ ), et les paramètres d'analyse comme la taille du voisinage ( $m$ ). Cet article vise à examiner les effets du choix des paramètres  $m$  et  $p$  sur la méthode des ILDC, en plus d'étudier les effets de l'étendue spatiale sur*

impacts the prevalence and interpretation of significant results. Increasing data grain ( $p$ ) had varying effects on significant LICD results. When implementing LICD the choice of  $m$  and  $p$  impacts results. Exploring multiple combinations of  $m$  and  $p$  will provide insight into selection of ideal parameters for analysis.

**Key words:** spatial analysis, fragmentation, spatial pattern, composition, configuration, mountain pine beetle, *Dendroctonus ponderosae*

*l'analyse, notamment la mesure de la composition locale. Utilisant les données de 1996 sur le couvert forestier d'une région de la Colombie-Britannique, Canada, où ont été élaborées des stratégies d'atténuation des populations de dendroctone du pin ponderosa ainsi que des opérations de récupération, nous mettons en lumière les effets des modifications apportées à  $m$  et  $p$ . Le choix de la taille du périmètre d'observation ( $m = 3, 5, 7$ ) a des incidences sur l'identification et l'interprétation de ce qui constitue un résultat statistiquement significatif. À cela s'ajoutent les multiples effets sur le niveau de signification des résultats en ajustant à la hausse le grain des données ( $p$ ). Il est clair que le choix de  $m$  et de  $p$  a des incidences sur les résultats obtenus avec les ILDC. Des études ultérieures portant sur les combinaisons possibles de  $m$  et  $p$  permettront de préciser la définition des paramètres d'analyse.*

**Mots clés :** Analyse spatiale, fragmentation, forme spatiale, composition, configuration, dendroctone du pin ponderosa, *Dendroctonus ponderosae*

## Introduction

With advances in methods for obtaining and analysing spatial data, geographers and ecologists have demonstrated the importance of the interaction between landscape spatial patterns and a number of different ecological processes including: sediment transport to streams (Jones *et al.* 2001); the spread of natural disturbances such as wildfire (Turner *et al.* 1989a), and forest insects (Barclay *et al.* 2005); as well as wildlife habitat occupancy (Helzer and Jelinski 1999) and movement (Wiens *et al.* 1997). There currently exist a number of statistical methods for examining spatial patterns of spatially referenced geographic data such as global measures of spatial autocorrelation (Cliff and Ord 1981), local measures of spatial autocorrelation (Getis and Ord 1992; Anselin 1995), geostatistics (Cressie 1993), geographically weighted regression (Fotheringham *et al.* 2002) and density estimators for point data (Ripley 1976; Silverman 1986). To avoid breaking statistical assumptions most methods generally require interval/ratio data. However, often it is necessary to examine the spatial pattern of qualitative (categorical) variables. For instance, landscapes are of

ten represented by categorical land cover classes. Methods for quantifying spatial pattern in categorical data have emerged primarily within the field of landscape ecology, which explores relationships between spatial patterns of ecological variables and ecological processes (Turner 1989). Pattern analysis methods for categorical data currently do not follow any statistical testing framework (Fortin *et al.* 2003; Csillag and Boots 2005). However, methods incorporating a statistical framework do exist for comparing maps of categorical variables (Csillag *et al.* 2006; Remmel and Csillag 2006). An improved ability to formulate statistical tests on categorical datasets will enable stronger inferences on observed spatial patterns.

Landscape pattern indices (LPIs) are the primary method for quantifying spatial pattern in spatially referenced categorical data. In general, LPIs measure spatial pattern properties with the aim of understanding the connection between pattern (e.g., the number and size of patches, or amount of edge) and ecological process (e.g., wildlife movements). The continued use of LPIs can be attributed to their ease of implementation (McGarigal and Marks 1995; Mladenoff and DeZonia 2004), broad applicability



(Cardille and Turner 2002) and computational simplicity (McGarigal and Marks 1995). Unfortunately, LPIs have a number of issues including: correlation among metrics (Riitters *et al.* 1995), scaling effects (Wu 2004), interpretability (Li and Wu 2004), lack of statistical testing (Turner *et al.* 2001; Fortin *et al.* 2003; Remmel and Csillag 2003) and difficulty linking quantified pattern with process (Li and Wu 2004). As well, research has advocated caution in the use of LPIs due to the dependence of configuration measures on landscape composition (Remmel and Csillag 2003), with some researchers using composition as a surrogate for density and related fragmentation (Wickham *et al.* 2008).

Local indicators for categorical data (LICD) (Boots 2003, 2006), provide a means for statistically quantifying local spatial pattern in categorical data. Local measures of spatial pattern can be clearly defined through a contrast with global measures. Global measures assume spatial stationarity and produce a single value of pattern for the entire study area. Geographic phenomena often exhibit spatial nonstationarity and necessitate local methods, which allow measures of spatial pattern to be generated for each data point (Fotheringham and Brunsdon 1999). LICD are spatially local and have overcome several of the problems associated with LPIs. LICD use conventional metrics often considered with LPIs. However, unlike LPIs, LICD consider the dependence of configuration on composition and include a form of statistical testing. Using a given configuration index, the distribution of all the possible configuration values with each possible composition level are generated. LICD denote significance when the measured configuration value falls in the outer tails of this distribution at a given  $\alpha$  value. LICD also attempt to minimize correlation among metrics by choosing mutually exclusive measures of spatial pattern based on previous studies (see Boots 2006).

The objective of this research is to outline the impacts of scaling decisions when implementing LICD for quantifying local spatial pattern in binary geographic data. Scaling effects centre on the choices of: local neighbourhood, data grain and spatial extent. As an example, LICD are applied to a land cover dataset generated from Landsat imagery representing forest conditions for 2006 in a study area in British Columbia,

Canada. We use LICD to evaluate the local spatial pattern characteristics of this region.

## Local Indicators for Categorical Data

The spatial pattern of landscape elements can be considered in two components: *composition* and *configuration* (Li and Reynolds 1994; Gustafson 1998). The six LICD measures are separated into these two components: a singular measure of local *composition*, and five measures of local *configuration*. Here, we provide a brief overview of LICD and their derivation (for additional details see Boots 2003, 2006). We will use black/white (B,W) to indicate binary classes for introduction of metrics.

### Composition

In binary data, global composition can be exclusively defined using a single measure: the proportion of one class (number of B or W cells). Similarly, local composition can be defined exclusively for a local spatial neighbourhood (window). Therefore, the LICD compositional measure (see Table 1) tests whether the composition of any specified  $m \times m$  local neighbourhood ( $m = 3, 5, 7, \dots$ ), is significantly different from the global composition (Boots 2003).

### Configuration

Five measures have been chosen as representative of the major components of spatial configuration (see Table 1). These measures quantify join counts within and between classes (examining the spatial clustering of classes, which also relates to class edge amounts), the number of patches, sum of squared area (SSA) of patches, patch proximity and how class cells are arranged in the local neighbourhood. Configuration of landscape elements is known to be dependent on landscape composition (Gustafson and Parker 1992); accordingly LICD consider configuration in the context of composition. An appropriate null hypothesis when using LICD is: given the composition of a local neighbourhood, is the configuration different than what would be expected by a random spatial process. With LICD, significance is defined as those LICD

**Table 1**  
Formulation of the six LICD

LICD measure	Computed for	Formulation	Notes
Local composition <sup>†*</sup>	B	$\Pr(X = x) = \binom{r}{x} P_b^x (1 - P_b)^{r-x}$	$x = \text{number of black cells}$ $r = \text{cells in neighbourhood}$ $P_b = \text{global proportion}$
Local joins <sup>††*</sup>	B, W, BW	$g_{jk} \rightarrow (j = B, W; k = B, W)$	$g = \text{join between adjoining cells}$
Number of patches <sup>††*</sup>	B, W, Total	$n_j \rightarrow (j = B, W) \text{ and } n$	$n = \text{number of patches}$
Sum of squared area <sup>††*</sup>	B, W, Total	$\sum_{i=1}^{n_j} d_{ij}^2 \rightarrow (j = B, W) \text{ and } \sum_{j=1}^k \sum_{i=1}^{n_j} d_{ij}^2$	$a = \text{area of patch}$ $k = \text{number of classes}$
Patch proximity <sup>††</sup>	B, W	$\sum_{i=1}^{n_j} h_{ij} \rightarrow (j = B, W)$	$h = \text{distance between patches}$
Eccentricity <sup>††</sup>	B	$\sum_{i=1}^{x_j} d_i \rightarrow (j = B)$	$d = \text{distance from each black cell to focal cell of neighbourhood}$

Illustrative references: <sup>†</sup>Boots (2003), <sup>††</sup>Boots (2006).

\*Indicates measure used in our study.

values that fall in the upper and lower tails of their corresponding distributions at a given significance level (Boots 2006). Therefore, LICD portray significance where local configuration values are higher or lower than what would be expected by a random spatial process. When the proportion of one class in the local neighbourhood is high, the ability of LICD to detect statistically significant spatial pattern is reduced. As a result, homogeneous (and near homogeneous) local neighbourhoods are portrayed by LICD as not-significant. This result is the same as a heterogeneous local neighbourhood that does not show significant deviation from a random spatial process. One must account for these situations when interpreting the not-significant values of LICD analysis.

For some of the five previously indicated configurational LICD, it is possible to compute the measure for both individual classes (B,W) and overall, thus a total of 12 configurational LICD can be derived. LICD, their formulation and notes, are displayed in Table 1.

### Limitations

LICD have been designed as a spatially local method for evaluating pattern in binary categorical data represented in a raster (regular lattice) format. LICD are limited to binary data ( $k = 2$ ), where  $k$  is the number of classes. Local analy-

sis limits the number of observations in the local neighbourhood. When  $k > 2$  the analytical complexity of multiple classes, combined with low number of observations does not facilitate meaningful scientific questioning using statistical methods. Also, if the probability of one of the classes is low, interpretation issues also arise for  $k = 2$  (Boots 2003). While LICD are implemented for raster (regular lattice) data, they may also be implemented with other representations, but the statistical distributions for nonlattice neighbourhoods have not been generated.

### Scaling Issues and Methods

When implementing spatial analysis methods, researchers often make a series of decisions to define the parameters used in analysis. Often, as well, these decisions are complicated by issues of data characteristics such as classification scheme and scale, and analysis considerations such as the definition of local neighbourhood. For instance, the choice of spatial scale at which data are represented defines the size of resolvable features in subsequent analysis. We focus on the effects of scaling decisions required when using LICD, specifically the local neighbourhood ( $m$ ) and the data grain ( $p$ ). For all analysis,  $\alpha = 0.05$  was used to denote significant high and low values using the two-tailed LICD method.



### Local neighbourhood

Local analysis requires the definition of a local neighbourhood (see Nelson and Boots 2008). In raster datasets this is generally a square moving window of size  $m \times m$  ( $m = 3, 5, 7, \dots$ ). The window is centred on each individual cell in the dataset and the local statistic calculated for the window and stored in the centre cell. Currently, distributional properties for LICD have been generated for windows of  $m = 3, 5, 7$ . Due to the small number of observations that exist when  $m$  is small (e.g., when  $m = 3$ , there are 9 data sites in the local window, while when  $m = 7$ , there are 49), the smallest window sizes may be least useful, as significant results are rare leading to the suggestion that larger window sizes should be explored, and may provide improved distributional properties for LICD (Boots 2006). Given the problems when  $m$  is small ( $m = 3$ ), we suggest local windows of size  $m \geq 5$ . The largest window will need to be selected based on the nature of the phenomena under study.

### Spatial scale

Data grain represents the smallest unit of observation in analysis (quadrats for a field ecologist or pixels for a remote sensor; Wiens 1989). Spatial extent refers to the overall area encompassed by a study (Wiens 1989). Data grain and spatial extent can be thought to define the upper and lower limits of resolution in a study. Together, data grain and spatial extent combine to convey the spatial scale of a study (see Dungan *et al.* 2002 for a comprehensive discussion of scale in spatial analysis).

Data grain is the most important scalable data property to consider when implementing LICD. LICD are currently developed for use with lattice (raster) datasets where grain is the size of individual units, often referred to as cells or pixels. In ecology, there has been much discussion on how to determine an effective grain size (Turner *et al.* 1989b; Wiens 1989; Wu 2004). It is particularly important to consider grain when data are represented using lattice cells that are not sized in relation to the ecological process of interest. For instance, lattice data are commonly derived from remotely sensed imagery where pixel size is determined based on sensor specification

rather than ecological process (e.g., forest cover can be derived from 30 m Landsat or 250 m MODerate resolution Imaging Spectroradiometer (MODIS) sensors). LICD are most effective when the data grain is comparable in size to the land cover objects (e.g., forest patches) in question (see Gergel 2007 for a discussion on data grain).

The second scalable data property when implementing LICD is spatial extent. An issue related to the spatial extent of a study area is edge effects. Edge effects are an inherent property of local spatial analysis (Fortin and Dale 2005). With LICD, this problem manifests itself such that the area which results are reported for is slightly smaller than the original study area. Near the study area edge a complete local neighbourhood cannot be defined. Edge effects are magnified as the size of local neighbourhood is increased in relation to the study area extent. When implementing LICD we recommend definition of a spatial extent that is slightly larger than the area of the study area. Results are then reported for the study area, while data beyond the study area boundary is used only in the computation of local measures. Spatial extent also directly affects the local composition measure. The local composition measure tests local composition against global composition. It is necessary to choose a spatial extent that exemplifies a typical composition level in order to derive meaningful interpretations from local composition results.

### Case Study

We apply LICD to an area of managed forest in British Columbia's interior region that has been impacted by the mountain pine beetle (*Dendroctonus ponderosae*) and resulting pine mortality and related salvage harvest and mitigation activities. Through the mountain pine beetle example we demonstrate how LICD results are impacted by scaling parameters. In our example we use the number of patches metric with  $p = 75$  to demonstrate the effects of changing  $m$  on LICD results (Table 2). We use  $m = 7$  and four data grains ( $p = 25, 75, 175, 275$ ) to demonstrate the effects of changing data grain on four selected LICD (local composition, forest/nonforest joins, number of forest patches and SSA forest patches) in the mountain pine beetle example (Table 2). A

**Table 2**

Implementation parameters used for analysis of the local spatial pattern of forest in an area impacted by mountain pine beetle related salvage harvest and mitigation activities

Implemented		Rationale	Selected reference(s)
Local neighbourhood	$m^{\dagger} = 3, 5, 7$	Large $m$ may be most useful due to lack of observations in local neighbourhood when $m$ is small. $m = 7$ is used to demonstrate impacts of changing data grain.	
Spatial scale <i>Data grain</i>	$p^* = 25, 75, 175, 275$	Multi-scale approach is warranted Medium resolution sensors represent applicable data grain for resolving forest harvests An aggregation procedure was used to create the three larger data grains originating from an original 25 m pixel dataset	Wu 2004 Coops <i>et al.</i> 2007 Healey <i>et al.</i> 2007
<i>Spatial extent</i>		Known mountain pine beetle impacts Maximize study area size, while minimizing impacts of urban development	
Selected LICD	Local composition	Forest composition has been used as a surrogate for forest fragmentation	Wickham <i>et al.</i> 2008
	F/N Joins	Edge habitat plays important ecological role in forest ecosystems Can also be used to explore spatial dependence in categorical spatial data	Ranney <i>et al.</i> 1981 Cliff and Ord 1981
	# Patches ( $F$ )	Useful in measuring landscape configuration in forested landscapes	Haines-Young and Chopping 1996
	SSA ( $F$ )	Common indicator used for measuring fragmentation	Fahrig 2003

\* $p$  is pixel size in metres.

† $m$  represents local neighbourhood size (e.g., an  $m \times m$  window).

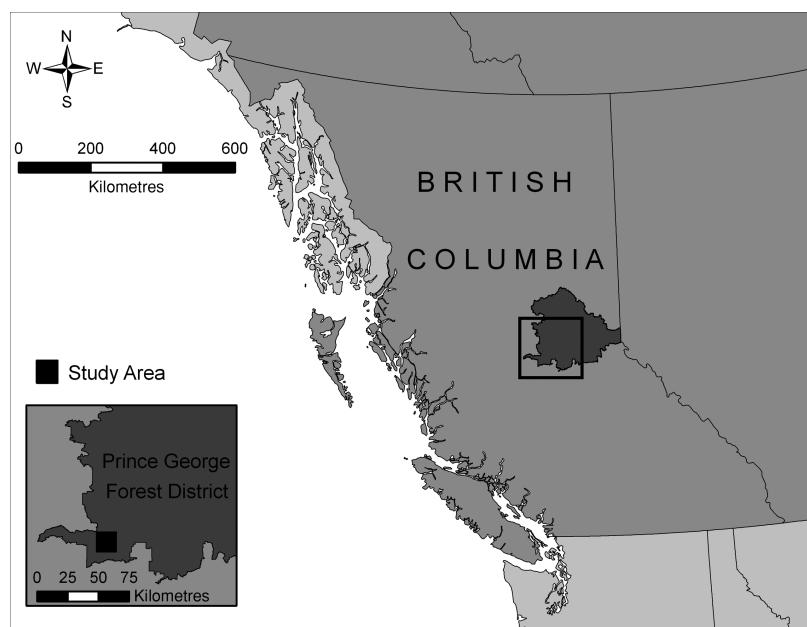
17 km × 17 km spatial extent corresponding to a landscape that has undergone mountain pine beetle salvage and mitigation harvesting was selected for analysis, and is relevant for interpretation of these landscapes.

### Study area

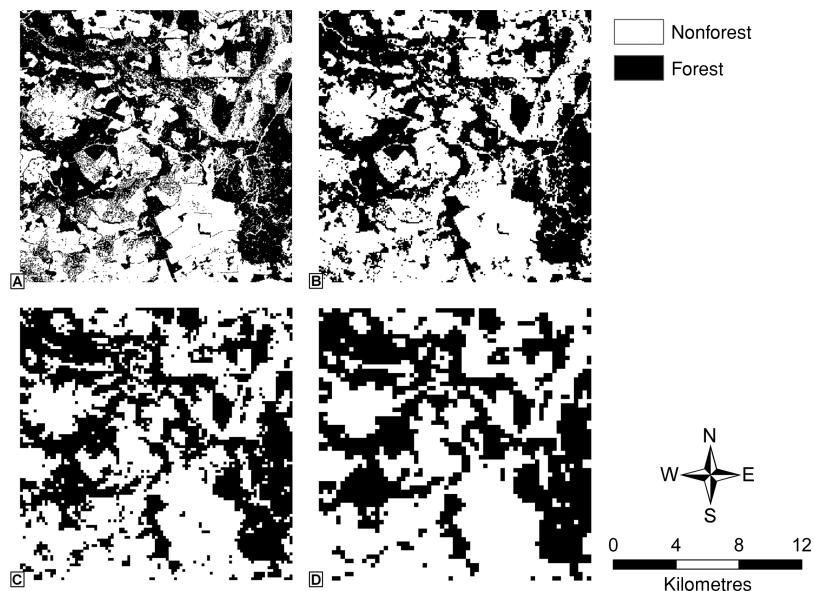
The study area (Figure 1) is located within the interior plateau of British Columbia and resides within the sub-boreal spruce biogeoclimatic zone (Meidinger and Pojar 1991). This region is characterized by extreme weather fluctuations across seasons, with hot, moist summer months, followed by long, dry and cold winters. White spruce (*Picea glauca*), subalpine fir (*Abies lasiocarpa*) and lodgepole pine (*Pinus contorta*) are the dominant forest species within this region. We chose a study area in a region that has been subject to an outbreak by mountain pine beetle, with relatively homogenous forest conditions and minimal urban development.

### Data

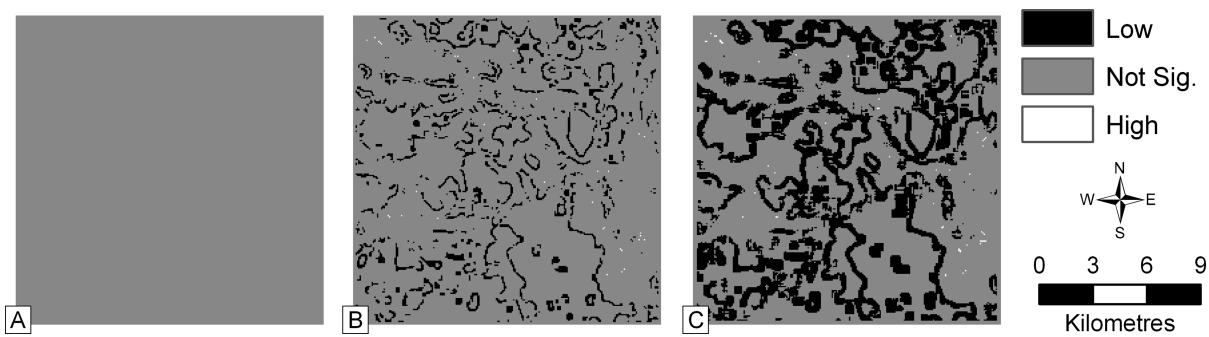
Data used for this study are derived from a land cover dataset created by the Earth Observation for Sustainable Development of Forests (EOSD) program in Canada (see Wulder *et al.* 2003, 2008). The EOSD data consists of a 23 class hierarchy which can be aggregated to a number of class levels (including forest/nonforest, as in this study) depending on the user's necessity for categorical detail (Wulder and Nelson 2003). Using a change detection algorithm (following Han *et al.* 2007), we updated the 2000 EOSD dataset to represent 2006 land cover conditions using appropriate 2006 Landsat Thematic Mapper (TM) data. The 2006 land cover reflects forest conditions post-mountain pine beetle infestation. From the original 25 m pixel resolution ( $p = 25$ ) EOSD data we generated three larger data grains ( $p = 75, 175$  and  $275$  m) using a modal aggregation procedure. The composition of the study area for each data grain is presented in Figure 2.

**Figure 1**

Study area, a 17 km × 17 km region located in the Prince George Forest District, British Columbia, Canada

**Figure 2**

The initial composition of the study area using (A)  $p = 25$ , (B)  $p = 75$ , (C)  $p = 175$ , (D)  $p = 275$

**Figure 3**

The effect of changing local neighbourhood size— $m$ , number of forest patches ( $p = 75$ ): (A)  $m = 3$ , (B)  $m = 5$ , (C)  $m = 7$

Fine-scale heterogeneity and the ability to detect small features are reduced as grain size is increased.

## Results

### Local neighbourhood

When  $m = 3$ , there are no cases where significant results are found (Figure 3). With  $m = 5$ , 10.7 percent of the study area are significant (0.1 percent high, 10.6 percent low). When  $m = 7$ , the number of significant results for forest patches is 31.3 percent (0.1 percent high, 31.2 percent low). The trend of increasing amount of significant results from  $m = 3$  to  $m = 7$  is consistent across the other configurational metrics. Windows of  $m = 3$  do not identify significant patterns. We use  $m = 7$  in subsequent analysis to examine the impact of changing data grain.

### Spatial scale

**Local composition.** LICD test local composition by comparing the composition of the local neighbourhood to that of the entire study area. The local composition measure is affected by global composition. Reported values for global composition (Table 3) demonstrate that the aggregation procedure has only a marginal affect on global composition in this case.

We interpret local composition results (Table 4, Figure 4) relative to the global composition. Thus, significantly high areas represent local regions where forest composition is greater than

**Table 3**

Impact of changing data grain (through aggregation) on global forest proportion

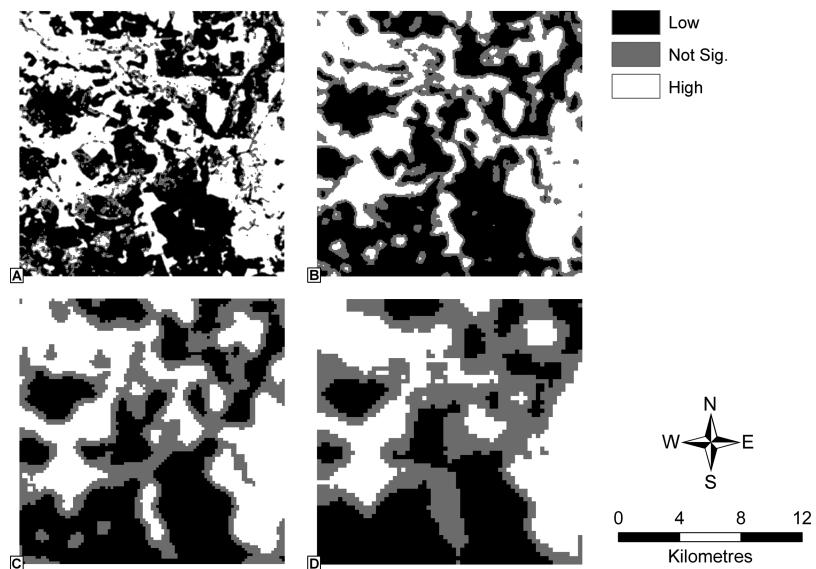
Data grain (m)	Global forest proportion (%)
$p = 25$	46.5
$p = 75$	44.4
$p = 175$	43.7
$p = 275$	43.3

**Table 4**

Area, in percent of study area, of each of the high, low and not significant (N.S.) classes

	$p =$	Data grain			
		25 m	75 m	175 m	275 m
Local composition	High	44.4	36.5	33.6	30.7
	N.S.	16.1	21.5	29.5	35.1
	Low	39.5	42.0	36.9	34.2
Forest/non-forest joins	High	0.0	0.0	0.0	0.0
	N.S.	48.9	39.7	33.0	32.7
	Low	51.1	60.3	67.0	67.3
Number of forest patches	High	0.1	0.1	0.1	0.0
	N.S.	76.0	68.7	66.1	62.4
	Low	23.9	31.2	33.8	37.6
SSA forest patches	High	23.8	32.0	35.2	36.9
	N.S.	75.8	67.3	64.4	62.6
	Low	0.4	0.7	0.4	0.5

would be expected given the overall composition of the study area. Significantly low areas represent the local regions where forest composition is less than would be expected given the overall composition. Ecologically we can interpret

**Figure 4**

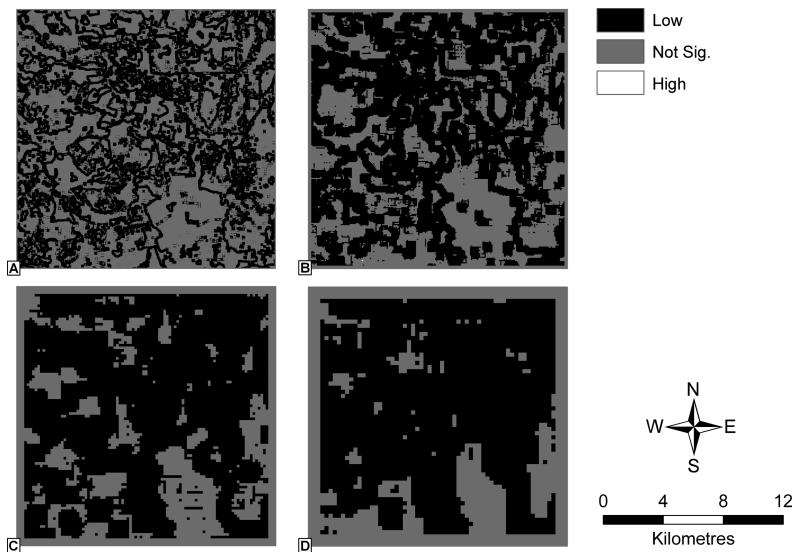
Local composition measure: (A)  $p = 25$ , (B)  $p = 75$ , (C)  $p = 175$ , (D)  $p = 275$

significant results as regions of abundance (high) or absence (low) of expected forest quantity. Areas where results are not significant contain the level of forest we would expect given overall composition. Using a fine data grain this measure may function as an edge detecting procedure and roughly replicates the initial dataset ([A] in Figure 4). However, when using a coarser data grain, local proportion may function to define forest as edge (not significant), core (high significance) and open (low significance) (see [C] in Figure 4). When the global landscape proportion of one class is high, inference on local composition is limited.

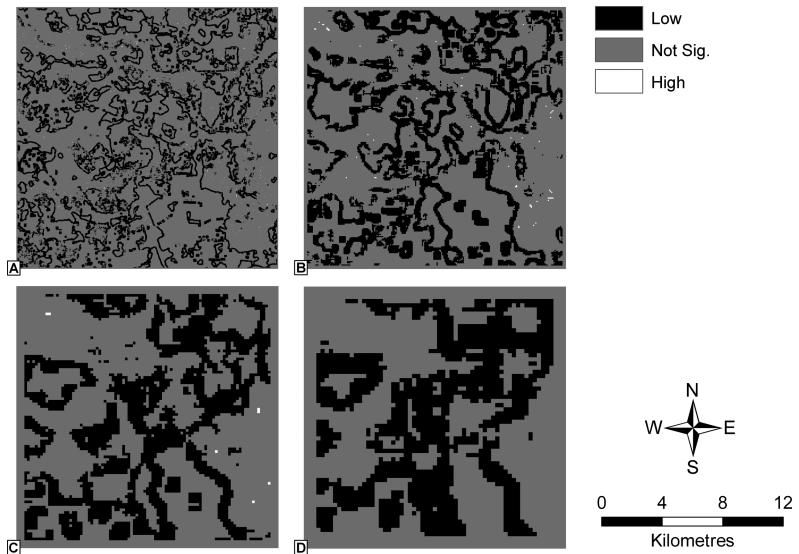
**Forest/nonforest joins.** Using forest/nonforest joins, we examine the amount of edge habitat, as well as look for spatial dependence in the data (see Table 4, Figure 5). Significantly low areas represent locations where forest is clustered or patchy with respect to a pattern generated by a random process (positive spatial autocorrelation). Clustering is prevalent across our study area. Locations with significantly high values indicate regions where forest patterns are dispersed with respect to patterns generated from a random process (negative spatial autocorrela-

tion). Our study area portrays predominantly regions of positive spatial autocorrelation, and not significant spatial pattern as detected by the join count LICD. Regions that do not have significant pattern represent one of two scenarios; either (a) the local neighbourhood exhibits a configurational pattern (with respect to forest/nonforest joins) that is not significantly different than what would be expected from a random spatial process, or (b) the proportion of the local window is dominated by one class and no meaningful inference on spatial pattern can be made. In Figure 5(A) and (B), the significantly low areas follow the edges of the forest components, with large portions of interior forest and nonforest being not-significant. As grain size becomes coarser (Figure 5(C) and (D)) edge dependency becomes reduced and results change. We interpret these results by saying that in our study area, forest components exhibit patchy characteristics (positive spatial autocorrelation).

**Number of forest patches.** The number of forest patch measures quantifies similar spatial patterns to those characterized by the join count measure (Figure 6). One difference is the existence of both significantly low and high regions

**Figure 5**

Forest/nonforest joins: (A)  $p = 25$ , (B)  $p = 75$ , (C)  $p = 175$ , (D)  $p = 275$

**Figure 6**

Number of forest patches: (A)  $p = 25$ , (B)  $p = 75$ , (C)  $p = 175$ , (D)  $p = 275$

in the two intermediate grain sizes (Figure 6 (B)  $p = 75$  and (C)  $p = 175$ ; Table 4). Significantly low regions are located mainly along forest edges. Scattered throughout are a few

instances of significantly high number of forest patches. These areas represent where forest is fragmented more than we would expect if this amount of forest were located there by a random



spatial process. Significantly high numbers of forest patches (an indication of highly fragmented patches) represent interesting anomalies in the overall trend of forest patchiness.

**Sum of squared area (forest patches).** The SSA is a measure of forest patch size. Significantly high areas represent where forest patches are larger than we would expect by a random spatial process. Significantly low areas represent those where forest patches are smaller than we would expect by random spatial process. In Figure 7, we portray the results for the SSA forest patches across our study area. Similar to the significantly high values for number of forest patches, SSA forest patches shows significantly high values along forest edges. Significantly low values are found in many of the same regions as significantly high values for number of forest patches but are more prevalent (Table 4), demonstrating regions with small patch areas. We would expect there to be some relationship between these two measures as they are related to patch characteristics.

### Analysis

Landscape configuration is known to be impacted by landscape composition (Gustafson and Parker 1992; Remmel and Csillag 2003) and LICD address this directly in their derivation. The usefulness of LICD is greatest in regions where local composition does not heavily favour one class. Along forest edges and where continuous forest is interrupted by nonforest patches, local spatial pattern effects deemed low, high or not significant provide insight into spatial processes being observed. In local regions where composition heavily favours one class, we expect results to be predominantly not significant, given how LICD are defined.

The lone compositional measure provides information on local forest composition relative to the global level. Results from this study indicate that the composition LICD may be suitable for delineating core, open and edge habitat; where core habitat are those regions with significantly high composition, open habitat is those regions with significantly low composition, and edge being the regions of no significance. This will be appropriate when the level of global com-

position can be related to core, open and edge habitat.

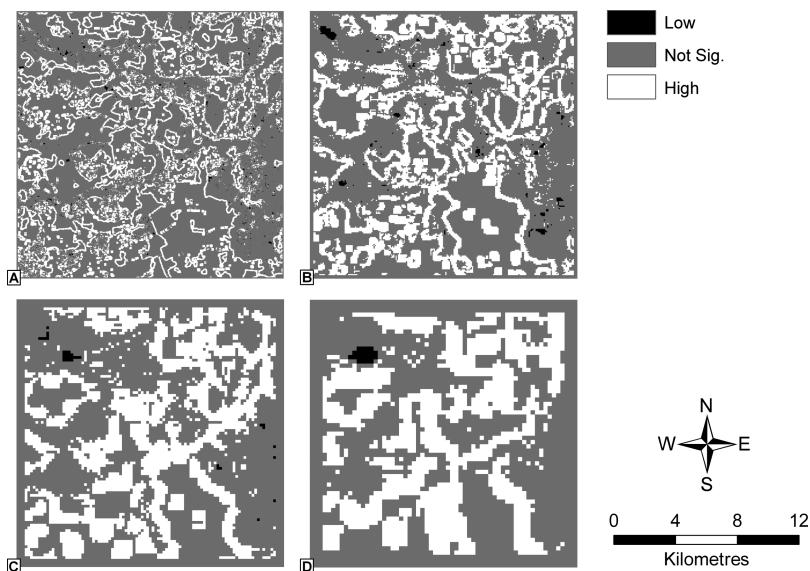
We chose three configurational LICD to evaluate forest trends at the spatially local level. We found results from forest/nonforest joins to be useful in describing the nature of spatial dependence in the data. Number of forest patches and SSA forest patches provide useful information on forest fragmentation and area properties of forest patches which affect biodiversity (Fahrig 2003).

### Discussion

Implementation of spatial analysis methods requires the careful consideration of input parameters used in the model (Fortin and Dale 2005). In a 2005 article, Csillag and Boots develop a framework for implementing statistical inference into spatial pattern analysis. The inclusion of statistical testing in spatial analysis of categorical data has been desired for some time (Turner *et al.* 2001) and new research methods attempt to overcome this problem (e.g., Fortin *et al.* 2003; Remmel and Csillag 2003; and LICD). Here we provide researchers with an example and discussion of the scaling effects when using LICD for geographical studies.

LICD allow for testing the significance of results against random spatial process. However, in many cases the assumption of randomness may not be valid. First-order heterogeneity, or spatial variation in the probability of occurrence (e.g., of forest) may influence observed spatial patterns. Accounting for first-order heterogeneity in the generation of test statistics improves the validity and interpretation of results (Kabos and Csillag 2002). With LICD, as with many common spatial statistics, first-order heterogeneity is not accounted for. Thus, we adopt an exploratory role and use LICD to define meaningful patterns as those with statistically significant deviations from some expectation (in this case randomness).

The use of small windows (i.e.,  $m = 3$ ) should be avoided, as LICD are unlikely to detect unusual or significant patterns. To understand why small neighbourhoods are problematic consider the number of patches metric in a  $3 \times 3$  window with a composition of 44 percent (i.e., 4/9 cells black). There are  $C(9,4) = 126$  ways to



**Figure 7**  
SSA of forest patches: (A)  $p = 25$ , (B)  $p = 75$ , (C)  $p = 175$ , (D)  $p = 275$

arrange the four black cells, however these result in only four possible number of black patches results:  $NP = (1,2,3,4)$ . With  $m = 5$  and a composition of 44 percent (i.e., 11/25 cells black) there are  $C(25,11) = 4,457,400$  ways to arrange the black cells and 11 possible number of black patches results,  $NP = (1,2,3,\dots,11)$ . When  $m = 7$  with composition 45 percent there are  $C(49,22) \approx 5.0 \times 10^{13}$  arrangements and 22,  $NP = (1,2,3,\dots,22)$ , possible LICD results. The wider distributions associated with  $m = 5, 7$  make these two window sizes most useful for LICD analysis. Boots (2006) has suggested that of the three analysis neighbourhoods for which LICD are currently defined ( $m = 3, 5, 7$ ),  $m = 7$  may be most appropriate, and we demonstrate this with our example. Boots (2006) also suggested that use of larger analysis neighbourhoods be explored.

The ability to capture small features is reduced as grain size is increased. Consider the mountain pine beetle case study, using the largest grain size ( $p = 275$ , (D) in Figure 3), remnant forest patches in the large nonforest openings are not represented. Conversely, the smallest grain size ( $p = 25$ , (A) in Figure 3), portrays such fine spatial heterogeneity that it is inappropriate for evaluating broad spatial trends using local anal-

ysis. The intermediate grain sizes chosen ( $p = 75$  and  $p = 175$ , (B) and (D), respectively, in Figure 3) may represent the most appropriate grain sizes for utilising LICD in these forested landscapes. The choice of grain size will impact results and subsequent interpretation. Proper grain sizes will be just small enough to resolve key features for analysis, such as the remnant forest patches in the case study.

By changing the data grain ( $p$ ) and analysis neighbourhood ( $m$ ) we have produced several maps that show differing LICD results for the same study area. The discrepancies result from what is known as the modifiable areal unit problem—MAUP (Openshaw and Taylor 1979; Openshaw 1984), a common problem when analysing data stored in areal units. Two distinct but connected issues arise from MAUP: the scale problem and the zoning (or aggregation) problem (Openshaw 1984). Changing the data grain ( $p$ ) or analysis neighbourhood ( $m$ ) are examples of the scale problem. Alternatively, changing the shape or arrangement of the analysis units (while keeping the number of units constant) is known as the zoning (or aggregation) problem. As shown in this example the scale problem associated with the MAUP will impact LICD results.



## Conclusion

LICD provide a spatially local method for detecting spatial pattern in categorical geographic data that includes the use of a statistical testing framework. In this article we identify the impacts of changing scaling parameters on LICD analysis. A study area that has undergone substantial forest harvesting following infestation from the mountain pine beetle is used as an example for implementing LICD to detect local spatial patterns. We were able to detect local regions that had higher than expected number of forest patches, and smaller than expected patch areas, as examples of a local spatial forest pattern that is highly fragmented.

Increasing the local neighbourhood parameter ( $m = 3, 5, 7$ ), leads to increasing detection of significant spatial pattern. We identified  $m = 7$  as the most suitable local neighbourhood for analysis of forest patterns. Changing data grain with LICD influences results. Larger data grains that are still able to resolve important features (e.g., remnant forest patches) are most useful. The scaling issues with LICD can be linked to the scale problem associated with the MAUP, a common challenge encountered with spatial analysis.

## Acknowledgements

Support for this research was provided by NSERC and the Government of Canada through the Mountain Pine Beetle Program, a three-year, \$100 million program administered by Natural Resources Canada, Canadian Forest Service. Thanks to Dr. Barry Boots for discussion on this work. We also thank the anonymous reviewers whose helpful comments improved this manuscript. Those interested in utilising LICD for their own research should consult the Spatial Pattern Analysis and Research (SPAR) Lab webpage for computational tools to assist with implementing LICD: <http://www.geog.uvic.ca/spar/tools.html>.

## References

- ANSELIN, L. 1995 'Local indicators of spatial association—LISA' *Geographical Analysis* 27, 93–115
- BARCLAY, H., LI, C., BENSON, L., TAYLOR, S. W., and SHORE, T. 2005 'Effects of fire return rates on traversability of lodgepole pine forests for mountain pine beetle (Coleoptera: Scolytidae) and the use of patch metrics to estimate traversability' *Canadian Entomologist* 137, 566–583
- BOOTS, B. 2003 'Developing local measures of spatial association for categorical data' *Journal of Geographical Systems* 5, 139–160
- . 2006 'Local configuration measures for categorical spatial data: binary regular lattices' *Journal of Geographical Systems* 8, 1–24
- CARDILLE, J. and TURNER, M. G. 2002 'Understanding landscape metrics I' in *Learning Landscape Ecology: A Practical Guide to Concepts and Techniques*, eds. S. E. Gergel and M. G. Turner (New York, NY: Springer-Verlag), 85–111
- CLIFF, A. D. and ORD, J. K. 1981 *Spatial Processes: Models and Applications* (London: Pion)
- COOPS, N. C., WULDER, M. A., and WHITE, J. C. 2007 'Identifying and describing forest disturbance and spatial pattern: data selection issues and methodological implications' in *Understanding Forest Disturbance and Spatial Pattern: Remote Sensing and GIS Approaches*, eds. M. A. Wulder and S. E. Franklin (Boca Raton, FL: Taylor & Francis Group), 31–61
- CRESSIE, N. 1993 *Statistics for Spatial Data* (New York: John Wiley & Sons Inc.)
- CSILLAG, F., and BOOTS, B. 2005 'A framework for statistical inferential decisions in spatial pattern analysis' *The Canadian Geographer* 49, 172–179
- CSILLAG, F., KABOS, S., and REMMEL, T. K. 2006 'Confidence in coincidence' *International Journal of Remote Sensing* 27, 1269–1276
- DUNGAN, J. L., PERRY, J. N., DALE, M. R. T., LEGENDRE, P., CITRON-POUSTY, S., FORTIN, M. J., JAKOMUSKA, A., MIRITI, M., and ROSENBERG, M. S. 2002 'A balanced view of scale in spatial statistical analysis' *Ecography* 25, 626–640
- FAHRIG, L. 2003 'Effects of habitat fragmentation on biodiversity' *Annual Review of Ecology and Systematics* 34, 487–515
- FORTIN, M. J., BOOTS, B., CSILLAG, F., and REMMEL, T. 2003 'On the role of spatial stochastic models in understanding landscape indices in ecology' *Oikos* 102, 203–212
- FORTIN, M. J., and DALE, M. R. T. 2005 *Spatial Analysis: A Guide for Ecologists* (Cambridge, UK: Cambridge University Press)
- FOTHERINGHAM, A. S., and BRUNSDON, C. 1999 'Local forms of spatial analysis' *Geographical Analysis* 31, 340–358
- FOTHERINGHAM, A. S., BRUNSDON, C., and CHARLTON, M. 2002 *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships* (Chichester, England: Wiley)
- GERGEL, S. E. 2007 'New directions in landscape pattern analysis and linkages with remote sensing' in *Understanding Forest Disturbance and Spatial Pattern: Remote Sensing and GIS Approaches*, eds. M. A. Wulder and S. E. Franklin (Boca Raton, FL: Taylor & Francis Group), 173–208
- GETS, A., and ORD, J. K. 1992 'The analysis of spatial association by use of distance statistics' *Geographical Analysis* 24, 189–206
- GUSTAFSON, E. J. 1998 'Quantifying landscape spatial pattern: What is the state of the art?' *Ecosystems* 1, 143–156
- GUSTAFSON, E. J., and PARKER, G. R. 1992 'Relationships between land-cover proportion and indices of landscape spatial pattern' *Landscape Ecology* 7, 101–110
- HAINES-YOUNG, R., and CHOPPING, M. 1996 'Quantifying landscape structure: a review of landscape indices and their application to forested landscapes' *Progress in Physical Geography* 20, 418–445
- HAN, T., WULDER, M. A., WHITE, J. C., COOPS, N. C., ALVAREZ, M. F., and BUTSON, C. 2007 'An efficient protocol to process Landsat images for change detection with tasseled cap transformation' *IEEE Geoscience and Remote Sensing Letters* 4, 147–151

- HEALEY, S. P., COHEN, W. B., YANG, Z., and KENNEDY, R. E. 2007 'Remotely sensed data in the mapping of forest harvest patterns' in *Understanding Forest Disturbance and Spatial Pattern: Remote Sensing and GIS Approaches*, eds. M. A. Wulder and S. E. Franklin (Boca Raton, FL: Taylor & Francis Group), 63–84
- HELZER, C. J., and JELINSKI, D. E. 1999 'The relative importance of patch area and perimeter-area ratio to grassland breeding birds' *Ecological Applications* 9, 1448–1458
- JONES, K. B., NEALE, A. C., NASH, M. S., VAN REMORTEL, R. D., WICKHAM, J. D., RIITTERS, K. H., and O'NEILL, R. V. 2001 'Predicting nutrient and sediment loadings to streams from landscape metrics: a multiple watershed study from the United States Mid-Atlantic Region' *Landscape Ecology* 16, 301–312
- KABOS, S., and CSILLAG, F. 2002 'The analysis of spatial association on a regular lattice by join-count statistics without the assumption of first-order homogeneity' *Computers & Geosciences* 28, 901–910
- LIU, H., and REYNOLDS, J. F. 1994 'A simulation experiment to quantify spatial heterogeneity in categorical maps' *Ecology* 75, 2446–2455
- LIU, H., and WU, J. 2004 'Use and misuse of landscape indices' *Landscape Ecology* 19, 389–399
- MCGARIGAL, K., and MARKS, B. J. 1995 *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure* (Portland, OR: General Technical Report PNW-GTR-351, USDA, Forest Service, Pacific Northwest Research Station)
- MEIDINGER, D., and POJAR, J., eds. 1991 *Ecosystems of British Columbia*, Special Report Series Vol. 06 (Victoria, BC: Research Branch British Columbia Ministry of Forests)
- MLADENOFF, D. J., and DEZONIA, B. 2004 *APACK 2.23 Analysis Software User's Guide*, Version 4-13-04
- NELSON, T., and BOOTS, B. 2008 'Detecting spatial hot spots in landscape ecology' *Ecography* 31, 556–566
- OPENSHAW, S. 1984 *The Modifiable Areal Unit Problem* (Norwich, England: Geo Books)
- OPENSHAW, S., and TAYLOR, P. J. 1979 'A million or so correlation coefficients: three experiments on the modifiable areal unit problem' in *Statistical Applications in the Spatial Sciences*, ed. N. Wrigley (London: Pion)
- RANNEY, J. W., BRUNER, M. C., and LEVISON, J. B. 1981 'The importance of edge in the structure and dynamics of forest islands' in *Forest Island Dynamics in Man-Dominated Landscapes*, eds. R. L. Burgess and D. M. Sharpe (New York: Springer-Verlag), 67–96
- REMMEL, T. K., and CSILLAG, F. 2003 'When are two landscape pattern indices significantly different?' *Journal of Geographical Systems* 5, 331–351
- . 2006 'Mutual information spectra for comparing categorical maps' *International Journal of Remote Sensing* 27, 1425–1452
- RIITTERS, K. H., O'NEILL, R. V., HUNSAKER, C. T., WICKHAM, J. D., YANKEE, D. H., TIMMINS, S. P., JONES, K. B., and JACKSON, B. L. 1995 'A factor analysis of landscape pattern and structure metrics' *Landscape Ecology* 10, 23–29
- RIPLEY, B. D. 1976 'The second-order analysis of stationary point processes' *Journal of Applied Probability* 13, 255–266
- SILVERMAN, B. W. 1986 *Density Estimation for Statistics and Data Analysis* (London: Chapman and Hall)
- TURNER, M. G. 1989 'Landscape ecology: the effect of pattern on process' *Annual Review of Ecology and Systematics* 20, 171–197
- TURNER, M. G., GARDNER, R. H., DALE, V. H., and O'NEILL, R. V. 1989a 'Predicting the spread of disturbance across heterogeneous landscapes' *Oikos* 55, 121–129
- TURNER, M. G., GARDNER, R. H., and O'NEILL, R. V. 2001 *Landscape Ecology in Theory and Practice* (New York: Springer-Verlag)
- TURNER, M. G., O'NEILL, R. V., GARDNER, R. H., and MILNE, B. T. 1989b 'Effects of changing scale on the analysis of landscape pattern' *Landscape Ecology* 3, 153–162
- WICKHAM, J. D., RIITTERS, K. H., WADE, T. G., and HOMER, C. 2008 'Temporal change in fragmentation of continental US forests' *Landscape Ecology* 23, 891–898
- WIENS, J. A. 1989 'Spatial scaling in ecology' *Functional Ecology* 3, 385–397
- WIENS, J. A., SCHOOLEY, R. L., and WEEKS, R. D. JR. 1997 'Patchy landscapes and animal movements: do beetles percolate?' *Oikos* 78, 257–264
- WU, J. 2004 'Effects of changing scale on landscape pattern analysis: scaling relations' *Landscape Ecology* 19, 125–138
- WULDER, M. A., DECHKA, J. A., GILLIS, M. D., LUTHER, J. E., HALL, R. J., BEAUDOIN, A., and FRANKLIN, S. E. 2003 'Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program' *The Forestry Chronicle* 79, 1075–1083
- WULDER, M. A., and NELSON, T. 2003 *EOSD land cover Classification Legend Report: Version 2* (Victoria, BC: Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre)
- WULDER, M. A., WHITE, J. C., CRANNY, M., HALL, R. J., LUTHER, J. E., BEAUDOIN, A., GOODENOUGH, D. G., and DECHKA, J. A. 2008 'Monitoring Canada's forests—Part 1: Completion of the EOSD land cover project' *Canadian Journal of Remote Sensing* 34, 549–562