RESEARCH   **Open Access**

Refining spatial neighbourhoods to capture terrain effects

Trisalyn A Nelson[1*] and Colin Robertson[2]

Abstract

Introduction: Spatially explicit ecological research has increased substantially in the past 20 years. Most spatial approaches require the definition of a spatial neighbourhood or the region over which spatial relationships are modelled or assessed. Spatial neighbourhood definitions impact analysis results, and there are benefits in considering neighbourhood definitions that better capture ecological processes. The goal of this research is to present a simple and flexible approach in constraining ecological spatial neighbourhoods using terrain data.

Methods: Using watershed boundaries, we can restrict spatial neighbourhoods from combining populations or processes that should be separated by terrain effects. We demonstrate the need for ecological constraints by way of a simulation study and highlight our approach with a case study examining mountain pine beetle (*Dendroctonus ponderosae*, Coleoptera; Hopkins) infestation hot spots.

Results: Our results demonstrate how failure to constrain neighbourhoods can lead to errors when the spatial signals from unrelated populations are mixed. Also, unconstrained spatial neighbourhoods can unintentionally detect spatial relationships across many scales.

Conclusions: There will be benefits to studies that develop new, ecology-based approaches in defining spatial neighbourhoods that better illuminate ecological function of phenomena under study.

Keywords: spatial weights, spatial analysis, spatial ecology, hot spots, topography

Introduction

Spatial ecology has developed over the past two decades as spatial research questions have intersected with advancing, geographically explicit technology, such as global positioning systems [GPS], geographic information systems [GIS] and remote sensing. Spatial perspectives to ecological research have proven beneficial (e.g. Liebhold and Gurevitch 2002), but they have also presented researchers with new methodological challenges (e.g. Legendre and Fortin 1989; Legendre et al. 2002). In response, many good references have been written to guide ecologists as they collect and analyse spatial data to explore or generate geographically explicit ecological hypotheses (e.g. Fortin and Dale 2005; Dormann et al. 2007; Nelson and Boots 2008). Most spatial methods require an *a priori* definition of a spatial neighbourhood,

also referred to as a spatial weight matrix or kernel (Bavaud 1998; Getis and Aldstadt 2002), and results of analyses are dependent on the spatial neighbourhood definition (e.g. Moilanen and Nieminen 2002). While ecologists are aware of the importance of spatial neighbourhood selections (see Fortin and Dale 2005, pages 113 to 118, for discussion of standard approaches), few definitions capture the ecological function.

Spatial neighbourhoods are commonly categorised as binary or weighted (O'Sullivan and Unwin 2003, pages 44 to 45). If binary, locations are either part of the same neighbourhood or not. Binary neighbourhoods, which are computationally simple and require the least *a priori* knowledge, are the most common. Weighted neighbourhoods allow the amount of interaction between two locations within the same neighbourhood to vary based on proximity or some other measure. Weighting is most often based on linear distance decay (O'Sullivan and Unwin 2003, pages 44 to 45). Spatial neighbourhoods can also be fixed or adaptive (e.g. Aldstadt and Getis

* Correspondence: trisalyn@uvic.ca
[1]Spatial Pattern Analysis and Research Lab, Department of Geography, University of Victoria, PO Box 3060, Victoria, British Columbia, V8W 3R4S, Canada
Full list of author information is available at the end of the article

2006). Fixed neighbourhoods have the same definition regardless of location in the study area, whereas adaptive neighbourhoods enable the definition of neighbourhood to vary over the study area. It may be appropriate to use adaptive neighbourhood definitions when a study area is known to have fine scale processes in one region and coarse scale processes in another. Due to ease of use and interpretability, most spatial studies use fixed neighbourhoods.

Within the aforementioned categories of spatial neighbourhoods, there are many possible definitions. A complete enumeration of spatial neighbourhoods is beyond the scope of this paper. Here, we focus on neighbourhood definitions that are most commonly used with point and areal data, and we refer the reader to O'Sullivan and Unwin (2003, pages 41 to 52) for more details. The majority of studies use definitions based on distance, *k* neighbours or contiguity (Griffith 1996).

Distance definitions can be employed by buffering each location in the study with a radius of a specific distance (Figure 1a). A binary distance neighbourhood would assign all locations within the buffer to the same neighbourhood. Distance weighting could also be employed, and distances used to buffer locations may be variable or fixed in terms of size (Figure 1.1 in O'Sullivan and Unwin 2003). When distance neighbourhoods are of fixed size and when densities vary over space, the number of locations in each neighbourhood will change.

k-Neighbour definitions are used to create neighbourhoods with a constant number of locations in each neighbourhood (Griffith 1996; Figure 1b). *k*-Neighbour statistics are used to identify the nearest data locations or quantify the distance to a specific number of nearest neighbours (see Bailey and Gatrell 1995 for further discussion and equations). For instance, *k* could equal 10, and all locations would be assigned neighbourhoods that
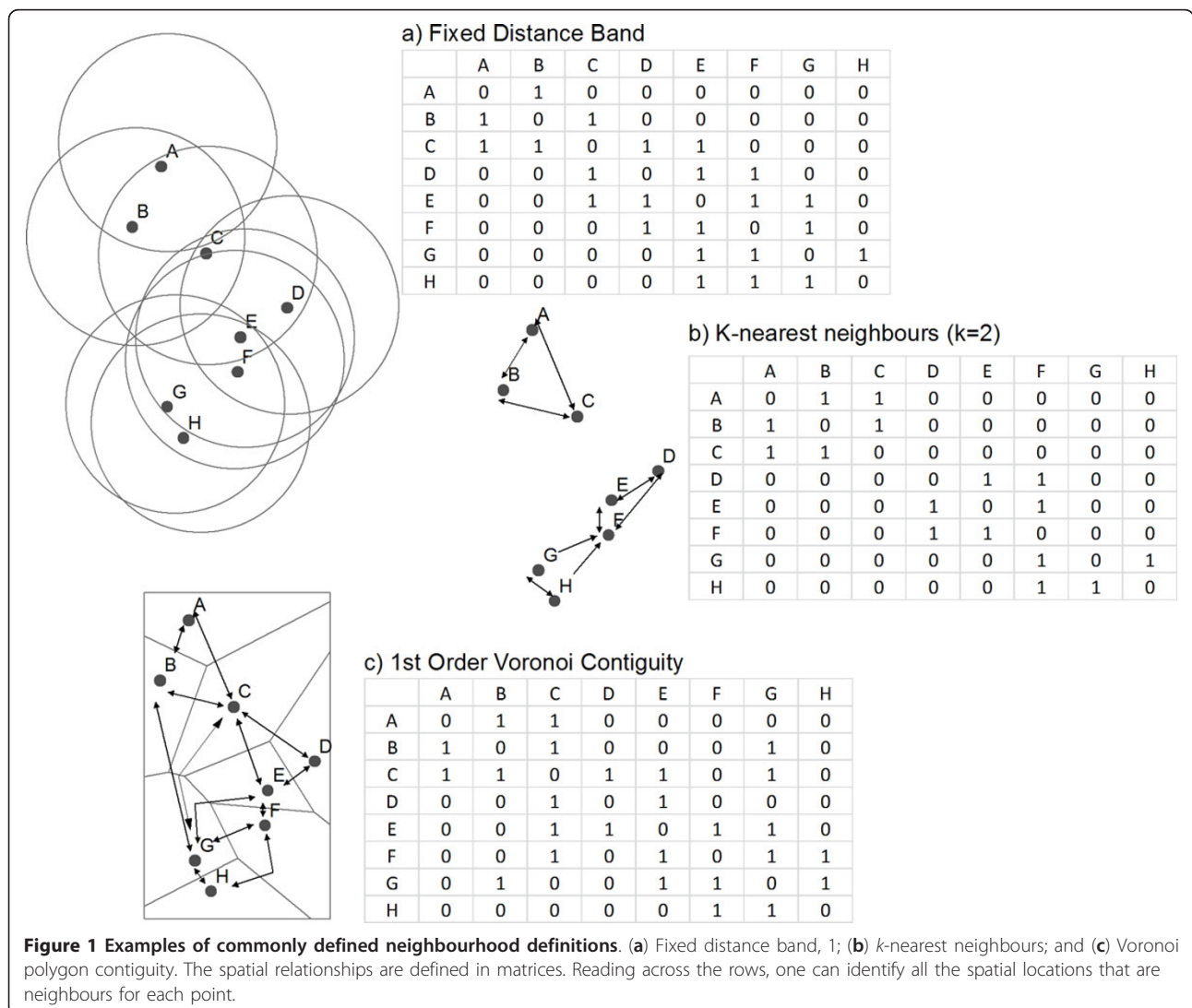

a) Fixed Distance Band

	A	B	C	D	E	F	G	H
A	0	1	0	0	0	0	0	0
B	1	0	1	0	0	0	0	0
C	1	1	0	1	1	0	0	0
D	0	0	1	0	1	1	0	0
E	0	0	1	1	0	1	1	0
F	0	0	0	1	1	0	1	0
G	0	0	0	0	1	1	0	1
H	0	0	0	0	1	1	1	0

b) K-nearest neighbours (k=2)

	A	B	C	D	E	F	G	H
A	0	1	1	0	0	0	0	0
B	1	0	1	0	0	0	0	0
C	1	1	0	0	0	0	0	0
D	0	0	0	0	1	1	0	0
E	0	0	0	1	0	1	0	0
F	0	0	0	1	1	0	0	0
G	0	0	0	0	0	1	0	1
H	0	0	0	0	0	1	1	0

c) 1st Order Voronoi Contiguity

	A	B	C	D	E	F	G	H
A	0	1	1	0	0	0	0	0
B	1	0	1	0	0	0	1	0
C	1	1	0	1	1	0	1	0
D	0	0	1	0	1	0	0	0
E	0	0	1	1	0	1	1	0
F	0	0	1	0	1	0	1	1
G	0	1	0	0	1	1	0	1
H	0	0	0	0	0	1	1	0

Figure 1 Examples of commonly defined neighbourhood definitions. (**a**) Fixed distance band, 1; (**b**) *k*-nearest neighbours; and (**c**) Voronoi polygon contiguity. The spatial relationships are defined in matrices. Reading across the rows, one can identify all the spatial locations that are neighbours for each point.

include the 10 nearest neighbouring locations. If data locations are close together in one part of the study area and far apart in another region, the area associated with each neighbourhood will change. Some statistical, spatial measures are more stable when k is constant or larger than a threshold; for instance, the distribution of the Getis statistics is known when $k \geq 8$, and statistical considerations may make k neighbours a useful neighbourhood definition (Ord and Getis 1995).

Contiguity- or adjacency-based neighbourhoods are based on topology (O'Sullivan and Unwin 2003, pages 43 to 44) and are best described in an application to spatial data represented as uninterrupted polygons (i.e. satellite, census or forest inventory data; Figure 1c). Contiguity is defined based on polygon adjacency or shared boundaries (Dubin 2009). First-order contiguity assigns to the same neighbourhood all locations that share a boundary with a single polygon. Higher-order contiguity can be defined by including polygons that share boundaries with any first-order polygon. If data are represented as points, contiguity can be used to define neighbourhoods if points are first tessellated using triangulation or Voronoi polygons (O'Sullivan and Unwin 2003, pages 50 to 52). Voronoi polygons tessellate space so that all locations within a polygon are closer to the point where the polygon is generated from than all other points (for further details and equations, we refer the reader to Okabe et al. 2000). Properties of neighbourhood structures defined by contiguous Voronoi polygons can be a link between the point pattern and the generating spatial process. For example, complete spatial randomness processes will typically produce neighbourhood structures with six first-order neighbours (Okabe et al. 2000). For regular lattices, contiguity definitions can include all neighbours (queen) or can be directional, using only neighbours in the north, south, east and west directions (rook) or only in corner locations (bishop) (Dubin 2009). Contiguity-based weighted neighbourhoods can also be weighted, for instance, by length of shared boundaries.

Exploration of spatial neighbourhood definitions is often insufficiently explored when conducting analysis. There are many examples where the definition of spatial neighbourhood is not even stated in spatial literature. In part, it can be wise to avoid drawing attention to the selection of spatial neighbourhood, which is inherently subjective and easy fodder during the review process. However, and more importantly, the impact of spatial neighbourhood has not been fully explored in spatial literature, and relatively little attention has been given to appropriate selection of spatial neighbourhoods in ecology specifically (Fortin and Dale 2005, pages 113 to 118). The absence of spatial neighbourhood selection discussions is problematic in ecology given that most

standard definitions were developed outside the discipline. Ecology has unique spatial issues to consider in selecting spatial neighbourhoods. For example, if the objective of analysis was to model the spatial diffusion of mountain pine beetle across a heterogeneous landscape, a fixed distance definition of spatial neighbourhoods may produce misleading results. Mountain pine beetle populations disperse along valleys, and mountains are typically a barrier to dispersion (Robertson et al. 2009a). As such, when modelling diffusion or detecting hot spots, spatial neighbourhoods should be restricted within valleys or along the sides of mountains and should not reach across mountain peaks or ridges. The underlying issue is that the conceptualisation of space in most spatial analysis methods is in reference to a two-dimensional plane when, in reality, topographic relief structures much of the ecological variation in plant and animal communities. Neighbourhoods defined by distance will ignore topography and may aggregate the spatial signals from multiple, unrelated sub-populations.

When study areas are small, it is reasonable to modify spatial neighbourhoods manually. However, as with many applied areas of spatial analysis, ecological data sets are increasingly large, and automated approaches are required for appropriate spatial neighbourhood delineation. The goal of this research is to present a semi-automated approach to the delineation of spatial neighbourhoods that captures impacts of terrain features on ecological function. To meet this goal, we use terrain data to delineate watersheds and demonstrate how watersheds can be used to constrain spatial neighbourhoods. Watersheds are useful because they are often divided by mountain peaks and can be used to partition terrain that separates species' populations. Our general approach is demonstrated through the application of a local measure of spatial autocorrelation, Moran's I_i (Anselin 1995), useful for detecting spatially explicit hot spots or locations where values are similar and extreme relative to the mean. We apply our topographic approach in constraining spatial neighbourhoods to simulated data and a case study identifying infestation hot spots in epidemic mountain pine beetle populations in British Columbia, Canada. The methods and results for the simulation and case study are presented separately. First, however, we will outline our terrain approach to spatial neighbourhood definition and outline the spatial statistic used in our demonstrations.

Approach to spatial neighbourhood delineation

Our approach in modifying spatial neighbourhoods is to constrain standard definitions of spatial neighbourhoods, like distance, by watershed boundaries. The topographic approach to spatial neighbourhood selection will be most appropriate where sub-populations are separated by topographic barriers, when dispersal is constrained

by mountains, valleys or water bodies, or when samples from different topographic regions are known *a priori* to have unique characteristics. Also, for spatial research of hydrological applications, a watershed approach has been shown to be beneficial (Laffan 2002). Our method's simplicity is its strength, in which creating more ecologically meaningful weight matrices is a relatively straightforward extension to standard techniques, and constraints are easily implemented in standard GIS software packages. Using an elevation model, watershed boundaries can be delineated. It may be necessary to remove watershed boundaries in low-lying areas that do not have terrain high enough to constrain ecological function. Spatial neighbourhoods can then be defined by distance, k neighbours or contiguity, but the watersheds provide a barrier forcing neighbours to be selected from only within watersheds. More specific steps will be provided.

There are many good textbooks that outline how to delineate watersheds using elevation data, and it is not our aim to replicate these (see Chang 2006, Chapter 15). Briefly, there are four conceptual steps. First, the direction that water will flow out of each location (cell) in a digital elevation model [DEM] is calculated (flow direction). Second, flow accumulation is tabulated as the number of cells that water will flow through to accumulate in each cell. Third, a stream network is defined by applying a threshold to the flow accumulation layer. Selecting the accumulation threshold value is largely subjective. Larger thresholds will generate smaller watersheds, and threshold selection enables users to control the scale of watershed delineation. The accumulation threshold should be set to ensure that 'impassable' terrain is separated by the watershed boundaries. Also, it is important that each watershed is large enough to have a sufficient number of data points to ensure the spatial analysis or that statistic is robust. In our study, we set the accumulation threshold to ensure that large terrain features such as mountain ridges were spatially separated. For further discussion of the impact of watershed accumulation threshold selection, we refer to Band (1993). Fourth, using inputs generated on flow, watersheds are defined around each stream.

Once watersheds are defined, it is best to assess the characteristics of the terrain within each watershed. If the watershed has low variability or maximum elevation, variation or height of terrain may be insufficient to impact ecological function, and watershed boundaries should be dissolved. The remaining watershed boundaries may be used to constrain the definition of spatial neighbourhood in combination with any standard approach in assigning spatial neighbours.

Moran's I_i

To show the impact of spatial neighbourhood definition on spatial analysis and the benefit of the terrain approach to neighbourhood definition, we undertake hot spot detection using local Moran's I_i, a measure of spatial autocorrelation. Spatial autocorrelation is the first law of geography which states that all things are related and that near things more so than far (Tobler 1965). Positive spatial autocorrelation exists when nearby events are similar. Negative spatial autocorrelation exists when nearby events are dissimilar. Local Moran's I_i quantifies spatial autocorrelation for every location in the study region, differing from its global counterpart, which generates one summary value for an entire study area (Anselin 1995).

If the observed values x, $i \in \{1, ..., n\}$ of a random variable x are recorded at a set of n data sites, then local measures of spatial autocorrelation take the general form of a cross-product statistic:

$$\Gamma_i = \sum_j w_{ij} \gamma_{ij} \tag{1}$$

where w_{ij} is a measure of the spatial relationships of data sites i and j at a given time, and y_{ij} is a measure of their relationship in attribute space (Getis and Ord 1992; Boots 2002).

For local Moran's I_i, the attribute relationship in the cross-product statistic is defined as $\gamma_{ij} = (x_i - \overline{x})(x_j - \overline{x})$ and

$$I_i = \left(\frac{z_i}{\left(\frac{\sum_i z_i^2}{n} \right)} \right) \sum_j w_{ij} z_j, \tag{2}$$

where $z_{ij} = (x_i - \overline{x})$. Positive values of Moran's I_i indicate positive spatial autocorrelation in values that are extreme relative to the mean. Negative Moran's I_i values indicate a negative spatial autocorrelation in values that are extreme relative to the mean. When Moran's I_i approaches 0, it could be that there is no spatial autocorrelation or that spatial autocorrelation is present in values near the mean.

For hot spot detection, a Moran's scatter plot is useful to categorise positive and negative spatial autocorrelations based on the attribute value of a location in relation to the attribute value of its neighbours. On a Moran's scatter plot, the x axis is the attribute value in deviation form and the y axis is a standardized average of the neighbour values also in deviation form (Anselin 1995). The upper right

quadrant indicates high values surrounded by high values (high-high); the upper left quadrant indicates low values surrounded by high values (low-high). The lower right quadrant indicates high values surrounded by low values (high-low), and the lower left quadrant indicates low values surrounded by low values (low-low) (see Nelson and Boots 2008 for details and figure of scatter plot). Moran's scatter plot can be combined with statistical testing to determine the significance of spatial autocorrelation. Expected values of I_i can be derived under the hypothesis of no local spatial autocorrelation (Sokal et al. 1998); however, the distribution does not follow a known distribution (Boots and Tiefelsdorf 2000), and a randomization approach must be adopted for statistical testing. It should be mentioned that all local spatial statistics are complicated by the presence of global spatial autocorrelation and multiple and correlated tests (Boots 2002). As such, statistical significance is usually considered exploratory rather than confirmatory. In our study, hot spots are defined as high-high values that are statistically different from configurations of data expected by chance.

Methods

Simulation

A simulation study was conducted to examine the theoretical impacts of using spatial neighbourhoods constrained by terrain. When a broad phenomenon, such as terrain, influences the mean of a spatial process, landscapes exhibit first-order non-stationarity and violate the assumption of most spatial statistics (Kabos and Csillag 2002). The aim of the simulation is to examine the impact of accounting for terrain constraints when doing spatial analysis on point patterns that exhibit first-order non-stationarity.

To create a simulated landscape with first-order non-stationarity, a spatially autocorrelated process was generated onto a 30 × 30-km grid. Each grid cell value was drawn from a normal distribution with a mean of 0 and a spatially structured standard deviation. Standard deviations were estimated with the spatial covariance function (Ripley 1981) up to a maximum distance of 10 km (i.e. range parameter) with a standard error of 1 km. The spatially correlated surface was categorised into 5 quantiles (Figure 2a), representing distinct theoretical landscape types whereby ecological processes of interest within each landscape were functionally independent. A 3 × 3 modal-spatial filter was run over the surface to remove unrealistically small landscape polygons. Separate autocorrelation processes were simulated for grid cells in each of the landscape polygons, thereby creating first-order homogeneity within each polygon and non-stationarity across polygon boundaries (Figure 2b). Points (n = 1,000) were randomly
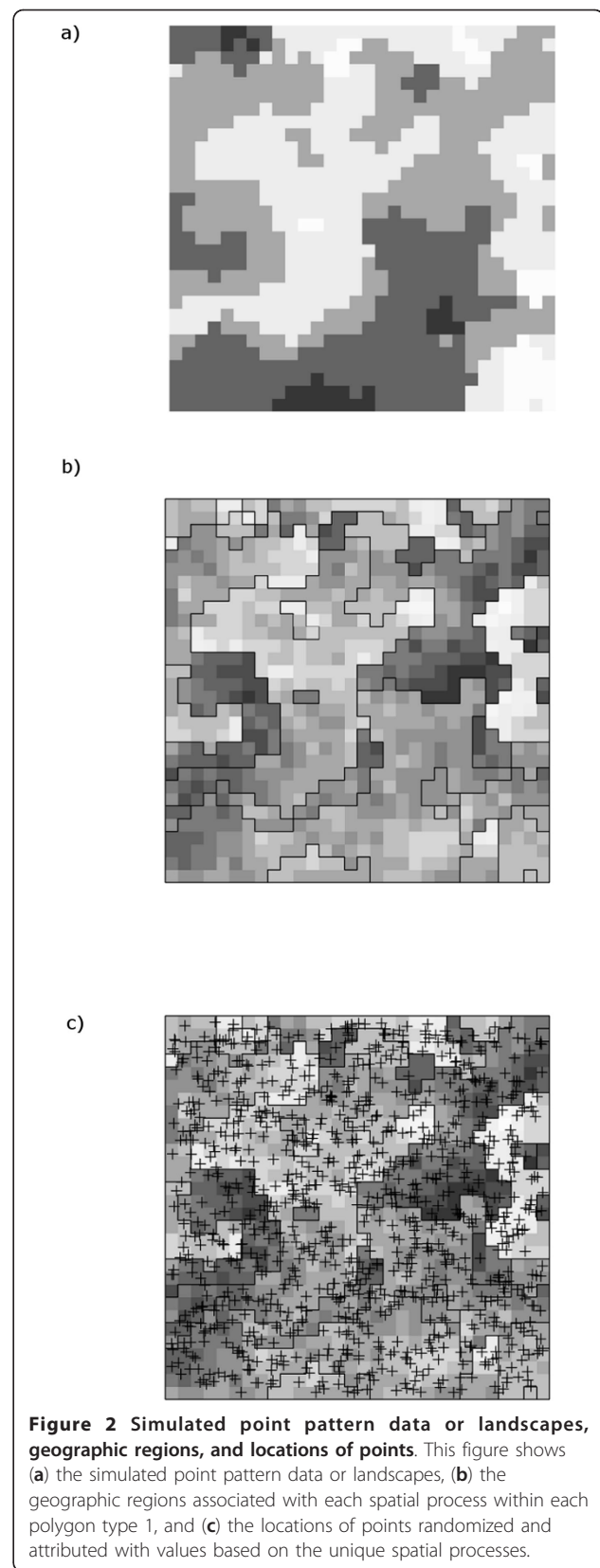

Figure 2 Simulated point pattern data or landscapes, geographic regions, and locations of points. This figure shows (**a**) the simulated point pattern data or landscapes, (**b**) the geographic regions associated with each spatial process within each polygon type 1, and (**c**) the locations of points randomized and attributed with values based on the unique spatial processes.

allocated to the study area (Figure 2c), and assigned values were extracted from the grid cells. This process was replicated 99 times to create a total of 99 simulated point patterns with different realisations of the non-stationary process. These point patterns were used for subsequent hot spot analysis.

Hot spots were quantified for the simulated data using local Moran's I_i. Pseudo-significance was determined via randomisation using 999 permutations. The spatial neighbourhoods defined were binary, fixed and based on distance. Neighbourhoods were defined using constraints similar to topographic barriers (i.e. only using points within a landscape polygon to calculate local Moran's I_i) and also a standard unconstrained definition (using all points within the spatial neighbourhood). For both constrained and unconstrained neighbourhoods, several distances were used to explore the impact of the neighbourhood scale on analysis results. Distances of 2, 5, 10, 15, and 20 km were examined, and the number of hot spots was tabulated for each of the 99 point patterns.

Case study
Beginning in the mid 1990s, British Columbia became home to the largest recorded infestation of mountain pine beetle (Robertson et al. 2009a). The infestation continues to have substantial impact on pine (*Pinus* spp.) forests. Due to infestation severity, many spatial studies of mountain pine beetle have been conducted, and standard spatial neighbourhood definitions are typically employed. Given that mountain pine beetle tends to spread via valleys and that mountains and large water bodies can be barriers to dispersal (Robertson et al. 2009b), standard neighbourhood definitions are not always appropriate. Ideally, neighbourhood relationships are defined based on knowledge of the ecosystem functioning in the study area. However, landscape scale studies require large data sets, making manual definition of spatial neighbourhoods problematic. We hypothesize that watershed boundaries are ideal for restricting spatial neighbourhoods to represent the ecological characteristics of mountain pine beetle spatial processes as terrain impediments to dispersal can be captured.

Study area and data
The case study area is the southern portion of the Vanderhoof Forest District located in central British Columbia. This area of Vanderhoof was chosen due to its topographic variability, ranging from being flat in the north to being mountainous in the south. The Vanderhoof Forest District has experienced epidemic levels of mountain pine beetle over the last decade (Nelson et al. 2006a).

Vanderhoof Forest District monitors the mountain pine beetle infestation using point-based, GPS-based aerial surveys (Nelson et al. 2006a). Aerial surveys of mountain

pine beetle infestations use indicators of pine mortality, mainly changes in crown foliage colour, to monitor mountain pine beetle activity. During aerial surveys, clusters of visually infested trees are identified, typically those with yellow and red crowns, indicating mortality 1 to 2 years prior to infestation, and a GPS is used to map cluster centres with a point. For each cluster, the number of infested trees is estimated and the infesting insect species, recorded. Attributes have been shown to be accurate to ± 10 trees for 92.6% of points (Nelson et al. 2006b). In the current study, there are a total of 14,117 GPS points representing 170,639 infested trees.

A DEM was used to derive watershed boundaries. The elevation model had 25-m^2 grid cells and was created from the 1:20,000 scale of Terrain Research Information Management data (Province of British Columbia 1996). The data were interpolated using a linear interpolation process, and the DEM is reported to be accurate within 10 m (Province of British Columbia 1996).

Methods
For the mountain pine beetle example, the first step was to delineate the watersheds. The elevation model was clipped to an area that extended 50 km beyond the study area boundary to ensure watershed boundaries were defined without edge effects. Watersheds were defined in both the flat and mountainous terrain throughout the study area. However, to adequately represent ecological function, spatial neighbourhoods only need to be constrained in mountainous regions; low-lying terrain will not impact the spatial processes of mountain pine beetles. To select mountainous watersheds, each watershed was attributed with a maximum elevation. All watersheds with a maximum elevation ≥ 1,800 m were identified as regions where the neighbourhood should be adjusted. The maximum elevation threshold is also subjective, but in this case, it was based on mountain pine beetle biology (Safranyik and Carroll 2006).

Binary, fixed spatial neighbourhoods were calculated using unconstrained and watershed-constrained distance neighbours. Spatial neighbourhoods were generated using distances of 1, 5, 10, 15, and 20 km. Using all spatial neighbourhood definitions, local Moran's I_i statistics were calculated, and the number of hot spots identified compared across neighbourhood definitions.

Results and discussion
Simulation
At all distance lags, the number of hot spots detected is higher for constrained compared with unconstrained neighbourhoods (Figure 3). A lower number of hot spots for unconstrained neighbourhoods are the result of including multiple processes within the same neighbourhood. When processes from multiple landscapes are included in the same neighbourhood and simultaneously
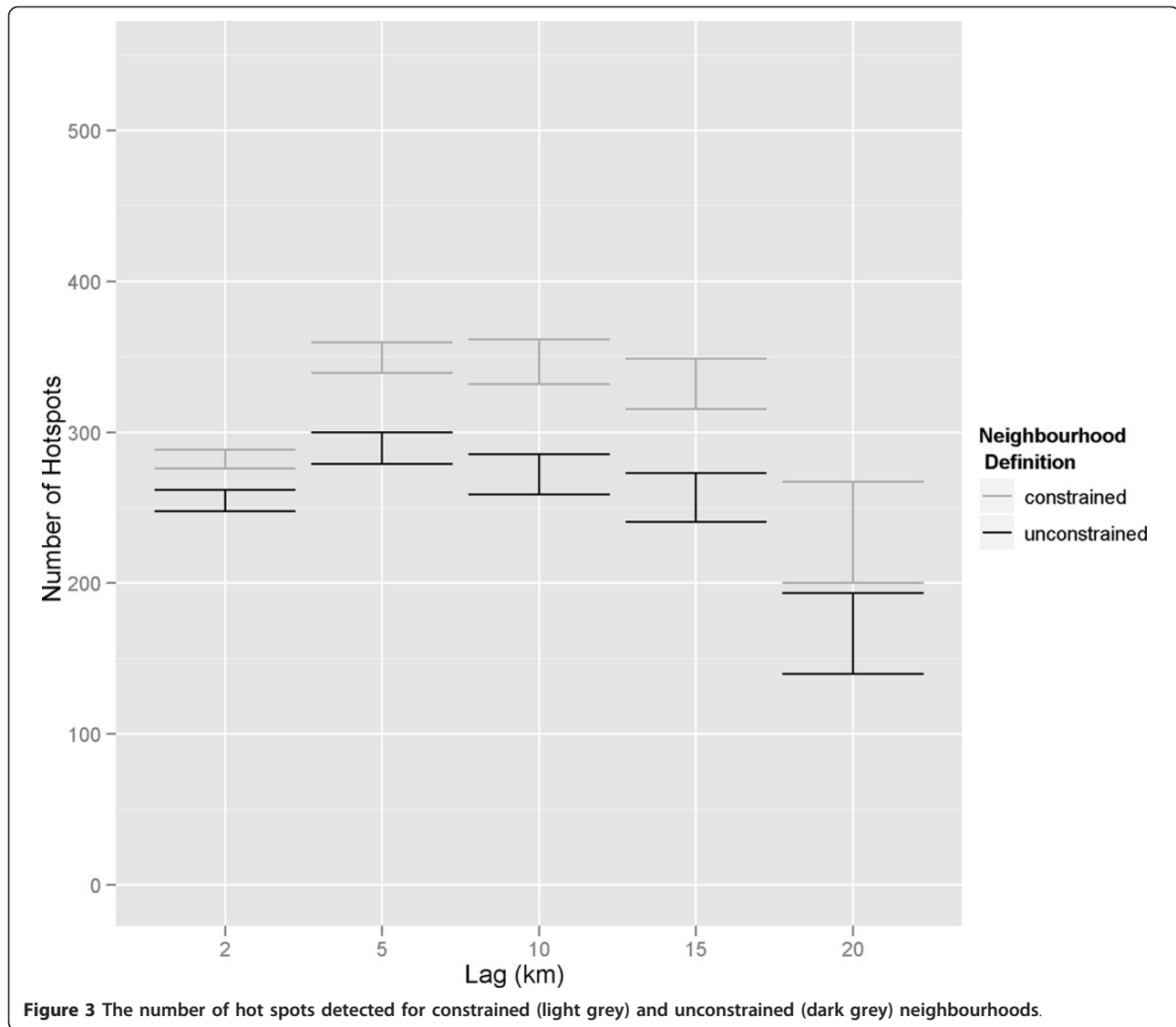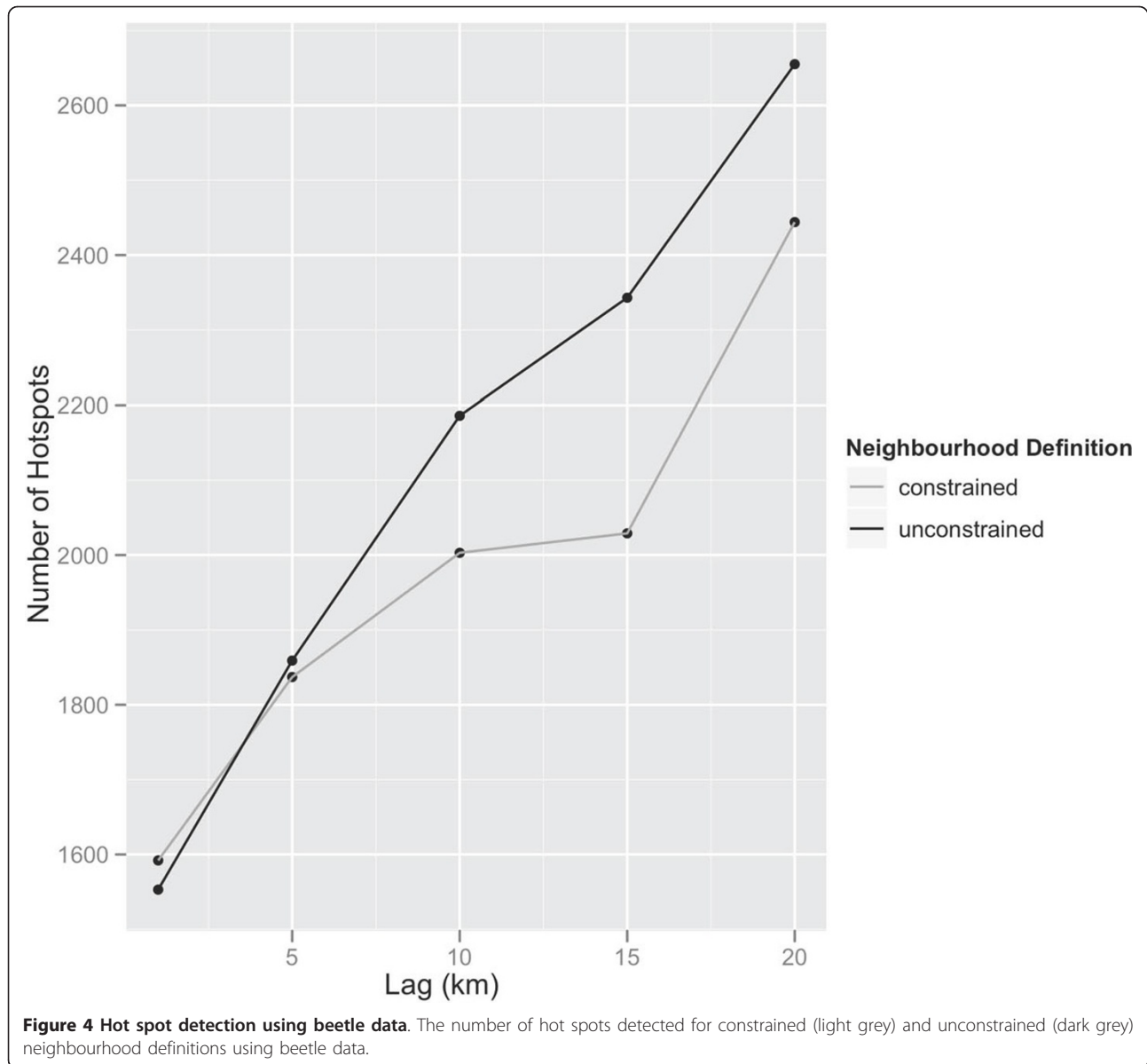
Figure 3 The number of hot spots detected for constrained (light grey) and unconstrained (dark grey) neighbourhoods.

considered by Moran's I_i statistic, spatial patterns from two processes are mixed, and the spatial pattern signal, masked. This simulation represents an extreme case, where processes change markedly across polygon boundaries. In real landscapes, ecological boundaries are generally fuzzy, and measurements taken in separate landscape types may still be correlated. Spatial patterns are always generated from a mixture of processes. However, even when patterns generated from multiple processes exhibit correlation, results of analysis that have inadvertently mixed processes, for instance multiple dispersal processes, may be misleading and lead to an erroneous hypothesis generation.

Case study
In the mountain pine beetle example, more hot spots are detected with unconstrained neighbourhoods for all

spatial neighbours defined by distances larger than 1 km (Figure 4). Figure 5 shows how the locations of hot spots change when neighbourhoods are constrained. Of particular note are the linear features along the mountain sides which are identified as hot spots when unconstrained and becoming insignificant when constrained. These results reflect the unique biology of the mountain pine beetle. Beetles are generally poor fliers and will only disperse locally up from 1 to 5 km (Barclay et al. 2005). As such, if local dispersal was the single mechanism for beetle movement, only locations within 5 km from one another that are in the same valley should be related. However, a small portion of mountain pine beetles are also dispersed over longer distances by wind, which may cause autocorrelation at longer distances. In addition, while the mechanism is not fully understood, mountain pine beetle populations, as well as other

Figure 4 Hot spot detection using beetle data. The number of hot spots detected for constrained (light grey) and unconstrained (dark grey) neighbourhood definitions using beetle data.
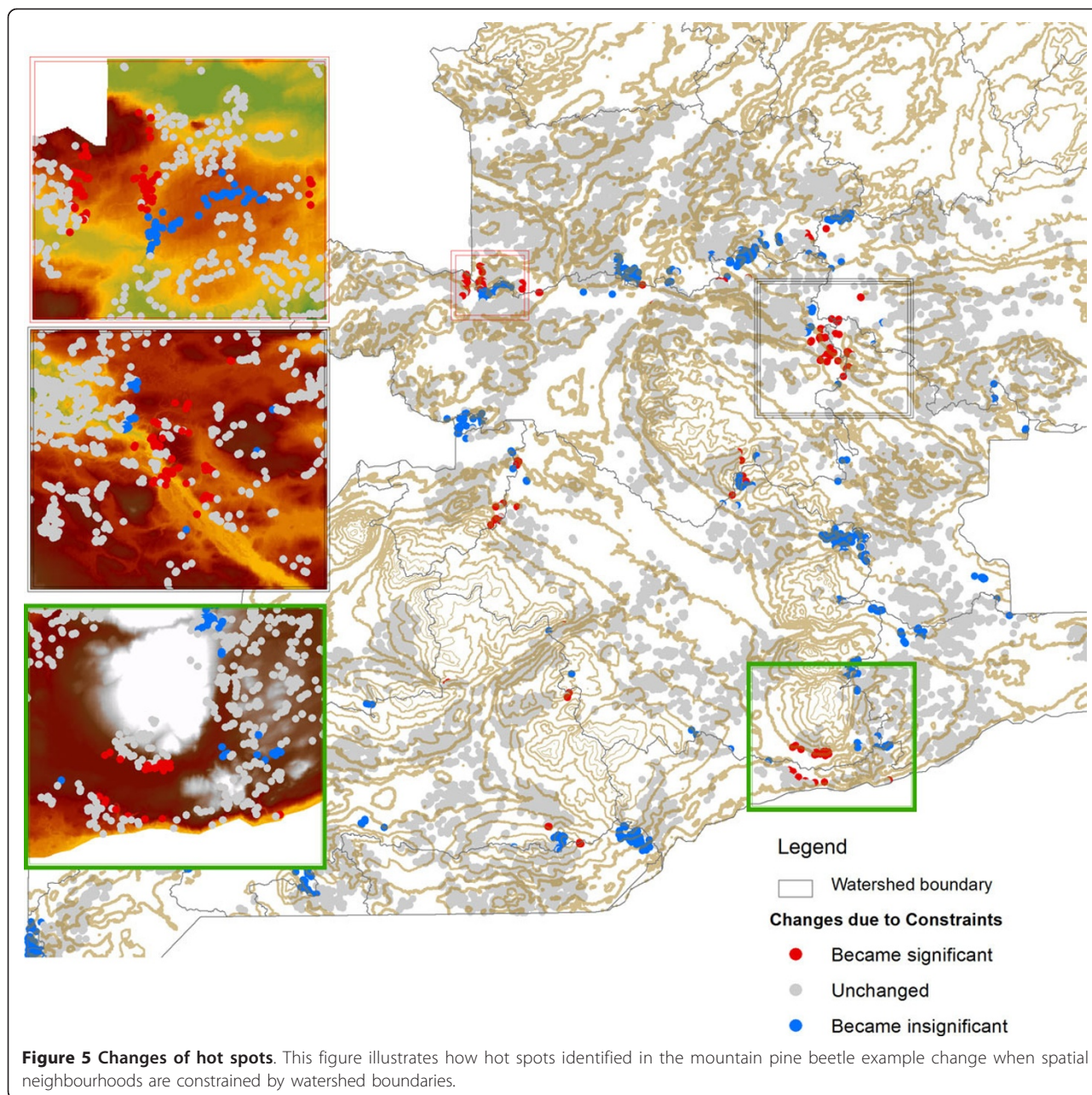
insects, exhibit spatial synchrony at very large distances (Aukema et al. 2006). In the case of the mountain pine beetle, synchronous populations are likely linked to regional weather patterns as the mountain pine beetle life cycle is largely temperature dependent. Thus, while constrained hot spots likely represent local interactions associated with below-canopy dispersal, unconstrained hot spots reflect synchronous population fluctuations due to environmental factors and potentially some longer-range dispersal (Robertson et al. 2007). From a management perspective, identifying dispersal hot spots is likely of most interest in mitigating the spread of beetles. Careful selection of spatial neighbourhoods enables different aspects and scales of beetle spatial dynamics to be investigated in the case study.

Conclusions

Neighbourhood definitions will impact results of spatial analyses. There is an opportunity to develop spatial neighbourhoods that better consider ecological processes. We present an approach to semi-automated spatial neighbourhood definition, which is appropriate when terrain, particularly mountains and valleys, influence ecological spatial processes. Once again, the strength of the terrain-based approach is its simplicity as it requires only a DEM to implement.

In a simulated environment, we demonstrate that multiple uncorrelated processes and spatial correlations will be fewer if spatial neighbourhoods are defined *across* process boundaries. Perhaps even more concerning would be a less extreme scenario than what we present in our

Figure 5 Changes of hot spots. This figure illustrates how hot spots identified in the mountain pine beetle example change when spatial neighbourhoods are constrained by watershed boundaries.

simulation. If processes are distinct but spatial patterns happen to be correlated, the spatial patterns identified as significant may generate misleading hypotheses. Given that there are usually a diverse range of processes impacting the generation of spatial patterns, correlated data across landscapes with different processes are expected. The case study provides a more realistic example of the importance of spatial neighbourhood selection. In the mountain pine beetle example, patterns are correlated across topographic barriers. Given multiple dispersal processes, the spatial neighbourhood definition influences the scale of spatial processes that are

investigated. Spatial neighbourhoods constrained by topography are best used for investigating a local scale within stand dispersal. Larger unconstrained neighbourhoods explore patterns associated with long-distance dispersal and spatial synchrony.

Edge effects are another challenge impacting most spatial analyses. Edge effects that arise as data locations near the centre of the study area can have neighbourhoods defined in all directions, whereas data sites near the study area extent will only have neighbours towards the study area centre (O'Sullivan and Unwin 2003, pages 40 to 41). Unless additional data are collected to create

a buffer of data outside the study area, edge effects occur. A few methods (*k*-function) have sophisticated edge corrections, but many do not. Edge effects are beyond the scope of this paper. However, it is important to point out that creating internal boundaries, as we do with topographic constraints, creates internal edge effects.

Further research on appropriate spatial neighbourhoods for spatial ecological research will be beneficial. We hope we have illuminated the issue and provided an example solution that is flexible. The approach we have presented to spatial neighbourhoods is primarily appropriate when analysing point or areal data. We have not dealt explicitly with networks or spatial neighbourhoods for assessing connectivity (i.e. Moilanen and Nieminen 2002).

Thoughtful neighbourhood selection is imperative to ensuring robust results. Perhaps the most unnerving result of our research is that regardless of the spatial neighbourhood used, we were able to describe a spatial pattern; yet, we have shown how easily one might examine a process operating at a different scale than the study or hypothesis is designed to test. At a minimum, explicit statement and justification of spatial neighbourhood selection should be standard in peer-reviewed ecological literature using spatial approaches.

Author details
[1]Spatial Pattern Analysis and Research Lab, Department of Geography, University of Victoria, PO Box 3060, Victoria, British Columbia, V8W 3R4S, Canada [2]Department of Geography and Environmental Studies, Wilfrid Laurier University, Waterloo, Ontario, N2L 3C5, Canada

Authors' contributions
TN led the writing and analysis of the application. CR led the simulation analysis and development of figures. Both authors contributed to the development of goals. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

References
Aldstadt J, Getis A (2006) Using AMOEBA to create a spatial weights matrix and identify spatial clusters. Geogr Anal 38(4):327–343

Anselin L (1995) Local indicators of spatial association–LISA. Geogr Anal 27(2):93–115

Aukema BH, Carroll AL, Zhu J, Raffa KF, Sickley TA, Taylor SW (2006) Landscape level analysis of mountain pine beetle in British Columbia, Canada: spatiotemporal development and spatial synchrony within the present outbreak. Ecography 29:427–441

Bailey TC, Gatrell AC (1995) Interactive spatial data analysis. Longman, Essex

Band LE (1993) Effect of land surface representation on forest water and carbon budgets. J Hydrol 125(2-4):749–722

Barclay HJ, Li C, Benson L, Taylor S, 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. Can Entomol 137(5):566–583

Bavaud F (1998) Models for spatial weights: a systematic look. Geogr Anal 30(2):153–171

Boots B (2002) Local measures of spatial association. Ecoscience 9:168–176

Boots B, Tiefelsdorf M (2000) Global and local spatial autocorrelation in bounded regular tessellations. J Geog Syst 2:319–348

Chang KT (2006) Introduction to geographic information systems. McGraw Hill, New York

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30(5):609–628

Dubin R (2009) Spatial weights. The SAGE handbook of spatial analysis. SAGE Publications, London

Fortin MJ, Dale MRT (2005) Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge

Getis A, Aldstadt J (2002) Constructing the spatial weights matrix using a local statistic. Perspectives on spatial data analysis. Springer, New York

Getis A, Ord J (1992) The analysis of spatial association by use of distance statistics. Geog Anal 24:189–206

Griffith D (1996) Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. Practical handbook of spatial statistics. CRC Press, Boca Raton

Kabos S, Csillag F (2002) The analysis of spatial association on a regular lattice by join-count statistics without the assumption of first-order homogeneity. Comp Geosci 28(8):901–910

Laffan SW (2002) Using process models to improve spatial analysis. Int J Geogr Inf Sci 16(3):245–257

Legendre P, Fortin MJ (1989) Spatial pattern and ecological analysis. Plant Ecol 80:107–138

Legendre P, Dale MRT, Fortin M-J, Gurevitch J, Hohn M, Myers D (2002) The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography 25:601–615

Liebhold AM, Gurevitch J (2002) Integrating the statistical analysis of spatial data in ecology. Ecography 25(5):553–557

Moilanen A, Nieminen M (2002) Simple connectivity measures in spatial ecology. Ecology 83(4):1131–1145

Nelson T, Boots B, Wulder MA (2006) Large-area mountain pine beetle infestations: spatial data representation and accuracy. Forest Chron 82:243–252

Nelson T, Boots B, Wulder MA, Shore T, Safranyik L, Ebata T (2006) Rating the susceptibility of forests to mountain pine beetle infestations: the impact of data. Can J Forest Res 36:2815–2825

Nelson TA, Boots B (2008) Detecting spatial hot spots in landscape ecology. Ecography 31(5):556–566

Okabe A, Boots B, Sugihara K, Chiu SN (2000) Spatial tessellations: concepts and applications of Voronoi diagrams. Wiley, West Sussex

Ord JK, Getis A (1995) Local spatial autocorrelation statistics: distributional issues and an application. Geogr Anal 27(4):286–306

O'Sullivan D, Unwin DJ (2003) Geographic information analysis. Wiley, New Jersey

Province of British Columbia (1996) Gridded DEM specifications. Ministry of Sustainable Resource Management, Victoria

Ripley BD (1981) Spatial statistics. Wiley, New Jersey

Robertson C, Nelson TA, Boots B (2007) Mountain pine beetle dispersal: the spatial-temporal interaction of infestations. Forest Sci 53(3):395–405

Robertson C, Farmer CJ, Nelson TA, Mackenzie IK, Wulder MA, White JC (2009) Determination of the compositional change (1999-2006) in the pine forests of British Columbia due to mountain pine beetle infestation. Environ Monit Assess 158:593–608

Robertson C, Nelson TA, Jelinski DE, Wulder MA, Boots B (2009) Spatial-temporal analysis of species range expansion: the case of the mountain pine beetle, *Dendroctonus ponderosae*. J Biogeogr 36(8):1446–1458

Safranyik L, Carroll AL (2006) The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. In: Safranyik L, Wilson WR (ed) The mountain pine beetle: a synthesis of biology, management, and impacts on lodgepole pine. Natural Resources Canada, Victoria

Sokal RR, Oden NL, Thomson BA (1998) Local spatial autocorrelation in a biological model. Geog Anal 30:331–354

Tobler WR (1965) Computation of the correspondence of geographical patterns.
Papers in Regional Science 15(1):131–139