NOTE / NOTE

Quantifying bias in pattern indices extracted from spatially offset landscape samples

Guillermo Castilla, Julia Linke, Adam J. McLane, and Gregory J. McDermid

Abstract: Modern ecological models often account for the influence of the surrounding environment by using landscape pattern indices (LPIs) as measures of landscape structure. Ideally, the landscape samples from which these LPIs are extracted should be centered on the locations where the response variable was measured. However, in situations where this is not possible due to a lack of adequate full-coverage landcover data, the question arises as to what degree this circumstance creates a bias in the value of the LPIs, thereby obscuring their relation with the response variable. To address this question, we extracted four representative LPIs from 30 rectangular (3 × 6 km) landscape samples evenly distributed across a 10 000 km^2 boreal forest study area. These rectangles were subjected to systematic displacements across a range of distances (0.5 to 2.5 km) and directions, after which we recomputed the LPIs. We found that a 1 km spatial offset led to an average of 15% deviation of original LPI values. Unfortunately, as the offset increased, the range of resulting deviations also widened, making it difficult to predict this effect. Our findings fill a gap in the literature on landscape pattern analysis and suggest that researchers should avoid LPIs extracted from spatially offset landscape samples.

Résumé : Les modèles écologiques modernes tiennent souvent compte de l'influence de l'environnement avoisinant en utilisant des indices paysagers (IP) comme mesure de la structure du paysage. Les régions pour lesquelles ces IP sont extraits devraient idéalement provenir du centre des endroits où la variable étudiée a été mesurée. Cependant, dans les cas où cela n'est pas possible dû au manque de données qui couvrent adéquatement toute la couverture terrestre, la question surgit à savoir jusqu'à quel point cette situation crée un biais dans la valeur des IP, obscurcissant ainsi leur relation avec la variable étudiée. Pour aborder cette question, nous avons extrait quatre IP représentatifs de 30 échantillons rectangulaires (3 × 6 km) de paysage également répartis dans une zone d'étude de 10 000 km^2 de forêt boréale. Ces rectangles ont été soumis à des déplacements systématiques sur une série de distances (0,5 à 2,5 km) et dans différentes directions après quoi les IP ont été recalculés. Nous avons trouvé qu'un décalage spatial d'un kilomètre entraînait une déviation moyenne de 15 % par rapport aux valeurs d'IP originales. Malheureusement, à mesure que le décalage augmente, la dispersion des déviations qui en résulte s'accroît, ce qui rend difficile la prédiction de cet effet. Nos résultats comblent une lacune dans la littérature et indiquent que les chercheurs devraient éviter d'extraire des IP à partir d'échantillons décalés de paysage.

[Traduit par la Rédaction]

Introduction

Landscape pattern indices (LPIs; or landscape metrics, as they are also commonly known) are quantitative measures of the composition (assortment of habitat types) and configuration (spatial distribution of habitat types) of mosaics representing forested or other "patchy" environments (Turner 2005). LPIs are often used as covariates in ecological models (i.e., as predictors of the response variable), wherein for each observation of the response variable (e.g., the density of cedar saplings within field plot *x*; see Forester et al. 2008), there is one or more LPIs indicative of the spatial context in which the observation is embedded (e.g., ratio of deciduous to coniferous forest within a 2 km radius of plot *x*).

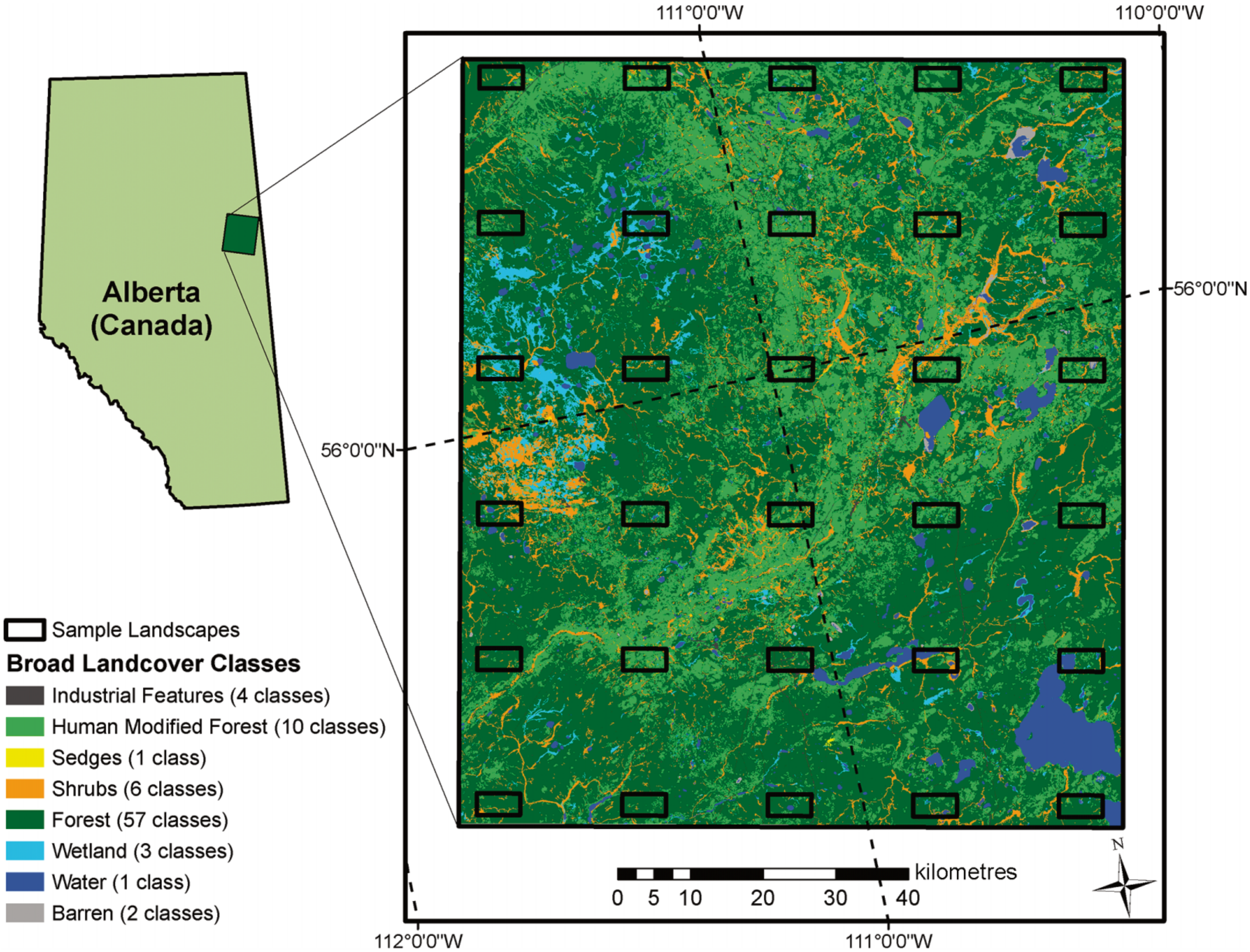
LPIs are typically extracted from areas of fixed size and shape, hereafter referred to as landscape samples. Each landscape sample is normally represented by a mosaic of non-overlapping patches in which no two adjacent patches share the same habitat or landcover type. Ideally, landscape samples should be centered on the location where the ecological variable of interest was observed. This location could be a focal patch of irregular shape and size, a field plot of fixed shape and size, a count station with no formal boundaries, or a transect. However, when there is a lack of adequate full-coverage habitat or landcover data, situations can arise wherein the centers of the landscape samples are spatially offset from the field data locations. For example, an ecological project could require that several species be surveyed

Received 5 May 2011. Accepted 19 July 2011. Published at www.nrcresearchpress.com/cjfr on 4 October 2011.

G. Castilla, J. Linke, A.J. McLane, and G.J. McDermid. Foothills Facility for Remote Sensing and GIScience, Department of Geography, University of Calgary, 2500 University Drive N.W., Calgary, AB T2N 1N4, Canada.

Corresponding author: Guillermo Castilla (e-mail: gcastill@ucalgary.ca).

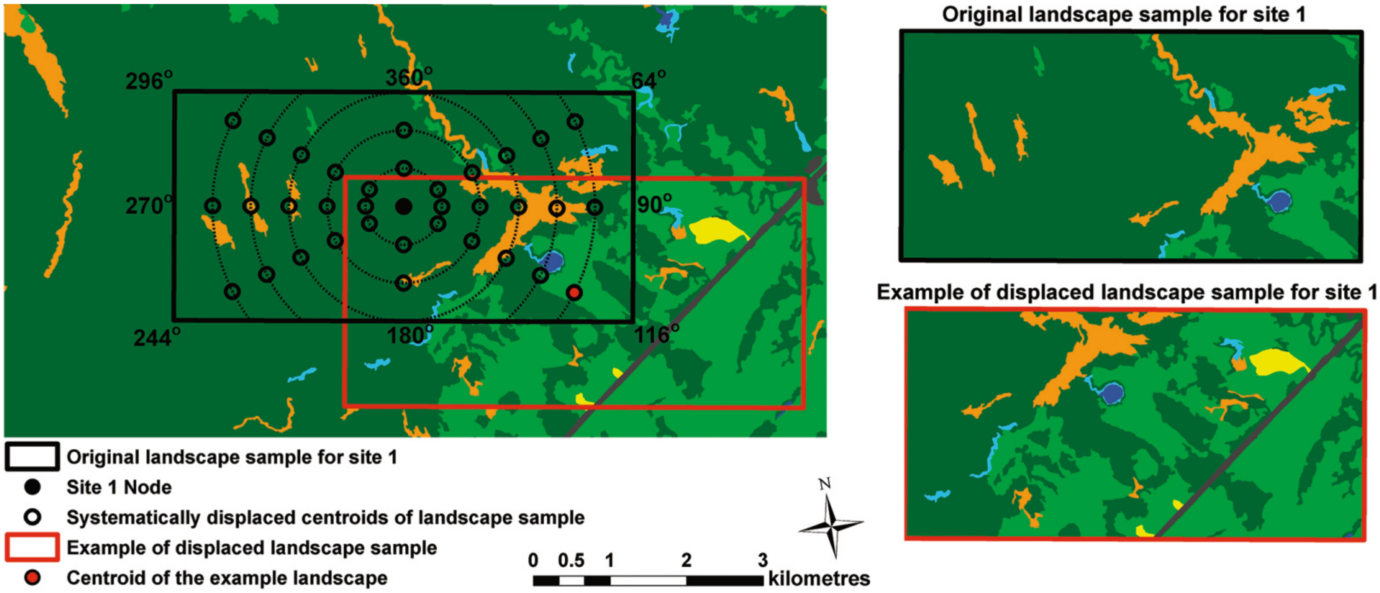
Fig. 1. Location of the thirty 3 × 6 km landscape samples within the ALPAC Forest Management Area, in the boreal forest region of Alberta, Canada.



within a given field site, each requiring a different sampling unit (e.g., focal patches of different habitat type) and consequently a different location within the site. If the project has to create its own landcover data, the LPIs could be extracted from a common bounding box that may, for financial reasons, happen to be smaller than what would be desirable from a modeling point of view. Under these conditions, spatial offsets between the response variable (species data) and predictor variables (LPIs) could arise. A similar situation would be created by a systematic sampling design in which landcover is only mapped at predefined areas centered at the nodes of a grid and where there are no adequate data outside of these areas. If for some reason (e.g., concerns of confidentiality or constraints on the biophysical characteristics of the plot) the ecological data were collected at variable distances away from the nodes and the LPIs needed to be extracted at the full extent of the mapped area, then spatial offsets could arise again. A real-world example of the above scenario is provided in the next section.

In each of the situations described above, the centroid of the landscape sample is offset a certain distance from the centroid of the focal patch, plot, or count station in which the ecological response variable was observed. Under these conditions, the question arises: to what degree does this offset influence the value of the LPIs? More specifically, how different would they be had they been extracted from landscape samples centered at the correct location? How do these differences change with distance and direction of the spatial offset? Would these differences affect the usefulness of the extracted LPIs as covariates? The goal of this note is to answer these questions and fill an existing gap in the landscape pattern analysis literature. To pursue this, we completed a case study in a 10 000 km^2 study area located in the Boreal Plains ecozone of northern Alberta, Canada, containing 30 rectangular landscape samples 3 × 6 km in size. We subjected these rectangles to 34 systematic displacements across five different distances (from 0.5 to 2.5 km) and eight different directions (along the main axes and diagonals of the rectangles) and analyzed the resulting differences in four representative LPIs. To the best of our knowledge, this issue has not been addressed yet in the literature, though it is related to others problems that are known to affect the LPIs, e.g., a change in the extent of the landscape samples. Even if our problem is avoidable in many instances, we are raising

Fig. 2. Spatially offset landscapes are generated through the systematic displacement of centroids of a landscape across eight directions (64, 90, 116, 180, 244, 270, 296, and 360 degrees) and five distances (0.5, 1, 1.5, 2, and 2.5 km). The example shows the original landscape at site 1 and one of the spatially offset landscapes.



awareness about pitfalls that can occur when researchers find themselves in one the aforementioned situations.

Materials and methods

We selected a set of thirty 3 × 6 km landscape samples across a 100 × 100 km study area located in the Alberta–Pacific Forest Industries Forest Management Agreement area in northeastern Alberta, Canada (Fig. 1). The study area is mainly populated by trembling aspen (*Populus tremuloides* Michx.), jack pine (*Pinus banksiana* Lamb.), and black spruce (*Picea mariana* Mill.). The landscape samples are part of the monitoring network of the Alberta Biodiversity Monitoring Institute (ABMI, www.abmi.ca), an organization that provides biodiversity monitoring for more than 2000 species and habitats across Alberta. The ABMI terrestrial network is based on a systematic sampling design consisting of a sample site at each node of a 20 km grid covering the entire province, which is coincident with the grid of the National Forest Inventory of Canada (NFI; Gillis et al. 2005). As part of this program, terrestrial biodiversity data on vascular and nonvascular plants, fungi, and soil arthropods are collected in 1 ha field plots located within 3 × 6 km rectangles centered at each grid node. In addition to these ground data, detailed information on habitat (e.g., vegetation floristic and structural attributes, modifiers about natural disturbances) and anthropogenic features (e.g., roads, pipelines, seismic cutlines, wellpads) within the rectangles is derived from softcopy interpretation of 1 : 30 000 scale color aerial photography. The motivation behind the rectangular shape and slightly tilted orientation of the 3 × 6 km landscape samples was to enable more-efficient image acquisition (only one flight line per site is required, which is aligned with the neighboring nodes to the west and east). Another characteristic of this program, which initially motivated the research questions posed in this work, is that at each site, the field plot can be offset up to 3 km away from the corresponding grid node. Using this strategy, the precise location of the field site can be kept confidential to avoid management bias and ensure landowner anonymity. However, this may lead to problems in situations in which the landscape context (expressed via LPIs) for a given ecological variable has to be extracted across the full extent of the rectangle, as its center is spatially offset from the location at which the response variable was observed.

To evaluate the effect of this spatial offset, we used the Alberta Vegetation Inventory (AVI; Alberta Environmental Protection 1991) to create a wall-to-wall landcover raster map of the study area, partially displayed in Fig. 1. The AVI is a forest inventory that, unlike the ABMI compilation, covers the entire area (however, it is only available for forests under a management agreement, which is why ABMI cannot rely on AVI for its provincial needs). In its native polygon-vector file format, the AVI contains complex labels that portray information on structural forest attributes such species composition, canopy height, crown closure, and stand age. To facilitate the extraction of LPIs, we converted the AVI to a raster layer with 10 m pixel size and a 1 ha minimum patch size. We translated the structural-attribute information to an 80-class landcover legend designed to preserve the original AVI information as much as possible (notwithstanding, we note that the actual number of classes is not as high as the legend may suggest: the average number of classes present in a landscape sample was less than 30, compared with 80 possible classes). Parallel to this, for each of the 30 sites, we created 34 spatially offset rectangles through the systematic displacement of the centroids of each original 3 × 6 km rectangle across eight directions (major and minor symmetry axes, plus the diagonals) and five distances (0.5, 1, 1.5, 2, and 2.5 km) (Fig. 2; note that the two largest distances were not computed along the minor axis, as the resulting new centroids would lie outside the original rectangle). The idea behind this scheme was to assume that there is a hypothetical ecological response variable observed exactly at the NFI

Fig. 3. Range of absolute percentage deviations (minimum, lower quartile, median, upper quartile, and maximum) between original and displaced landscapes for each of the 30 sites (n = 34/site) in (A) percentage area of human-modified forest (PAHF), (B) interspersion (IJI), (C) class richness (CR), and (D) patch density (PD). The horizontal broken line indicates the mean absolute percentage deviation across all sites.

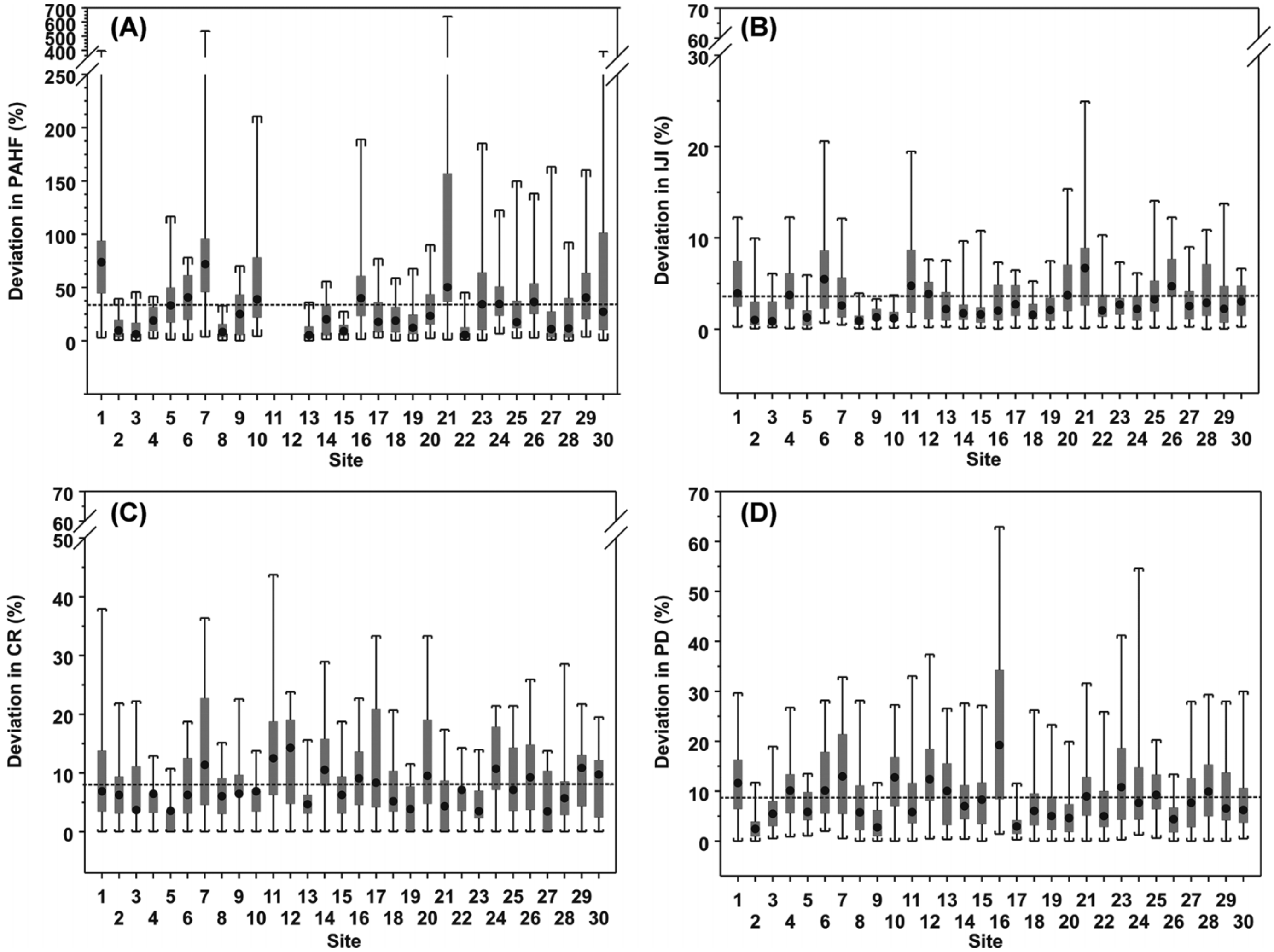


Table 1. Mean, coefficient of variation, and minimum and maximum values of percentage area of human-modified forest (PAHF), interspersion (IJI; %), class richness (CR; number of classes, out of 84), and patch density (PD; number of patches/km^2), each calculated across the 30 original landscape samples.

Summary statistics	PAHF	IJ	CR	PD
Mean	21.2	66.6	28.7	14.2
Coefficient of variation (%)	91.8	6.6	20.8	30.7
Minimum	0	55.6	16	6.2
Maximum	64.0	76.9	43	22.6

node. We then displaced the landscape samples around it and measured how the different spatial offsets affected the value of the LPIs relative to the correct (centered at the node) location.

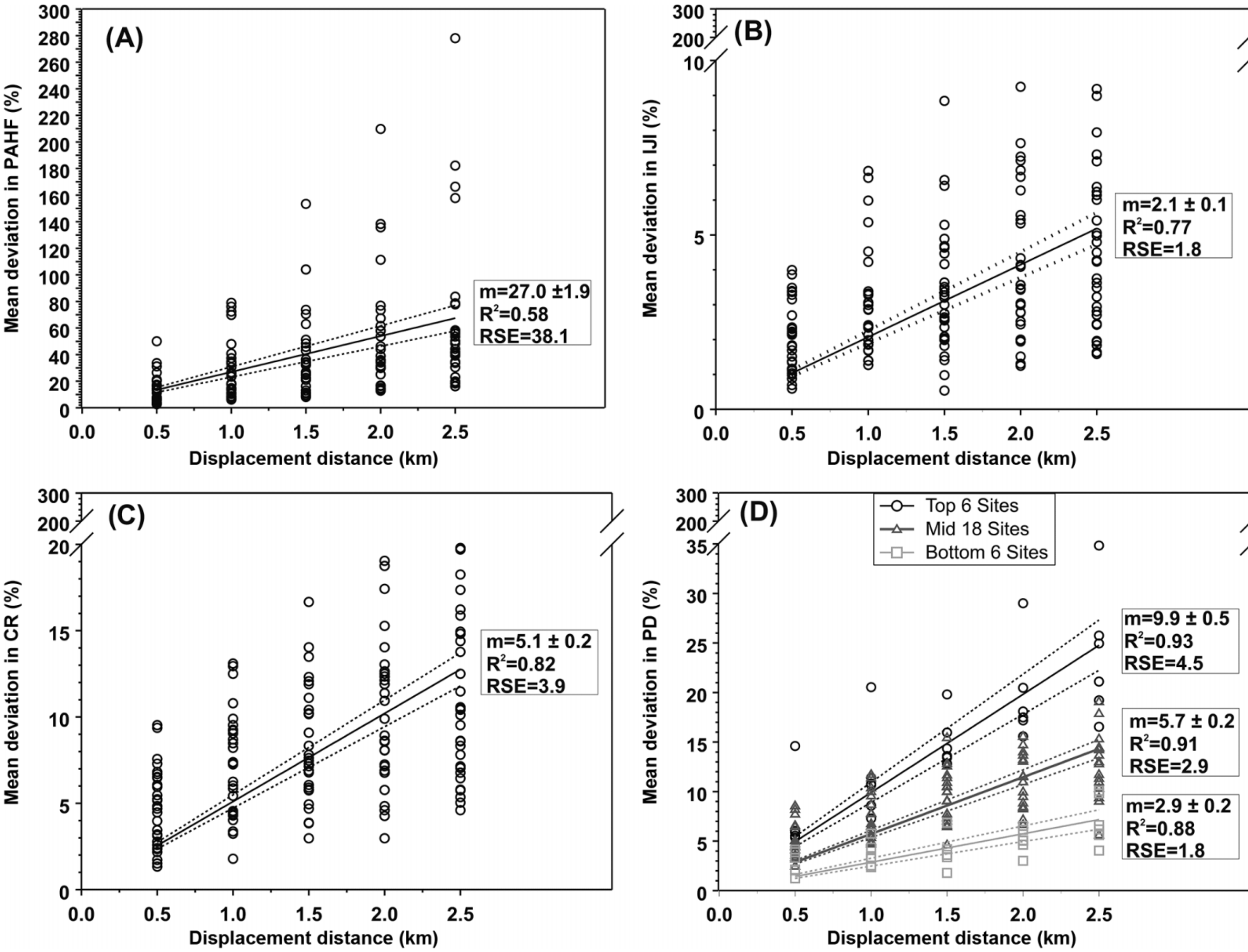
To assemble an empirical data set upon which to evaluate the effect of landscape-sample displacement on selected LPIs, we generated 1050 landscape samples (35 per site) by clipping the AVI-derived landcover map by both the original and displaced rectangles. Each landscape sample was subsequently processed using FRAGSTATS 3.3 build 5 (McGarigal et al. 2002). The following set of LPIs, which are representative of the typical indices that may be used as covariates in an ecological model, were computed: (*i*) percentage area of human modified forest (PAHF); (*ii*) landscape interspersion (IJI), a measure of the spatial intermixing of different landcover patches; (*iii*) class richness (CR), or total number of different landcover classes existing within the landscape sample; and (*iv*) patch density (PD), expressed as the total number of patches per square kilometre.

For each LPI and displacement, we calculated the absolute percentage deviation (APD_{LPI}), i.e., the percentage difference between the value of a given LPI at the displaced location (χ_d) and the one at the original location (χ_o), as

$$[1] \qquad APD_{LPI} = \frac{|(\chi_d - \chi_o)|}{\chi_o} \times 100$$

For each LPI, we graphically summarized the APDs at each site using box-and-whisker plots (n = 34 displacements/site). For the analysis of the relation between displacement

Fig. 4. Linear regressions of mean absolute percentage deviation across 30 sites between original and displaced landscapes as a function of displacement distance in (A) percentage area of human forest (PAHF), (B) interspersion (IJI), (C) class richness (CR), and (D) patch density (PD). Patch density was fitted using quantile regression separating the top 20%, the middle 60%, and the bottom 20% of the 30 sites according to their overall deviation ranks.

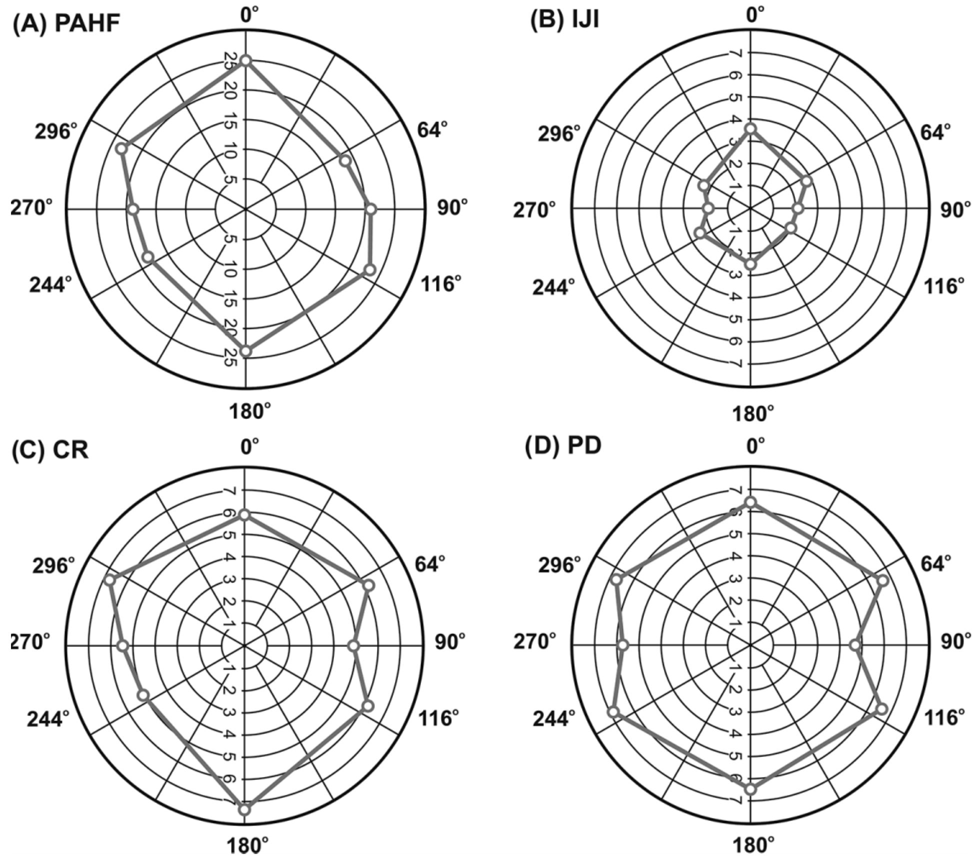


distance and LPI deviation, we computed the mean APD_{LPI} across all directions at each site for each of the five distances. We then fit a linear regression of the mean APD_{LPI} as a function of displacement distance (n = 30 per displacement distance and LPI), and computed the mean regression coefficient (m) with standard error, the coefficient of determination (R^2), and the residual standard error (RSE). To further investigate the variability of the displacement-induced LPI deviations, we applied quantile regression to one of the LPIs: patch density (PD). Quantile regression is a statistical method that can be applied when the full set of observations shows only a weak relationship between the response and predictor variables, whereas this relationship is improved when an ordered (by quantiles) subset is used instead (Cade and Noon 2003). For this, we ranked the 30 sites by descending value of mean APD_{PD}, assigned them to one of three subsets (top 20%, middle 60%, and bottom 20%; corresponding, respectively, to the six most affected sites, the 18 moderately affected sites, and the six least affected sites), and computed a separate linear regression for each subset. For the analysis of the relation between displacement direction and LPI deviation, we computed the mean APD_{LPI} for each of the eight displacement directions across the 30 sites ($n = 2 \times 30$ per displacement direction; note that only the first two displacement distances were considered, as the rectangular shape of the sample landscapes did not allow for including all five displacement distances with equal representation). Finally, we summarized the results by plotting the mean deviations using polar graphics.

Results and discussion

Most spatial displacements led to a change in the values of the LPIs. The LPI with the highest mean APD (broken horizontal lines in Fig. 3) was the PAHF (mean APD 38.5%), followed by PD (9.1%) and CR (8.1%), with IJI (3.3%) being the least-affected LPI. PAHF was at least four times more sensitive to the displacements than the other LPIs. This is not surprising, as unlike the other three LPIs, PAHF is a class-level metric computed on a mosaic consisting of only two classes (human-modified forest and all others). This mo-

Fig. 5. Mean absolute percentage deviation across 30 sites between original and displaced landscapes as a function of displacement direction in (A) percentage area of human forest (PAHF), (B) interspersion (IJI), (C) class richness (CR), and (D) patch density (PD). Note that angles are measured clockwise from the rectangle minor axis.



saic, when nonempty (there were two sites that lacked this class of forest), necessarily has a more uneven configuration than the mosaic including all classes and therefore tended to change more with a small displacement. Another interesting observation is that the range of absolute deviations differed considerably among sites (Fig. 3): the mean deviation in the 10% most affected sites was at least three times more than that of the 10% least affected sites for all LPIs. There was also a large variability within each site, even larger than the overall variations of LPI values across the 30 original landscape samples (coefficients of variation; Table 1).

On average, the deviations increased with the length of the displacement for all LPIs, i.e., for any given site and LPI, the larger the displacement, the greater the difference between the values obtained from the original and displaced landscape samples (Fig. 4). Again, the largest average increase corresponded to PAHF (27% mean absolute deviation per kilometre of displacement) and the least to IJI (2.1%) (Fig. 4). The range of deviations also widened with distance, which translates to large residuals. As a result, the regression fits were only moderate (R^2 ranging between 0.58 and 0.82), preventing any reliable prediction of the actual effect of a given displacement distance. Notwithstanding, the fit may be improved by using quantile regression (Cade and Noon 2003), which we applied to PD as an example. When separating the sites into three groups (top 20%, middle 60%, and bottom 20% quantiles of mean APD in PD), the three linear regression fits for PD were tighter compared with the overall fit for the 30 sites, yielding $R^2 > 0.88$ (Fig. 4D) and lower residuals. Overall, the slope of the regression lines appears to be related to the heterogeneity in patch size of the landscape mosaic of the area around each site encompassing all the displaced rectangles. That is, we found that sites with patches of uniform size were less affected by the displacement than sites with greater size variability. However, although this observation permits the identification of sites prone to stronger effects, it does not allow predicting the sign of the deviation and thus cannot be used to correct for the effect of the spatial offset.

The direction of the displacement does not seem to play a relevant role in the magnitude of the deviations, as evidenced by the low coefficient of variation (less than 20% for all four LPIs) of the set of eight directional values of mean APD. Notwithstanding, a slight trend is noticeable on the polar graphics (Fig. 5), wherein the north–south direction consistently yielded slightly larger deviations than any other direction. However, this is most likely an artifact of the rectangular shape of the landscape samples: the width of the rectangles is double their height; therefore a displacement along the north–south direction will yield twice as much nonoverlapping area between the original and displaced rectangles than the same displacement along the east–west direction. The larger deviations along the north–south direction are thus consistent with the expectation that differences between two overlapping landscape samples would increase as their area of overlap decreases.

These results fill a gap in the literature on landscape pattern analysis. The dependence of LPIs on several characteristics of the landcover maps from which they are derived has been well established with regard to grain and extent (e.g., Wu 2004), minimum mapping unit (e.g., Saura 2002), thematic resolution (e.g., Castilla et al. 2009), thematic accuracy (e.g., Langford et al. 2006), and spatial inconsistencies (Linke et al. 2009). However, to date, the effects on LPIs of slightly changing the location of the landscape sample have not yet been reported. This displacement problem is related to the well-known change-of-extent problem (e.g., Wu et al. 2002): in both, there is always an area of overlap between the original and changed (i.e., either displaced or expanded or shrunk) landscape samples and an area that is exclusive to one or both of them. In the first case, the mismatch is created by shifting the position of the original landscape, in the second, by enlarging or shrinking its extent. However, displacement is a distinct problem because the extent is kept constant, and therefore, it deserves the specific examination undertaken in this study.

Conclusions

Extracting LPIs from landscape samples spatially offset from the location of the ecological observations leads to considerably different values than those that would have been obtained from landscapes centered at those locations. In our case study in the boreal forest of Alberta, Canada, displacing the 18 km^2 landscape samples 1 km led, on average, to a 15% deviation with respect to the LPI values obtained from the original landscapes. Although the direction of the spatial offset had no predictable impact, the distance had a direct relationship with the magnitude of the deviations. Unfortunately, as the spatial offset increased, the range of resulting deviations also widened, making it difficult to predict or correct for this effect. This has an important implication if those LPIs are intended to be used as covariates of an ecological response variable: they will add noise to the model and thus obscure the relationship between landscape context and the response variable, as their value will be different than what would have been obtained from landscape samples centered at the location of the ecological observations. The specific impact on the model output will depend on the effect size of each particular covariate and on the heterogeneity around the landscape samples, which are case-specific.

In conclusion, caution should be used to avoid LPIs extracted from spatially offset landscape samples. More concretely, if the extent of the available landcover data does not allow researchers to extract the LPIs from landscape samples centered at the locations at which the response variable was observed, then it seems preferable to reduce the size of the landscape samples. The degree to which this recommendation holds depends on the relative strength of the effects of displacing landscape samples versus reducing their extent, a subject that we intend to address in future research.

Acknowledgements

This research was supported by the Alberta Biodiversity Monitoring Institute (ABMI), the University of Calgary, and the Natural Sciences and Engineering Research Council of Canada (NSERC) through project CRDPJ 401966-10. We also thank our colleague Jennifer Hird, the Associate Editor assigned by this Journal to this manuscript, and two anonymous reviewers for their helpful comments.

References

Alberta Environmental Protection. 1991. Alberta vegetation inventory standards manual. Version 2.1. Alberta Environmental Protection, Resource Data Division, Data Acquisition Branch, Edmonton, Alberta.

Cade, B.S., and Noon, B.R. 2003. A gentle introduction to quantile regression for ecologists. Front. Ecol. Environ. **1**(8): 412–420. doi:10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2.

Castilla, G., Larkin, K., Linke, J., and Hay, G.J. 2009. The impact of thematic resolution on the patch mosaic model of natural landscapes. Landsc. Ecol. **24**(1): 15–23. doi:10.1007/s10980-008-9310-z.

Forester, J.D., Anderson, D.P., and Turner, M.G. 2008. Landscape and local factors affecting northern white cedar (*Thuja occidentalis*) recruitment in the Chequamegon–Nicolet National Forest, Wisconsin (USA). Am. Midl. Nat. **160**(2): 438–453. doi:10.1674/0003-0031(2008)160[438:LALFAN]2.0.CO;2.

Gillis, M.D., Omule, A.Y., and Brierley, T. 2005. Monitoring Canada's forests: the National Forest Inventory. For. Chron. **81**: 214–221.

Langford, W.T., Gergel, S.E., Dietterich, T.G., and Cohen, W. 2006. Map misclassification can cause large errors in landscape pattern indices: examples from habitat fragmentation. Ecosystems (N.Y.), **9**(3): 474–488. doi:10.1007/s10021-005-0119-1.

Linke, J., McDermid, G.J., Pape, A.D., McLane, A.J., Laskin, D.N., Hall-Beyer, M., and Franklin, S.E. 2009. The influence of patch delineation mismatches on multi-temporal landscape pattern analysis. Landsc. Ecol. **24**(2): 157–170. doi:10.1007/s10980-008-9290-z.

McGarigal, K., Cushman, S.E., Neel, M.C., and Ene, E. 2002. FRAGSTATS: spatial pattern analysis program for categorical maps. Computer software program produced by the authors at the University of Massachusetts [online]. Available from www.umass.edu/landeco/research/fragstats/fragstats.html [accessed 5 May 2011].

Saura, S. 2002. Effects of minimum mapping unit on land cover data spatial configuration and composition. Int. J. Remote Sens. **23**(22): 4853–4880. doi:10.1080/01431160110114493.

Turner, M.G. 2005. Landscape ecology: what is the state of the science? Annu. Rev. Ecol. Evol. Syst. **36**(1): 319–344. doi:10.1146/annurev.ecolsys.36.102003.152614.

Wu, J. 2004. Effects of changing scale on landscape pattern analysis: scaling relations. Landsc. Ecol. **19**(2): 125–138. doi:10.1023/B:LAND.0000021711.40074.ae.

Wu, J., Shen, W., Sun, W., and Tueller, P.T. 2002. Empirical patterns of the effects of changing scale on landscape metrics. Landsc. Ecol. **17**(8): 761–782. doi:10.1023/A:1022995922992.