CHAPTER 42

Assessment of the Spatial Structure and Properties of Existing Ecoregionalization Systems of Ontario

D.J.B. Baldwin, L.E. Band, and A.H. Perera

INTRODUCTION

Ecoregionalization systems are widely recognized as critical tools in effective resource management (Hills, 1959; Burger, 1993; Bailey, 1988). Several existing systems contain many intuitive delineations that many experts would not dispute represent ecologically significant units. The problem lies in quantifying this significance. Many of these systems were developed prior to extensive, accurate spatial databases and include many subjective construction properties. Unfortunately, these construction methods are often not well documented and the specific properties of the units within a system are not known. To meet the growing need to include ecologically-based methods in resource management, reliance on these systems increases. Effective use of existing systems can only come from knowing what they mean quantitatively, including their strengths and weaknesses for stratifying the phenomena relevant to questions we want to answer.

This chapter summarizes the methodological aspects of work completed for Master's research to help provide this knowledge (Baldwin, 1997)..The goal of the thesis was to quantify three existing systems in Ontario using spatial surfaces of geoclimatic variables relevant to ecosystem function and composition. The main objective of this quantification was to compare the strengths and weaknesses of each system in delineating these variables, illustrating which phenomena are delineated most effectively by each system, at each location. In this chapter, we present the techniques developed to quantify and compare ecoregionalization systems, rather than the detailed results relating to the three Ontario systems. The results obtained for one of the systems (Hills, 1959) are presented to illustrate these methods. We also outline conclusions relevant to the development of new ecoregionalization systems.

METHODOLOGY

Spatial Data Set Development

Variables relevant to basic ecosystem composition and function were determined from the literature. Table 42.1 outlines the series of geoclimatic variables included in the analyses.

Climate Variables

The climate variables were extracted from the Ontario Climate Model (OCM) outlined in detail by Graham (1995). Each surface was registered and maintained as a 1-km resolution, quadtree compressed raster using SPANS GIS software (TYDAC Technologies, 1994). These rasters were resampled to 10 km resolution for the core analyses outlined later.

Terrain Variables

The elevation surface was obtained from Mackey et al., 1994 and registered at 1 km resolution using SPANS. A surface was derived to quantify terrain complexity relationships. This was accomplished using a boundary modeling function in SPANS (TYDAC Technologies, 1994). The computational algorithm returns the number of cell boundaries that have differ-

Table 42.1. Summary of Variable Surfaces Included in the Analyses.

Variable	Source	Abbreviated Name
Mean annual min. temperature	OCM[a]	annmin
Mean annual max. temperature	OCM[a]	annmax
Mean monthly min. temperature-Jan.	OCM[a]	minjan
Mean monthly max. temperature-Jul.	OCM[a]	maxjuly
Mean monthly temperature - Jan.	OCM[a]	jantemp
Mean monthly temperature - Jul.	OCM[a]	julytemp
Mean annual temperature	OCM[a]	anntemp
Growing season length	OCM[a]	growlen
Annual growing degree days > 5°C	OCM[a]	anndd5
Mean annual precipitation	OCM[a]	annpm
Mean monthly precipitation - Jan.	OCM[a]	janpm
Mean monthly precipitation - Jul.	OCM[a]	julypm
Elevation	Mackey et al., 1994	km1dem
Elevation complexity	Baldwin, 1997	dembound
NDVI	Band, 1994	psnraw
Geology mean patch size	Baldwin, 1997[b]	ge_mps
Geology edge density	Baldwin, 1997[b]	ge_ed
Geology diversity	Baldwin, 1997[b]	ge_sim
Geology contagion	Baldwin, 1997[b]	ge_con
Land-cover mean patch size	Baldwin, 1997[c]	vg_mps
Land-cover edge density	Baldwin, 1997[c]	vg_ed
Land-cover diversity	Baldwin, 1997[c]	vg_sim
Land-cover contagion	Baldwin, 1997[c]	vg_con

[a] Derived from the Ontario Climate model maintained by the Genetics Program, Ontario Forest Research Institute, Ministry of Natural Resources, Sault Ste. Marie, Ontario.

[b] Derived using classified LANDSAT TM developed by Spectranalysis (1992).

[c] Derived using surficial geology database developed by Perera et al. (1996).

ent adjacent values within an N × M neighborhood surrounding each cell. Only nonzero values in the neighborhood are used. For example, the following 3 × 3 neighborhood returns a boundary function value of 6 (bars and dashes indicate interfaces with different adjacent nonzero values).

```
1 | 3 | 2
–
5 | 3 | 2
–
2   0   2
```

The raw elevation surface was preclassified into 25 m intervals to eliminate excessive noise. The boundary function was applied to this elevation surface, using a 5 × 5 neighborhood to delineate complex vs. gentle terrain. Three-dimensional representations of the elevation model were inspected to confirm correlation of this surface to complex terrain areas.

Land Cover

Surfaces of normalized difference vegetation index (NDVI) derived from Advanced Very High Resolution Radiometer (AVHRR) data were obtained from Band (1994). An annual aggregated surface was used in this study to represent a broad measure of land cover conditions. This surface was obtained at a resolution of 1 km. This layer was imported and registered in the SPANS project study area.

Land Cover and Geology Spatial Structure Variables

The nominal classifications of the surficial geology and LANDSAT land cover layers were not appropriate for the statistical tests used to analyze the ecoregionalization systems. A series of spatial indices was developed based on these layers to provide variables defining the structure of these attributes. A program was written in TCL to combine the spatial data manipulation functions of SPANS with Fragstats. The latter is an application developed to calculate a series of spatial metrics for landscape ecological applications (McGarigal and Marks, 1993).

This program was used to calculate localized measures of the spatial characteristics of the landcover/geology for 10 km × 10 km "landscapes" centered on each of the points used to sample the other variables (defined later). The land cover and surficial geology data were examined at 125 m resolution within each landscape. Four of the many spatial metrics derived by Fragstats were chosen for this study to quantify the land cover and geology layers: mean patch size, edge density, Simpson's diversity index, and a contagion index. These indices are detailed by McGarigal and Marks (1993).

Multivariate Data Set Preparation

The values of each variable and the region identifiers were appended to points generated at 1 km intervals for the boundary gradient analyses and 10 km intervals for the region core comparisons outlined later. The variable surfaces were resampled to 10 km resolution for the region core comparisons. The points were exported to create a multivariate data set for analysis in the SAS statistical package (SAS, 1993). All of the variables were standardized to have a mean of 0 and a standard deviation of 1 to eliminate discrepancies in measurement scales.

Compositional statistics

Standard univariate measures were generated to characterize each surface and each regional unit. These values were generated as a reference and a quantitative basis for users of the ecoregionalization systems for decision-making and analysis

Region Core Analyses

Core Area Definition

The strength of the geoclimatic variable partitioning was examined. Inherent in the construction of each of these systems is a limited locational precision of the regional boundaries. Comparisons among the regions were confined to "core" areas. These core areas were modeled by imposing a buffer inward from the boundaries. A buffer distance of 40 km was selected to define robust region cores with viable region sizes for sampling.

The region lines were extended past the provincial extent to eliminate the effect of the arbitrary provincial limit. Ecologically, this administrative delineation has no true meaning except in cases where the boundary is defined by significant natural features such as the Great Lakes or Hudson Bay.

Variance Component Analysis

The within-unit to between-unit variance was examined. An effective classification should display greater between-unit variance than within-unit variance. A breakdown of the variance of each variable across the regional units of each ecoregionalization system into these two components was completed using the SAS Varcomp procedure.

Multivariate Distance

The internal structure of the system was examined by calculating the pairwise squared distance (Mahalanobis distance) between each of the regions:

$$D^2(i|j) = \left(\overline{X}_i - \overline{X}_j\right)' \ \mathrm{cov}^{-1}\left(\overline{X}_i - \overline{X}_j\right)$$

The distances between each region and all others were calculated and sorted to examine the multivariate similarity between the regions. This provides a method to examine the relationships between multivariate similarity and spatial proximity and illustrates the relative strength of the boundaries between units. Choropleth maps were generated for the total of all multivariate distances for each region as well as maps showing the pairwise values between an individual region and all of the other regions.

Surficial Geology Similarity

The regional core units were compared in terms of their surficial geology compositional similarity using the Morisita-Horn index of similarity (Turner and Gardner, 1991). This index was calculated as:

$$MH = \frac{2\sum p_{ij}p_{ik}}{\sum p_{ij}^2 + \sum p_{ik}^2}$$

where MH = Simplified Morisita-Horn index of overlap between region j and k
p_{ij}, p_{ik} = Proportion geology type i is of the total area in both regions ($i=1,2,3,\ldots,n$)
n = Total number of geology types

Choropleth maps were generated similar to the multivariate distance comparisons outlined earlier.

Principal Components Analysis (PCA)

PCA was performed to examine the internal structure among regions. PCA was conducted for each regional unit, using all variables. The eigenvector loadings of each variable for the first few principal components were sorted for each regional unit to identify the key variables related to the region's structure. Maps were created showing a bar chart in each region with the eight largest positive loadings color-coded by the relevant variable. These maps were generated for the first three principal components which together typically explained more than 70% of the variance in each region

Region Boundary Analyses

Database Preparation

The boundaries between regional units were examined to identify the trends and key variables associated with each interface. A gradient was generated across each regional interface by creating 10-km wide bands, 50 km in each direction from the boundary. Points were generated at 1-km intervals; the region identifiers, buffer distance, and geoclimatic variable values were attached to the points. The spatial metric variables for vegetation and geology structure generated earlier for the core analyses were not included in the data set. Some interfaces were not examined, particularly northern units where climate interpolation was weak due to low climate station density.

Means, Multivariate Distance, and Geologic Similarity Comparisons

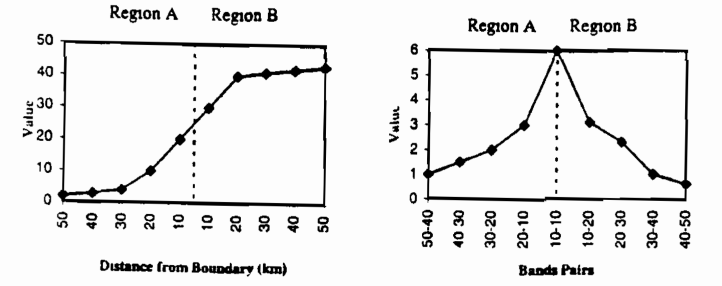
The means for each of variables were calculated for each band across each regional interface. The multivariate distance (Mahalanobis distance) and the geologic similarity (Morisita-Horn index) were calculated between adjacent band pairs. These values were plotted against an x axis of the 10-km bands, centered on the regional boundary. This allowed examination of trends across the region interface and the sharpness of the gradient between regions.

Trend Comparison Statistics

A series of indices was developed to examine the slopes at different points over the gradient to quantify the trends across the region interface. The gradient was broken into three sections: one "core" area (30 km on each side of the boundary) and two "tail" areas (remaining 20 km in each region). The slopes of the lines were calculated for each variable between successive bands. Each section's slopes were compared as proportions of the entire trend, in order to standardize the measure for the many different scales of the various variables.

This relationship was calculated and interpreted differently depending on the type of analysis. For example, comparing means, a strong boundary trend is illustrated by a gradual slope in the first region tail, followed by a sharp increase or decrease across the boundary and then more gentle slopes away from the boundary into the second region. In contrast, strong multivariate distance trends are shown by a rapid increase in dissimilarity approaching the boundary from each region. Examples of strong trends for different comparison types are shown in Figure 42.1. Graph (a) shows a strong trend for the mean comparison (index value is 0.9024 calculated as the sum of slopes in the core area as a proportion of all slopes). Graph (b) illustrates a strong multivariate distance comparison trend (index = 0.7404 calculated as the sum of the absolute values of slopes in the core area as a proportion of all slopes). A strong trend for geologic similarity would be a vertical mirror image of Graph (b).

The shapes of all curves were examined in conjunction with the index values, to create a ranking scheme for the variables within each region. A ranking of strong, medium, weak, or no trend (scored 3, 2, 1, and 0, respectively), was determined for each variable, for each interface using threshold values of the index in combination with line shape. These scores were tabulated to provide a summary of the trends across each interface, for each ecoregionalization system. Total scores, standardized by the number of interfaces in each ecoregionalization system, were calculated to compare ecoregionalization systems. Summary maps were prepared, showing each bound-



(a) Mean comparison

(b) Multivariate distance comparison

Figure 42.1. Examples of strong trend patterns for means comparison and multivariate distance comparison methods.

ary in a line width proportional to the strength of the boundary, based on the scores. The top five variables for each interface were graphed as bars showing their index values to illustrate which variables display the strongest trends for each interface.

RESULTS AND DISCUSSION

Region Core Analyses

The variance component analysis showed lower within-unit variance than between-unit variance for almost all variables except the spatial structure variables for geology and vegetation. This suggests that all of the system partition the geoclimatic variables well. This test, however, may be overstating this result. In Ontario, the variance of many of the climate variables, is linear and almost unidirectional, either latitudinal (temperature) or longitudinal (precipitation) This makes drawing 10 or 20 regions that partition most of the variance of this multivariate data set, a fairly rudimentary exercise. This does not diminish the merits of the systems studied: the authors of these systems could do this better than anyone, with knowledge of the subtleties of the ecological composition of the landscape well beyond the data. This does, however, limit the usefulness of this type of analysis for quantifying ecoregionalizations in Ontario or other locations where the variance patterns are similar. The differentiation of regions is better illustrated by the multivariate distance and geologic similarity analyses.

Figure 42.2 illustrates the maps used to examine the trends in multivariate distances among regions. Map (a) shows the total of all pairwise multivariate distances between a region and all others. Maps (b) and (c) show the pairwise multivariate distances between a focus region and the other regions. The southernmost and northernmost regions are the most strongly defined: the southern units were more distinct among themselves than the northern ones. This is logical because they are separated by more isotherms and have no counterparts at the similar latitudes. The strength of the differentiation of southern units could also be related to stronger climate interpolations resulting from higher climate station frequency.

Trends in geologic similarity among regions were illustrated using maps like those in Figure 42.3. These proved very useful in illustrating that this geology data set is likely most useful for defining units at a spatial level above the region, such as ecozones or provinces. The regions approximating the Hudson Bay Lowland showed very low similarity values to other regions except among themselves. In addition, many of the central regions are not well differentiated among themselves. This suggests that the regions, as they are defined, may form complexes of geologic units which are broken down subsequently by other variables.

The regional PCAs were performed primarily as a reference for detailed study of individual regional units and are not presented here. In combination with the eigenvalues for each region, they proved very useful for examining the importance of each variable to each region.

Region Boundary Analyses

Figure 42.4 provides a summary of the boundary analyses results. A standardized score for all interfaces was used to rank the three systems in terms of their overall boundary strength: however, the boundary analysis results are most useful for examining specific interfaces. This information is critical to place the strength of the boundaries in a specific context. For example, if a system is needed to stratify a sampling program for a temperature-based question, information about boundary strength is needed in the context of temperature gradients.

One interesting trend these results show is that regions may stratify one type of variable well but their

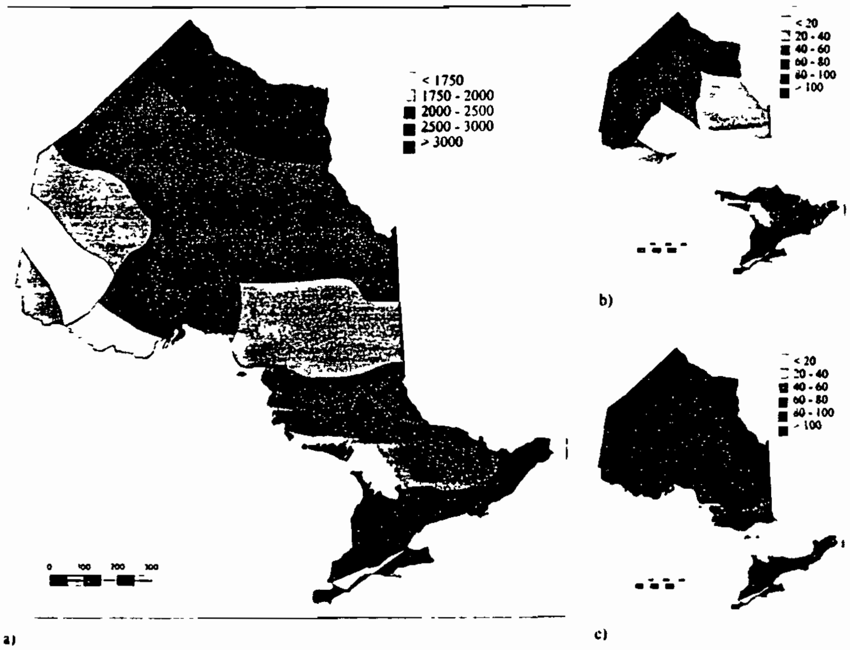


Figure 42.2. Multivariate distance values among regions. Map (a) shows the sum of pairwise multivariate distances for each region. Maps (b) and (c) show the pairwise values between a focus region (in white) and all other regions.

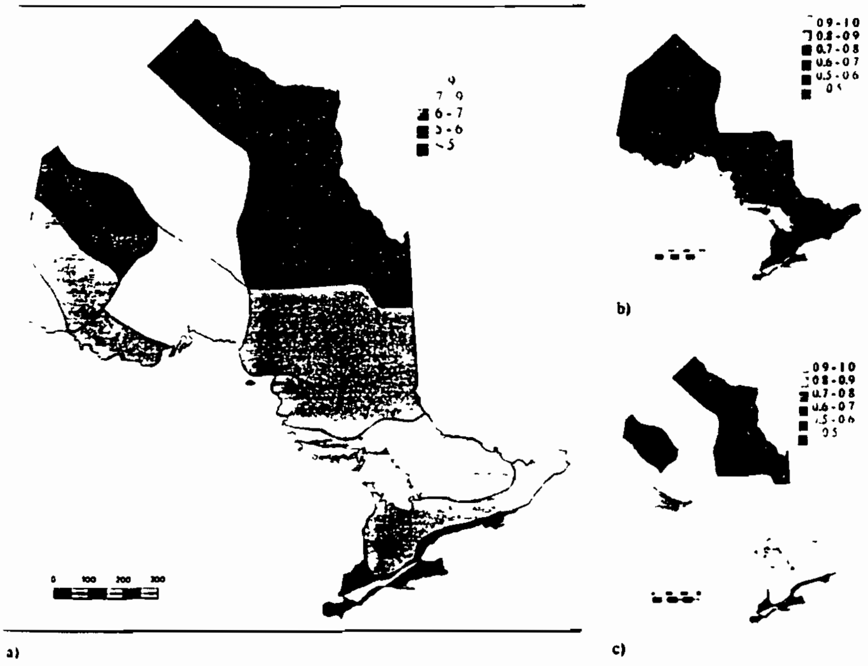


Figure 42.3. Geologic similarity values among regions. Map (a) shows the sum of Morisita-Horn similarity for each region. Maps (b) and (c) show the pairwise values between a focus region (in white) and all other regions.

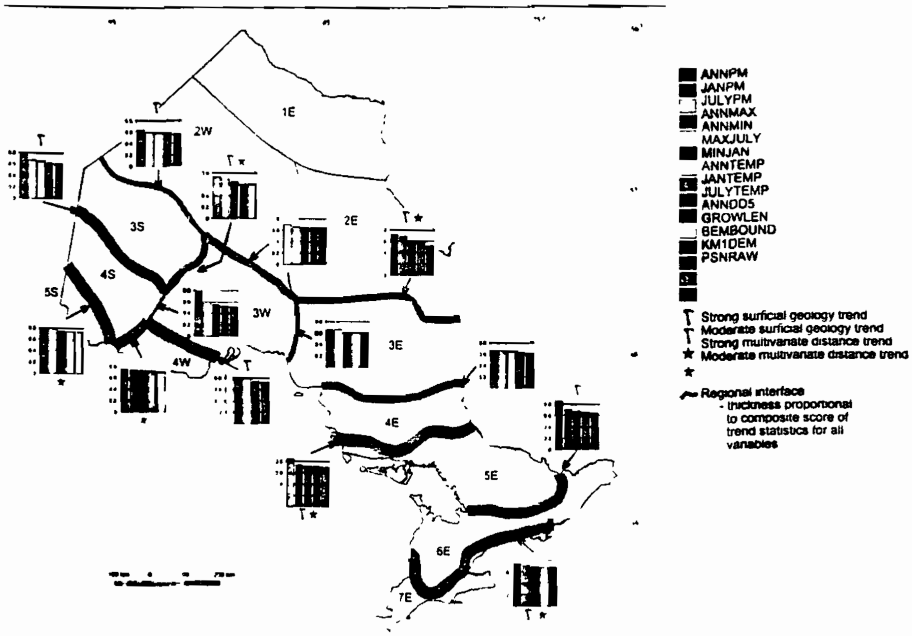


Figure 42.4. Regional interface characteristics. Region lines are drawn proportional to their total ranking scores. Bars indicate value for trend statistic of the top five variables for each interface.

exact boundary may be defined by another type. For example, we found in many cases that regions were most related to temperature variables (from the PCA results) but the exact boundaries were defined by a strong geology trend. This is consistent with the construction of many of these systems that defined regions based on landform subdivisions of climatically homogeneous areas.

CONCLUSIONS

This study has generated a number of tools to examine and compare differences in ecoregionalization systems at specific region and interface levels. In Ontario, and likely in other areas where climate variation patterns are similar, simple examinations of within-unit to between-unit variance is not very useful. On the other hand, mapping the multivariate distance and geologic similarity between regions provides an excellent visualization tool to examine the system's internal structure. The PCA results provide a useful reference to study the variables most relevant to specific regions. The methods developed in this study to quantify boundary strength, particularly in the context of variable type (i.e., climate, geology, terrain) provide an excellent quantitative basis for using existing systems more intelligently.

The selection of a particular ecoregionalization systems or specific regions within an ecoregionalization systems must be made considering both the variables that define the region(s) as a whole as well as those that define the specific boundaries of the region(s). Choosing among many regions, however, amounts to a subjective combination of delineations to define a system for the specific phenomena of study. This results in the same limitation of using many of these existing systems—a subjective process that may not be reproducible.

This concept must be considered when new systems are developed. An effective ecoregionalization system should allow the user to interactively decide which variables are relevant to the phenomena, weight these variables according to their relative importance to this phenomena, and then generate a system by a reproducible, explicit methodology. Subjective improvements could be incorporated later, much like satellite classification. These changes, however, would be relevant to a specific analysis and would not limit stratification in subsequent analyses or for different phenomena. This rationale is being adopted in the development of the Hierarchical Ecoregional Framework (Perera et al., 1995).

REFERENCES

Bailey, R.G. *Ecogeographic Analysis: A Guide to the Ecological Division of Land for Resource Management*, USDA Forest Service Miscellaneous Publication 1465, 1988.

Baldwin, D.J.B. *Quantification of Existing Eco-Regionalizations of Ontario*, thesis presented to the University of Toronto in partial fulfillment of the requirements for the degree of Master of Science, 1997.

Band, L.E. *Development of a Landscape Ecological Model for Management on Ontario Forests*, Rep. No. 7, Forest Fragmentation and Biodiversity Project, Ont. Min. Nat. Resour., Ont. For. Res. Inst., Sault Ste. Marie, Ont., 1993.

Band, L.E. *A Pilot Landscape Ecological Model for Forests in Central Ontario*, Rep. No 17, Forest Fragmentation and Biodiversity Project, Ont Min. Nat. Resour., Ont. For. Res. Inst., Sault Ste. Marie, Ont., 1994.

Burger, D. *Revised Site Regions of Ontario: Concepts, Methodology and Utility*, For. Res. Rep. No. 129, Ont. Min. Nat. Resour., Ont For. Res. Inst., Sault Ste. Marie, Ont., 1993.

Graham, B.J. *Modelling Distributions of Rare Carolinian Tree Species in Southwestern Ontario*, thesis presented to the University of Western Ontario, London, in partial fulfillment of the requirements for the degree of Master of Science, 1995.

Hills, G.A. *A Ready Reference to the Description of the Land of Ontario and Its Productivity*, Prelim. Res. Report, Ont. Dep. of Lands and Forests, Maple, Ont., 1959.

Hills, G.A. *The Ecological Basis for Land-Use Planning*, Research Report No. 46, Ont. Dep. of Lands and Forests, Maple, Ont., 1961.

Mackey, B.G., D.W. McKenney, C.A. Widdifield, R.A. Sims, K. Lawrence, and N. Szcyrek. *A New Digital Elevation Model of Ontario*, NODA/NFP Tech. Rep. TR-6, Dep. of Nat. Resour. Can., CFS—Ont., Sault Ste. Marie, Ont., 1994.

McGarigal, K. and B.J. Marks. *Fragstats: Spatial Pattern Analysis Program for Quantifying Landscape Structure*, Oregon State University, Corvallis, 1993.

Perera, A.H., J.A. Baker, L.E. Band, and D.J.B. Baldwin. A Strategic Framework to Eco-Regionalize Ontario, *Environ. Assess. Monitor.*, 39, pp 85–96, 1995.

Perera, A.H., J. Laguna, D.J.B. Baldwin, R. Bae, H. Godschalk, and M. Ouellette. *Surficial Geology of Ontario: A Digital Database*, Rep. No. 22, Forest Fragmentation and Biodiversity Project, Ont. Min. Nat. Resour., Ont. For Res. Inst., Sault Ste. Marie, Ont., 1996.

SAS Institute. *Procedures Manual for the SAS System*, 1993.

Spectranalysis Inc., *Development of a Digital Database of Red and White Old-growth Pine Forests in Ontario—East*, Rep. No. 22, Forest Fragmentation and Biodiversity Project, Ont. Min. Nat. Resour., Ont. For. Res. Inst., Sault Ste. Marie, Ont., 1992.

Turner, M.G. and R.H. Gardner. *Quantitative Methods in Landscape Ecology*, Springer, New York, 1991.

TYDAC Technologies. *SPANS Reference 5.3 Reference Manual*, Ottawa, 1994

Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources

Edited by

Kim Lowell
Annick Jaton

Ann Arbor Press
Chelsea, Michigan

Library of Congress Cataloging-in-Publication Data

Spatial accuracy assessment : land information uncertainty in natural resources / edited by Kim Lowell and Annick Jaton.

p. cm.

Based on papers presented at the Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences in Quebec City, Canada, May 20–22, 1998.

ISBN 1-57504-119-7

1. Geographic information systems—Congresses. 2. Spatial analysis (Statistics)—Congresses. 3. Natural resources—Management—Statistical methods—Congresses. I. Lowell, Kim. II. Jaton, Annick. III. International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (3rd : 1998 : Quebec City, Quebec)

G70.212.S636 1999
910'.285—dc21 99-18424
CIP

ISBN 1-57504-119-7

ANN ARBOR PRESS
121 South Main Street, Chelsea, Michigan 48118
Ann Arbor Press is an imprint of Sleeping Bear Press, Inc.

PRINTED IN THE UNITED STATES OF AMERICA
10 9 8 7 6 5 4 3 2 1