Ecology, 95(9), 2014, pp. 2434–2442

Spatially explicit structural equation modeling

Eric G. Lamb,[1,5] Kerrie L. Mengersen,[2] Katherine J. Stewart,[3,4] Udayanga Attanayake,[1] and Steven D. Siciliano[4]

[1]*Department of Plant Sciences, University of Saskatchewan, 51 Campus Drive, Saskatoon, Saskatchewan S7N 5A8 Canada*
[2]*School of Mathematical Sciences, Queensland University of Technology, 2 George Street, Brisbane 4000 Australia*
[3]*Yukon Research Centre, Yukon College, 500 College Drive, Whitehorse, Yukon Territory Y1A 5K4 Canada*
[4]*Department of Soil Science, University of Saskatchewan, 51 Campus Drive, Saskatoon, Saskatchewan S7N 5A8 Canada*

Abstract. Structural equation modeling (SEM) is a powerful statistical approach for the testing of networks of direct and indirect theoretical causal relationships in complex data sets with intercorrelated dependent and independent variables. SEM is commonly applied in ecology, but the spatial information commonly found in ecological data remains difficult to model in a SEM framework. Here we propose a simple method for spatially explicit SEM (SE-SEM) based on the analysis of variance/covariance matrices calculated across a range of lag distances. This method provides readily interpretable plots of the change in path coefficients across scale and can be implemented using any standard SEM software package. We demonstrate the application of this method using three studies examining the relationships between environmental factors, plant community structure, nitrogen fixation, and plant competition. By design, these data sets had a spatial component, but were previously analyzed using standard SEM models. Using these data sets, we demonstrate the application of SE-SEM to regularly spaced, irregularly spaced, and ad hoc spatial sampling designs and discuss the increased inferential capability of this approach compared with standard SEM. We provide an R package, sesem, to easily implement spatial structural equation modeling.

Key words: lag distance; spatial correlation; spatial ecological analysis; spatial environment–ecological response relationships; structural equation modeling; variance–covariance matrices.

Introduction

Structural equation modeling (SEM) is a powerful statistical approach for the testing of hypotheses about networks of direct and indirect theoretical causal relationships in complex data sets with intercorrelated dependent and independent variables (Shipley 2000*a*, Pugesek et al. 2003, Grace 2006, Kline 2011). The many advantages of SEM include its mathematical and statistical rigor, flexibility for describing complex hypotheses about relationships between variables, scientific inferential capacity and visually intuitive representation of networks among ecological factors. These features have ensured that SEM is now widely applied to an array of questions in ecology and related fields. However, notwithstanding these advantages, the standard SEM implementation is not well suited to some common features of ecological data. Recent methodological advances have opened broad possibilities for the modeling of nonlinear relationships and other complex data structures (e.g., Grace and Bollen 2008, Shipley 2009, Clough 2012, Grace et al. 2012), but data with a spatial component remains difficult to model in a SEM framework. This is a significant gap as most ecological data have underlying spatial structure, yet standard spatial statistics are limited to univariate and bivariate analyses (Fortin and Dale 2005). Just as spatial scale can influence individual ecosystem components, spatial scale likely influences the networks of relationships among ecosystem components (e.g., Levin 1992). Here we propose a simple method for spatially explicit structural equation modeling (SE-SEM) for the analysis of spatial structure in complex ecological networks. SE-SEM has numerous potential applications. It may provide a tool, for example, to link the common application of standard SEM in biodiversity, ecosystem function studies (e.g., Grace et al. 2007*a*), and the spatially explicit models needed to link experimental data to landscape-level modeling and decision making (Cardinale et al. 2012).

Most ecological data contain underlying spatial structure that needs to be accounted for during analysis (Legendre and Fortin 1989, Legendre et al. 2002, Fortin and Dale 2005, Legendre and Legendre 2012). Spatial structure in ecological data can include spatial dependence, where spatially structured environmental variables influence ecological response variables, and spatial autocorrelation, where the relationships among variables is a function of distance among samples (Legendre et al. 2002). Spatial dependence in environment–ecological response relationships likely influences many of the networks of causal relationships at the heart of most ecological applications of SEM, yet standard SEM

Manuscript received 23 October 2013; revised 6 February 2014; accepted 12 February 2014. Corresponding Editor: B. D. Inouye.

[5] E-mail: eric.lamb@usask.ca

methodology cannot directly incorporate that spatial dependence. Extending the SEM approach to spatial data requires expansion of the SEM causal framework (Shipley 2000*a*, Grace 2006) to more readily incorporate spatially explicit dependent causal relationships. In this paper, we argue that direct spatially dependent causal relationships produce indirect spatial dependencies in subsequent pathways in a SEM.

A variety of approaches for spatial SEM have been proposed. One common approach is to incorporate a distance measure as an observed or latent variable in a standard SEM model (Bailey and Krzanowski 2012). Other methods have largely been developed for health and sociometric data aggregated by administrative districts such as counties or city wards such as the modeling of spatially structured residuals accounting for relationships between adjacent districts (Congdon et al. 2007, Oud and Folmer 2008, Congdon 2010). Also proposed are an extension of the common factor model to include neighborhood information (Wang and Wall 2003) and a hierarchical extension for simultaneous modeling of relationships between latent variables while accounting for spatial relationships (Liu et al. 2005). While some SEM software packages can accommodate hierarchical or blocked sampling designs to a limited degree (e.g., M-Plus; Muthén and Muthén 2010), in most cases these methods have not achieved widespread use and are not available in widely used SEM software packages.

Common geostatistical techniques analyze spatial structure by dividing the data into a series of lag distances (i.e., by calculating all of the pairwise distances among sample points and allocating sample pairs into bins encompassing a particular range of distances apart). Properties of the sample pairs within each lag distance bin are then calculated and plotted against lag distance (Fortin and Dale 2005, Si 2008). Typical values analyzed in this way include the semivariance, or the variance among sample pairs within a bin, and Moran's *I*, a measure of the autocorrelation among sample pairs (Moran 1948, Fortin and Dale 2005). This geospatial approach produces visually intuitive plots of the relationship between the ecological similarity among samples and the physical distance between those samples (Banerjee and Siciliano 2012). Geospatial techniques are primarily univariate, although cross-semivariance methods can be used to assess the joint spatial patterns of more than one variable (Fortin and Dale 2005), multiscale ordination and spectral decomposition methods can be used to describe spatial patterns in community composition (ver Hoef and Glenn-Lewin 1989, Borcard and Legendre 2002, Wagner 2003, 2004, Borcard et al. 2004). The highly visual output of the semivariogram, spectral decomposition, and standard SEM approaches have contributed strongly to the embrace of those methods by ecologists. However, these features are not readily available in current spatial SEM methods motivating the development of a spatial SEM methodology that merges SEM and geostatistics.

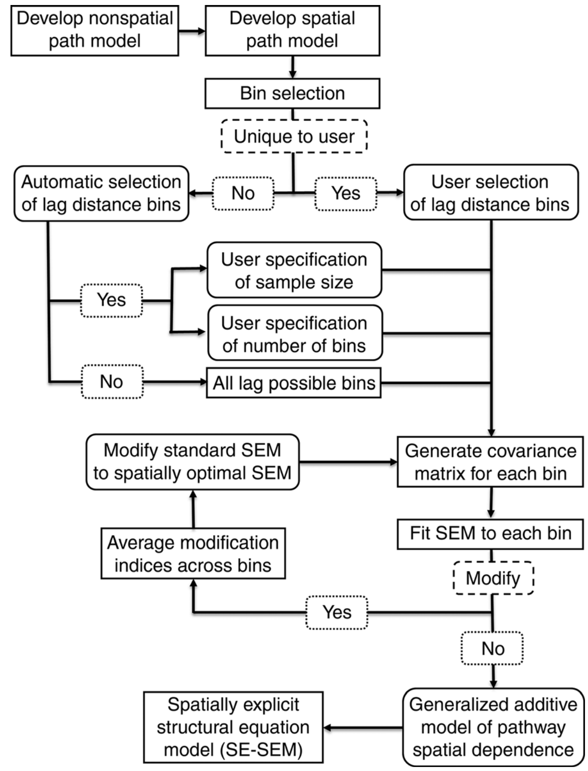

FIG. 1. Flowchart depicting the process and decision points in a spatial structural equation modeling (SEM) analysis.

Here we propose a simple method for spatially explicit SEM (SE-SEM) based on the analysis of variance–covariance matrices calculated for a range of lag distances. This method provides readily interpretable plots of the change in path coefficients across scale and can be implemented using any standard SEM software package. We demonstrate the application of SE-SEM using a data set with a spatial component previously analyzed using a standard SEM (Stewart et al. 2011*b*).

Description of the Method

We propose a method for spatially explicit structural equation modeling based on the analysis of a series of spatially explicit variance–covariance matrices from a range of lag distances (Fig. 1). Standard SEM is based on the analysis of variance–covariance matrices; the SE-SEM method we propose here is simply to fit the same SEM model to a series of variance–covariance matrices calculated for different lag distances. Functions, annotated examples, and example data to implement SE-SEM using the R package (R Development Core Team 2011) and the R laavan library (Rosseel 2012) to fit the SEM models is provided in the R sesem library and in the Supplement.

Specify and fit a standard SEM (nonspatial) model.—The first step in SE-SEM is to develop and fit an initial standard SEM model to the data, assuming spatially independent responses and without any explicit spatial terms in the model. We assume that the reader has a working knowledge of SEM and has developed and fit a standard SEM (i.e., nonspatial) model to their data. Numerous books and review papers targeted at a range of audiences and disciplines provide an accessible introduction to the theory and practice of structural equation modeling (e.g., Bollen 1989, Shipley 2000*a*, Pugesek et al. 2003, Grace 2006, 2008, Kline 2011, Lamb et al. 2011). In short, development of the nonspatial model requires the specification of an initial path model summarizing a causal hypothesis in the form of paths representing causal relationships among the variables in the model. Model fitting involves the estimation of parameters (path coefficients) for each relationship included in the initial model, followed by tests to evaluate model fit. In cases where the fit of the model is not adequate, modification indices are often used to identify additional paths that, when added to a model, achieve adequate fit.

In SE-SEM, not only does the nonspatial model specify causal hypotheses among individual variables (Shipley 2000*a*), the model also specifies a series of *spatial* causal hypotheses about how variables affect one another at a distance. When developing the nonspatial model it is therefore critical to articulate for each path a causal mechanism that can plausibly be expected to operate at a given distance. Some spatial mechanisms result from obvious interactions among community components; a tall shrub in one plot, for example, can reasonably shade lichens on a treeless plot a few meters away. Many if not most spatial causal mechanisms are going to be driven by spatially dependent environmental factors, however. Spatial dependence, where a spatially structured environmental variable influences one or more ecological response variables (Legendre et al. 2002), is likely to be very common in ecological applications of SE-SEM. A simple spatially dependent mechanism may be captured by a direct path from the spatially structured environmental variable to a biotic one. Complex spatially dependent mechanisms following indirect paths originating with a spatially dependent variable and passing through one or more subsequent variables are likely to be encountered. Spatial dependence in soil moisture, for example, may in turn have important direct and indirect spatially dependent causal influences on components of the plant community. It is these indirect spatial mechanisms, easily incorporated into SE-SEM, that are most difficult to model with alternative geostatistical methods.

Calculate spatially explicit variance–covariance matrices for a series of lag distances.—The second step is to select the lag distance bins and calculate a series of spatially explicit variance–covariance matrices. Typically, geospatial analysis is limited to 50% of the maximum and minimum lag distances, so field designs and the maximum bin size should reflect a spatial pattern suitable for the inference space under study. Selection of the number of distance bins and the size of each bin are important considerations that depend on the specific goals of a study. Bin sizes should have reasonable ecological interpretation(s), and the sample size (number of sample pairs within the bin should be adequate for SEM analysis). Small sample sizes may lead to bias or inaccuracy in the variance–covariance matrix, further resulting in both unreliable parameter estimates and tests of model fit (Shipley 2000*a*, Kline 2011). Recommendations for minimum sample size in the SEM literature vary widely, with 100–200 samples or 5–10 times the number of parameters in the model often suggested (Grace 2006, Kline 2011).

Within the implementation of the methodology, three options for lag distance bin size selection are provided: (1) direct selection of bin sizes by the user, (2) automatic generation of bins based on a minimum sample size (number of sample pairs) per bin with a default sample size of 100 and (3), automatic generation of a number of bins of equal distance range and all observations from that range in that bin. Option 1 is to be used when the investigator has a priori reasons to investigate particular lag distances associated with particular processes with a known scale, and has developed an initial path diagram representing spatial hypotheses specific to those lag distance(s). Option 2 is useful when the primary interest is how path strengths change as a function of space. Setting a minimum sample size allows the maximum number of statistically reasonable bins to be used. Option 3 is useful when an investigator wants a precise estimate of path coefficients but still provides enough lag distance bins to be useful for investigating changes in path coefficients with scale. Options 2 and 3 require an initial path diagram representing spatial causal hypotheses that are applicable across the full range of lag distances to be investigated. Following lag distance bin selection, a spatially explicit variance–covariance matrix is calculated using the sample pairs that fall within each bin. These matrices are the basis of the subsequent SE-SEM models.

Fit and evaluate SEM models for each lag distance bin.—An initial path model equivalent to the nonspatial SEM model is then fit to the variance–covariance matrix for each lag distance. Following model fitting, indices of model fit should be plotted against lag distance. The standard advice on the evaluation of model fit, such as nonsignificant χ^2 tests and acceptable values for other indices of model fit (Grace 2006, Kline 2011), applies to the evaluation of model fit across lag distance with some caveats. If the model fit is adequate across all, or most, lag distances then the modeler should proceed directly with the comparison of individual path coefficients across lag distances. If not, the investigator can proceed along a number of different routes. If the primary goal is the direct comparison of individual path coefficients across different lag distances, it is important that the same structural model be used for each lag distance. The inclusion of additional direct or indirect paths at some lag distances but not others may affect other path coefficients in the model, making subsequent comparisons of path coefficients across distances difficult. If different models at different lags are acceptable, three approaches are available. (1) If the primary goal is to examine processes at a particular lag distance, the use of modification indices to improve the fit of a single model at that lag distance is a reasonable strategy. (2) If the primary goal is to make cross-lag-distance comparisons, then modification indices averaged across all lag distances can be used. This approach will identify paths that when added will result in the largest drops in χ^2 values across all lag distances, even if such paths are nonsignificant at particular distances. (3) If the primary goal is more flexible, or if different causal relationships are hypothesized for different lag distances, separate models can be fit at different distances. This approach, however, will preclude direct comparison of the model parameters across distances. In all cases, model modification should be done cautiously with only biologically reasonable paths added, and appropriate concern for overfitting (Grace 2006, Kline 2011).

Path coefficients vs. lag distance.—A primary goal of spatially explicit SEM is the comparison of particular path coefficients across a range of lag distances. These comparisons should be made using unstandardized path coefficients as the standardization process makes it difficult to directly compare standardized coefficients across different data sets (Grace 2006). Unstandardized coefficients can be directly comparable across lag distances since those variables are scaled in the same units across lag distances. The trend across lag distances can then be evaluated graphically and analyzed using, for example, a polynomial or generalized additive model (Wood 2006) with significant trends identified and distances where the paths are significantly different from zero identified. Generalized additive models (GAM) models are implemented in our method using function gam in the mgcv library (Wood 2011).

Model testing: spatial dependence of nitrogen fixation in a high Arctic community

We illustrate the application of SE-SEM using the "Truelove" data set, a study that used a regularly spaced sampling design to study links between nitrogen fixation and plant community structure in a high Arctic plant community (Stewart et al. 2011*b*). Raw data, SE-SEM functions, and the annotated R-code (R Development Core Team 2011) to carry out these examples using the laavan library (Rosseel 2012) are provided in the Supplement. In the Appendix, we also provide two additional example data sets, "Alexandra Fiord" and "Fescue Grassland," (Lamb and Cahill 2008, Stewart et al. 2011*b*) that demonstrate the use of SE-SEM for irregular lag distance (Alexandra Fiord), and ad hoc (Fescue Grassland) spatial sampling designs.

Truelove Lowland sampling

The Truelove study examined the relationships between vascular plant cover and the cover of common N_2-fixation associations (bryophytes, lichens, and biological soil crusts) and N_2-fixation rates in a high Arctic polar oasis (Stewart et al. 2011*b*). The study site was on Truelove Lowland, a 43-km^2 polar oasis on the north shore of Devon Island, Nunavut, Canada (75°67′ N, 84°58′ W; see Plate 1). The lowland is bordered by shoreline to the north, west and part of the south and by steep cliffs (~300 m) to the east and remaining south (Muc and Bliss 1977). The lowland has a distinct spatial structure, with raised beach crests occurring at regular intervals (Lev and King 1999; see also Appendix: Fig. A1). Ridges were dominated by cushion-plant–lichen communities and the intervening lowlands by Hummocky sedge–moss meadows; names follow Muc and Bliss (1977). Sampling was conducted using a regular sampling design with 129 points located every 4 m on a 512-m transect over a series of beach ridges.

Nonspatial SEM Model

The nonspatial structural equation model for Truelove Lowland was based on the final path model used by Stewart et al. (2011*b*). In that study, the SEM was fit as a multi-group model spanning four sites; here we focus on only the Truelove site using a single-group model. The SEM consisted of directly observed measures of soil moisture, shrubs, gramminoids, forbs, bryophytes, lichens, biological soil crusts, and their link to N_2 fixation. Briefly, Stewart et al. (2011*b*) developed a path model to describe how soil moisture and functional vascular plant community composition (i.e., proportion of graminoids, forbs, and shrubs) directly influenced cover of bryophytes, soil crust, and lichen abundance and hence indirectly influenced N_2 fixation. Direct paths from potential N_2-fixing cyanobacteria associations (bryophyte, lichen, and bare ground) to N_2 fixation were included. Soil crusts were included as a potential N_2-fixing association because "bare ground" as recorded in the field supported communities composed of bacteria, cyanobacteria, algae, mosses, liverworts, fungi, and small lichens (Stewart et al. 2011*b*). Paths from soil moisture to all plant components were included because (1) soil moisture is a key environmental factor determining the distribution of vegetation types in Arctic environments (Oberbauer and Dawson 1992) and (2) interactions between plant communities and soil moisture can be important in determining the operating environment of N_2-fixing associations (Zielke et al. 2002, 2005, Stewart et al. 2011*a*).

In specifying the nonspatial model for SE-SEM, it is critical that all paths represent spatial causal hypotheses. In the Truelove nonspatial SEM, each path represents a spatially dependent ecological mechanism linked back to soil moisture. Variation in soil moisture at Truelove Lowland, the key underlying environmental factor in the SEM, is driven by changes in topography and soil

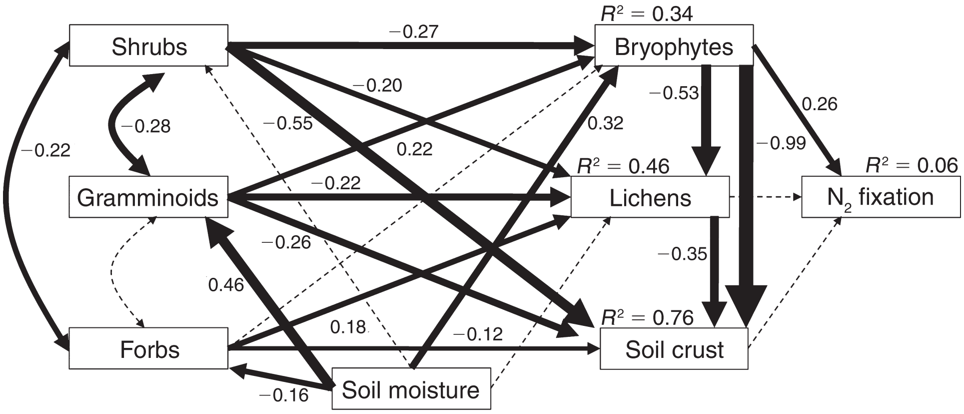

FIG. 2. Fitted nonspatial SEM model for Truelove Lowland. Standardized path coefficients are shown; dotted lines indicate nonsignificant paths ($P > 0.100$).

drainage patterns along the beach ridge cantinas occurring at scales of tens to hundreds of meters (Lev and King 1999). Numerous studies at this site and others have demonstrated moisture driven spatial dependence in arctic bryophyte, vascular plant, and soil communities (e.g., Peterson and Billings 1980, Muc and Bliss 1977, Ostendorf and Reynolds 1998, Banerjee et al. 2011). There are both direct and indirect spatially dependent ecological mechanisms linking vascular plants (shrubs, gramminoids, and forbs) to cryptogams (bryophyte, lichen, soil crust) in this community. The direct mechanisms, such as shading and competitive displacement, are likely most important at very small scales. The indirect mechanisms linking moisture effects on the vascular plants to bryophytes likely occur at scales of tens to hundreds of meters in Truelove Lowland (Muc and Bliss 1977, Lev and King 1999, Banerjee et al. 2011) and thus are appropriate for testing in an SE-SEM context across scales.

Model fitting

Given the regular sampling design, we chose to model a separate SE-SEM for each lag distance from 4 m (smallest distance measured) to 248 m (one-half of the full transect length). Using all possible bin sizes maximizes the spatial resolution of the SE-SEM (62 separate SEM models), with the drawback in this case that lag distance bins greater than 118 m had fewer than 100 samples (with a low of 67 in the 248-m bin).

RESULTS

The nonspatial Truelove structural equation model had an adequate fit ($\chi^2_5 = 10.83$, $P = 0.05$; Fig. 2). The model explained a relatively low percentage of the variation in N_2 fixation ($R^2 = 0.06$), but the abundance of bryophytes ($r^2 = 0.34$), lichens ($r^2 = 0.46$), and bare ground ($r^2 = 0.76$) were relatively well explained. The spatially explicit models had reasonably good fit across most lag distances, although there was a trend toward poorer model fit at lag distances greater than 175 m (Fig. 3, Appendix: Fig. A2). There were no average modification indices greater than 3.7, further demonstrating that the path model was valid across all lag distances.

The importance of spatial scale on the moisture–bryophyte–N_2-fixation relationship in this system is evident in the SE-SEM, as is the positive influence of moisture on gramminoids and subsequently bryophyte abundance (Fig. 4). Path coefficients describing the direct influence of moisture on gramminoids were significant at all lag distances examined and all but two of the 62 lag distances for bryophytes. Both the proximity of permafrost to the soil surface and moisture retention by bryophytes create a moist environment in these low-lying areas, which is likely to be crucial in maintaining higher rates of N_2 fixation (Stewart et al. 2011*b*). The SE-SEM demonstrates that, while moisture appears to influence the presence of bryophytes at many

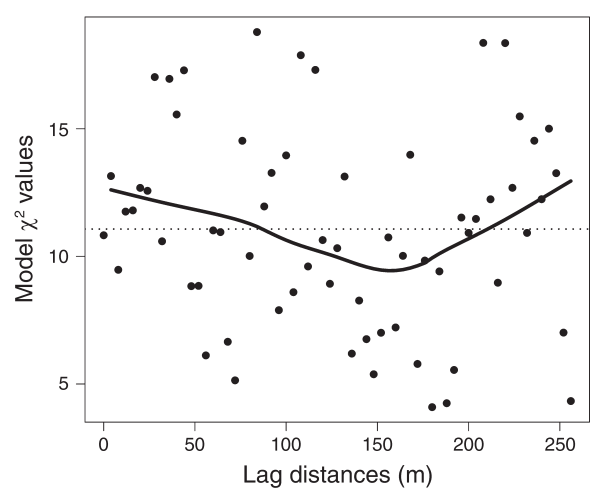

FIG. 3. Change in model fit (χ^2 values) with lag distance for the Truelove Lowland. The solid line is a smoothed curve fit using function lowess. Dotted lines indicate the critical χ^2 value corresponding to $P = 0.05$ for the model.

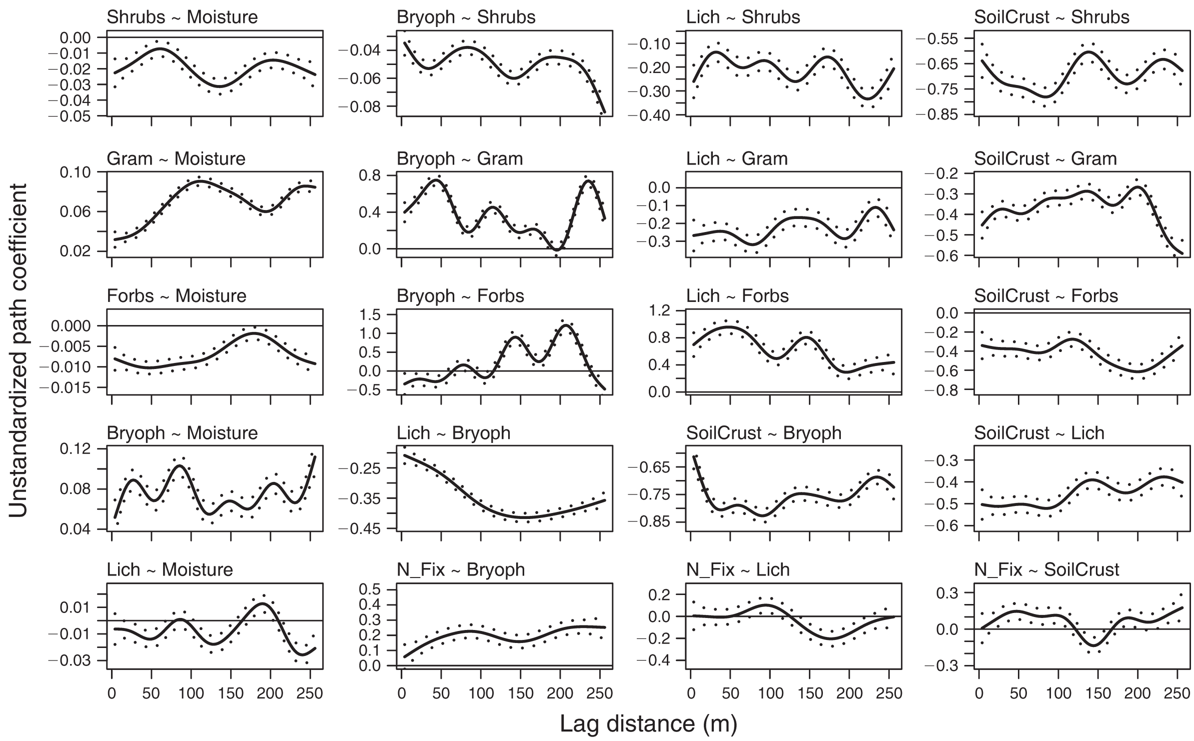

FIG. 4. Changes in unstandardized path coefficients with lag distance for Truelove Lowland. Predicted lines are plotted for each generalized additive model (GAM) relationship (solid lines); dotted lines represent ±SE. Path coefficients should be considered nonsignificant when the standard error lines cross the horizontal lines at zero. Parameter abbreviations are Shrubs, shrub cover; Moisture, gravimetric soil moisture; Bryoph, bryophyte cover; Lich, lichen cover; SoilCrust, biological soil crust; gram, graminoid cover; Forbs, forb cover; N_Fix, nitrogen fixation rate measured as mean acetylene reduced ($\mu mol \cdot m^{-2} \cdot h^{-1}$). In the panel labels, "A ~ B" indicates "A as predicted by B."

scales larger than 4 m, the indirect influence of moisture on N_2 fixation via bryophytes is strongest at intermediate lag distances (i.e., 50–100 m). These lag distances correspond with the dominant microtopographical feature of the landscape: the alternation of raised beach crests dominated by cushion-plant–lichen communities and intervening lowlands with hummocky sedge–moss meadows (Muc and Bliss 1977, Lev and King 1999). Neither biological soil crusts nor lichens had a significant influence on N_2 fixation in the nonspatial model. However, significant path coefficients from bare ground to N_2 fixation were observed at intermediate distances (31–39 m, 43–47 m, 143–147 m, and 151–155 m), likely reflecting indirect moisture effects driven by the higher abundance of biological soil crusts on the exposed tops of beach ridges. Significant path coefficients from lichens to N_2 fixation were also only found at larger intermediate lag distances (103–107 m and 159–171 m), likely capturing similar topographical spatial dependencies influencing N_2-fixing lichens. Increased rates of N_2 fixation under drier conditions have been observed in Arctic landscapes where lichens are abundant (Hobara et al. 2006, Stewart et al. 2011*a*), while in lower landscape positions lichens can be excluded by competitive displacement or intolerance to prolonged hydration episodes (Moser and Nash 1978, Joly et al. 2009).

Discussion

The spatially explicit SEM methodology described in this paper was effective at evaluating the broad spatial relationships in the Truelove data set. Importantly, a number of paths that were nonsignificant in the nonspatial model captured important spatial dependencies driven by moisture and topographic position. The primary challenge in SE-SEM is specification of spatial causal hypotheses in the nonspatial path diagram. In the absence of a clear spatial causal hypothesis, SE-SEM becomes a multivariate description of patterns of spatial autocorrelation. This in itself may be a useful application, but does not harness the full power behind SEM. Throughout this paper we have argued that a bivariate spatial dependency, or a direct spatial causal relationship between two variables (Legendre et al. 2002), will propagate through the structural model as a series of indirect spatial causal relationships. It is this scenario that makes modeling of the changes in path coefficients with lag distance useful. In cases where clear spatial causal relationships cannot be justified at all scales, it may be preferable to optimize the SE-SEM for a single lag distance as described in *Description of the method: Fit and evaluate SEM models for each lag distance bin*.

PLATE 1. The transect at Truelove Lowland, Devon Island, Nunavut, Canada, extends across an alternating series of raised beach crests, back slopes, and moist depressions. Sample points are marked with collars for a greenhouse gas emission study. Photo Credit: Martin E. Brummell.

The SE-SEM methodology outlined in this paper differs substantially from the alternative spatial SEM approaches available in the literature (Wang and Wall 2003, Liu et al. 2005, Congdon et al. 2007, Oud and Folmer 2008, Congdon 2010). Those methods have not achieved widespread usage in the natural sciences. There are at least two reasons for the lack of uptake. First, these methods are primarily aimed at data aggregated by administrative districts such as counties or city wards, while ecologists are more likely to be interested in fully geo-located data. Second, software for the existing methods is not widely available in commonly used SEM software packages. The methodology we propose has a number of advantages over the existing spatial SEM approaches. First, the flexible choice of lag distance bins puts the analyst in full control of the type of spatial questions they wish to address. Second, the visually intuitive combination of SEM path diagrams and geostatistical plots of path coefficients across lag distances are likely to be attractive to ecologists because they can be broadly interpreted by readers who do not have a detailed understanding of the underlying methods. Third, the methodology can be rapidly implemented using any standard SEM software package.

The SE-SEM methodology proposed here has links with the multi-scale ordination and spectral decomposition methods developed to describe spatial patterns in community composition and plant–environment relationships (ver Hoef and Glenn-Lewin 1989, Borcard and Legendre 2002, Wagner 2003, 2004, Borcard et al. 2004, Legendre and Legendre 2012). Multi-scale ordination will likely remain a preferred method for examining spatial patterns of community composition, as species-level community composition can be very difficult to bring directly into an SEM. Typical SEM applications use ordination axes as a proxy variable for composition (e.g., Grace et al. 2007*b*, Lamb and Cahill 2008). The principle advantage of SE-SEM is the ability to specify and then test a multivariate hypothesis relating biotic and environmental variables.

Further development and testing of SE-SEM is needed on a number of fronts. The examples provided here all involve observed variable models; further testing to validate SE-SEM for latent variable models and other common SEM methods such as multi-group models and composite variables is needed. Further, we have restricted our implementation of SE-SEM to standard maximum-likelihood methods for fitting the models to each variance–covariance matrix. An alternative approach that could be used to test the spatial causal hypotheses in the path diagram for each lag distance bin is the d-separation (d-sep) test (Shipley 2000*a*, *b*, 2012). The d-sep approach may be particularly useful in cases where model identification or convergence present problems. In addition, methods need to be developed to separate spatial autocorrelation and spatial causal relationships, and to detect whether a data set has little or no spatial signal. A useful approach here may be to formulate path models describing spatial autocorrelation (Legendre and Legendre 2012) as well as the spatial causal models used here; model comparison could allow the relative strength of the causal processes to be more clearly evaluated. In this paper, we follow general practice in geostatistics and only report results for lag distances up to 50% of the maximum distance separating samples. In initial testing, however, we utilized larger lag distances and observed a general trend for decreasing model fit with increasing lag distance. This pattern may represent either a weakness of the method or a real pattern of weakening spatial control at larger scales. A decline in the strength of spatial pattern is observed in many autocorrelation studies, i.e., it is found that at some scale spatial autocorrelation becomes negligible and samples can be treated as independent (Legendre et al. 2002, Banerjee et al. 2011). If spatial pattern is indeed weakening at larger scales, then there are implications for the use of averaged modification indices for model fit improvement. In particular, the averaged modification indices may become dominated by noise from the larger lag distance bins leading to the danger of model overfitting. Alternatively, the nonspatial model may be inappropriate to describe the processes occurring at larger spatial scales (i.e., nonstationarity). If local spatial processes dominate in the formulation and testing of the nonspatial model, then it should be expected that model fit will decline with lag distance. This suggests that different path models may need to be developed for different lag distance ranges.

The SE-SEM we present here is sufficient for the analysis of relatively simple spatially structured data collected without SE-SEM in mind. Future applications will benefit, however, from a careful selection of sampling designs to ensure samples for lag distance bins covering the range of important spatial processes. The Alexandra Fiord data set described in the Appendix, for example, utilized a sampling design intended for spatial analysis. In hindsight, it is clear that important spatial processes are occurring at that site at the 0.2–0.5 m scale, yet sample size limitations precluded a lag distance bin of that size. In cases where there is insufficient information to establish an appropriate regularly or irregularly spaced transect, sampling designs, such as, a partial Fibonacci spiral may be effective (Fortin and Dale 2005).

In summary, SE-SEM provides a simple and visually intuitive method to incorporate spatial data into structural equation models. Here we demonstrate the application of SE-SEM using a data set previously analyzed using standard SEM methods. Important spatial patterns and processes hidden in the standard SEM were evident in the SE-SEM results. The power to reveal spatial dependencies among multiple variables makes SE-SEM a significant advance in the application of SEM to ecological questions.

Acknowledgments

This project was supported by NSERC Discovery Grants to E. G. Lamb and S. D. Siciliano, and a Canadian Foundation for Innovation (CFI) grant to E. G. Lamb. We thank two anonymous reviewers for comments that greatly improved the manuscript.

Literature Cited

Bailey, T. C., and W. J. Krzanowski. 2012. An overview of approaches to the analysis and modelling of multivariate geostatistical data. Mathematical Geosciences 44:381–393.

Banerjee, S., B. C. Si, and S. D. Siciliano. 2011. Evidence of high microbial abundance and spatial dependency in three Arctic soil ecosystems. Soil Science Society of America Journal 75:2227–2232.

Banerjee, S., and S. D. Siciliano. 2012. Factors driving potential ammonia oxidation in Canadian Arctic ecosystems: does spatial scale matter? Applied and Environmental Microbiology 78:346–353.

Bollen, K. A. 1989. Structural equations with latent variables. John Wiley and Sons, New York, New York, USA.

Borcard, D., and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological Modelling 153:51–68.

Borcard, D., P. Legendre, C. Avois-Jacquet, and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. Ecology 85:1826–1832.

Cardinale, B. J., et al. 2012. Biodiversity loss and its impact on humanity. Nature 486:59–67.

Clough, Y. 2012. A generalized approach to modeling and estimating indirect effects in ecology. Ecology 93:1809–1815.

Congdon, P. 2010. A spatial structural equation model with an application to area health needs. Journal of Statistical Computation and Simulation 80:401–412.

Congdon, P., M. Almog, S. Curtis, and R. Ellerman. 2007. A spatial structural equation modelling framework for health count responses. Statistics in Medicine 26:5267–5284.

Fortin, M. J., and M. R. T. Dale. 2005. Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge, UK.

Grace, J. B. 2006. Structural equation modeling and natural systems. Cambridge University Press, Cambridge, UK.

Grace, J. B. 2008. Structural equation modeling for observational studies. Journal of Wildlife Management 72:14–22.

Grace, J. B., et al. 2007*a*. Does species diversity limit productivity in natural grassland communities? Ecology Letters 10:680–689.

Grace, J. B., and K. A. Bollen. 2008. Representing general theoretical concepts in structural equation models: the role of composite variables. Environmental and Ecological Statistics 15:191–213.

Grace, J. B., H. D. Safford, and S. Harrison. 2007*b*. Large-scale causes of variation in the serpentine vegetation of California. Plant and Soil 293:121–132.

Grace, J. B., D. R. Schoolmaster, G. R. Guntenspergen, A. M. Little, B. R. Mitchell, K. M. Miller, and E. W. Schweiger. 2012. Guidelines for a graph-theoretic implementation of structural equation modeling. Ecosphere 3:art73.

Hobara, S., C. McCalley, K. Koba, A. Giblin, M. Weiss, G. Gettel, and G. Shaver. 2006. Nitrogen fixation in surface soils and vegetation in an Arctic tundra watershed: a key source of atmospheric nitrogen. Arctic, Antarctic, and Alpine Research 38:363–372.

Joly, K., R. R. Jandt, and D. R. Klein. 2009. Decrease of lichens in Arctic ecosystems: the role of wildfire, caribou, reindeer, competition and climate in north-western Alaska. Polar Research 28:433–442.

Kline, R. B. 2011. Principles and practice of structural equation modeling. Third edition. Guilford Press, New York, New York, USA.

Lamb, E. G., and J. F. Cahill. 2008. When competition does not matter: grassland diversity and community composition. American Naturalist 171:777–787.

Lamb, E. G., S. J. Shirtliffe, and W. E. May. 2011. Structural equation modeling in the plant sciences: an example using yield components in oat. Canadian Journal of Plant Science 91:603–619.

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

Legendre, P., and M. J. Fortin. 1989. Spatial pattern and ecological analysis. Vegetatio 80:107–138.

Legendre, P., and L. Legendre. 2012. Numerical ecology. Elsevier, Amsterdam, The Netherlands.

Lev, A., and R. H. King. 1999. Spatial variation of soil development in a high arctic soil landscape: Truelove Lowland, Devon Island, Nunavut, Canada. Permafrost and Periglacial Processes 10:289–307.

Levin, S. A. 1992. The problem of pattern and scale in ecology: The Robert H. MacArthur award lecture. Ecology 73:1943–1967.

Liu, X., M. M. Wall, and J. S. Hodges. 2005. Generalized spatial structural equation models. Biostatistics 6:539–557.
Moran, P. A. P. 1948. The Interpretation of statistical maps. Journal of the Royal Statistical Society B 10:243–251.
Moser, T. J., and T. H. Nash. 1978. Photosynthetic patterns of *Cetraria cucullata* (Bell.) Ach. at Anaktuvuk Pass, Alaska. Oecologia 34:37–43.
Muc, M., and L. C. Bliss. 1977. Plant communities of Truelove Lowland. Pages 143–154 *in* L. C. Bliss, editor. Truelove Lowland, Devon Island, Canada: a high Arctic ecosystem. University of Alberta Press, Edmonton, Alberta, Canada.
Muthén, L. K., and B. O. Muthén. 2010. Mplus users guide. Muthén and Muthén, Los Angeles, California, USA.
Oberbauer, S. F., and T. E. Dawson. 1992. Water relations of arctic vascular plants. Pages 259–275 *in* F. S. Chapin, R. L. Jefferies, G. R. Shaver, and J. Svoboda, editors. Arctic ecosystems in a changing climate. Academic Press, New York, New York, USA.
Ostendorf, B., and J. Reynolds. 1998. A model of arctic tundra vegetation derived from topographic gradients. Landscape Ecology 13:187–201.
Oud, J. H. L., and H. Folmer. 2008. A structural equation approach to models with spatial dependence. Geographical Analysis 40:152–166.
Peterson, K. M., and W. D. Billings. 1980. Tundra vegetational patterns and succession in relation to microtopography near Atkasook, Alaska. Arctic and Alpine Research 12:473–482.
Pugesek, B. H., A. Tomer, and A. von Eye, editors. 2003. Structural equation modeling: applications in ecological and evolutionary biology. Cambridge University Press, Cambridge, UK.
R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.r-project.org
Rosseel, Y. 2012. lavaan: an R package for structural equation modeling. Journal of Statistical Software 48:1–36.
Shipley, B. 2000*a*. Cause and correlation in biology. Cambridge University Press, Cambridge, UK.
Shipley, B. 2000*b*. A new inferential test for path models based on directed acyclic graphs. Structural Equation Modeling 7: 206–218.
Shipley, B. 2009. Confirmatory path analysis in a generalized multilevel context. Ecology 90:363–368.
Shipley, B. 2012. The AIC model selection method applied to path analytic models compared using a d-separation test. Ecology 94:560–564.
Si, B. C. 2008. Spatial scaling analyses of soil physical properties: a review of spectral and wavelet methods. Vadose Zone Journal 7:547–562.
Stewart, K., D. Coxson, and P. Grogan. 2011*a*. Nitrogen inputs by associative cyanobacteria across a low Arctic tundra landscape. Arctic, Antarctic, and Alpine Research 43:267–278.
Stewart, K. J., E. G. Lamb, D. S. Coxon, and S. D. Siciliano. 2011*b*. Bryophyte-cyanobacterial associations as a key factor in N_2-fixation across the Canadian Arctic. Plant and Soil 344: 335–346.
ver Hoef, J. M., and D. C. Glenn-Lewin. 1989. Multiscale ordination: a method for detecting pattern at several seales. Vegetatio 82:59–67.
Wagner, H. H. 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. Ecology 84:1045–1057.
Wagner, H. H. 2004. Direct multi-scale ordination with canonical correspondence analysis. Ecology 85:342–351.
Wang, F., and M. M. Wall. 2003. Generalized common spatial factor model. Biostatistics 4:569–582.
Wood, S. N. 2006. Generalized additive models: an introduction with R. Chapman and Hall, Boca Raton, Florida, USA.
Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society B 73:3–36.
Zielke, M., A. S. Ekker, R. A. Olsen, S. Spjelkavik, and B. R. Solheim. 2002. The influence of abiotic factors on biological nitrogen fixation in different types of vegetation in the high Arctic, Svalbard. Arctic, Antarctic, and Alpine Research 34: 293–299.
Zielke, M., B. Solheim, S. Spjelkavik, and R. Olsen. 2005. Nitrogen fixation in the high Arctic: role of vegetation and environmental conditions. Arctic, Antarctic, and Alpine Research 37:372–378.

Supplemental Material

Appendix

Additional results for the Truelove Lowland spatially explicit structural equation model (SE-SEM), and example applications of SE-SEM to two additional data sets (*Ecological Archives* E095-216-A1).

Supplement

R functions to perform SE-SEM and R scripts and data for the examples used in this paper (*Ecological Archives* E095-216-S1).