The University of Maine
DigitalCommons@UMaine

University of Maine Office of Research and Sponsored Programs: Grant Reports

Special Collections

6-7-2001

Hierarchical Modeling for Integrated Environmental Assessments

Raymond O'Connor
Principal Investigator; University of Maine, Orono

Deidre Mageean
Co-Principal Investigator; University of Maine, Orono

Final Report for Period: 01/1998 - 08/1999
Submitted on: 06/07/2001
Principal Investigator: O'Connor, Raymond .
Award ID: 9711623
Organization: University of Maine
Title:
Hierarchical Modeling for Integrated Environmental Assessments

Project Participants

Senior Personnel

Name: O'Connor, Raymond

Worked for more than 160 Hours: Yes

Contribution to Project:

As P.I. supervising overall work and providing technical guidance on biological and quantitative ecological issues. Presented multiple talks about the project results and drafted manuscripts

Name: Mageean, Deirdre

Worked for more than 160 Hours: Yes

Contribution to Project:

As Co-P.I. responsible for socio-economic data analysis and interpretation. Presented multiple talks about project results and assisted in writing manuscripts for publication.

Name: White, Denis

Worked for more than 160 Hours: Yes

Contribution to Project:

A principal collaborator for the subcontrast with Oregon State University. Led the effort on methods for visualization and for evaluation of cross-validation methods. Wrote papers and made presentations about that component of the work.

Name: Urquhart, N

Worked for more than 160 Hours: Yes

Contribution to Project:

Statistician on project, at Oregon State University. Designed and led effort to incorporate spatial elements into regression tree analysis.

Post-doc

Name: Lawler, Joshua

Worked for more than 160 Hours: Yes

Contribution to Project:

Undertook major statistical analyses of socioeconomic data and biological data, particularly on methodological issues. Made presentationa and drafted manuscripts.

Graduate Student

Name: Bartlett, John

Worked for more than 160 Hours: Yes

Contribution to Project:

Undertook major statistical analyses of socioeconomic data and biological data. Prepared a thesis, made presenations, and drafted manuscripts.

Undergraduate Student

Name: Wagner, Tansey

Worked for more than 160 Hours: Yes

Contribution to Project:

Undertook an Honors thesis in wildlife ecology. Collated independent data from extant Internet files with which to evaluate predicitons of a model central to the project, processed these data and compared results to a list of predictions from central model. Wrote up the results for an undergraduate thesis.

Research Experience for Undergraduates

Organizational Partners

USDA Forest Service Southern Research Station

Joint collaboration on a confrence presentation

Patuxent Wildlife Research Center

Collaborated in developing research questions and in implementing analysis, and co-wrote a manuscript in press

Other Collaborators or Contacts

Activities and Findings

Project Activities and Findings: (See PDF version submitted by PI at the end of the report)

Training and development activities involved the provision of research experience to one undergraduate, Ms. Tansey Wagner, who used the work as the basis for a successful Honors thesis at University of Maine (supervised by R. J. O'Connor); partial support for Ph.D. research by John G. Bartlett, leading to successful completion of a thesis in wildlife ecology (supervised by R. J. O'Connor and D. M. Mageean); and postdoctoral experience for Joshua J. Lawler, a doctoral graduate of Utah State University (supervised by D. M. Mageean and R. J. O'Connor). In addition, Ms. Laura Hayes, a Master's Student in Spatial Information Science and Engineering at Maine, undertook research with our methodology as part of a requirement in SISE that she apply her spatial training coursework to an applied problem.

Project Training and Development: (See PDF version submitted by PI at the end of the report)

Research Training:

For all graduate personnel, experience in the analysis of large datasets with spatial information about the cases attached. Also either training in, or advanced experience with the use of S-plus in statistical analysis.

For Lawler and Bartlett, significant exposure to interdisciplinary work across the biological and social sciences, both as to diversity of datasets involved and the different backgrounds and approaches within the specific disciplines. For Bartlett the experience extended his gaduate training to doctoral level.

For Wagner, involvement in the project provided research experience as an undergraduate, plus first experience in writing a thesis.

Outreach Activities:

Substantial effort was put into promoting this research and its outcomes to the diverse communities reflected in its interdisciplinary nature. In particular we spent considerable effort on presentations to conferences of the scientific and demographic societies and to the human dimensions research community. This effort gave us a sense of how reviewers from the respective disciplines might view the research and how the results needed to be cast to be acceptable to both disciplines. A list of these presentations is below.

The project provided for research in, inter alia, bird-habitat relationships, human and biodiversity interactions, and in ecological risk assessment. The project made extensive use of the national Breeding Bird Survey organized by the USGS Biological Resources Division and by the Canadian Wildlife Service and O'Connor was asked to chair a programmatic review of that program. The resulting Report is available as:

O'Connor, R. J., E. Dunn, D. H. Johnson, S. L. Jones, D. Petit, K. Pollock, C. R. Smith, J. L. Trapp, and E. Welling. 2000. A Programmatic Review of the North American Breeding Bird Survey: Report of a Peer Review Panel. Report to Patuxent Wildlife Research Center, Laurel, MD. (http://www.pwrc.usgs.gov/bbs/bbsreview/bbsfinal.pdf)

In addition, O'Connor served on the Partners in Flight Landbird Monitoring Strategy Working Group and was a member of the NSF-sponsored Workshop on Biodiversity Observatories held at Blandy Farm and Arboretum, University of Virginia, Boyce, VA. September 9-12. The research in ecological risk assessment supported by the grant was also relevant to O'Connor's membership on the USEPA ECOFRAM Working Group on probabilistic risk assessment methods for pesticide effects.

Journal Publications

Bartlett, J. G., D. M. Mageean, and R. J. O'Connor, "Residential expansion as a continental threat to U.S. coastal ecosystems", Population & Environment, p. 429, vol. 21, (2000). Published

O'Connor, R. J., Boone, R. B., Jones, M. T., and Lauber, T. B., "Linking continental climate and land use patterns with grassland bird distribution in the conterminous United States", Studies in Avian Biology, p. 45, vol. 19, (1999). Published

White, D., and J. C. Sifneos, "Regression tree cartography", Journal of Computational and Graphical Statistics, p. , vol. , ( ). Accepted

Hayes, L., and R. J. O'Connor, "The distribution of forest-patch sensitivity over the range of a neotropical migrant bird", Forest Science, p. , vol. , ( ). Submitted

Sifneos, J.C., White, D., and Urquhart, N. S., "A comparison of pruning methods for regression trees: evidence from simulation and published studies", Biometrical Journal, p. , vol. , ( ). Submitted

Books or Other One-time Publications

Hahn, D.C., and R. J. O'Connor, "Contrasting determinants of the abundance of an invasive species in its ancestral and colonized ranges", (2001). Book, Accepted
Editor(s): Scott, J. M., P. J. Heglund, F. Samson, J. Haufler, M. Morrison, M. Raphael, and B. Wall
Collection: Predicting Species Occurrences: issues of scale and accuracy
Bibliography: Island Press

Bartlett, J. G., "Anthropogenic Stressors on the Environment: Identifying Interactions Between Demographic and Environmental Factors in the Conterminous United States", (1999). Thesis, Published
Bibliography: University of Maine, Orono, Maine

Wagner, T., "An evaluation of an avian diversity model", (1999). Thesis, Published
Bibliography: Honors Thesis, University of Maine, Orono, Manie

Web/Internet Site

URL(s):

Description:

Other Specific Products

Contributions

Contributions within Discipline:

The contributions of our research to the core disciplines of the project can be summarized as a set of methodological improvements and a set of demonstration studies illustrating their application to ecological and human dimensions studies.
We developed a new and more efficient method of identifying confounding variables within regression tree analysis. In multiple linear regression one can screen out confounded variables by excluding variables that are highly corrleated. In regression tree analysis, however, the data sample is recursively partitioned into smaller and smaller subsets, and if cross-correlations arise within one of these there is a problem. Previous methods of identifying confounding variables identified the potential 'surrogate' and 'competing' variables at each node about to be

split and in principle one can investigate the problems their cross-correltaion will induce. However, if one changes one of these variables for its correlated competitor, all the subsequent splits may also change, so the possibilities multiply. Our technique involves introducing a small perturnation only to those variables that are in what will be our final regression tree if no correlation is revealed. Hence the number of possibilities to explore if correlation is detected is smaller. The perturbation works because it effectively 'fuzzes' the target variable, so if there is an (unchanged) variable around that is well correlated with the target one, it emerges because it does a better job of prediction than the perturnbed one can.

The relative importance of variables in a regression tree have always been quantifiable as the the equivalent of a coefficient of determination. One can develop the analogue of a regression coefficient for the variable, characterizing by how much the response variable changes on making a change in the predictor variable of interest. One problem with regression trees is that the effects of variables are inherently nested by virtue of the tree structure, so the apparent effect of most variables actually contains effects from other variables. We developed a method of separating out these interacting effects, essentially by introducing analogues of dummy variables in multiple linear regression. Future users of regression tree analysis can readily apply the same thinking to their particular analysis.

In the study of human-environment interactions one of the fundamental problems is to know how much of a pattern in the natural world is due to humans and how much to environment, especially when part but not all of either effect may be correlated with the other. Bocard and his colleagues in Quebec elegantly solved a related problem, how to partition effects that have a spatial component from those of an environmental factor that is also in part spatial. We modified this to solve the human-environment interaction problem. Basically we proceed by first computing all effects as though they were purely huma in origin, by omitting the environmental predictor: this analysis will therefore also contain those effects that are environmental but linked to the human predictor variable. Then we re-do the analysis as though it were all environmental, omitting the human predictors: this contains the effects of human predictors that are linked to the environmental factor. Finally we re-do the analysis with both predictors in, thus measuring the solely human, the solely environmental, and the coupled human-environment effects. Treating the results as a set of simultaneous equations allows one estimate the various components, including the error trm.

Our fourth result provide empirical evidence as to which of several ways of optimizing the fit of a regression tree works best. By testing competing criteria for optimizing regression tree analyses on test data that had known structure we could determine which of the various alternatives did the best job of correctly identifying the structure present. Statistician previously had several possible candidates for the task. We showed that the 1SE rule was the most reliable, thus resolving a source of uncertainty when using regression tree analysis. Regression trees are gainig rapidly in popularity as a tool for environmntal data analysis, so resolving this uncertainty will be of considerable value to the discipline.

We developed an improved method for the scientific visualization of classification and regression tree results that exploits the properties of the color wheel to explicate the relationships among predictor variables in the tree in a spatial map. The idea is simple: cases that differ only in the effect of one variable should be more similar in color on a map of their locations than are cases that differ in more than one variable but perceptual issues make these decisions non-trivial, particularly for thos enot expert in the display of color data. Our findings will make this technique more accessible to such users.

Contributions to Other Disciplines:

Contributions to Human Resource Development:

Both graduates working at the University of Maine successfully obtained positions on leaving the project. Dr John G. Bartlett obtained a permanent position with the USDA Forest Service Southern Global Change Program in Rayleigh, NC, where he focuses on the socioeconomic drivers of environmental change, apparently as a direct result of the specific raining he received while working on the project. Dr Josh Lawler was successful in obtaining a competitive National Research Council Fellowship to work with USEPA in spatial modeling of biodiversity at their Environmental Research Laboratory in Corvallis, OR.

Contributions to Science and Technology Infrastructure:

The databases and methods developed within this project are now being applied in new areas, notably in research into the ecological and economic costs and benefits of carbon sequestration through afforestation. A manuscript has been submitted to Ecological Economics.

Contributions: Beyond Science or Engineering:

The research results on population and environment linkage have been picked up and used by several 'popular' sources. These include the Sierra Club's web site on the effects of population and sprawl; an article in University of Maine Alumnus magazine (Houtman, N. 2001. Understanding human pressures on the environment. Maine 82(1): 8-9); and an article in Forestry Source ('Rural areas are also affected by population growth and development', Forestry Source, June 2001).

Categories for which nothing is reported:

Any Product

Contributions: To Any Other Disciplines

Major Research Activities

Our research activity consisted primarily of conceptual development leading to improved ways of computing statistical effects of interest when modeling human-environment relationships. Activity was undertaken in seven areas:

1) Determining the Accuracy of Current Model Predictions

We tested the accuracy of predictions from our earlier statistical model by comparing our model predictions of species richness against data from the Breeding Bird Census (BBC). The BBC is also a national bird census scheme (performed by effectively a mapping method rather than by the transect-like BBS method). The data - published annually in "American Birds" - have been compiled into digital format by the National Biological Service. We started with the 1990 BBC data, corrected for survey effort, and compared the estimate of species richness on the plot with our model prediction for the hexagon in which the plot was located.

2) Incremental Value of Agricultural Statistics Service Data in Biodiversity Modeling

Earlier models of national species distribution in relation to environmental factors and land cover had used either cropping and related data from the National Agricultural Statistics Service or remotely sensed land cover data derived from the NOAA weather satellite AVHRR signals. We investigated the relative value of these two data sources by allowing them compete with each other as to predictive power.

3) Relative Effectiveness of Incorporating Socioeconomic Data from the Bureau of the Census in the Models

We examined the utility of variables from the Bureau of the Census in determining human-environment interactions. We tested a spectrum of variables chosen to capture key attributes of the human population across the U.S.. We used these either individually or as multivariate complexes both as predictors of bird distribution and as response variables that might reflect environmental correlates generating effective eco-regionalization of the country.

4) Developing Ways of Incorporating the Effects of Spatial Autocorrelation of Predictor Variables and Identifying Relevant Scales

Regression tree models split nodes to get the best deviance partition between sub-nodes but when spatial data are involved, such splitting may be contaminated by spatial autocorrelation of data points, thus changing the statistical significance of results. Our Oregon State University collaborators took the lead in trying to implement the method of Bochard et al. (1992) to partition variance (deviance) into a purely spatial pattern, a purely environmental pattern, and a correlated (inseparable) spatial and environmental covariation of the two.

As an alternative approach to the problem we explored the effects of "pre-whitening" the data before use in regression tree analysis. This ensured that the regression tree generated was not contaminated by spatial autocorrelation but clearly could not yield the sort of regression tree analysis based on spatial coherence we envisaged originally. However, we could use this approach to determine where spatial autocorrelation influenced the results, by generating two regression trees, one with the original data and the other with the pre-whitened data. Since all spatial units fell into a unique end-node in each of the two trees, cross-tabulation of hexagons with respect to which end-node they were in shows which of the original end-nodes were altered by the removal of the spatial autocorrelation.

A related problem needing investigation was how to determine the spatial scale of effects manifest within a regression tree. We used data on the distribution of humans across the U.S. to investigate this, since this dataset provided response variable values for every unit in the country. We investigated the utility of mapping residuals from the regression tree predictions as the basis for scale identification.

5) Adapting Regression Tree Analysis to the Determination of Environmental Stressor Effects

One of the goals of using regression tree analysis in environmental work is to identify the location and magnitude of environmental stressors. A particular problems is that multiple stressors may be at work or may be present only in areas where natural factors - the availability of habitat or the presence of a particular climate space - are also present. Given the nesting of effects inherent in the recursive partitioning of regression trees, we investigated ways to model the separate effects variables with overlap in domain of influence. We also undertook several case studies to demonstrate the variety of patterning of stressor effects.

6) Improving Regression Tree Analysis Methods

Unconstrained fitting of data by regression trees leads to over-fitting. Initially over-fitted trees are typically pruned back to an optimal size by cross-validation to trade variance explained against model complexity but there are differences of opinion among statisticians as to the best criteria to use in cross-validation. Our Oregon State University collaborators undertook a program of evaluating the relative performance of the different criteria in recovering the known structure of artificial datasets. They also asked how the criteria performed in the face of the variability introduced by the internal cross-validation sampling when applied to typical environmental datasets.

The University of Maine team addressed the issue of identifying confounded variables in regression tree analysis. For prediction of response values confounding is unimportant but when regression trees are used to guide a search for causal predictors the identification of confounding variables is critical. In regular regression analysis an initial inspection of the data matrix reveals all relevant cross-correlations among variables but in regression tree analysis the recursive partitioning of the data means that confounding of variables must be investigated repeatedly. Extant methods identify cross-correlated variables at every node split but each choice of one of these variables can change the subsequent regression tree structure, each of which in turn must be investigated for cross-correlation. We explored the possibility that a better computational strategy might be to generate an initial regression tree without regard for cross-correlation and then to re-run the analysis with modified datasets in which the predictor variables that entered that initial tree were perturbed (numerically) to reduce their cross-correlation enough to allow cross-correlated variables to appear. This approach is linear in variables investigated for cross-correlation, against the multiplicative nature of node-by-node consideration. As this approach did emerge as superior, we also determined the level of perturbation needed to control cross-correlation to the equivalent of the correlation of 0.9 or 0.8 typically used as a screening threshold in conventional linear multiple regression.

We also developed the use of this perturbation strategy - but now applied to the response variable - as a means of ensuring stability of regression tree results. This was prompted by the finding of the Oregon team that even the optimal cross-validation criterion could in some circumstances lead to quite variable regression tree structures from the same data. Perturbing the response variable reveals such instability.

7) Development of Visualization Methods Facilitating Interpretation and Communication of Model Outputs

Regression trees have a structure resembling an upturned tree (though always with binary splits) generated by the recursive partitioning of the data. The final end-nodes or "leaves" of the tree contain (between them) all the individual cases in the sample. Where the data are also spatial, the locations of these cases are typically mapped using some form of color or symbol code, typically arbitrarily selected. The Oregon team investigated the value of implementing a scheme for color selection that would allow the maps more readily display the relationships between cases e.g., whether they were sibling nodes, cousins, etc. In particular they explored how a color wheel schema might be weighted to optimize visual display of the recursive relationships otherwise obtainable only from the separate display of the regression tree.

Presentations

Presentations arising from the research were as follows:

Bartlett, J. G., R. J. O'Connor, D. M. Mageean, and S. G. McNulty. 1999. Potential impacts of climate change and residential expansion on water resources in the United States. Annual meeting, Ecological Society of America, Spokane, WA, August 1999.

Bartlett, J. G., R. J. O'Connor, and D. M. Mageean. 1999. Residential expansion as a threat to desert ecosystems in the United States. Session on Spatial Demographics, 30th Annual Meeting of the Southern Demographics Association, San Antonio, TX, October 1999.

Bartlett, J. G., J. Moore, D. M. Mageean, and R. J. O'Connor. 2000. Modeling human-environment interactions at the local scale. Symposium on *The Human Dimensions of a Spatially-Explicit Environment* at US-IALE 15th Annual Symposium "Integrating Societal and Landscape Heterogenity: Problems and Solutions" Fort Lauderdale, FL, 18 April 2000.

Hahn, D. C., and R. J. O'Connor. Scaling of host and environmental patchiness in the distribution of Brown-headed Cowbirds. Conference on Predicting Species Occurrences: issues of scale and accuracy, Snowbird, UT, October 1999.

Hayes, L., and R. J. O'Connor. The continental distribution of habitat-sensitivity in birds. Annual Meeting of the Society for Conservation Biology, University of Maryland, College Park, MD, June 18, 1999.

Lawler, J. J., R. J. O'Connor, and J. G. Bartlett. 2000. Natural and anthropogenic correlates of butterfly and bird species richness. Symposium on *The Human Dimensions of a Spatially-Explicit Environment* at US-IALE 15th Annual Symposium "Integrating Societal and Landscape Heterogenity:Problems and Solutions" Fort Lauderdale, FL, 18 April 2000.

Mageean, D. M., J. G. Bartlett, and R. J. O'Connor. 1998. Indexing the human dimensions of global environmental change. Poster and presentation GCTE-LUCC Open Science Conference on "The Earth's Changing Land", Barcelona, Spain. March 14-18, 1998

Mageean, D. M., J. G. Bartlett, and R. J. O'Connor. 1998. Indexing the human dimensions of global environmental change. Eastern U.S. Forest Modelling and Analysis Workshop, Raleigh, NC. May 13-15, 1998

Mageean, D. M., J. G. Bartlett, and R. J. O'Connor. 1999. Operationalizing the human dimensions of the environment: a case history of continent-wide interactions. Session on Population and the Environment: linking levels of analysis. Annual meeting of the American Sociological Association, Chicago, IL, 7 August 1999.

O'Connor, R. J. 1998. Hierarchical models of human-environment relations. Biodiversity Research Meeting, USEPA Environmental Research Laboratory, Corvallis, OR. May 27-28, 1998.

O'Connor, R. J. 1999. The spatio-temporal structuring of avian distributions. Plenary Speaker, Conference on Predicting Species Occurrences: issues of scale and accuracy, Snowbird, Utah, October 1999.

O'Connor, R. J. 2000. The spatio-temporal structuring of avian distributions. Seminar, Department of Wildlife and Fisheries Sciences, Texas A&M, College Station, TX. April 2000.

O'Connor, R. J., J. G. Bartlett, and D. M. Mageean. 1999. Ecoregions as confluence of ecosystem process, biodiversity, and environmental stressors. British Ecologicl Society Winter Meeting, University of Leicester, England, January 1999.

O'Connor, R. J., J. G. Bartlett, and D. M. Mageean. 1999. Congruence of continental productivity, biodiversity, and anthropogenic stressors under climate and landscape constraints. International Association for Landscape Ecology Congress, Snowmass, CO, August 1999.

Major Research Findings

1) Determining the Accuracy of Current Model Predictions

We found that the predictions of our national model were significantly correlated with the number of species on the BBC plot in that hexagon, provided one controlled for habitat (Wagner 1999). That is, predictions for a heavily forested hexagon were reflected in the number of species present in a woodland plot in that area, but the prediction for a largely agricultural hexagon was naturally different than the number of species present in an isolated woodlot in the area.

2) Incremental Value of Agricultural Statistics Service Data in Biodiversity Modeling

We found that much of the predictive power of the survey-based National Agricultural Statistics Service data in respect of species richness distribution was also present in remotely sensed land cover data derived from AVHRR, but not conversely (Bartlett 1999). Since the AVHRR data are in principle updated annually, the NASS dataset are of less value in modeling changes in the constraints on national biodiversity. For some individual species, however, the NASS data have added value (O'Connor et al. 1999, Hahn and O'Connor in press).

3) Relative Effectiveness of Incorporating Socioeconomic Data from the Bureau of the Census in the Models

Socio-economic data were of little value as simple predictors of the national distribution of species richness. However, when various functions of these variables were used as *respons*e variables, their analysis proved very productive. Among the major findings were that the spread of housing impacting natural resources was not solely in the form of exurban sprawl but had a second, independent dimension of green field building concentrated into coastal and barrier islands and desert edge habitats in which threatened and endangered species were differentially concentrated (Bartlett et al. 2000). We were also able to use the human settlement data to illustrate how to identify different scales within a continental distribution (Bartlett 1999), with one interesting finding being that humans are differentially located along the boundaries between Omernik ecoregions. Finally, we also found that although models of ecosystem functioning, biodiversity, and human density (as index of stressors) can in principle generate independent regionalization of the conterminous U.S., in practice there is substantial congruence of regions in the three models at a level far beyond chance and not explicable in terms of shared driving variables (Bartlett 1999).

4) Developing Ways of Incorporating the Effects of Spatial Autocorrelation of Predictor Variables and Identifying Relevant Scales

We found that the Bocard et al (1992) method of partitioning variance into spatial, environmental, and coupled components was not feasible to implement within normal regression tree calculation (computational costs are excessive). We were successful, though, in adapting the approach to partition variance in a response variable (here butterfly richness in the eastern U.S.) into components due to anthropogenic factors, to natural environmental factors, and to their correlated variation (Symposium presentation by Lawler et al. 2000).

We developed code to implement pre-whitening of data for regression tree analysis, to remove spatial autocorrelation. This work halted with the retirement of our statistician collaborator Dr N. S. Urquhart and will need to be documented more thoroughly than at present..

As part of our investigation as to how to determine the spatial scale of effects manifest within a regression tree, we used data on the distribution of humans across the U.S. to investigate this, since this dataset provided response variable values for every unit in the country. By mapping residuals from the regression tree predictions as the basis for scale identification we found that there is a distinct continent-wide spatial gradient in densities (corrected for regional influences), as well as higher than expected densities of people along eco-region boundaries.

5) Adapting Regression Tree Analysis to the Determination of Environmental Stressor Effects

We found that for several of the neotropical migrant bird species known to be affected by forest fragmentation population losses nationally may exceed 20 per cent, largely concentrated into the southeast Symposium presentation by Hayes and O'Connor 1999). We also showed that habitat fragmentation is an issue for several grassland species and that poor quality of the habitat in a patch requires the patch to be larger than otherwise necessary if it is to be used by the birds (O'Connor et al. 1999). We developed a least squares estimation approach to estimating the individual effects of stressors that overlap others (or natural factors) in domain of influence, and applied it to a case study of the hooded warbler (Hayes and O'Connor submitted).

6) Improving Regression Tree Analysis Methods

We investigated how effective the different cross-validation criteria used in regression tree analysis were in recovering the known parameter structure of artificial data. One of the less frequently used criteria for cross-validation, the one standard error criterion, gave best results (Sifneos et al. submitted). We also found that the internal sampling used in cross-validation could yield far more variability of tree structure for typical environmental datasets than is generally appreciated but that this variation is also minimized by use of the same criterion. We also developed a method of perturbing the response variable data to test whether this problem is likely to be of concern with any given dataset, it being easier to run our perturbation approach in answering this question than it is to undertake the multiple analyses of the target regression tree needed to approach the problem directly.

We developed code to routinely apply a perturbation-based method of identifying confounded variables within a regression tree. This method is quicker than conventional methods of identifying confounding. We also developed criteria as to the intensity of perturbation needed to ensure the equivalent of correlation matrix screening against cross-correlations exceeding 0.8 or 0.9 in conventional multiple linear regression.

These perturbation methods were used in all the regression tree papers published from our work.

7) Development of Visualization Methods Facilitating Interpretation and Communication of Model Outputs

White and Sifneos developed a method to allow use of a color wheel schema for the visual display in a spatial map of the recursive relationships otherwise obtainable only from the separate display of the regression tree (White and Sifneos in press).