Multi-scale Statistical Approach to Critical-area Analysis and Modeling of Watersheds and Landscapes

G. P. Patil[1], G. D. Johnson[1], W. L. Myers [2], and C. Taillie [1]

[1]Center for Statistical Ecology and Environmental Statistics
Department of Statistics
[2]School of Forest Resources and
Environmental Resources Research Institute

Penn State University, University Park, PA 16802 USA

Prepared with partial support from the NSF/EPA Water and Watersheds Program, National Science Foundation Cooperative Agreement Number DEB-9524722. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

Abstract

Environmental and ecological statistics is poised for dramatic growth both for reasons of societal challenge and information technology. It is becoming clear that environmental and ecological statistics is demanding more and more of non-traditional statistical approaches. This is partly because environmental and ecological studies involve space, time and relationships between many variables, and require innovative and cost-effective monitoring, sampling and assessment. Also, environmental and ecological statistics methodology must satisfy environmental policy needs in addition to disciplinary and interdisciplinary environmental and ecological research imperatives. And all of this is true of research and policy issues involving water and watersheds for disciplinary and cross disciplinary research, such as: (i) a need to develop an enhanced predictive understanding of the processes and mechanisms that govern the dynamics and properties of surface and subsurface water and watershed ecosystems; (ii) a need to identify and research indicator variables, analytical methods, and other tools for determining waters and watersheds at risk and reducing the uncertainties of extrapolating information across broad spatial and temporal scales; and (iii) a need to interpret relationships between populations and communities of organisms and the quality and quantity of water, particularly as these relate to ecosystem processes, land use patterns, and landscape structure.
Keywords: Adaptive cluster sampling, Composite sampling, Environmental sampling, Hierarchical modeling, Landscapes, Multiscale assessment, Ranked set sampling, Spatial modeling, Watersheds.

1 Motivation, Issues and Background

1.1 Motivation

Many motivations arise for assessing the relationships between abiotic and biotic characteristics of a watershed or landscape, the concern normally being about biological effects of environmental changes. All inferences about such relationships are not only relative to specified spatial and temporal scales, but also depend on measurement scales used to obtain the necessary data for drawing inferential information.

Most ecological studies have been performed at very localized scales, either for academic research or for highly detailed site investigations. In recent times, there has been a rapid expansion of large regional ecological assessments. This is a result of increased awareness of regional environmental impacts, as well as development of regional measurements from remote sensing technologies and growing ability to manage spatial information through geographic information systems.

A thorough ecological assessment for a given region at a given time would encompass multiple scales. On the one hand, small-scale local investigations may reveal healthy communities, while a regional perspective reveals that habitat is sufficiently fragmented that the localized communities are threatened from lack of migratory routes and insufficient genetic mixing. On the other hand, what may appear to be only subtle changes from a large scale regional perspective can translate to potentially dramatic changes at smaller scale locales. For example, Graham, *et al.*(1991) show through spatial stochastic modeling of the Adirondack region that even though overall forest cover may not change much regionally, small areas of altered forest can have a dramatic effect on water quality in headwater watersheds.

There is a long standing awareness of regional impacts from airborne pollutants such as acidic deposition and ozone (Graham, *et al.*1991). Furthermore, the loss of species diversity through regional habitat modification is considered a top rated risk in the United States (Kiester, et al. 1993). Following a catastrophic event, large-scale regional assessments may need to be made rapidly in order to locate smaller scale areas that are most impacted. Furthermore, nationwide assessments of ecological and environmental resources are under way through the Environmental Monitoring and Assessment Program (EMAP) of the EPA and the Gap Analysis Program of the U. S. Geological Survey's Biological Resources Division. All of these motivations are directed toward an initially broad scale assessment that is used to identify local critical areas for more detailed higher resolution studies. It soon becomes clear that both data sources and inference objectives involve multiple scales, and therefore the relationships among scales must be understood if we are to reliably combine data from different measurement scales as well as make predictions at scales other than those for which we have sufficient data.

Issues: Environmental and ecological data are available at a variety of spatial scales, arising from different sources which range from satellite imagery to field plots. Meanwhile, we need to infer about characteristics and processes at other scales, such as may be delineated by watershed boundaries, political subdivisions, or a predefined unit of fixed size and shape like an EMAP 635 km^2 hexagon.

For example, various indicators may need to be computed for assessing ecological risk relative to a large watershed, followed by assessment for tributary sub watersheds, then for sub tributary sub-sub watersheds, and so on. For the purpose of resource allocation by a federal or state agency, we need summary assessments of large watersheds that correspond to higher order rivers. Within the domain of county planning, assessments may be needed for smaller sub watersheds corresponding to tributaries of the larger rivers. Still, ultimate management and mitigation efforts are applied at a very local scale which requires assessments of fine scale watersheds corresponding to first or second order streams.

A question then arises: Since we simply can not afford to perform a detailed ecological risk assessment for every watershed of every size, to what extent can we infer about the quality of smaller watersheds from composite information about their 'parent' watersheds and, vice versa, how reliably can we infer about a parent watershed from information obtained from a sample of its 'children' watersheds? As an example, if we are concerned with acid deposition impacts within a large watershed, all first order streams may be considered vulnerable due to low annual flows. However, the 'local' spatial distribution of limestone may be so heterogeneous that one headwater stream may have a high acid buffering capacity associated with high alkalinity while a neighboring watershed has very little buffering capacity. To what extent can we predict local variation and decide which first order watersheds should have field survey? Meanwhile, with what accuracy and precision can we compute estimates for larger watersheds based on a sample of their component watersheds?

Much of the source data are not only obtained at different measurement scales, but are also not strictly hierarchical, in that smaller scale sample units may not necessarily be nested exactly within larger scale units. In other words, the boundaries of sample units among the different measurement scales may be non-aligned. Also, the shapes of primary sample units may differ among various measurement scales.

Besides having to address both continuous and discrete response variables, some measurements may be additive as measurement scale increases, such as "number of trees" or "acreage in wetland", while other measurements may be non-additive, such as "species richness".

Of all the metrics that may be used for an ecological assessment, we desire those which not only contain the most information for a given scale of measurement, but also extrapolate with the least uncertainty to other scales. Conventional metrics should be tested and new metrics should be considered. The need for research into new metrics is especially relevant for characterizing landscape structure and pattern in such a way that allows sound, defensible statistical comparisons among landscape units such as watersheds or within a landscape unit over time.

1.2 Background

The issue of measurement scale was addressed early by agronomists, as Smith (1938) derived a regression relationship between the variance of crop yield and the size of quadrats used to sample the crops. This provided an 'empirical' law for relating variability to measurement scale which Smith claimed could be used to decrease the need for pilot studies.

Plant ecologists have also long recognized that measurement scale affects one's observations of species dispersion. In response, Grieg-Smith (1952) introduced an agglomerative approach to analyze contiguous quadrat data using nested analysis of variance. Measurement scale challenges mining engineers and soil scientists also. A need for representative sampling of heterogeneous discrete material prompted Gy (1982) to write a treatise on the subject. Pitard (1989) shows that it is a matter of maintaining a minimum mass or volume of each primary sample unit for a given amount of internal heterogeneity.

Although it seems to be often overlooked, the effect of measurement scale on statistical inference with ecological studies has been rigorously questioned (Schneider, 1994; Weins, 1989). Turner (1994) highlights the importance of defining the scale at which an ecological researcher desires to make inference. As stated by Turner, Dale and Gardner (1989), the primary question now appears to be "What, if any, are the rules of extrapolating across scales?".

When measurement scales are regular and strictly hierarchical, geostatisticians refer to changing scale as the change of support (Isaaks and Srivastava, 1989; Cressie, 1991). For additive variables, when only fine scale data are available for drawing inference based on larger scales, corrections need to be made for the change of support. Isaaks and Srivastava (1989) show two methods for adjusting an estimated distribution to account for the support effect, whereby the mean is unchanged and the variance is adjusted.

Progress has been made through fractal-like interpretations of predictability as a function of sampling resolution (Costanza and Maxwell, 1994). Through analysis of rasterized maps, these authors evaluated the spatial *auto-predictability* which is the reduction in uncertainty about the state of a pixel in a scene, given the states of adjacent pixels in that scene, and spatial *cross-predictability* which is the reduction in uncertainty about the state of a pixel in a scene, given the state of corresponding pixels in other scenes. While *auto-predictability* measures the ability of local data to predict, *cross-predictability* measures the ability of a model to predict. Their results clearly showed that *auto-predictability* increases while *cross-predictability* decreases as the resolution (number of pixels / unit area) increases. Since the predictability changed in a regular way, reflecting self-similarity regardless of scale, a "fractal dimension" was calculable that permits easy conversion of predictability taken at one resolution to other resolutions. The authors conjecture that other spatial measures may exhibit such self-similarity, perhaps leading to a generalized theory of scaling.

Scaling through a fractal dimension can also be used for estimating the spatial measure of an object, such as the length of a geographic boundary or the area of a geographic region. If the relationship between the log of estimated total length or area and the log of measurement scale is linearly decreasing, then an estimate of the slope of this linear trend is an estimate of the fractal dimension of the object in question where fractal dimension increases with spatial complexity (Sugihara and May, 1990;

Maurer, 1994). With such an estimate in hand, the spatial measure of an object can be predicted at scales other than the scale of actual measurement.

When modeling is the objective, Rastetter *et al.*(1992) provide several methods for aggregating fine-scale data for coarser-scale model predictions.

So far, results appear to have been obtained for additive variables where smaller-scale measurement/inference units are regular and hierarchically nested within larger scale units. Irregular, non-aligned multi-scale units present the greatest challenge; however, irregular hierarchically nested multi-scale units, such as the structure of landscape units, have promise of being a tractable challenge.

2 Multi-scale Critical Area Assessment of Watersheds and Landscapes

2.1 Overview

For both hydrological and ecological aspects of the environment, differing sensitivity across landscapes is the norm rather than the exception. Certain localities of watersheds may be highly erodable due to localized factors and/or factor combinations such as steep slopes, soil type, sparse vegetative cover, etc. Particular areas contribute inordinately to nonpoint pollution due to highly generative land-use, impermeable soils, sparse cover, and so on. Wetlands, ephemeral ponds, old growth, corridors, and the like are especially important determinants of habitat suitability in a landscape context. Determination of degree and extent of risk therefore requires spatially specific reconnaissance to assess the need for remedial/preventative programs. When overall programmatic needs for remediation/conservation have been determined, spatially specific evaluation is again required to allocate resources and assistance in cost-effective manner. Spatial coincidence of risk and hazard are crucial in both respects.

Contemporary spatial information technologies such as remote sensing and geographic information systems provide coverage on several determinants of ecosystem function at different scales/resolutions with commensurate costs. Broad-area coverage with limited resolution is now available for most of the more important factors, but acquisition of highly detailed information is much more costly and must necessarily be limited to particular areas. As funding constraints for environmental management efforts become increasingly evident, it becomes correspondingly more important to develop efficient strategies for augmenting coarse scale information with finer-scale data in the more crucial areas. Each level of information must be exploited to predict more accurately the distribution of critical areas for further attention at the next level of detail. Statistical inference provides the only proven approach to such problems of sequential allocation, but has been underutilized for spatial problems and particularly so for situations involving polygonal mappings of categorized environmental variables.

Models are central to modern hydrologic environmental assessment, and may take several forms according to specific purposes (Myers, 1994). Although somewhat less well developed than those for hydrology, models are also of growing importance for ecological environmental assessment. Finite-difference flow and transport models use a rectangular tessellation of space, whereas finite-element models use a triangular network. In either case, tessellations impose parameter lumping over space in accordance with size of spatial partitions (Novotny and Chester, 1989). Coarse tessellations involve greater lumping and produce more generalized views of the modeled process over space. Thus configuration of spatial partitions and consequent spatial generalization of output constitutes a modeling version of scale. Hydrologic models are computationally intensive and data requirements increase exponentially with degree of scale refinement. Practicality thus dictates that scaling of models be commensurate with investigative intent. The more sophisticated models make provisions for varying intensity of partitioning over space. Configuration of the spatial framework for a modeling effort therefore substantially determines its cost and effectiveness.

Whereas the scientific modeling community has been most concerned with the mathematical/mechanistic aspects of models, cost-effective environmental application of models requires that the spatial logistics of modeling be accorded greater attention. The latter is largely the province of statistics. It may be interesting and worthwhile to develop generic multi-phase (multi-scale) statistical strategies for model-based environmental assessment whereby progressively more spatially distributed (and perhaps more mathematically sophisticated) models are exercised at each stage to generate expectations of locally more variable and/or extreme model output at the next stage. Spatial configuration for the next stage of modeling will then be formulated in accordance with expectations from the prior stage. It should be noted that previous modeling experience over a region also becomes an important source of expectation for subsequent modeling work. We therefore have in view a statistical learning strategy for environmental critical-area assessment in a particular region.

2.2 Sources of Geographic Data Coverages

Pennsylvania provides an ideal setting for conducting this kind of research by virtue of the level of maturity in its synoptic spatial data resources and the prior hydrologic modeling work which has been conducted in the context of Chesapeake Bay watershed concerns. The Office for Remote Sensing of Earth Resources (ORSER) in the Environmental Resources Research Institute (ERRI) houses an extensive library of physiographic spatial coverages. ORSER serves as a statewide spatial information repository and analysis center for the Pennsylvania Department of Environmental Protection. Nonpoint-source pollution potential has been modeled for major watersheds in Pennsylvania in conjunction with the multi-state Chesapeake Bay nutrient reduction program (Petersen *et al.*1991). Pennsylvania likewise shares in the MAHA (Mid-Atlantic Highlands Assessment) multi-agency water quality thrust. The Pennsylvania Gap Analysis work is providing synoptic biotic and land-cover information to complement the existing physiographic information base, along with extensive satellite level digital information. The Pennsylvania Gap Analysis effort has likewise generated a new echelon theory of regionalization for spatial variables of ordinal and interval strength (Myers, Patil and Taillie, July 1997).

2.3 Types of Variables

When searching for finer-scale hotspots within larger-scale spatial units, mathematical properties of the variable of concern must be considered. Whether we are measuring discrete or continuous variables, we must consider whether the variable is additive or non-additive. Variables like "acreage of wetland" or "number of individuals in a population" are additive; whereas, variables like "species richness" are non-additive. In the case of additive variables, spatial correlation at one scale can be related to the correlation at both finer and coarser scales because of the bilinearity of the covariance operator. Establishing such relationships is more difficult for non-additive variables and, in general, has to be done empirically. However, there are particular classes of non-additive variables, such as species richness and presence/absence indicators, that obey specific subadditive rules under spatial aggregation; for these, it may be possible to relate spatial dependence at different scales.

2.4 Types of Scaling Hierarchies

Nested regular tessellations occur when smaller-scale units constitute a regular tessellation (systematic grid) whose elements are nested within larger-scale units. Such hierarchies may occur when a grid structure is imposed on a region for the purpose of modeling a spatial process such as sediment transport. Nested irregular tessellations occur when smaller-scale units have irregular shaped boundaries, but are still nested within larger-scale units, which may also have irregular boundaries. Such hierarchies arise naturally for watersheds, and may also occur when spatial units have political boundaries such as for townships within counties. Non-aligned hierarchical tessellations occur when small-scale units are not necessarily nested within larger-scale units but the tesselations can be hierarchically ranked according to scale despite imperfect boundary alignment across scales.

2.5 Prospective Approach

An extensive statewide raster database such as that developed by Petersen *et al.*(1991) at 100-meter resolution, provides a point of departure. Basic parameter information includes factors such as slope and soil which were used in weighted indexing models to generate potential loadings of sediment and nutrients.

A prospective research approach might consist of four major components.

Component A. A multi-scale assessment of spatial dependence structures and relationships among parameters. This is an important empirical component since its outcome determines the appropriate depth of regionalized data augmentation which is anticipated to be parameter-specific. Regionalized variograms and cross-variograms would be the primary assessment vehicle; however, marginal and hierarchical cumulative distribution functions would also be compared at differential scales of resolution for their assessment potential.

Finer scale data acquisition is needed in areas of high local variability, but this is expected to correlate with large values of model parameters (i.e., homogeneous coefficient of variation rather than homogeneous variance). Data augmentation would be then directed toward the extremes (typically, the large values) of process intensity and would therefore be a form of biased sampling. Accordingly, this component would also examine relationships between (i) the extremes of large-scale process intensity and (ii) smaller-scale spatial dependence and cross-dependence.

Component B. Development of large-scale to small-scale prediction algorithms for model parameters: While the predictions are of potential interest as model inputs, the primary application here envisioned is to drive the search algorithms for promising regions of data augmentation. Linear least squares methods would be employed in conjunction with the empirical results of Component A as well as available covariate information. Algorithms of this type have already been developed by Aragon, Patil, and Taillie (1994) in the context of two-way compositing but with two important differences. For two-way compositing, the correlation structure could be derived mathematically instead of obtained empirically. Secondly, the database size was comparatively small. Efficient and parsimonious data handling techniques are therefore needed in relation to the spatial referencing of the data.

Component C. Development of search strategies for localized data augmentation: "Optimal" strategies are considered to be impractical on both computational and theoretical grounds. Corresponding to each of the primary definitions of critical areas, the investigation might identify promising strategies based on the predictions of Component B. These strategies would be evaluated and compared with both real and simulated data acquisition and modeling protocols pertinent to water and watersheds. The evaluation would require the determination of realistic cost functions for information acquisition. An important question that would be addressed is the cost/benefit tradeoff corresponding to enhanced data resolution versus diminishing information returns.

Experience within the context of hotspot identification for composite sample measurements (Gore and Patil, 1994; Gore, Patil, and Taillie, 1995a,b) might offer some insights into developing such strategies.

Component D. Measurement error consequences: Model parameters often are estimated numbers that differ from true values. Adjustments for such errors are available at the prediction stage (non-interpolatory kriging, e.g., Vecchia, 1992). More important is the inflationary impact of error upon the assessment of local spatial variability due to the presence of this additional and short range component of variation. The resulting inflated assessment of variability would suggest that data be collected at finer scale and higher cost than is actually necessary. Sensitivity analyses could ascertain and document the magnitude of this effect. If necessary, simulated errors can be added to model inputs.

The methodology is intended to be applicable to broad classes of models and input variables. Effects of nonlinear models and of nonadditive input variables both need to be examined. Nonadditive variables of particular importance include classification variables, presence/absence variables, and species richness. The latter, for example, is of interest in connection with the Pennsylvania Breeding Bird Survey (proprietary database, Pennsylvania Academy of Natural Sciences; cf. Brauning, 1992).

3 Sample-Based Validation of Spatial Modeling

3.1 Overview

Efficacy of the foregoing ideas for a multi-stage spatial modeling strategy requires observational economy in validating each stage of actual modeling to determine the degree and localization of departure in model projections from field conditions. Without such validation, modeling errors can be propagated and magnified from one stage to the next. Since field measurements are expensive, we need to incorporate the most cost effective sampling methods that are possible. Such methods should exploit any available information about the area that can aid in maximizing sampling efficiency through the balance of cost with necessary precision, while maintaining unbiasedness.

Sampling for validation of spatial models should distribute effort according to apparent localization of criticality in order to detect false positives. It should also, however, provide estimates of extent for false negatives and give indications of probable location for the more substantial instances of critical-area non-detection on the part of the model. This phase of the investigation may be directed at adapting and refining available sampling methodologies of observational economy to serve the needs of validation for spatial modeling in the context of critical area analysis on watersheds at varying scales. The following methods have particular appeal in this validation context.

3.2 Composite Sampling

When risk is characterized by exposure to chemical or pathogenic contaminants, ground truthing may require the measurement of such contaminants in soil, water, air and biological tissue. Analytical chemistry can become the most expensive component of a sampling and analysis plan. Composite sampling offers a way to obtain the needed information while remaining within budgetary constraints. Specifically, compositing consists of physically mixing k sample units, such as soil cores, followed by analysis of the single composite instead of all k individual samples.

Lovison, Gore and Patil (1994) have identified the three primary applications of composite sampling as follows:

1. Estimate a population mean. Here, the increased efficiency that results from compositing allows one either to reduce the number of analytical tests while achieving the same level of precision, or to increase precision with the same number of analytical tests.
2. Classify each of the constituent sample units with respect to being above or below a numerical criterion, or with respect to having the presence or absence of a trait. Here, Johnson and Patil (1994) have found that compositing can be very cost effective, even after including the additional effort and costs associated with compositing and retesting, provided the prevalence (or exceedance probability) is less than 30 percent for most situations or less than 10 percent for all situations studied.
3. Identify those constituent samples that comprise a specified upper percentile of the distribution. Here, a sweep-out retesting protocol has been developed by Gore and Patil (1994) and has been shown to be quite efficient, particularly for situations of high sample skewness or of high sample heterogeneity.

3.3 Ranked Set Sampling

The design of sampling plans for ground truthing should exploit available auxiliary information to enhance sampling efficiency and reduce survey cost. Several such methods are available, including ratio estimation, regression estimation, and stratification (Thompson, 1992). But, these classical methods depend upon assumptions that limit their applicability or their effectiveness.

A method known as Ranked Set Sampling (RSS), originally proposed by McIntyre (1952), and recently reviewed by Patil, Sinha and Taillie (1994), provides a type of double sampling estimator that is extremely flexible and robust in its use of covariate information. While RSS can utilize quantitative covariates, a strength of the method is that covariate information can also be categorical or mixed categorical-quantitative. In addition, judgment and expert opinion can be components of the auxiliary information without introducing bias into the final estimates.

Ranked set sampling works by randomly allocating m^2 prospective sample units into m sets, each of size m. The units are then rank ordered within each set with respect to the perceived relative magnitude of the variable of interest. The lowest ranked sample unit may be chosen from the first set, the second lowest rank chosen from the second set, and so on until the m^{th} ranked unit is chosen from the m^{th} set. This whole process is repeated r times in order to obtain the desired sample size $n = rm$. From measurements obtained in this manner, one can obtain unbiased estimates of the mean and of any other characteristic that is representable as a mathematical expectation.

A generalized form of ranked set sampling allocates different numbers of measurement to different rank orders and can achieve high efficiency for skew distributions while focusing sampling intensity upon targeted portions of the population.

Myers, Johnson and Patil (1995) have suggested that information from sources such as aerial photography, satellite imagery, and other spatially referenced databases could be successfully exploited for improved sample design through ranked set sampling.

3.4 Adaptive Cluster Sampling

Many natural populations are spatially distributed in a clumped fashion, and are not sampled efficiently by conventional methods. Adaptive cluster sampling has been developed to improve sampling efficiency for such situations (Thompson, 1992).

An initial sample is obtained in a manner that may be either probability based or systematic. After obtaining this initial set of measurements, $\{y_i : i = 1, 2, \cdots, n\}$, a chosen criterion is applied to each measurement to determine if neighboring units should be sampled. A common criterion is to include neighboring units if $y_i > c$ for some constant c. The procedure is repeated sequentially until no new units satisfy the criterion. An additional stopping rule is often imposed to keep total sampling intensity within limits.

While the simple arithmetic mean of the initial sample of n units is unbiased, the arithmetic mean of the final sample obtained through adaptive cluster sampling is biased upwards. Thompson (1992) describes two estimators that are unbiased for the population mean or total. One is a modification of the Hansen-Hurwitz estimator when sampling with replacement, and the other is a modification of the Horvitz-Thompson estimator.

One may expect to have fairly extensive information about an area, once it is identified as a critical area, and actually exploit such information, much of which may come from GIS analysis. On one hand, GIS-based information may indicate where the clusters are, thus reducing the need for "adaptive" cluster sampling. On the other hand, even when clusters are identified from an initial sample, the GIS-based information can be used to help select neighbors for subsequent sampling.

3.5 Prospective Approach

Each stage of modeling involves two spatial scales and, for ease of exposition, we refer to the larger scale units as "blue units" and the smaller scale units as "green units." For this stage, the critical area assessment proposed earlier has identified a set of blue units as likely candidates for containing critical areas. Model projections have been obtained for each of the green cells within these candidate blue units. In addition, the spatial predictions from Component B are computable for each green cell, although computer resources will place limits on the number of such predictions. Finally, on the basis of the model projections, the green cells have been partitioned into two regimes: the P-regime (for "positive") consists of those green cells currently declared as hotspots, while the N-regime (for "negative") contains the remaining green cells, declared to be not hotspots.

The false positive rate for the hotspot declarations is the conditional probability

$$\Pr(\text{ green unit is not a hotspot} \mid \text{P-regime}) , \tag{1}$$

and, similarly, the false negative rate is

$$\Pr(\text{ green unit is a hotspot} \mid \text{N-regime}) . \tag{2}$$

These rates are to be estimated from field measurements, for which an appropriate sampling protocol involves two steps. The first step is the selection of green cells from each of the two regimes for field measurement, and the second step is the actual field measurements within the selected green cells. In most instances, the field measurements yield an *estimate* of the true value, i.e., ground "truthing" is really ground "estimating." An inaccurate ground estimate can produce a classification error in determining the true hot-spot status for the green cell and induce a corresponding inaccuracy in determining the conditional probability given by expression (1) or (2). Efficient sample design for step 2 is intended to yield ground estimates that are close to ground truth and reduce the foregoing source of inaccuracy. Any of the sampling designs described earlier in this section are potentially available for achieving observational economy for step 2. However, there is a point of diminishing returns at which the above classification error becomes negligibly small.

Precise determination of the two error rates is largely controlled by the design of step 1. Efficiency for this step can reduce the number of green cells needing field measurement and hold down overall costs. Each of the three sampling methods (compositing, ranked set sampling, and adaptive cluster sampling) described earlier might be evaluated for their effectiveness in step 1 sampling. Ranked set sampling, particularly with unequal allocations, appears particularly promising since model projections and spatial predictions are both available for ranking purposes. Adaptive sampling could be useful in delineating regions with high error rates. The use of compositing is more problematic since it would involve aggregating green cells. There is an extensive literature on estimating individual sample proportions from composite measurements. However, these methods deal with true categorical responses (e.g., presence/absence rather than dichotomized continuous responses) and assume that the individual sample responses are independent rather than spatially correlated.

The preceding approach does not make full use of available spatial information. One might also examine the possibility of a Bayesian approach, treating the model projections and spatial predictions with suitable error distributions, as the prior. After conditioning on the field measurements, the posterior provides an area-wide estimate of ground truth that is fairly accurate in neighborhoods of the field measurements and less so in more distant regions. The posterior allows assessment of the *magnitude* of error in model projections as well as delineating probable locations of false positives and of false negatives.

4 Hierarchies in Watershed Assessment

4.1 Overview

The spatial-organizational views of watersheds for purposes of management, analysis, and modeling have been substantially simplistic and subjective. The lineage of such views arises from manual/analog technologies of mapping and airphoto interpretation which prevailed in prior decades. Digital successors to the manual/analog technologies are now well established, but unfortunately have been incarnated essentially as computer-assisted mapping systems to facilitate generation of conventional information structures that have gained little in sophistication relative to their predecessors. The organizational depth of informational representations of watersheds in contemporary (GIS) is still determined primarily by the visual spatial insights of human analysts and rendered as digital layer equivalents of conventional maps. Such layers are organizationally flat despite their derivation from complex surface data and intrinsically hierarchical environmental factor influences.

Simplistic views geared to human visual comprehension cannot add substantially to the efficacy of analysis just by virtue of being computer-facilitated. Generalized delineations of watersheds and/or hydrologic units according to expert intuitive judgment are essentially informational dead ends. Aggregate characterizations of internal variability for heuristically defined units can be compiled, but such compilations only indicate possibilities for improvement – not how optimal organization at a given level of detail can be progressively achieved in working across a spectrum of scales.

A future capability component the prospective approach might look toward intelligent development of information structures from digital data pertaining to watershed processes in a progressive sequence of generalization that optimizes information value of representations at successive levels of detail. This may be motivated by the nature of watersheds as gravity-induced tree structures which may or may not be self-similar, but it also recognizes that factor influences for management/remediation may span watershed boundaries at multiple hierarchical levels. The hierarchical information structures thus generated are conducive to formulation and application of tree metrics as generalized computer analogs of classical photointerpretive recognition and differentiation of drainage patterns.

This approach has a foundation of exploratory theoretical investigation in the Environmental Resources Research Institute at Penn State University under the USGS state water center allotment program, wherein it was pursued under the acronym ESCALATR (Expressing Space Complexes As Logically Abstracted Thematic Regions). Arising from that research is an object-oriented concept of meta-networks as nodal linkages jointly spanning complexes in both coordinate and relational spaces (Myers, 1993). The full concept encompassing parallel computing capability of nodal objects with methods is still beyond the horizon of even working prototype software. However, realization of the structuring capability in meta-networks has recently been determined to be within the scope of suitably enhanced geographic information systems. Future research may build on this initial effort in terms of both software and statistical theory. The goal will be to achieve capability for computer-generated hierarchical representations of watersheds, while also exploring statistical theory that would support comparative analysis of the resulting structures as stochastic networks in the graph theoretic domain.

References

[1] Aragon, M. E. D., Patil, G. P., and Taillie, C. (1994). Use of best linear unbiased prediction for hotspot identification in two-way compositing. Technical Report Number 94-0416, Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University.

[2] Brauning, D. W. (1992). Atlas of Breeding Birds in Pennsylvania. University of Pittsburgh Press, Pittsburgh. 484 pp.

[3] Costanza, R. and Maxwell, T. (1994). Resolution and predictability: an approach to the scaling problem. *Landscape Ecol.*, **9**, 47–57.

[4] Cressie, N. (1991). *Statistics for Spatial Data*. Wiley, New York. 900 pp.

[5] Gore, S. D. and Patil, G. P. (1994). Identifying extremely large values using composite sample data. *Environmental and Ecological Statistics*, **1**, 227–245.

[6] Gore, S. D., Patil, G. P., and Taillie, C. (1995a). Identification of the largest individual sample values using composite sample data and certain modifications of the sweep-out method. Technical Report Number 94-0701, Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University.

[7] Gore, S. D., Patil, G. P., and Taillie, C. (1995b). Identifying the largest individual sample value from a two-way composite sampling design. Technical Report Number 95-0102, Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University.

[8] Graham, R. L., Hunsaker, C. T., O'Neill, R. V. and Jackson, B. L. (1991). Ecological risk assessment at the regional scale. *Ecological Applications*, **1**, 196–206.

[9] Grieg-Smith, P. (1952). The use of random and continuous quadrats in the study of the structure of plant communities. *Annals of Botany*, **16**, 293–316.

[10] Gy, P. M. (1982). Sampling of Particulate Materials–Theory and Practice. Second edition. Elsevier, Amsterdam.

[11] Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, New York. 561 pp.

[12] Johnson, G. D. and Patil, G. P.. (1994). Cost Analysis of Composite Sampling for Classification. Technical Report Number 95-0202, Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University.

[13] Kiester, A. R., *et al.*(1993). Research Plan of the Biodiversity Research Consortium.

[14] Lovison, G., Gore, S. D. and Patil, G. P. (1994). Design and analysis of composite sampling procedures: a review. In *Handbook of Statistics, Volume 12: Environmental Statistics*, G. P. Patil and C. R. Rao, eds. North-Holland, Amsterdam. pp. 103-166.

[15] Maurer, B. A. (1994). *Geographical Population Analysis: Tools for the Analysis of Biodiversity*. Blackwell Scientific Publications. 130 pp.

[16] McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*, **3**, 385–390.

[17] Myers, W. L. (1993). A meta-network approach to higher-order spatial constructs and scale effects. Chapter 18. In *Multivariate Environmental Statistics*, G. P. Patil and C. R. Rao, eds. North Holland.

[18] Myers, W. L. (1994). Environmental remote sensing and geographic information systems-based modeling. In *Handbook of Statistics, Volume 12: Environmental Statistics*, G. P. Patil and C. R. Rao, eds. North Holland, Amsterdam. pp. 615-642.

[19] Myers, W. L., Johnson, G. D., and Patil, G. P. (1995). Rapid mobilization of spatial/temporal information in the context of natural catastrophes. *COENOSES*, **10**, 89-94.

[20] Myers, W. L., Patil, G. P. and Joly, K. (1997). Echelon approach to areas of concern in synoptic regional monitoring. *Environmental and Ecological Statistics*, **4**, 131–152.

[21] Novotny, V. and Chester, G. (1989). Delivery of sediment and pollutants from nonpoint sources: a water quality perspective. *Soil Water Cons.*, **44**, 568–576.

[22] Patil, G. P., Sinha, A. K., and Taillie, C. (1994). Ranked set sampling. In *Handbook of Statistics, Volume 12: Environmental Statistics*, G. P. Patil and C. R. Rao, eds. North-Holland, Amsterdam. pp. 167–200.

[23] Petersen, G., Hamlett, J., Baumer, G., Miller, D., Day, R. and Russo, J. (1991). Evaluation of agricultural nonpoint pollution potential in Pennsylvania using a geographic information system. Report ER-9105, Environmental Resources Research Institute, The Pennsylvania State University, University Park, PA.

[24] Pitard, F. (1993). Pierre Gy's Sampling Theory and Sampling Practice. Second edition. CRC Press, Ann Arbor. 488 pp.

[25] Rastetter, E. B., King, A. W., Cosby, B. J., Hornberger, G. M., O'Neill, R. V. and Hobbie, J. E. (1992). Aggegating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems. *Ecological Applications*, **2**, 55–70.

[26] Schneider, D. C. (1994). *Quantitative Ecology: Spatial and Temporal Scaling.* Academic Press, San Diego. 395 pp.

[27] Smith, H. F. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Science*, **28**, 1–23.

[28] Sugihara, G. and May, R. (1990). Applications of fractals in ecology. *Trends in Research in Ecology and Evolution*, **5**, 79–87.

[29] Thompson, S. K. (1992). *Sampling.* John Wiley and Sons Inc., New York. 343 pp.

[30] Turner, M. G., Dale, V. H. and Gardner, R. H. (1989). Predicting across scales:theory development and testing. *Landscape Ecology*, **3**, 245–252.

[31] Turner, S. J. (1994). Scale, observation and measurement: critical choices for biodiversity research. (Paper presented at: Measuring and Monitoring Biodiversity in Tropical and Temperate Forests, August 28, 1994. submitted for jury review for publication of conference proceedings.)

[32] Vecchia, A. V. (1992). A new method of prediction for spatial regression models with correlated errors. *Journal of the Royal Statistical Society, Series B,* **54**, 813–830.

[33] Wiens, J. A. (1989). Spatial scaling in ecology. *Functional Ecology*, **3**, 385–397.