Finding Upper-Level Sets in Cellular Surface Data Using Echelons and SaTScan

MYERS, Wayne L.
Environmental Resources Research Institute, The Pennsylvania State University, University Park, PA 16802, USA, wlm@psu.edu
KURIHARA, Koji
Faculty of Environmental Science and Technology, Okayama University, 3-1-1 Tsushima-naka Okayama 700-8530, Japan, kurihara@ems.okayama-u.ac.jp
PATIL, Ganapati P.
Department of Statistics, Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University, University Park, PA 16802 USA, gpp@stat.psu.edu
VRANEY, Ryan
Department of Statistics, Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University, University Park, PA 16802 USA, rvraney@stat.psu.edu

ABSTRACT

Across a spectrum of contemporary contexts from public health to landscape ecology and natural resources, there is need for objective determination of elevated occurrence in phenomena such as disease incidence and biodiversity. Occurrences of such phenomena constitute response surfaces, but data regarding the surface is typically acquired in a cellular framework. The cells may comprise a regular grid, or may be of irregular shapes such as counties in which statistics are collected. Echelons are a topologically based approach to systematic determination of spatial structure in a step surface. Spatial scan statistics are a probability-based approach to the same issue when interest lies in a rate variable. Here we examine the use of echelons both separately and in conjunction with the SaTScan implementation of spatial scan statistics for purposes of determination and visualization of upper-level sets. Consideration is given to both conventional geographic space and to the cellular pseudo-space of contingency tables for ordered categorical variables.

1. Introduction

Objective determination of spatial structure is key to regional assessment for ecological and environmental condition in terms of quantitative measures and numerically scaled indicators based on remote sensing and Geographic Information Systems (GIS). We focus attention on spatial data having the information as values of surface variables pertaining to specified locations. Geospatial cellular data are synoptic observations covering the entirety of a partitioned region, like cancer rates corresponding to each county in a state. Data can be collected directly within each region, and thereby ascribed to a particular region. In a more abstract sense, algorithmic and statistical interpolations can impart a synoptic character to point data by imputing values for reference points such as centroids in respective partitions. In a still more abstract but useful way, cellular arrays such as contingency tables can be considered as pseudo-spaces and the numerical values therein as forming step surfaces of generalized response.

Investigation would often benefit by building of a model to clarify the structure of geospatial cellular data. Conventional geographic information systems are powerful analytical platforms, but do not provide directly for describing spatial structure of cellular data. Likewise, statistically mechanistic spatial scan methods (Kulldorff, 1997) draw attention to the uppermost sector(s) of a surface, but do not fully exploit topology in revealing progressive slopes and bridging between locally (but not necessarily globally) distinctive features. Echelons (Myers et al., 1997; Myers and Patil, 2002) are useful constructs for studying the topological structure of a surface in systematic and objective fashion that can augment and enhance the effectiveness of GIS, SaTScan, statistical visualization, and even analysis of contingency tables. Echelon analysis has broad utility in prospecting for areas of interest in regional monitoring of surface variables.

Echelons are based on areas of relative high and low values for response variables of spatial data. The echelon approach aggregates the surface elements over areas having the same topological structure, and specifies hierarchically related organization of these areas. Echelons are derived from occurrence of peaks, slopes, and saddles as topological elements when a surface is scanned progressively from uppermost to lowermost levels. Saddles are critical structures that distinguish echelons in relation to each other. A saddle delimits overlying echelons from an underlying foundation echelon. Foundations likewise have foundations in a hierarchical recursion that lends itself to expression as a dendrogram. Each echelon component comprises a particular subset of the surface that can be mapped in topview and characterized with regard to a suite of both conformational and relational attributes.

Several investigations of spatial structure in data have been based on echelon analysis. One such area of application is change detection from repeated coverage of remotely sensed multi-spectral image data (Myers et al., 1999; Smits et al., 2000). Another context is the relationship between density of human residence and sparseness of vegetation computed from remote sensing data (Kurihara et al. 2000). Regional features such as hotspots (Patil et al., 2000; Patil et al., 2003) and trends are shown in echelon maps and detailed in dendrograms and tables. Echelons complement kriging and related spatial statistics by providing a unified framework for handling quantitative and estimated variables in cellular formats that lends itself naturally to data mining. In this paper, we explore echelon analysis for different types of environmental information and indicators of status or change. We illustrate capabilities for identifying regional features in both automation and visualization modes.

2. Formation of echelons

Conceptualization of echelons is facilitated by first considering a one-dimensional case, and then progressing to mappings in two dimensions. Intuitive understanding can be promoted by imagining the surface as being inundated by water that progressively drains away, and visualizing how locally elevated portions of the surface will emerge and grow as islands that eventually join to form larger compound islands.

2.1 One-dimensional schematic

A simple case is the hypothetical set of hillforms for one horizontal dimension shown in Figure 1. The horizontal axis shows the position (x) of spatial data, and the vertical axis shows the value (*h*) or height of the response variable for the specified horizontal position. Such a profile would correspond to a cross-sectional view for a topographical map. The nine numbered topological components of structure in the figure are called echelons. These echelons consist of peaks, foundations of peaks, foundations of foundations, and so on recursively. Numbering begins with the peaks (1 … 5) in order of summit height, which are termed *first-order* echelons. Numbering then proceeds with foundations, again in order of top (saddle) height. A foundation supporting simple peaks is of second order, and a foundation supporting both first and second orders remains also of second order. A foundation supporting two (or more) second orders increases to third order. Thus, foundation order increases at a 'crotch' supporting like orders. The echelons supported by a foundation are its *ascendants*. These genealogical foundation relationships give rise to a 'family tree' dendrogram as shown in Figure 2, in which the letters A, B. and C are used to show first, second, and third order echelons, respectively.

2.2 Two-dimensional extents for surfaces

A topographical map is a classic example of spatial data having the value (h) as the 'level' of a response variable at the two-dimensional position (x, y). Such a topographical map generally represents a continuous function $h=f(x,y)$, as illustrated by contour lines in Figure 3. This constitutes the topview of a structure to which the dendrogram of Figure 2 also pertains. The topographical map typifies a smooth response function defined over a differentiable manifold as studied by Morse (1934) in terms of higher-order connectivity for upper-level sets as the level decreases (Milnor, 1963; Matsumoto, 2002). Remote sensing intensity values for image data and digital elevation models (DEM) for GIS are conventionally recorded on a grid mesh (raster) of cells or 'pixels' over the n-by-m array area $D_{ij}= \{(x,y)|x_{i-1}<x<x_i , y_{j-1}<y<y_j\}$, $i=1,2,\ldots,n$, $j=1,2,\ldots,m$. For such meshed data, the function $h=f(i,j)$ is fairly complicated and is best approached algorithmically in terms of sets.

The emergence of islands relative to descent of a virtual water level motivates formalization as *insular sets*, with an insular set constituting a conceptual island. A *capping set* is a simple insular set. Capping sets are joined by a *bridging set* to form a *compound insular set*. Recursive compounding takes place as compound insular sets are joined together by bridging sets.

Let a *flat* be a set of contiguous cells having the same level, where cells are contiguous if they touch in any manner. An *insular set* (I_j) is a contiguous set in which all edge elements (cells or tiles) of the set are higher than any of the border elements external to the set, and the highest (external) border element has or is a member of a flat having at least one higher external neighbor that is not a neighbor of the set. Such a set is an island in the intuitive sense at the submersion level of its highest external border element, although the island may contain holes (lakes). Note that elements of 'lakes' are also considered external to the set. At any lower submersion level, the contiguity would abruptly encompass (an) entire external insular set(s). In other words, this is the lowest level of island expansion as opposed to island merger. For logical consistency, the full extent of the surface is also defined to be an insular set. Starting with any contiguous set of tiles, the encompassing insular set can be obtained by successively adding the highest neighboring element until the insularity conditions are encountered.

An insular set may or may not be compound by virtue of containing two or more internal insular subsets. A compound insular set contains at least two insular subsets and at least one *bridging set* (B_l). An insular set undergoes a first stage of decomposition by successively removing the lowest element until there is segregation of insular subsets. The subset of elements so removed is a bridging set. The remaining insular sets may or may not be compound. If a current stage decomposition process of removal yields a null set, then that insular set is a *capping set* (K_p) that constitutes a simple island or local peak.

In terms of insular sets, capping sets are first-order echelons. Bridging sets are higher-order echelons. An insular set is a *family* of echelons corresponding to a sub-tree in the echelon dendrogram. Higher-order echelons are, therefore, connectivity sets or natural ‘corridors’ of ‘communication’ between higher-level echelons. Note carefully that an insular set or family of echelons is always spatially contiguous, but that a bridging set or higher-order echelon is not necessarily so by itself. It should also be noted that this view of insularity does not directly distinguish ‘cupping sets’ as hollow components of a surface. Cupping sets can, however, be determined as capping sets of the complementary surface.

With reference to Figures 1-3, components 1-5 are capping sets and thus first-order echelons. Components 6-9 in these figures are bridging sets or higher-order echelons. The contiguous character of echelons 6-9 in Figure 3 is not a general property of higher-order echelons for cellular surfaces.

3. Echelons of contingency tables

Relatively small cellular arrays of data are convenient for initial acquaintance with echelon application. Although they constitute a pseudo-space rather than a geo-space, contingency tables have small array size and can offer opportunities for innovative use of echelons in a statistical context.

The cross-tabulation of factors shown in Table 1 is taken from a thesis by Joly (1996) investigating mammalian biodiversity in Pennsylvania. The geospatial context for this work involved compilation of data on habitat and species occurrence in hexagonal sub-regions encompassing 635 km^2 each, with 211 such hexagons comprising the Commonwealth of Pennsylvania. One of the habitat factors considered in the study was forest cover within the hexagon. Influence of forest cover was studied in five ordered categories as follows: 1) dense = 95% or more; 2) moderately dense = 75% to 95%; 3) moderate = 55% to 75%; 4) moderately sparse = 36% to 55%; 5) sparse = less than 36%. Testing by chi-square for the simple null hypothesis of independence between broad-scale mammalian species richness and forest cover given the observed marginal occurrences showed significance ($P<0.001$). Joly then proceeded interpretively to consider the apparent pattern of interaction, concluding that there is positive association between increasing forest cover and increasing mammalian biodiversity. Here, however, we go beyond subjective interpretation to investigate the pattern of interaction more objectively using echelons in conjunction with scan statistics (Kulldorff, 1997).

In using echelons to study the co-occurrence of mammal species and forest cover, we will treat column numbers as being X-coordinates and row numbers as being Y-coordinates. One could proceed to determine echelons directly for the occurrence frequencies as given in the body of the table, but this would address overall pattern of co-occurrence rather interaction relative to observed marginal occurrences. To investigate patterns of interaction, there is need to tabulate rate of occurrence relative to expectation on the basis of marginals. Accordingly, Table 2 is a recasting of Table 1 with ratios of observed to expected shown in the body of the table and salient echelon structure marked in different brackets.

Since interest lies in elevated rates of occurrence relative to expectation, the upper echelons are highlighted in Table 2. The peaks as first-order echelons or capping sets are marked by angle brackets as <>. It can be seen that the pattern is somewhat noisy in this regard, with several single-cell peaks scattered across the table. Major features of the pattern emerge when high-level bridges connecting peaks are designated in Table 2 by {} brackets. These high-level bridges are second order echelons with only first-order ascendants, thus being outer branches of the tree analogous to twigs. Two major features become apparent that support the interpretation of positive association for the two ordered factors. One feature represents substantial occurrence of moderately low diversity with lesser forest cover. The other shows substantial occurrence of moderately high diversity with greater forest cover. It is notable that designation of third-order, mid-level bridges marked as () serves to combine these two major features into one elongate feature that spans the table in a slant manner. There is little consistency for very low species richness where occurrences are rare. There are no instances of very high diversity with very low forest cover. It is perhaps worth noting that the mapping of forest cover also had relatively low resolution that ignored imbedded openings less than 1 km^2 in size. Thus, even the ‘dense’ forest cover could contain sufficient openings to meet the habitat needs of most ‘edge’ species.

Echelons offer an objective structural view of a cellular surface variable that helps to elucidate possible causal phenomena, but the echelons by themselves do not provide a statistical basis in probability for declaring that any particular patch of cells is significantly elevated relative to the general population of cells. The spatial scan statistic (Kulldorf, 1997) was developed as a surveillance tool to detect and test the significance of local cell clusters with regard to rate of disease incidence. The spatial scan statistic does not entail prior assumptions about the location or size of clusters, and it adjusts the inference for multiple testing inherent in the multiplicity of potential size and locations it considers. Conceptually it is defined in terms of a scanning window of any shape, but implementation in SaTScan software developed at the National Cancer Institute has been done in terms of circular zones (Kulldorff et al., 1998). The scanning window roves over the map region in a manner that centers it at cell centroids. The window size is varied at each location under the constraint that it not encompass more than half the population of cells. For each circular window, the observed and expected number of cases inside and outside is determined and the corresponding likelihood calculated. This likelihood is maximized over all circles to identify the particular window that indicates the greatest likelihood of being elevated. The likelihood ratio for this zone of maximum likelihood is the test statistic. A Monte Carlo simulation of 9999 random replications of the dataset is used to obtain distribution under the null hypothesis and corresponding *p*-value. Two distributional contexts are commonly used for SaTScan purposes: binomial and Poisson.

We use the Poisson model for the contingency table data, with the ratio of observed to expected being termed the *relative risk*. SaTScan identified a primary cluster of 7 cells having p = 0.001, and a secondary cluster of 11 cells also having p = 0.001. The primary cluster is designated in Table 2 by a pair of leading plus signs, and the secondary cluster is designated by a pair of leading minus signs. The primary SaTScan cluster substantially overlaps the echelon formation that is less heavily forested and less diverse. The secondary SaTScan cluster partially conforms to the echelon formation that is more heavily forested and more diverse. Thus, the SaTScan analysis provides statistical validation to the insights gained from the echelon analysis. However, in and of themselves, the SaTScan clusters would not elucidate the pattern with the clarity of the echelons.

The important thing here is that echelons and SaTScan used in combination provide much better insights regarding the data than would either method alone. Echelons are statistically inconclusive due to lack of a probability basis for determining significance. SaTScan is rather severely constrained for interpretive purposes by its restriction to considering circular zones that does not take full account of topological structure. The latter limitation can be appreciated more fully by examining the primary and secondary SaTScan clusters in relation to the surrounding echelon structure. There are elements having low relative risk in both clusters that can be substituted by adjacent elements having higher relative risk in a manner that unifies the primary and secondary SaTScan clusters so that the shape is in substantial accordance with the echelon form. The shape of the unified cluster is, however, decidedly non-circular. Because of the non-circularity, SaTScan would never have located a single cluster of this nature.

4. Echelon-Refined Hotspot Surveillance

The geometrically mechanistic and statistically incisive SaTScan is the contemporary exploratory tool of choice for detecting geographic clusters of disease and other incidence rates that are significantly elevated with respect to the regional setting. However, the foregoing study of a contingency table context serves to illustrate etiological advantage in topologically sensitive approaches to cellular surface information. An obvious strategy, then, is to combine the two approaches in ways that are analytically expedient.

One strategy that is immediately available uses echelon-based inspection for topological modification of SaTScan-based clusters, without attempting to determine the increased significance due to the spatial reconfiguration. This strategy is basically one of substitution in the SaTScan-derived cluster. The highest external border element is progressively substituted for the lowest internal border element of the SaTScan cluster. This opportunity arises by way of the compromise that SaTScan makes in expanding its circle to pick up high-level elements while accepting some dilution by also including lower-level elements. If the inherent topology of the cluster is non-circular, there will be opportunity for some such substitution in the SaTScan result while retaining the ability to declare significance with the SaTScan probability as a lower bound. The restrictions on substitution are first that an interior border element be traded for an exterior border element, and secondly that the exterior trades not be sparsely populated elements. The first restriction implies that the reconfigured cluster will have a different shape but the same number of elements (size) as the SaTScan cluster. This restriction arises from the inherent interior versus exterior nature of the SaTScan comparison. The logic of the second restriction is evident from the contingency table context where elements having very low expectation tend to induce extreme variability. Such elements occur in the corners of Table 1. This is also the reason for the usual advice to collapse cells having low expectations in usual contingency table analyses. There is, of course, no reason not to go beyond the ‘significant’ size in a speculative mode for the purpose of trying to better understand possible relationships of the cluster to other environmental and/or demographic features.

This strategy can be examined in terms of the SIDS mortality data for North Carolina counties during 1974-84 as analyzed by Kulldorff (1997) using the spatial scan statistic. Figure 4 shows a map of North Carolina with the numbers in the counties being SIDS incidence rates per 1000 births. The primary and secondary hotspots of incidence as determined by SaTScan are highlighted by heavy boundary lines and associated labels. The primary hotspot is seen to be compact and well distinguished by SaTScan with no opportunities for substitution. The secondary hotspot, however, has a problematic configuration with respect to neighboring Bertie County. Bertie County having an incidence rate of 37.5 is not included by SaTScan, whereas Hertford County with an incidence rate of 36.5 is included. In terms of cases, Hertford County with 12 does have one more than Bertie County with 11. One the basis of incidence rates, then, a substitution of Bertie County for Hertford County would be indicated. Since the two counties are so close with respect to SIDS and the secondary hotspot is smaller than the primary, the apparent conclusion is that inclusion of both counties would not lower the probability.

Although reconfiguration of SaTScan clusters by inspection is not difficult, it would be preferable to move toward automation in this regard while retaining specificity of probabilities. A first move in this direction might be to augment the SaTScan software so that a supplemental boundary scan is conducted after determining the primary and secondary hotspots as usual. This scan could determine the greatest of the nearest neighbor distances for centroids in a cluster. This distance would then be used around the centroid of each cell in the cluster to find neighbors not presently included as prospects for augmenting the cluster and reducing the rigidity of the shape constraint.

A third combinatorial strategy would obtain the best of both approaches with topologically intelligent clusters and exact probabilities, but entails longer-term developmental efforts with respect to both SaTScan and echelon software. This would involve an echelon-based progression in selection of candidate zones for assessment in the Monte Carlo mode of SaTScan. The current rendition of echelon software operates only on rectangular or hexagonal grids of cells. Quantities ascribed to irregular cellular areas can be analyzed, but only by first ‘rasterizing’ with GIS software. The identity of the original irregular cells is lost in the rasterization process. To recover the identity of the original irregular cells in terms of echelons would require ‘vectorization’ of the echelon raster, but programming of a data format conversion would be necessary to enable this. After vectorization, a new echelon software module would be needed for structuring the tree in terms of irregular units and indexing the elements within echelon families on the basis of surface level. Due to the requirement of contiguity, the echelon progression of candidate areas would necessarily be based on echelon families as insular sets. The SaTScan software would then need to be substantially modified to use the file of ordered elements in echelon families instead of centroid-based scanning with variable-radius circles. All of this is non-trivial undertaking.

5. Purposive visualization for upper-level echelons

There are a multitude of circumstances for environmental analysis in which it is appropriate to define upper-level sets purposively rather than in terms of probability. Current echelon software is configured to support innovative visualization when upper-level sets are defined in terms of range of surface response. In this sense, an upper-level echelon is defined as one having its highest element (peak or saddle level) in a specified top percentage of the range. For example, echelons might be considered as being upper-level if their top reaches into the upper 25% of the range of surface response. The innovative aspect of echelon-based visualization is that progressive shading can be used to portray descent of upper-echelons to a specified lower limit stated again as a percentage of the range. Consider, for instance, questions of distribution of high species richness across an area and the manner in which there is progressive as opposed to precipitous decline as one moves away from hotspots. This could be addressed by specifying portrayal only of echelons that penetrate the upper 25% of the range, but that elements of such echelons be progressively shaded with respect to decline in level until the midrange (upper 50%) is reached. Note that echelons that reach above 50% but not to 75% would then not appear in the display at all. This is very different from setting conventional thresholds at 50% or 75%. It focuses attention on upper-level structures not only in terms of occurrence, but also connectivity or bridging that occurs at lower but still elevated levels.

These displays are given for richness of habitat suitability relative to number of species in Pennsylvania as Figure 5 for mammals and Figure 6 for amphibians. It is readily seen from such visualization that habitat suitability for large numbers of mammal species is much more pervasive in Pennsylvania than it is for amphibians. Likewise, connectivity is better and decline more gradual for mammals than for amphibians. This is due both to the predominance of upland environments in Pennsylvania and also to the greater mobility and habitat latitude of many mammals as compared to amphibians.

A further advantage of visualization with purposively determined upper echelons is that much larger datasets can be handled through echelons in conjunction with geographic information systems than could be practically analyzed via the SaTScan approach. This is true even though extraction of echelons is a multi-step procedure involving extended computer processing times. Echelon extraction could, however, be made faster by several orders of magnitude though recasting of the algorithms to be more memory intensive rather than disk intensive. Given the rate of evolution in computer capacity and lower cost, such recasting is becoming ever more attractive. It should also be noted that rendition of different visualizations is quite rapid when the one-time task of echelon extraction has been accomplished.

6. References

Joly, K. 1996. Mammalian biodiversity in Pennsylvania at the USEPA 635 square kilometer hexagonal scale. Master of Science thesis in Ecology, The Pennsylvania State University, Univ. Park, PA 16802.

Kulldorff, M. 1997. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26:1481-1496.

Kulldorff, M., K. Rand, G. Gherman, G. Williams and D. Defrancesco. 1998. SaTScan v2.1: Software for the spatial and space-time scan statistics. Bethesda, MD: National Cancer Institute.

Kurihara, K., W. Myers and G. P. Patil. 2000. Echelon analysis of the relationship between population and land cover pattern based on remote sensing data. *Community Ecology*, 1(1):103-122.

Matsumoto, Y. 2002. An introduction to Morse theory. *Translations of Mathematical Monographs,* Vol. 208. Providence, RI: American Mathematical Society.

Milnor, J. 1963. Morse theory. Princeton, NJ: Princeton University Press.

Morse, M. 1934. The calculus of variations in the large. *Colloquium Publications*, Vol 18. New York: American Mathematical Society.

Myers, W. and G. P. Patil. 2002. Echelon analysis. Pp. 583-586 in *Encyclopedia of Environmetrics,* Volume 2. A. El-Shaarawi and W. W. Piegorsch, eds. UK: John Wiley & Sons.

Myers, W., G. P. Patil and K. Joly. 1997. Echelon approach to areas of concern in synoptic regional monitoring. *Environmental and Ecological Statistics*, 4:131-152.

Myers, W., G. P. Patil and C. Taillie. 1999. Conceptualizing pattern analysis of spectral change relative to ecosystem status. *Ecosystem Health*, 5:285-293.

Patil, G. P., J. Bishop, W. Myers, C. Taillie, R. Vraney and D. Wardrop. 2003. Detection and delineation of critical areas using echelons and spatial scan statistics with synoptic cellular data. *Environmental and Ecological Statistics* (to appear).

Patil, G. P., G. Johnson, W. Myers and C. Taillie. 2000. Multiscale statistical approach to critical-area analysis and modeling of watersheds and landscapes. Pp. 293-310 in *Statistics for the 21st Century*, C. R. Rao and G. J. Szekely, eds. New York: Marcel Dekker, Inc.

Smits, P. and W. Myers. 2000. Echelon approach to characterize and understand spatial structures of change in multitemporal remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5):2299-2309.

Table 1. Cross-tabulation of mammalian species richness (rows) against forest cover class (columns) for 211 hexagonal sub-regions of Pennsylvania.

Species	**Cover 1**	**Cover 2**	**Cover 3**	**Cover 4**	**Cover 5**	**Row total**
30	0	0	0	0	1	**1**
31	0	0	1	1	1	**3**
32	0	0	1	0	0	**1**
33	1	1	0	0	0	**2**
34	0	2	1	1	0	**4**
35	2	3	2	7	1	**15**
36	2	0	1	2	1	**6**
37	0	5	1	4	4	**14**
38	0	1	3	7	2	**13**
39	0	6	3	7	5	**21**
40	3	1	6	6	2	**18**
41	0	17	10	2	2	**31**
42	2	5	4	0	0	**11**
43	4	12	2	3	0	**21**
44	8	10	6	0	0	**24**
45	3	6	3	0	0	**12**
46	2	4	0	0	0	**6**
47	4	0	1	0	0	**5**
48	1	1	0	0	0	**2**
49	1	0	0	0	0	**1**
ALL	**33**	**74**	**45**	**40**	**19**	**211**

Table 2. Cross-tabulation of mammalian species richness (rows) against forest cover class (columns) for 211 hexagonal sub-regions of Pennsylvania showing rate of occurrence relative to expectation in the body. First-order echelons are marked by <>. Second-order echelons bridging between first-order echelons are marked by {}. Highest third-order echelon is marked by []. Second highest third-order echelon is marked by (). Primary SATSCAN hotspot is marked by ++ and secondary hotspot by --.

Species	**Cover 1**	**Cover 2**	**Cover 3**	**Cover 4**	**Cover 5**	**Row freq.**
30	0.00	0.00	0.00	0.00	<11.11>	**1**
31	0.00	0.00	{1.56}	{1.75}	<3.70>	**3**
32	0.00	0.00	<4.76>	0.00	0.00	**1**
33	<3.23>	1.43	0.00	0.00	0.00	**2**
34	0.00	1.43	1.18	(1.32)	0.00	**4**
35	0.85	0.57	0.62	<2.46>	0.74	**15**
36	<2.13>	0.00	0.78	{1.75}	{1.85}	**6**
37	0.00	1.02	0.33	{1.51}	++ <3.17>	**14**
38	0.00	0.22	1.08	++<2.85>	++{1.71}	**13**
39	0.00	0.82	0.67	++{1.76}	++<2.65>	**21**
40	1.06	0.16	(1.56)	++{1.76}	++ (1.23)	**18**
41	0.00	--(1.56)	(1.51)	0.34	0.72	**31**
42	--1.12	--(1.30)	--<1.70>	0.00	0.00	**11**
43	--(1.22)	--{1.63}	-- 0.45	0.75	0.00	**21**
44	--<2.13>	--1.19	-- 1.17	0.00	0.00	**24**
45	[1.60]	--(1.43)	1.17	0.00	0.00	**12**
46	<2.13>	<1.90>	0.00	0.00	0.00	**6**
47	<5.12>	0.00	0.93	0.00	0.00	**5**
48	{3.22}	(1.43)	0.00	0.00	0.00	**2**
49	<6.25>	0.00	0.00	0.00	0.00	**1**
Col. freq.	**33**	**74**	**45**	**40**	**19**	**211**

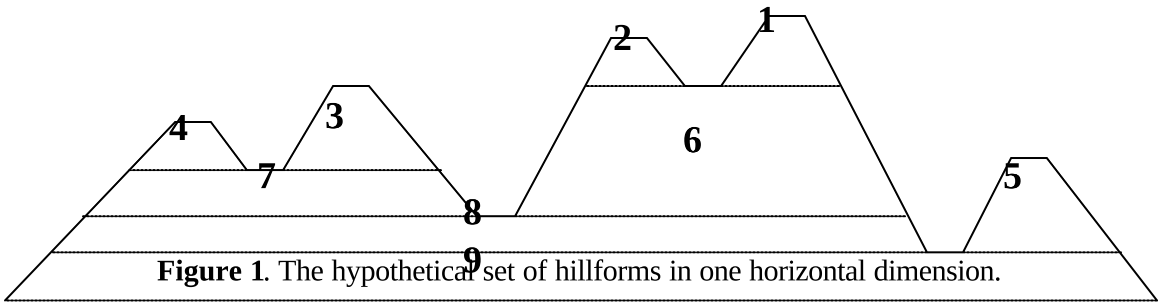
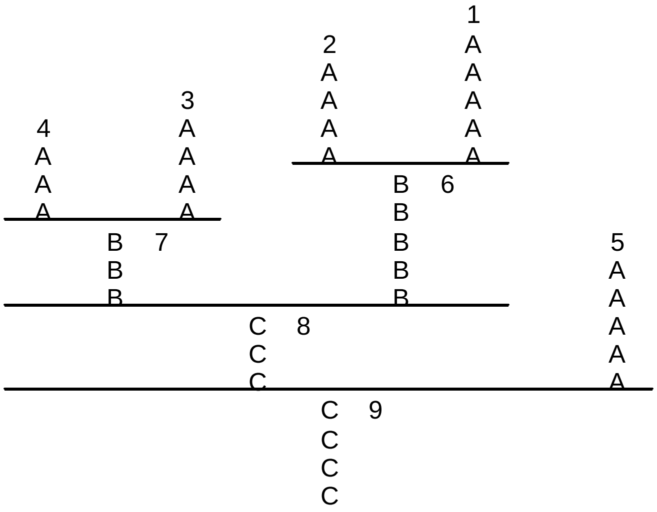


Figure 1. The hypothetical set of hillforms in one horizontal dimension.

Figure 2. The echelon dendrogram.for the hillforms of Figure1.



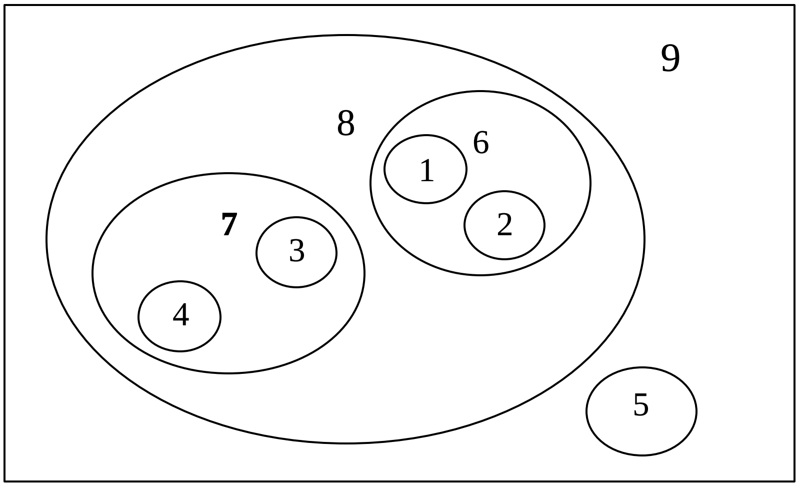


Figure 3. Spatial diagrammatic of continuous surface data on a plane (two dimensions of horizontal position) to which the dendrogram of Figure 2 also pertains.

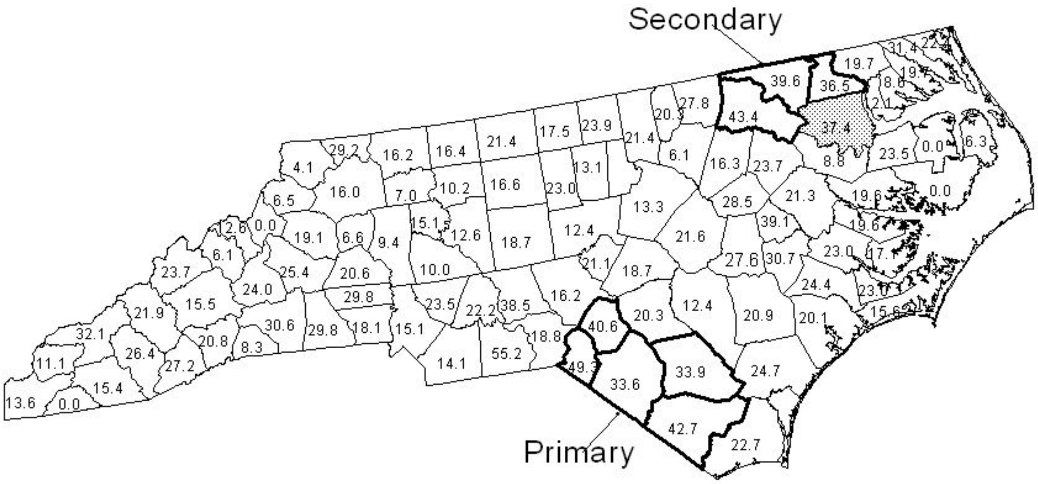


Figure 4. Rates of SIDS incidence per 1000 births for North Carolina counties during 1974-1984 showing primary and secondary SaTScan hotspots. The secondary hotspot has a neighboring candidate for substitution or augmentation.

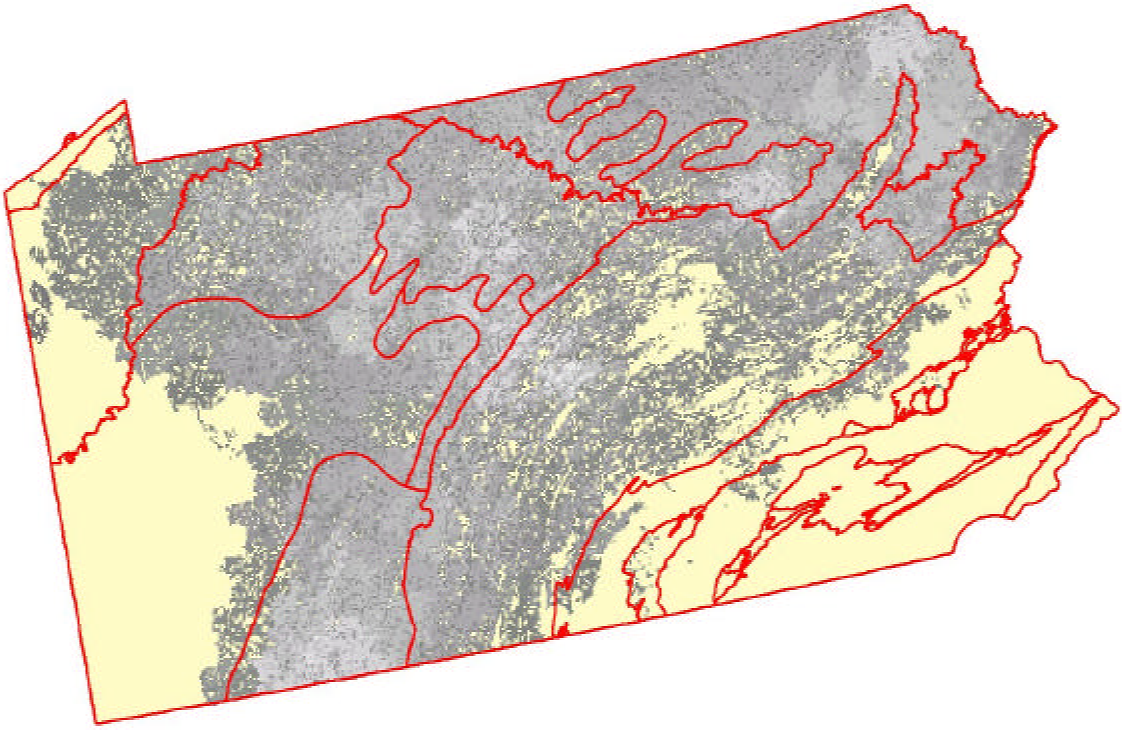


Figure 5. Upper-echelon mapping for number of mammalian species having suitable habitat in Pennsylvania. Mapping is for echelons of richness reaching into upper 25% of range in levels, with progressively darker shading down to midrange (50%) level. Heavy lines are boundaries of ecoregions.

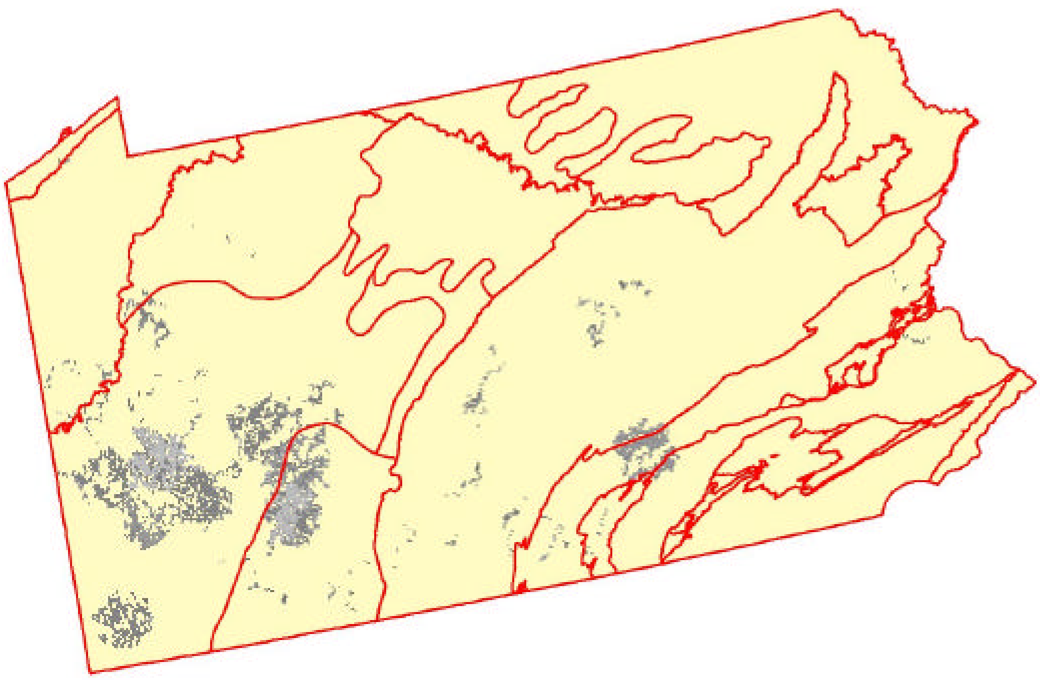


Figure 6. Upper-echelon mapping for number of amphibian species having suitable habitat in Pennsylvania. Mapping is for echelons of richness reaching into upper 25% of range in levels, with progressively darker shading down to midrange (50%) level. Heavy lines are boundaries of ecoregions.

.