CARTOGRAPHIC MODELING: COMPUTER-ASSISTED ANALYSIS OF SPATIALLY DEFINED NEIGHBORHOODS

J.K. Berry
Yale University
School of Forestry and Environmental Studies
205 Prospect Street
New Haven, CT 06511

C.D. Tomlin
Harvard University
Graduate School of Design
Cambridge, MA 02138

ABSTRACT

Cartographic models addressing a wide variety of applications are composed of fundamental map processing operations. These primitive operations are neither data base nor application-specific. By organizing the set of operations into a mathematical-like structure, the basis for a generalized cartographic modeling framework can be developed. Among the major classes of primitive operations are those associated with reclassifying map categories, overlaying maps, determining distance and connectivity, and characterizing cartographic neighborhoods. This paper establishes the conceptual framework of cartographic modeling and uses techniques for characterizing neighborhoods as a means of demonstrating some of the more sophisticated procedures of computer-assisted map analysis. A cartographic model for assessing effective roundwood supply is briefly described as an example of a computer analysis. Most of the techniques described have been implemented as part of the Map Analysis Package developed at the Yale School of Forestry and Environmental Studies.

INTRODUCTION

Most computer-oriented geographic information systems include processing capabilities which relate to the encoding, storage, analysis and/or display of cartographic data. The analytical operations used in many of the currently available systems (Calkins and Marble, 1980) are embedded within application-specific contexts. By extracting and organizing primitive operations in a logical manner, the basis for a generalized cartographic modeling structure, or "map algebra" can be developed (Tomlin, in preparation; Tomlin and Berry, 1979). In this context primitive map operations are analogous to traditional mathematical operations. The sequencing of map operations is similar to the algebraic solution of equations to find unknowns. In this case, however, the unknowns represent entire maps. The conceptual framework interrelating these primitive operations provides a basis for a modeling structure which accommodates a wide variety of computer analyses. This paper describes this conceptual framework and uses the techniques for characterizing cartographic neighborhoods to demonstrate some of the more sophisticated procedures and considerations of cartographic modeling.

DATA AND PROCESSING STRUCTURES

In order to use primitive operations in a modeling context a common data structure and a flexible processing structure must be used. The variety of mappable characteristics likely to be associated with any given geographic location may be organized as a series of spatially registered computer-compatible maps (Figure 1). In this way, a data base may be defined as a set of maps registered over a common geographic area; a map, or "overlay", may be defined as a set of mutually exclusive but thematically related categories; and a category, or "region", may be defined as a thematic value associated with a set of geographic locations, or "points" (Tomlin and Tomlin, 1981). While this is certainly not the only way to represent cartographic data (Chrisman and Peucher, 1975) it is one which relates directly and intuitively to traditional graphic techniques involving conventional geographic maps. It is also one which is common to many computer-oriented geographic information systems. Differences among these systems relate to either the way in which thematic attributes are represented (i.e. numerically, literally or in binary form) or to the way in which locational attributes are coded (i.e. rectangular cells, polygons, line segments, etc.). While these differences are significant in terms of implementation strategies, they need not affect the definition of fundamental cartographic techniques.

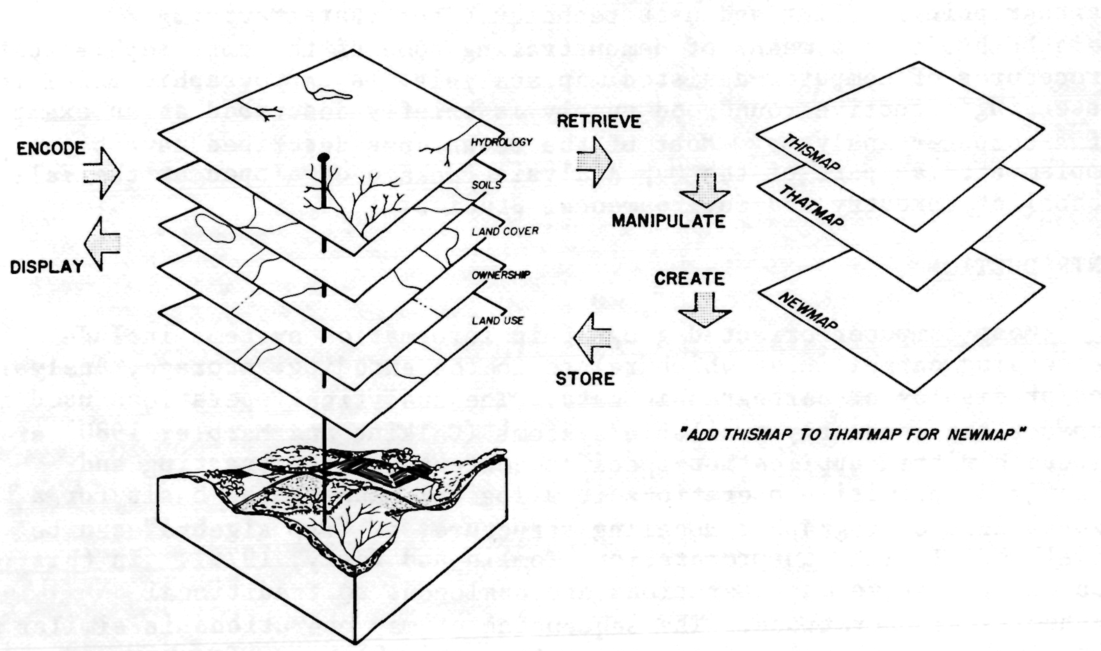


Figure 1. Cartographic Modeling Concept. A data base consists of spatially registered maps. Cyclical processing of these data involves retrieving one or more maps which are used to create a new map. The derived map then becomes available for subsequent processing.

If primitive operations are to be flexibly combined, each much accept input and generate output in the same format. Using a data structure as outlined above this may be accomplished by requiring that each analytic operation involve:

* retrieval of one or more maps from the data file;

* manipulation of that data;

* creation of a new map whose categories are represented by thematic values defined as a result of that manipulation; and,

* storage of that new map for subsequent processing.

The cyclical nature of this processing structure (Figure 1) is analogous to the evaluation of "nested parentheticals" in traditional algebra. The logical sequencing of primitive operations on a set of maps forms a cartographic model of a specified application. As with traditional algebra, fundamental techniques involving several primitive operations can be identified (e.g. a "travel-time" map) that are applicable to numerous situations. The use of primitive analytical operations in a generalized modeling context accomdates a variety of analyses in a common, flexible and intuitive manner. It also provides a framework for instruction in the principles of computer-assisted map analysis that stimulates the development of new techniques and applications (Berry and Tomlin, 1980).

<u>FUNDAMENTAL OPERATIONS</u>

Within the data and processing structures outlined above, each primitive operation may be regarded as an independent tool limited only by the general thematic and/or spatial characteristics of the data to which it is applied. From this point of view; four major classes of fundamental map analysis operations may be identified (Table 1). These involve:

* reclassifying map categories;

* overlaying maps;

* determining distance and connectivity; and,

* characterizing cartographic neighborhoods.

A brief discussion of these fundamental classes is presented below. More detailed discussions are presented in several of the references noted at the end of this paper (Berry and Tomlin, 1982; Berry, 1981; Tomlin and Berry, 1979).

The first of the four major groups of cartographic modeling operations is the simplest and, in many ways, the most fundamtental.

TABLE 1 FUNDAMENTAL MAP ANALYSIS OPERATIONS

FUNDAMENTAL CLASSES	FUNCTIONAL BASIS	EXAMPLE OPERATIONS
RECLASSIFYING MAP CATEGORIES- operations for reclassifying map categories involve reassigning thematic values to the categories of an existing map as a function of the initial value, the position, the size or the shape of the spatial configuration associated with each category.	+ INITIAL VALUE	+ ARBITRARY SCHEME (relabeling, isolating, aggregating) + ORDERING SCHEME (ranking, weighting) + MATHEMATICAL RULE (Isolating, arithmetics with constants)
	+ POSITION	+ SPATIAL LOCATION (reference coordinates, line orientation)
	+ SIZE	+ AREAL EXTENT + VOLUME
	+ SHAPE	+ BOUNDARY CONFIGURATION (edgeness, irregularity) + SPATIAL INTEGRITY (interior holes, fragmentation)
OVERLAYING MAPS-- overlay operations result in the creation of a new map where the values assigned to every location on that map is computed as a function of independent values associated with that location on two or more existing maps.	+ LOCATION-SPECIFIC	+ PERMUTATION (category combination) + DIVERSITY COUNTING + RELATIVE PROPORTION (frequency) + ORDINAL SELECTION (maximize, minimize, median, etc) + MASKING/SIEVING + LOGICAL COMBINATION (union, intersection) + ARITHMETIC COMBINATION (add, subtract, divide, etc) + WEIGHT AVERAGING
	+ CATEGORY-WIDE	+ PERMUTATION (category combination) + DIVERSITY COUNTING + RELATIVE PROPORTION (frequency, uniqueness, overlap) + ORDINAL SELECTION (maximize, minimize, median, etc) + LOGICAL COMBINATION (union, intersection) + ARITHMETIC COMBINATION (add, multiply, etc- commutative only)
	+ MAP-WIDE	+ SOLUTION OF MATHEMATICAL/STATISTICAL RELATIONSHIPS
DETERMINING DISTANCE AND CONNECTIVITY- operations for measuring cartographic distance involve the creation of new maps in which the distance and route between points can be expressed as simple Euclidean length or as a function of absolute and/or relative barriers.	+ SIMPLE DISTANCE/PROXIMITY	+ SHORTEST STRAIGHT LINE ("as-the-crow-flies")
	+ WEIGHTED PROXIMITY	+ SHORTEST ROUTE ("as-the-crow-walks") + ABSOLUTE BARRIERS (guide movement) + RELATIVE BARRIERS (expend units of movement)
	+ CONNECTIVITY	+ STRAIGHT LINE (simple distance, intervisibility) + OPTIMAL PATH (steepest downhill path, minimal "cost" route) + OPTIMAL PATH DENSITY (networks)
CHARACTERIZING CARTOGRAPHIC NEIGHBORHOODS- these operations involve the creation of a new map based on the consideration of "roving windows" of neighboring points about selected target locations.	+ SUMMARIZING THEMATIC ATTRIBUTES	+ STATISTICS (total, mean, maximum, etc) + ANOMALLY DETECTION (deviation, proportion similar) + INTERPOLATION (weighted average, nearest neighbor) + MAP GENERALIZATION (surface fitting)
	+ 2-DIMENSIONAL FEATURE ATTRIBUTES	+ NARROWNESS (shortest cord to opposing edges) + CONTIGUITY (individual clumps)
	+ 3-DIMENSIONAL SURFACE ATTRIBUTES	+ SLOPE (topographic slope, differentiation) + ORIENTATION (topographic aspect, direction of movement) + PROFILE (patterns along sequential cross-sections)

Each of the operations involves the creation of a new map by reassigning thematic values to the categories of an existing map. These values may be assigned as a function of the initial value, the position, the size, or the shape of the spatial configuration associated with each category. All of the reclassification operations involve the simple "repackaging" of information on a single map and results in no new boundary delineations.

Operations for overlaying maps begin to relate to the spatial, as well as to the thematic nature of cartographic information. Included in this class of operations are those which involve the creation of a new map such that the value assigned to every location is a function of the independent values associated with that location on two or more existing maps. In simple location-specific overlaying, the value assigned is a function of the spatially aligned coinidence of the existing maps. In category-wide compositing values are assigned to entire thematic regions as a function of the values associated with the regions contained on the existing maps. Whereas the first overlaying approach conceptually involves "vertical spearing" of a set of maps, the latter approach uses one map to identify boundaries from which information is extracted in a "horizontal summary" fashion from the other maps. A third overlay approach treats each map as a variable; each location as a case and each value as an observation in evaluating a mathematical or statistical relationship.

The third class of operations is one which relates primarily to the locational nature of cartographic information. Operations in this group generally involve the measurement of distance and the identification of routes between locations on a map surface. This class of operations served as the focus of a recent paper by the authors (Berry and Tomlin, 1982).

The simplest of these operations involves the creation of a map in which the value assigned to each location indicates the shortest distance "as the crow flies" between that location and a specified target area. The result is a map of concentric, equidistant zones around the target area. The traget area is not constrained to a single location and can be comprised of a set of dispersed points, lines or areal features.

If movement is implied in the measurement of distance the shortest route between two points may not always be a straight line. And even if it is straight, the Euclidean length of that line may not always reflect a meaningful measure. Rather, distance may be defined in terms of factors such as travel-time, cost, or energy which, unlike miles, may be consumed at rates which vary over space and time. Distance-modifying effects may be expressed cartographically as absolute and relative "barriers" located within the space over which distance is being measured. The resultant map identifies an effective proximity surface that characterizes movement from a target area over that space and through those barriers.

A distance-related set of operations determines the connectivity among specified locations. One such operation traces the steepest downhill path from a point on a three-dimensional surface. For a topographic surface, the path would indicate surficial water flow. For a surface represented by a travel-time map, this can be used to trace minimum-time (i.e. quickest) path. Another operation determines connectivity measured only for straight rays emanating from a target area over a three-dimensional surface to identify visual exposure.

The fourth and final group of operations includes procedures that create a new map in which the value assigned to a location is computed as a function of the independent values within a specified distance around that location (i.e., its neighborhood). This class of techniques will be discussed in detail in the remaining sections of this paper.

<u>CHARACTERIZING</u> <u>CARTOGRAPHIC</u> <u>NEIGHBORHOODS</u>

Most geograpahic information systems contain analytic capabilities for reclassifying and overlaying maps. These operations address the majority of applications that parallel conventional map analysis techniques (McHarg, 1969). However, to more fully integrate spatial considerations with contemporary analysis and planning, new techniques are emerging. The consideration of a location in context with its neighboring locations identifies a set of advanced operations. The summary of information within the neighboring locations can be based on the configuration of the surface (e.g. slope and aspect), the characterization of contiguous features (e.g. narrowness) or the satistical summary of thematic values (e.g. average value).

The initial step in characterizing cartographic neighborhoods is the establishment of neighborhood membership. A neighborhood, or "roving window," is uniquely defined for each target point as the set of all points which lie within a specified distance and direction around it. In most applications the window has a uniform geometric shape and orientation (e.g. a circle or square). However, as noted above that distance may not necessarily be Euclidean nor symetrical, such as a neighborhood of "down-wind" locations from a smelting plant.

The characterization of a neighborhood may be based on the relative spatial configuration of values that occur with the neighborhood. This is true of operations which measure topographic characteristics, such as slope, aspect or profile from elevation values. A frequently used techniques involes the "least squares fit" of a plane to adjacent

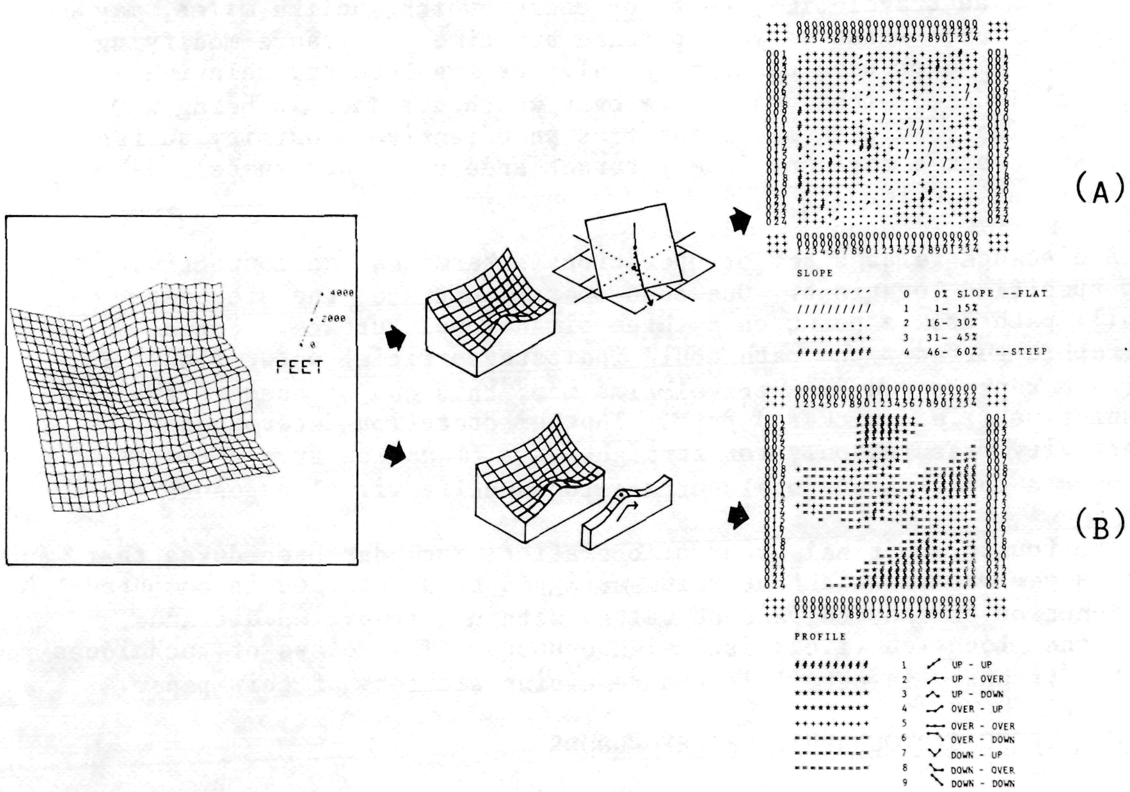


Figure 2. Characterizing Surface Configuration. Least squares fit of a plane to elevation values determines slope and aspect (a); computed cross-sectional profile as viewed toward the northeast (b).

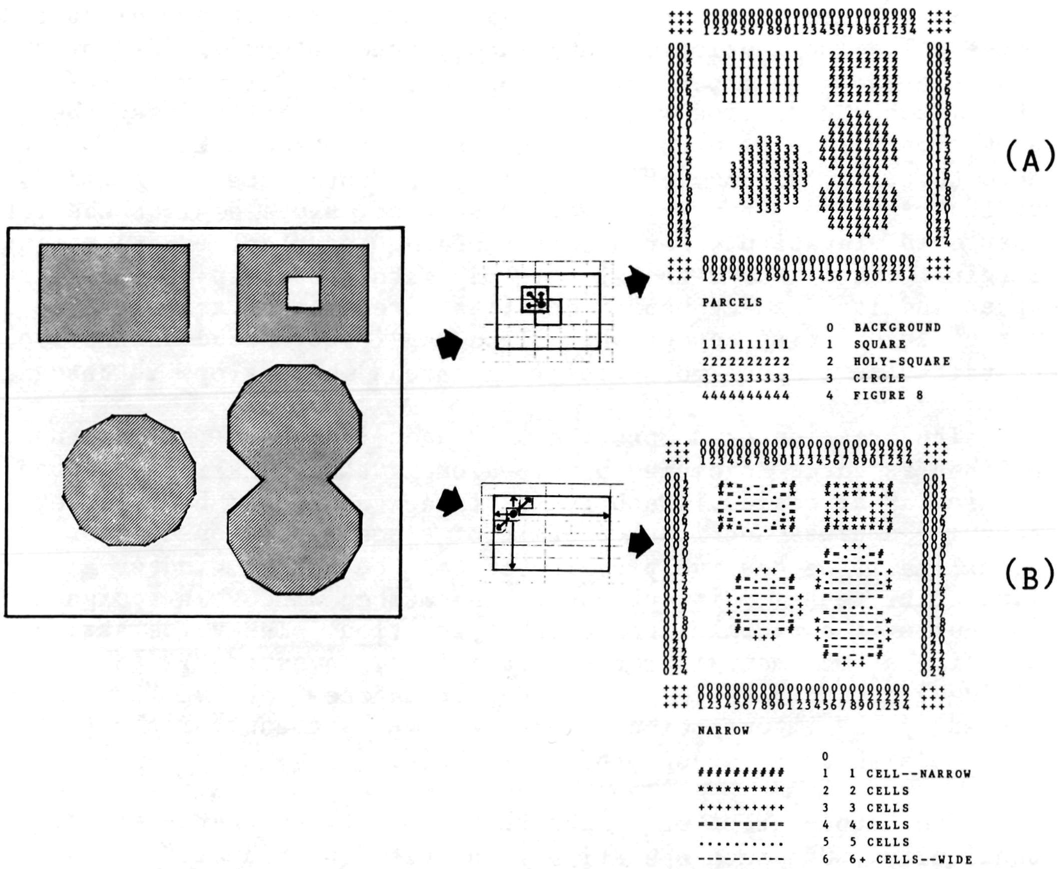
elevation values. This process is similar to fitting a linear regression line to a sereies of points expressed in two-dimensional space. The inclination of the plane denotes slope and its orientation characterizes the aspect within the immediate vicinity of the focus of the neighborhood. The window is successively shifted over the entire elevation map to produce a continuous slope (Figure 2a) or aspect map. Note that the "slope map" of any surface represents the first derivative of that surface. For an elevation surface, slope depicts the rate of change in elevation. For a cost surface, its slope map represents marginal cost. For a travel time map, its slope map indicates relative speed and its "aspect map" identifies direction of travel. The slope map of an existing topographic slope map (e.g. second derivative) will characterize surface roughness (e.g. areas where slope is changing).

The creation of a "profile map" uses individual neighborhoods defined as three adjoining points along a straight line oriented in a particular direction. Each set of three values can be regarded as defining a cross-sectional profile of a small portion of that surface. Each line is successively evaluated for the set of windows along the line. The center point of each three member window is assigned a value indicating the profile form at that location. The value assigned can identify a fundamental profile class (e.g. inverted "V" shape indicating a ridge) or indicate the magnitude, in degrees, of the "skyward angle" formed by the intersection of the two line segments of the profile. Figure 2b shows a map of profile changes.

The second group of neighborhood operations characterizes contiguity. One such operation identifies individual "clumps" of one or more points that are geographically connected. This involves noting the association between a "target" point and each point of similar thematic value which lies within its neighborhood. If this is done for all points of a given value on a map, spatially contiguous or near-contiguous subsets of those points can be identified. For example, given a map of many lakes this might be used to uniquely identify a particular lake.

The processing technique defines a window that includes neighboring points above and to the left of a location (Figure 3a). This window is successively moved from left to right beginning at the top of the map and proceeding to the bottom. The value assigned is determined according to the sequence in which individual clumps are encountered. If the thematic value of a target point is the same as a member of its neighborhood, it will be assigned the same clump number. If it is not the same, a new clump is indicated. The procedure assigns a common clump number to any groupings that are found to join in lower a portion of the map.

Another neighborhood characteristic which relates to spatial contiguity is narrowness. The narrowness at each point within a map feature is defined as the length of the shortest line segment which can be constructed through that point to diametrically opposing edges of the



<u>Figure 3.</u> Characterizing Feature Contiguity. Individual parcels of a common theme can be identified (a); narrowness of features is computed as the shortest cord through each point connecting the border (b).

feature. The state of Massachusetts, for example, is generally narrowest in the vicinity of Cape Cod. In establishing narrowness a window is defined in which the distance from a target location to each feature boundary location is computed (Figure 3b). The total length of each cord passing through the target point is the sum of the distance from that point to opposing boundary locations. The shortest of these cords identifies narrowness at that location. In order to avoid unnecessary processing for some applications, a window of maximum narrowness to be considered is specified.

The final class of neighborhood operations are those that summarize thematic values. Among the simplest of these involve the calculation of summary statistics associated with the map categories occuring within each neighborhood. These statistics might include, for example, the maximum income level, the minimum land value, the diversity of

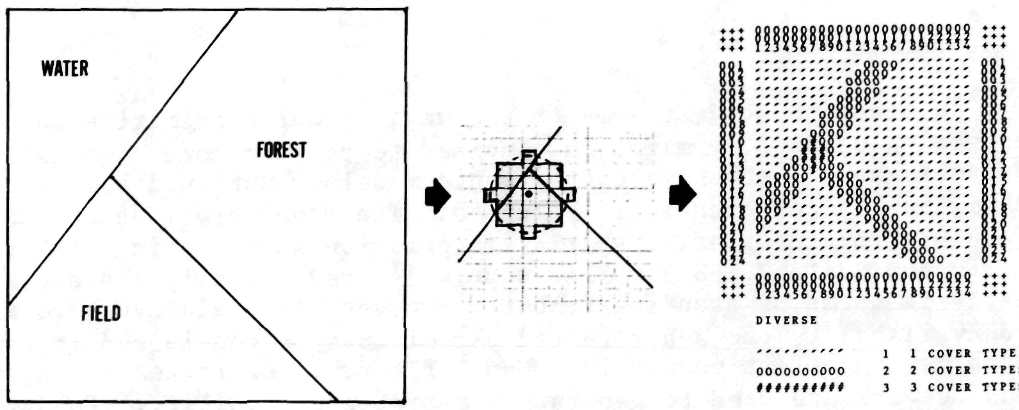


Figure 4. Summarizing Thematic Values. The diversity of cover types withn a specified distance can be computed. During processing a "roving window" is used to establish the set of neighboring points used in the summary.

vegetation within a half-mile radius, or perhaps a five-minute radius, of each target point (Figure 4). They might also include the total, the average, or the median value occurring within each neighborhood; the standard deviation or variance of those values; or the difference between the value occurring at a target point itself and the average of those around it.

Note that none of the neighborhood characteristics described so far relate to the amount of area occupied by the map categories within each neighborhood. Similar techniques might be applied, however, to characterize neighborhood values which are weighted according to spatial extent. One might compute, for example, total land value within three miles of each target point on a per-acre basis. This consideration of the size of neighborhood components also give rise to several additional neighborhood statistics including mode, the value associated with the greatest proportion of neighborhood areas; minority value, the value associated with the smallest proportion of neighborhood area; and uniqueness, the proportion of neighborhood area associated with the value occurring at the target point itself.

Another locational atrribute which might be used, in conjunction with size, to characterize a neighborhood is cartographic distance from the target point. While distance has already been described as the basis for defining a neighborhood's absolute limits, it might be also be used to define the relative weights of values within a neighborhood. Noise level, for example, might be measured according to the inverse square of the distance from surrounding sources. The azimuthal relationship between a neighborhood location and a target point may also be used to weight the value associated with that location. In conjunction with distance weighting, this gives rise to a variety of sampling and interpolation techniques. Azimuthal relationships may also be used to define absolute neighborhood limits.

CARTOGRAPHIC MODEL

In order to suggest some of the ways in which primitive map processing operations might be combined to perform more complex analyses, an illustrative cartographic model is outlined below and schematically represented in Figure 5. The model provides an analytical procedure for integrating spatial information in assessing timber supply (Berry and Sailor, 1982). This supply is traditionally characterized solely in terms of standing timber. However, the relative accessibility and availability of each forested parcel must be considered in establishing effective supply. Maps of terrain characteristics and the road network are used to generate a map of accessibility. Fundamental to this analysis is the use of a neighborhood operation for conversion of a map of elevation into a map of topographic slope. A distance measuring operation is then invoked to establish the relative proximity

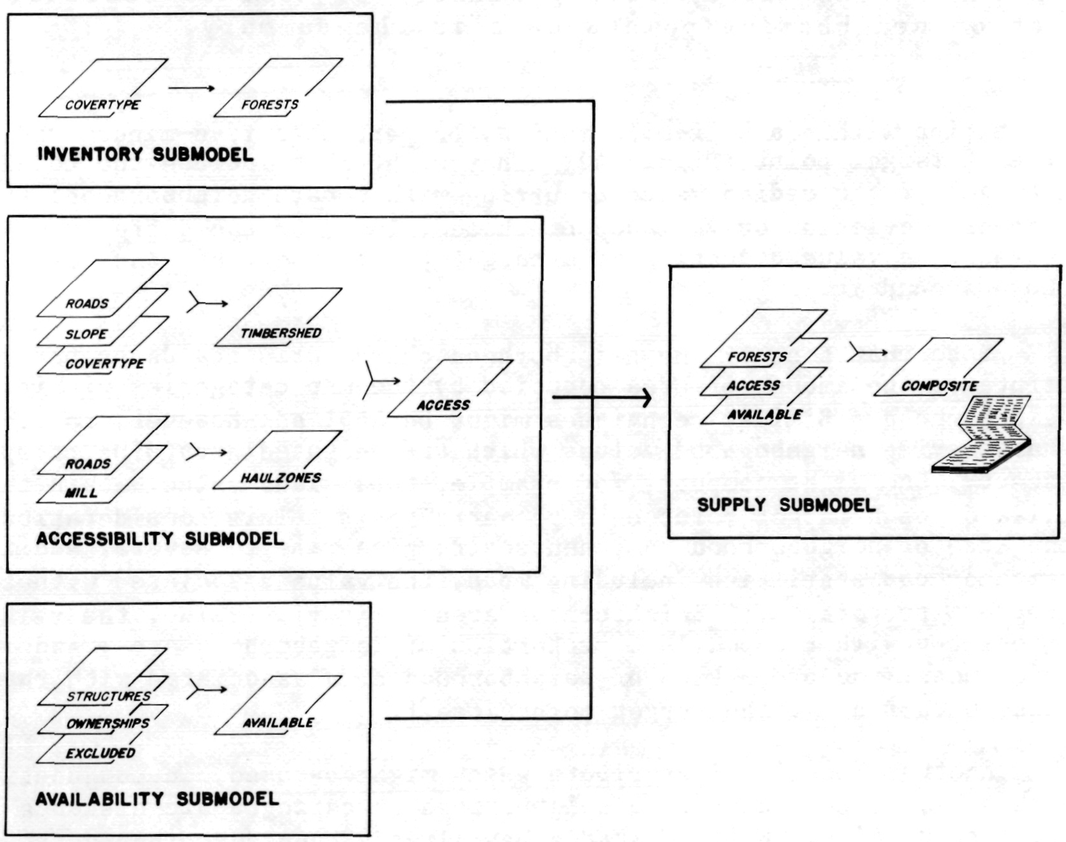


Figure 5. Flowchart of Effective Timber Supply Model. The cartographic model considers access and availability of forested areas, as well as physical inventory, in characterizing timber supply. Several primitive operations are logically sequenced to form the model.

of forest parcels to roads considering areas with steep slopes as harvesting barriers which must be circumvented. A distance operation is also used to establish "haul zones" from a mill based on the travel time along the road network.

The availability submodel uses a neighborhood operation to generate a map of housing density based on the total number of residential and commercial structures within a radius of 1/8 mile. A reclassification operation is used to establish the size of each ownership parcel. An overlay operation combines these two intermediate maps with one indicating areas excluded from harvesting to produce a map identifying the relative availability of areas for sale of stumpage. Areas of low housing density which are part of large ownership tracts are considered most likely to be available. The final submodel combines through an overlay operation, the maps of accessibility and availability to characterize effective timber supply. For selected combinations, a neighborhood operation is used to uniquely identify contiguous forest stands for management purposes.

Several other natural resource related models have been developed at the Yale School of Forestry and Environmental Studies using this approach as embodied in the Map Analysis Package software. These include:

* assessing deer habitat quality as a function of weighted proximity to natural and anthtropogenic factors;
* mapping outdoor recreation opportunity as determined by an area's remoteness, size and physical and social attributes;
* predicting storm runoff from small watersheds by spatially evaluating the standard Soil Conservation Service model;
* assesing the spatial ramifications of the comprehensive plan of a small town considering natural land use, preservation, growth and utility policies; and,
* characterizing spatial relationships among marine ecosystems factors to mode fish population dynamics.

In addressing these divergent applications a common set of fundamental map analysis operations were used. The logical sequencing of these operations on different sets of mapped data form the cartographic models of these different applications.

CONCLUSION

The modeling approach described in this paper can be used to extend the utility of maps for a variety of applications. A broad range of fundamental map analysis operations can be identified and grouped according to generalized characteristics. This organization establishes a framework for understanding of the analytic potential of computer-assisted map analysis.

ACKNOWLEDGEMENTS

The research described in this paper was supported internal Yale University developmental research funds.

REFERENCES

Berry, J.K. 1981. "Cartographic Modeling: Procedures for Extending the Utility of Remotely Sensed Data, "PECORA VII Symp. on Remote Sensing: An Input to Geographic Information Systems in the 1980's, American Society of Photogrammetry, 105 N. Virginia Avenue, Fall Church, Virginia, pp. 414-424.

Berry, J.K. and J.K. Sailor. 1982. "Spatially Characterizing Effective Timber Supply," Proc. of The National Conference on Energy resource Management, Amer. Planning Asso., Vienna, Virginia.

Berry, J.K. and C.D. Tomlin. 1982. "Cartographic Modeling: Computer-Assisted Analysis of Effective Distance, Proc. of Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana.

Berry, J.K. and C.D. Tomlin. 1980. "Cartographic Modeling in the Classroom: A Fundamental Approah to Computer-Assisted Map Analysis," Proc. of the American Congress on Survey and Mapping, 40th Annual Meeting, pp. 227-241.

Calkins, H. and D.F. Marble. 1980. Computer Software for Spatial Data Handling, volumes 1-3, International Geographical Union, Commission on Geographic Data Sensing and Processing, Ottawa, Ontario.

Chrisman, N. and T. Peucker. 1975. "Cartographic Data Structures," The American Cartograher, Volume 2, Number 1, pp. 56-69.

McHarg, I.L. 1969. Design with Nature, Doubleday/Natural History Press, Doubleday & Company, New York, 1969.

Tomlin, C.D., In preparation. "Digital Cartographic Modeling Techniques in Environmental Planning," doctoral disseration, School of Forestry and Environmental Studies, Yale University, New Haven, CT.

Tomlin, C.D. and J.K. Berry. 1979. "A Mathematical Structure for Cartographic Modeling in Environmental Analysis," Proc. of the American Congress on Surveying and Mapping," 39th Annual Meeting, pp. 269-283.

Tomlin, C.D. and S.M. Tomlin. 1981. "An Overlay Mapping Language Symposium on Regional Landscape Planning, American Society of Landscape Architects, Washington, D.C.