POLYGON OVERLAY TO SUPPORT POINT SAMPLE MAPPING: THE NATIONAL RESOURCES INVENTORY

Denis White
NSI Technical Services Corporation
200 SW 35th St.
Corvallis, Oregon 97333

Margaret Maizel
American Farmland Trust
1920 N Street, Suite 400, NW
Washington, DC 20036

Kelly Chan, Jonathan Corson-Rikert
Laboratory for Computer Graphics and Spatial Analysis
Graduate School of Design, Harvard University
48 Quincy St.
Cambridge, Massachusetts 02138

Problem

The construction of timely and pertinent policy for wise use and preservation of agricultural resources is predicated on an adequate knowledge of the status and extent of these resources. For nationwide policy development, it is therefore important to have comprehensive national surveys of natural resource information using uniform criteria. Such surveys of non-federal lands of the United States have been conducted by the Soil Conservation Service (SCS) at five different times in the last 30 years. The most recent of these was the 1982 National Resources Inventory (NRI).

The 1982 inventory consists of approximately 800,000 sample points in a statistical sampling design, each recording over 100 variables indicating soil, agricultural, and land use characteristics (Goebel and Dorsch, 1986). Each sample point is geographically referenced, in the distributed version of the survey, by the county, SCS Major Land Resource Area (MLRA), and US Geological Survey (USGS) hydrological cataloging unit in which it lies. (Latitude, Longitude coordinates for sample points are recorded by SCS.)

The statistical design of the inventory is such that the relative error associated with an estimate of area for a given crop or land use reduces as the area of the reporting unit (and hence the sample size) increases. So estimates for states have larger error bounds than those

for the entire country, and MLRA's and counties have successively larger bounds yet. Partly because of this characteristic, and partly because the design was optimized to report at the scale of MLRA's, little analysis of the 1982 NRI has been performed at scales of resolution finer than MLRA's.

Nevertheless, valuable information is contained in the NRI for even sub-county scales of resolution **if** interpretations are restricted to statements about the sample points, or aggregates of them. This paper describes a process for obtaining useful interpretations through geographic analysis techniques using geographic information system technology.

Method

One approach to NRI mapping is with raster methods. The georeferencing coverages (MLRA's, hydrological units, and counties) are rasterized separately into three index layers. An index layer contains pointers to (or geocodes for) individual MLRA's, hydrological units, or counties. NRI sample points are then grouped by the triplets of units of the three coverages that actually appear in corresponding cells of the rasterizations. Appropriate aggregations are then applied to NRI variables for the grouped sample points and the aggregated values are assigned to a raster layer.

We opted for the precision and flexibility of the analagous vector approach, that is, polygon overlay. In this case least common geographical units (lcgu's, Poiker and Chrisman, 1975) are created from the intersections of the MLRA, county, and hydrological unit coverages. Each lcgu is a polygon contained in a single MLRA, county, and hydrological unit. As with the raster approach, aggregated variables for sample points contained in each unique combination of the three coverages can then be mapped.

Federal lands can be excluded in the raster approach by rasterizing a binary coverage of federal/non-federal and excluding federal cells from NRI mapping. In the vector approach, the federal coverage can be overlayed upon the other three to obtain a similar exclusion.

Databases

Major Land Resource Areas (USDA SCS, 1981) are geographic regions that have similarity in natural resource characteristics as applied to agriculture, forestry, engineering, recreation, and other land uses. The characteristics used to define MLRA's are land use,

elevation and topography, climate, water resources, soils, and potential natural vegetation. There are 204 of these regions ranging in size from about 2,000 km^2 in the California Central Valley Delta to about 280,000 km^2 in the Northern Rocky Mountains (Figure 1). (For other multi-factor natural resource regionalizations see Bailey, 1976, and Omernik, 1987). The digital coverage of MLRA's was obtained from the SCS office in Fort Worth, Texas in DLG format. The published map of MLRA's is at a scale of 1:7,500,000.

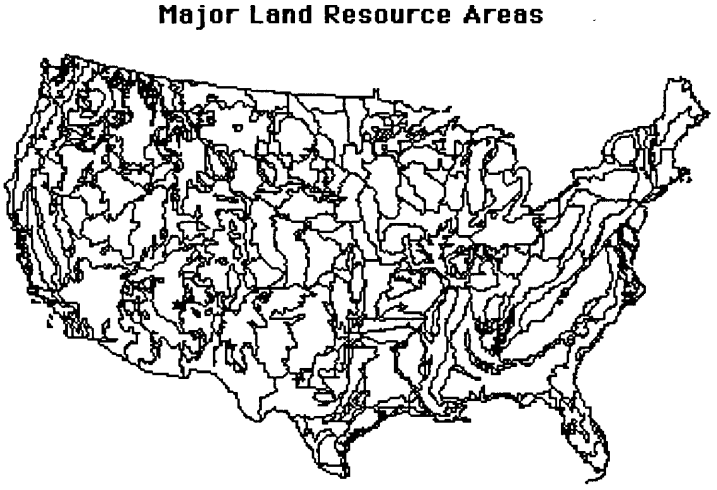

Figure 1

USGS Hydrological units are areal units that aggregate watersheds into areas of hydrographic similarity for research and resource management purposes. Cataloging units are the most detailed of a four level spatial hierarchy of units, totaling about 2100 polygons for the conterminous US (Figure 2). The digital coverage for these units was digitized from the two sheet USGS map of them at a scale of 1:2,500,000.

Figure 2

Federal land boundaries were obtained from the 1:2,000,000 scale DLG database distributed by USGS. A coverage for the conterminous US was assembled from the 15 sectional files in which this database is distributed by edge-matching along section boundaries. For purposes of this project, the various federal land ownership categories were aggregated to create a binary federal/non-federal coverage.

The coverage for counties of the conterminous US was a version of the Bureau of Public Roads county file digitized in the 1960's and distributed through the Bureau of the Census as the DIMECO file. This database had been maintained at the Lab for Computer Graphics for many years. The maximum suggested scale of use is 1:1,000,000 (Edson, 1984). The 1:2,000,000 scale DLG's were not used for this coverage because certain coastal county boundaries in this version follow legal definitions and lie in adjacent coastal waters, because edge-matching across sections would have been required, and because the resolution of the existing DIMECO file was adequate for this project.

All four coverages were (re)projected as necessary to the USGS standard Albers projection for the conterminous states and then coarsened (i.e., geometrically aggregated, Morehouse and Broekhuysen, 1982) to a resolution of 1.5 km. The resolution criterion derived from a consideration of the source scales of the coverages and expected final mapping scale.

Process

The polygon overlay of the four coverages (taken two at a time) resulted in a coverage of some 60,000 polygons using an overlay tolerance of 1 km. This file was coarsened to about 35,000 polygons using a tolerance sufficient to eliminate polygons smaller than about 1.5 km^2.

The overlaid geometric coverage was accompanied by a cross reference file relating the lcgu's to the containment units in the original coverages. This file was the basis for aggregating NRI data into appropriate attributes for mapping. The NRI sample point data file was sorted by its three georeferencing keys (MLRA's, hydrological units, and counties). The polygon cross reference file was sorted by these three and by the federal/non-federal land key.

The assignment of one or more mapping values to overlaid polygons consisted, conceptually, of synchronized passage through the two sorted files, aggregating attribute values for all sample points with the same unique combination of the three keys and assigning these values to all polygons cross referenced to the same unique combination (and in non-federal land). The actual implementation used an indexed sequential file for the polygon cross references to optimize performance.

A critical part of the process was the method of aggregation. Each NRI sample point includes an attribute called the "expansion factor" that records the number of acres the sample point represents. It is this factor, derived from the sampling process, upon which area estimates for land use and crop categories are based. In aggregating attributes, the expansion factor was used to calculate weighted averages of percent land in the various categories of the attributes represented by the sample points.

For example, one of the NRI attributes indicates whether the sample point is on land that meets prime farmland criteria. The calculated variable for overlay polygons is the percentage (ratio) of sampled land that meets prime farmland criteria and is computed as the ratio of the sum of each point in prime farmland times its expansion factor to the sum of all points, each weighted by its expansion factor. The title of this variable should then be something like "percent sampled land that meets prime farmland criteria".

Extrapolation of these percentages to produce an actual area estimate for an entire overlay polygon would be accompanied by relatively large confidence limits because of the relatively small sample sizes.

Since a choropleth mapping technique tends to imply uniformity of a statistic for a polygon, the qualifying titles for the maps are important.

Results

The Conservation Title of the Food and Security Act of 1985 (The 1985 Farm Bill), in instituting the Conservation Reserve Program (CRP), set up an important new mechanism to help conserve highly erodible or marginal cropland by restructuring price support payments for non-production on such lands into payments for placement of the same land into a conservation reserve for a ten year period.

The determination of eligible land for the CRP consists of a complex formula involving attributes sampled by the NRI. There are three criteria, the satisfaction of any one of which confers eligibility (7 CFR Part 704, Federal Register 2-11-87): cropland with an erodibility index greater than or equal to 8; or cropland in soils capability classes II through V with soil loss tolerance factor greater than 3T (three tons per acre per year tolerance); or cropland in soils capability classes VI through VIII.

The NRI provides a way to monitor the performance of the CRP, and the polygon overlay of the georeferenced databases allows a fine scale cartographic depiction of eligibility and performance (seen in coarse scale in Figure 3).

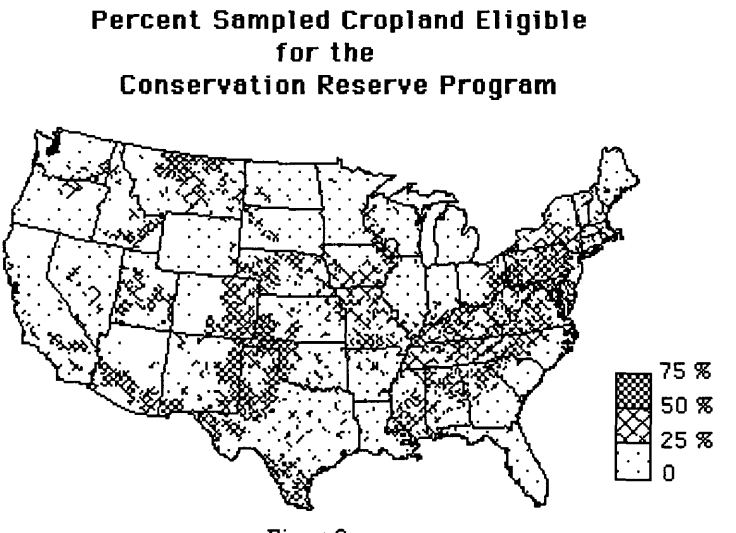

Figure 3

References

Bailey, R.G. 1976. Ecoregions of the United States. USDA Forest Service. Ogden, Utah.

Edson, D.T. 1984. Data Bases. U.S. National Report to the ICA, 1984. Special Issue of The American Cartographer.

Goebel, J.J., Dorsch, R.K. 1986. National Resources Inventory: A Guide for Users of 1982 NRI Data Files. USDA SCS.

Morehouse, S., Broekhuysen, M. 1982. Odyssey User's Manual. Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, Harvard University.

Omernik, J.M. 1987. Ecoregions of the Conterminous United States. Annals, Association of American Geographers, 77:118-125, and map supplement.

Poiker, T.K., Chrisman, N.R. 1975 Cartographic Data Structures. The American Cartographer, 2(1):55-69.

USDA SCS. 1981. Land Resource Regions and Major Land Resource Areas of The United States. Agriculture Handbook 296.