GIS-based Delineation of Soil Map Units:

titled

Systematic Application of Modeling, Remote Sensing, and Geographic Information Systems to Soil Surveys in National Parks and other Wildland Environments

for
the National Park Service,
Natural Resources Information Division

by

Henry Shovic, Ph.D.
Interagency Spatial Analysis Center
U. S. Forest Service, Gallatin National Forest
Bozeman, Montana

February 11, 1998 first Draft

Soil Map Unit Delineation Procedure

H. Shovic

Draft February 11, 1998

Introduction

We used three sets of computer-based systems in the Yellowstone National Park soil survey. The first (termed Taxonomic Description of Dominant Soils) was used to describe dominant taxonomic units, i.e., the important species of soil occurring in the area. These were developed under an analytical system using statistical site-specific data, literature, and field experience in the study area.

The second system was used to develop soil-forming concepts, i.e., what set of soil-forming factors produces a given soil taxon or related groups of taxa (termed Soil-forming Concept Development). The primary data sources for this analysis were statistical analysis of site specific data in databases, review of field transects, and experience of the soil mappers. The concepts represent the theoretical framework for prediction of soils.

We analyzed and grouped these concepts to produce rules that predict the distribution of soil map units. These rules are part of the third process, the pragmatic application of the concepts to the landscape (termed the Map Unit Delineation Process). Their character reflects availability of spatial data, its usefulness, and a number of experience-based judgements. All three of the above systems were used iteratively throughout the survey process as needed, until final maps, map unit descriptions, and taxonomic descriptions were completed.

The following text, figures, and tables illustrate how we used the third process to produce a spatial distribution of soil map units. Each illustration is exemplary and is not intended for use in replicating our results, as only portions of tables are shown to conserve space and each table has a certain level of "obsolescence" due to the iterative nature of our project. In other words, a few changes were made after these tables were developed, changes not always reflected in the tables. This is a necessary result of the process, that of balancing production of a soil survey against absolute accuracy of preliminary documentation. Records are of sufficient accuracy to allow an experienced professional to trace the process and results.

Though taxa and concept development are discussed only briefly here, a taxonomic list is included (Table A) which also shows coding for subsequent use. Table A1 contains a more descriptive list. These taxa were given soil family names (Figure B). An example of a soil-forming concept and its background data is given in Table C. The coding for site-specific background data is illustrated in Table D.

The Map Unit Delineation Process

Figure E shows a partial flow chart for our soil survey process. It includes only the portion after the taxonomy and concepts have been developed and proceeds through the electronic production of maps. In other words, it represents the spatial delineation process part of the soil survey. Only the highlighted steps are covered here.

There is an important difference between using a rule-based system to predict “soils” (a taxonomic unit), and a system to predict map units (which are combinations of soils.) Ideally each soil-forming concept refers to a single soil. Then, a map is created of individual soil, resulting in a very “pure” map having little variability within these taxons. This “purity” is sometimes implied in vegetation mapping projects using remote sensing techniques, where large areas on the map are defined as either one or another vegetation taxon or community. However, in the case of soil, an individual “soil” is on the order of a 1 x 1 meter area, and its location is impossible to predict with high purity given the much smaller scale of our data. Therefore we used the rule-base to predict soil “map units” which are combinations of soils having a defined level of variability and defined patterns of soils that repeat across the landscape of the map unit. Explicit in this process is the recognition and documentation of variability. This results in a level of “fuzziness” in the spatial distribution of soil types and the necessity of defining and documenting it. This “fuzziness” is controlled to keep it at a relatively low, but defined level. Hence our soil-forming concepts are relatively detailed, map unit names consist of more than one taxonomic unit name, and the map unit descriptions document ranges of properties usually crossing the boundaries of taxons.

Two systems were used to delineate these map units. The first is “reclassification and grouping.” The second is “rule-based” delineation. Both are essential to accurately model the soil mapping process to meet survey objectives in a timely manner.

Reclassification and Grouping

The reclassification system was necessary because the kinds of available data did not always meet our needs. We used a non-spatial process to group spatially related entities. In our case, the base layers we used (representing vegetative habitat types and landform/parent material) had far more classes than we needed for accurate prediction of soil properties at the scale of mapping and order of survey. The reduction of this data was an extensive and iterative process, requiring much technical discussion, review of data bases, and field excursions. We developed codes of these groups to use in the electronic portion of the study, illustrating them and their relationships to their parent types through a series of database tables. Table structure is shown in Appendix I. Please refer to the landform publication for definitions of landforms and matrix groups (Shovic, 1996) and to the soil survey publication (Rodman and Shovic, 1996) for definitions of habitat types.

We chose to separate the map units listed in the landform publication, giving about 80 landforms and about 60 matrix composition types. These were reclassified and grouped to about 20 landforms and 30 matrix composition types by clustering individual landforms (suffixed by a “3" code) and surficial material modifiers (again suffixed by “3") to groups suffixed by “4" codes. Table F shows definitions of the generalized landforms and Table G shows definitions and grouping parameters for matrix group modifiers. We used additional data from the landform publication to help us classify and error-correct our generalized landform groups and matrix modifiers. For example, in Table G1 average proportions of various surficial materials (e.g., columns under the headers of GlacF or glacio-fluvial alluvium and AlvC or coarse-textured alluvium) were used to help determine similarity between candidates for landform grouping. Since these data were in electronic form, we could track and evaluate the composition of each group (Figure H) for its usefulness in defining repeatable map units having acceptable variability. A similar, but separate procedure was followed for the matrix groupings (Tables G and G4). The final separate groupings are in Tables G2 for landforms and G3 for matrix groups. Though each landform group and matrix group existed separately in the landform publication, not all combinations of groups occurred on the landscape. We made a combined grouping table (Table K) using the data in Tables G2 and G3 including only those combinations that met that criterion. We then made a final numerical code for each of these combinations (Table L), resulting in the item SSNEW used in further analysis.

Groups of habitat types were developed the same way. Habitat types were grouped by similarity using Table I. This table includes descriptions and estimated properties used as criteria for grouping. The soil survey publication contains more information on habitat types. Table J contains a report based on these data, showing the new item (SOIL_HT) and the grouped types.

This somewhat elaborate coding/grouping process appears cumbersome, but since it is so formal, changes can be made in any step, with the relative certainty that these changes are reflected all the way through the process to the final maps, allowing quick error correction and repeatability. It also allows a great deal of flexibility since there are a large number of choices at any point in the grouping process.

Rule-Based Delineation

The second system involves the use of “rules” (statements of logical reasoning). Their application is based on theory of “expert systems” and uses some of that terminology. These are written to apply to spatial data layers, resulting in a spatial depiction of map units on maps. Of the many existing digital and hard copy spatial-data layers, we discovered through preliminary soil investigations and our concept development process that only two were needed to represent most local soil-forming factors used in the Jenny equation. We concluded an existing digital habitat type map adequately represented vegetation and climate and a landform/surficial material map portrayed topography, time, and parent material.

The landform/surficial material map did not exist prior to our study, so we found it necessary to create it during the survey process. Therefore, we were forced to use incomplete maps during the process. This would ordinarily be a stumbling block, as draft maps often have errors and incomplete legends. However, because this was an iterative GIS-based product and it was electronically “live-linked” to our survey modeling efforts through data bases and tables, the maps could be corrected and legend re-configured during the mapping process without requiring a complete re-evaluation of the soils legend each time a new version was produced. See Appendix II for details on how we made this layer. We added two derived layers toward the end of the process, a "frigid/cryic" boundary layer (termed "FRIGID") to solve the problem of predicting soil temperature with slope and aspect data, and a generalized rock type layer (called 'ROCKTYPE") for the presence/absence of andic materials in some geologic strata as well as other characteristics. These additions increased accuracy by accounting for data gathered in supplementary studies.

We wrote a rule for every soil-forming concept developed in the second process (about 300 of them). The soil-forming factors portrayed in each concept were converted to conditional statements relating to the layers described above. Figure M contains a partial list of rules. Each rule itself is fairly straightforward in format and logic. This example is in GRASS format, but is similar to that used in ARC/INFO. Each rule is preceded by a comment (prefixed by "!!" Or "!") and is in 'IF-THEN" format. The term "IFMAP" refers to an attribute of a layer, and the succeeding numbers refer to allowable values of that attribute. These values are separated by implied "OR." The term "ANDIFMAP" refers to other conditions that must be true. When a rule's conditions are all true and the resultant action occurs we term this "firing." When the rule fires, THENMAPHYP refers to a calculated value of another attribute that is set to a code and equivalent map unit name. For convenience, Table N contains an explanation of the map unit name codes. A limitation of the numerical coding process for SSNEW and SOIL_HT, FRIGID, and ROCKTYPE is that as written, rules have little descriptive value. The soil mapper must often refer to the coding tables L and J for meaning. However, because of advances in computer technology this limitation is no longer applicable. Word-based rules are now acceptable and computationally efficient.

For example, in Figure M the highlighted rule refers to MU (Map Unit) "27" with the name of "2924". If SSNEW is either "78", "80", "106" or "107" (rolling hills, covered with glacial till from a mixture of rock types, as described briefly in Table L) and SOILS_HT is either "2", "4", "5", "6", or "8"(a set of habitat types having sagebrush as a significant component in a non-forested environment derived from Table J) THEN the soil survey map unit will be code "27" or "2924", a unit including Mollisols formed in rolling hills of glacial till, in a non-forested shrubland environment with considerable sagebrush. Figure W contains a description of this unit. The detail contained in that description has been synthesized from the analysis of digital site data and information from the first two processes (Taxonomic Description of Soils and Soil Concept Development) briefly described above.

Each of the three hundred rules specifies a unique combination of attributes representing soil-forming factors, habitat type (representing vegetation and climate), landform (classified topography and parent material) and rules from other spatial data bases representing judgements of important factors' values (layers FRIGID and ROCKTYPE.) Note that the latter two layers were not used in the above example, but are used in other rules in Figure M. Because in many cases the occurrence of any one map unit could result from a variety of combinations of soil-forming factors, there are far more rules (300) then final map units (about 85). For example, see the two rules for Map Unit 2996 directly above the highlighted rule in Figure M.

The rule-base operates on a pre-developed interim polygon layer. This layer is the geometric intersection of all the important layers (in this case, an intersection of generalized habitat type, landform/matrix group, FRIGID, and ROCKTYPE). Each polygon has all attributes of these layers. The rule base is checked from its beginning to the first rule that fires, where the process exits the rule-base and starts again for every polygon in the layer. The computation is set up so that the first rule to fire stops the selection process for that polygon. Therefore the listed order of rules is sometimes important to the results. For example, if by mistake two identical sets of conditions were to give two different results, only the first one would ever fire, resulting in inconsistent results if that rule was changed later in the process, allowing the second to begin to fire.

After the rule-based process is complete, adjacent polygons having matching values for that attribute are combined (DISSOLVED) to produce a layer having a manageable number of polygons. It is then processed further to enhance its readability and utility (Figure E, lower ½ of the flow chart).

Though the reclassification and rule-based processes are generally too complex to visualize easily, a simple example given here shows the basic ideas. Please see Figures O, P, R, and T and V. These represent the north-central part of the Mammoth Map Sheet taken from the soil survey document. The scale matches publication scale of 1:62,500. The figures show respectively, reclassified habitat types, reclassified landform/parent material types, the resultant layer called "Soil Types," a legend for the first two figures, and a final soils map including labels for the "Soil Types." Making clear overheads of figures P and R will facilitate visualization. **For this document only, these overheads are included as Figures Q and S.**

First, observe Habitat Types (Figure O) and its legend (Figure T). These are generalized groups, as evidenced by presence of polygons surrounded or adjacent to polygons of the same color (e.g., the light brown color of nonforested, moist grassland.) Map units having a purple color represent areas of non-forested shrubland having considerable sagebrush. Lines in the Habitat Type map are relatively angular because this map was derived from raster (cell-based) data. We used operator-monitored, computer-based smoothing routines to produce a polygon-based map from the original raster data. We felt this was necessary to better approximate the patterns found in nature. Un-processed, raster-based polygons have a somewhat ragged appearance with many small arcs (Figure U) representing artifacts of the cell-based data.

The Landform/Parent Material Type map (Figure P) and its legend (Figure T) has been generalized from the landform/surficial material publication (Shovic, 1996). The lines represent the original data, and the patterns show the generalization. The light dot pattern represents rolling hills composed of glacial till. The lines on this map are relatively smooth because it was originally developed as a polygon layer with no intermediate raster layer.

These two layers and other data are used to create the Soil Types map (Figure S). This is a derived layer created with the rule-base as its core with some subsequent modification by the soil survey party to reflect field verification and matching requirements, and a number of cartographic revisions to improve its "ground truth."

The rule-base was applied to a layer representing the intersection of the two layers Habitat Types and Landform/Parent Material Types, modified to some extent by other layers. As an example the rule discussed above ("2924") fired in the two polygons labeled on the Soil Types map. We can approximate that process by overlaying the three maps (transparencies of the Landform/Parent Material and Soil Types over the colored Habitat Type map.) As is apparent from the overlays, these polygons are a combination of the non-forested sagebrush habitat type group (a group allowing habitat types 2, 4, 5, 6, or 8 and colored purple) and the glacial till landform/parent material type (a group allowing landform/parent material groups 78, 80, 106, or 107 with an open dot pattern), resulting in a map unit labeled "2924," described in the map unit description 2924 (Figure W). The 300 item rule-base was applied in a similar manner for each of the 25,000 intersected polygons in the survey area for every run of the model.

Extensive field verification provided data for additional changes. About 5 percent of the rule-based polygons were changed to reflect these recommendations. Also, edge-matching with adjacent administrative units required some modifications. In the case of Yellowstone National Park, soil map units from adjacent National Forests were used for some of the soils with high levels of volcanic ash. These were manually digitized at the end of the survey, but could have just as well have been placed in a set of rules that would make the necessary corrections every time the model is processed. Errors were discovered throughout the process in various layers. Changes were also made in hydrothermal areas and close to developed areas as supplementary data became available. These were corrected in the applicable layer itself, and re-intersected with the other coverages before each rule-base run.

Cartographic revisions were also required to better approximate the reality of the natural landscape and to improve readability. There is a good correlation between numerous lines when overlaying Landform/Parent Material Types on Habitat Types, implying there is in some places a good physical correspondence of both parameters. This reflects the nature of landscapes. However, the lines do not exactly match, as digitizing standards and methods, data sources, study objectives, and quality control procedures were different for the two layers. This is termed "coincidence" (the degree of exact overlayment of lines that represent the same physical entity in two different data layers). Overlay each map on the included topographic quadrangle (Figure X) to visualize these differences. Some, in particular those on steep narrow valleys do not correspond well. These concerns have lead to some interesting problems (and solutions) in computer-based revision (covered in a later publication). The publication map also has additional computer-generated smoothing, apparent from comparing the Soil Type map (Figure S) to an image of the Final Soils Map (Figure V.)

List of Tables and Figures

Table or Figure	Description
A	Taxonomic Subgroups, Names and Abbreviations
A1	Central Concepts for Yellowstone Soils
B	Soil Groups, Soil Family Names and Coding
C	Soil-forming Concept Example
D	Soil site data base definitions
E	Yellowstone National Park Map Production Flow Chart
F	Generalized landform/parent material groups and names
G	Matrix suffixes and groups
G1	Landforms and grouped landforms descriptions and characteristics
G2	Landforms and grouped landforms look-up table
G3	matrix suffixes and group codes
G4	matrix groups and descriptions
H	Landform groups and landform composition
I	habitat type groups
J	report of SOIL_HT and included habitat types
K	landform/parent material groups and landform/parent material types
L	landform/parent material groups and SSNEW codes
M	The rule base
N	Explanation of Map Unit names
O	Yellowstone National Park Habitat Type Groups- Mammoth Quadrangle
P	Yellowstone National Park Landform and Parent Material Types- Mammoth Quadrangle
Q	Yellowstone National Park Landform and Parent Material Types- Mammoth Quadrangle- Transparency
R	Yellowstone National Park Soil Types- Mammoth Quadrangle
S	Yellowstone National Park Soil Types- Mammoth Quadrangle- Transparency
T	Legends
U	National Forest Cover Types (from raster data)
V	Final Yellowstone National Park Soils Map- part of the Mammoth Map Sheet
W	Map unit description of 2924
X	Topography- part of the USGS Mammoth Quadrangle

References

Rodman, A., H.F. Shovic, and D. Thoma. 1996. Soils of Yellowstone National Park. Yellowstone Center for Resources, Yellowstone National Park, Wyoming, YCR-NRSR-96-2.

Shovic, H. 1996. Landforms and Associated Surficial Materials of Yellowstone National Park. Yellowstone Center for Resources, Yellowstone National Park, Wyoming, YCR-NRSR-96-3.

APPENDIX I Table Structures

APPENDIX II Landform publication methods

APPENDIX III Soil Survey Publication methods