THE GRAPHIC DISPLAY OF INVENTORY DATA

Philip J. and Carol A. Gersmehl
Department of Geography
University of Minnesota
Minneapolis, MN 55455

ABSTRACT

This paper distinguishes two mutually contradictory purposes for a regional information system: parcel classification and regional inventory. These two purposes ask different questions, call for different sampling and data storage procedures, yield different degrees of error at a given scale of analysis, and, most importantly, require different kinds of map display. An accurate classificatory system, with its focus on the dominant land uses within each data cell, will inevitably miscalculate regional totals. A valid inventory system, with its use of regular or random point sampling, will inevitably misclassify individual parcels of land. Inventory data are thus not appropriate for classificatory purposes such as tax assessment, land-use zoning, or prime-lands designation. It follows that inventory data should not be mapped in a way that allows identification of individual landholdings. This paper suggests three ways to map inventory data without inviting classificatory misuse: overlapping legend categories, probability blocks, and probability isolines. The selection of appropriate map techniques is a systems responsibility: the information system software should automatically adjust its graphic output to match the resolution and sampling logic of the input data.

INTRODUCTION

The story is perhaps too familiar. It begins with a request for a "reconnaissance study" of some resource -- prime farmland, forest volume, mineral prospects, wetlands, whatever. The project is an attempt to gain a general overview of the present state of a resource; for the purposes of this discussion, let us say that the study is supposed to determine the amount of prime farmland in the vicinity of an expanding metropolis. The objective is laudable, but the budget is small, and the result is a typical compromise between ideal research design and permissible cost. The constraints force us to gather data by some form of sampling rather than by a detailed survey. For a general picture of a resource, the sampling compromise is acceptable if the users are aware of its limitations.

The story, however, does not end here. The data are geocoded, stored on disk or tape, massaged by the computer, and finally output in graphic form for publication. The report is presented to the public, usually at a staged hearing with slides and a gallery of maps hung on panels around the speakers' table. During an intermission, the audience mills about, suitably impressed by the display, and then the balloon deflates: a few individuals crowd forward and peer intently at small parts of the maps. "They've got my back forty shown as swampland -- won't my cows be surprised!" "There hasn't been any timber on the Farley place since the War!" "Look, this map shows . . . " One by one, they find "mistakes" on the maps. The murmur spreads, and the audience after the intermission is more skeptical and much less receptive than before the close look at the maps. Their credibility is strained, and if the strain is too great, we may face a classic baby-with-bathwater situation as a valuable reconnaissance study is rejected along with the ill-conceived maps.

To repeat, the story is perhaps too familiar. The question is, what is the underlying cause and what can we do about it? We cannot accept the idea that the problem is merely ignorance on the part of the uninformed. That idea is seductive, but certainly too easy and probably wrong. A more likely culprit is the failure of many resource studies to distinguish between two mutually contradictory purposes of a resource information system: parcel classification and regional inventory. In this paper, we will do three things:

1) briefly describe the logical differences between regional inventory and parcel classification;

2) summarize some research on inventory and classification error rates; and

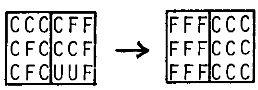
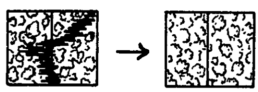
3) suggest ways to display inventory data without inviting classificatory misuse.

INVENTORY AND CLASSIFICATORY LOGIC

An ideal resource information system would grid the earth at half-meter intervals, storing instantly retrievable data on each tree and rock outcrop and intermittent waterway. Such an information system would allow the user to describe the resources at any specified location with a high degree of precision. It would also lend itself to very accurate regional inventories; the forest wealth of a county, for example, could be determined by counting trees in the computer memory.

In the real world, however, we can neither gather nor handle such a large volume of data, and therefore we are forced to choose between <u>parcel description</u> and <u>regional inventory</u>. These two end purposes of an information system are often incompatible because each has its own rules of sampling and data storage. In addition, the two require different kinds of map display, which is the substantive point of this paper. To make the distinction between the two types of logic as sharp as we can, we compare the two approaches point by point in Table 1:

TABLE 1. COMPARISON OF INVENTORY AND CLASSIFICATORY LOGIC

	REGIONAL INVENTORY	PARCEL CLASSIFICATION
Question being asked	How much resource X is in area Y ?	What is the resource in parcel Z ?
Typical answers	Area Y has 872 acres of mature forest. 21% of area Y is pastureland.	Parcel Z is a dense stand of pole firs. Parcel Z is mainly class II cropland.
Data type	Quantity measurement (ratio data)	Category assignment (nominal data)
Data cell* coding procedure	Each cell* is tabulated on the basis of the resource at the center (or some random point) within the cell.	Each cell* is categorized on the basis of the areally (or economically) dominant resource within the cell.
Trying to maximize	Statistical validity in the aggregate	Descriptive accuracy in the particular
Willing to sacrifice	Proper description of individual landholdings	Tabulation of small or odd-shaped resource areas
What can (often does, even should) happen	Sample point hits a tiny woodlot in a large open field and records the entire field as forested.	Narrow alluvial bottomland is lost from data record because it is too small to dominate any data cells.
		
Ideal procedure	Use enough sample points to make a valid estimate of even the least abundant resource.	Use small enough cells to keep them homogeneous for even the smallest resource areas.
Real-world compromise	Use sampling theory to limit error for major resources.	Use cell size that can capture most major resources.
Inevitable corollary	Inventories misclassify parcels.	Classifications miscalculate totals.

*The innocuous little word "cell" is the root of much confusion: it connotes, for a given discipline, some unique but vague blend of computer memory address, map location, grid intersection, sensor pixel, cadastral unit, or real-world observation site.

AN INHERENT INCOMPATIBILITY

A careful analysis of the premises and results of the two kinds of logic leads to an inescapable conclusion: if the data cells are larger in any dimension than the smallest homogeneous landscape units, the two approaches cannot be reconciled. A valid inventory system must inevitably misclassify individual parcels of land. Likewise, an accurate classificatory system must inevitably miscalculate regional totals. Reread those two sentences, please; they are the crux of the argument.

Several theoretical and empirical studies have shown that classificatory errors tend to grow much larger than inventory errors as data cell size increases (Figure 1; see also Wehde, 1979). For this reason, a cell or polygon that is far too large for accurate parcel classification may still yield acceptable regional inventories. Herein lies the root of the problem described in our introduction. Most resource information systems are proposed and justified (and therefore funded) for generalized inventory purposes. As such, they are valuable planning tools, but their inventory logic and coarse cell size renders them prone to serious classification errors. It follows, then, that inventory data should not be mapped in a way that allows identification of individual landholdings. Nor should inventory information be made available for zoning, tax assessment, prime lands designation, or corridor selection without explicit safeguards, preferably in the form of mandatory field checking (Yanggen, 1979).

Our track record on this score in not reassuring. In general, our maps have been SYMAP-type printouts or raster plots made directly from the data stored in the computer array (Nichols, 1975). The maps usually contain as much detail as we can show at the scale chosen for publication (Sometimes, we even use sophisticated software to recode coarse data cells or polygons into finer and more precise pixels, usually on the premise that such a procedure allows us to "capture" a map shape more accurately; that was the gist of the misguided criticism in Tomlinson et. al., 1976). Individual landholdings on these lineprinter maps are easy to identify, and our safeguards have generally consisted of vague legend notes about scale and generalization. In effect, we have been committing the informational equivalent of putting a turbocharged V-8 engine in a golf cart, with nothing more than a label on the dashboard to warn the user that the graphic power is too much for the data chassis.

A PROPOSED SOLUTION (OR THREE)

As we see it, the major remaining task is to find a way to display inventory data safely: to show the general pattern of a resource without specifying anything about individual landholdings. The problem is like that of a meteorologist, who knows that the air today is generally unstable, and therefore scattered thunderstorms are probable throughout a given region, but the data simply will not allow a yes-no prediction for a particular place within the region. The

ERRORS IN REGIONAL INVENTORY AND PARCEL CLASSIFICATION WITH DIFFERENT SIZES OF GRID CELLS

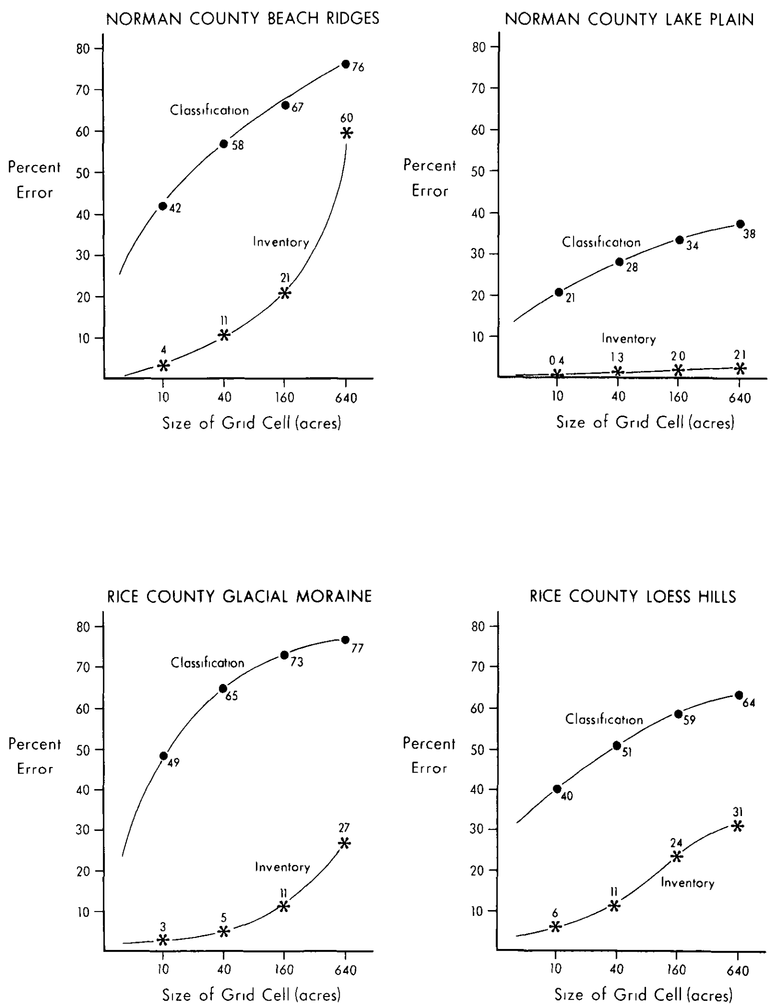


Figure 1. Inventory and classificatory errors in coding soil information from four representative Minnesota soil landscapes. "Percent error" is defined as the percentage of 2-acre polygons (digitized from the detailed County Soil Survey) that were miscategorized by the classificatory approach or were over- or under-estimated in the aggregate by the inventory approach (from Gersmehl and Napton, 1982).

conventional solution -- to report the probability of precipitation in percentage terms -- is reasonably well understood and widely accepted by lay people and experts alike. A similar approach should be equally useful in resource cartography. Operationally, this idea can take several forms:

Overlapping legend categories. The simplest way to show the uncertainty is to recode the map legend into overlapping or non-specific categories. Rather than using discrete and exclusive groups (like 0-10, 11-20, 21-35, etc.), we can adjust the group boundaries to reflect the degree of uncertainty (e.g. 0-18, 7-26, 16-35, etc., or Groups A and B, where A is 0-10 and B is 11-20). This approach makes it impossible to rank two parcels of land solely on the basis of the map, but it suffers from one fatal weakness: real-world variability of phenomena such as soil productivity or forest volume is often extremely great even within a small area. An accurate set of category boundaries to describe the land in a given small area can be nearly as wide as the entire range of variation throughout the map. The resulting map categories (e.g. 0-34, 4-35, 6-36) are not very helpful in communicating real patterns and differences in the region. Conventional statistical measures such as means and standard deviations may help, but they are both time-consuming and rather forbidding for lay readers.

Probability blocks. The second way demands a bit more computer time, but it preserves the essential traits of the natural landscape quite well. The process makes use of a "moving window," a widely used method of scanning and recoding data in a geocoded array.

The conventional use of a moving window is as a noise filter, to simplify a remotely sensed image by determining the majority type in a local area (Lillesand and Kiefer, 1979, pp. 573-577). The input to the process is a cluttered patchwork of small cells. The operation is a sequential placement of the "window" on a regular grid superimposed on the base; the window typically exposes a number of cells (say 25, for a 5-by-5 window) that are then classified and counted by type. If 16 of the 25 cells within the window at a given placement are of a particular type (as in the window shown on Figure 2), all cells within that window would be reclassified as that type. The output is a simplified map in which each windowed block is shown as a uniform expanse of whatever type dominates the most small cells within the window at each placement (Figure 2 left). A conventional moving-window procedure thus loses a great deal of information.

The right side of Figure 2 shows a variant of the moving-window technique, in which the windowed blocks are recoded by classifying each block according to the percentage of small cells in a given category. In this case, the output would be a series of rather coarse choropleth maps showing percentages of occurrence within blocks rather than actual occurrence within cells. The class boundaries, of course, can be changed or even omitted in favor of an unclassed map (Tobler, 1973; Brassel and Utano, 1979). The key idea is that the inevitable uncertainty at the scale of an individual data cell is masked by merging cells into larger blocks.

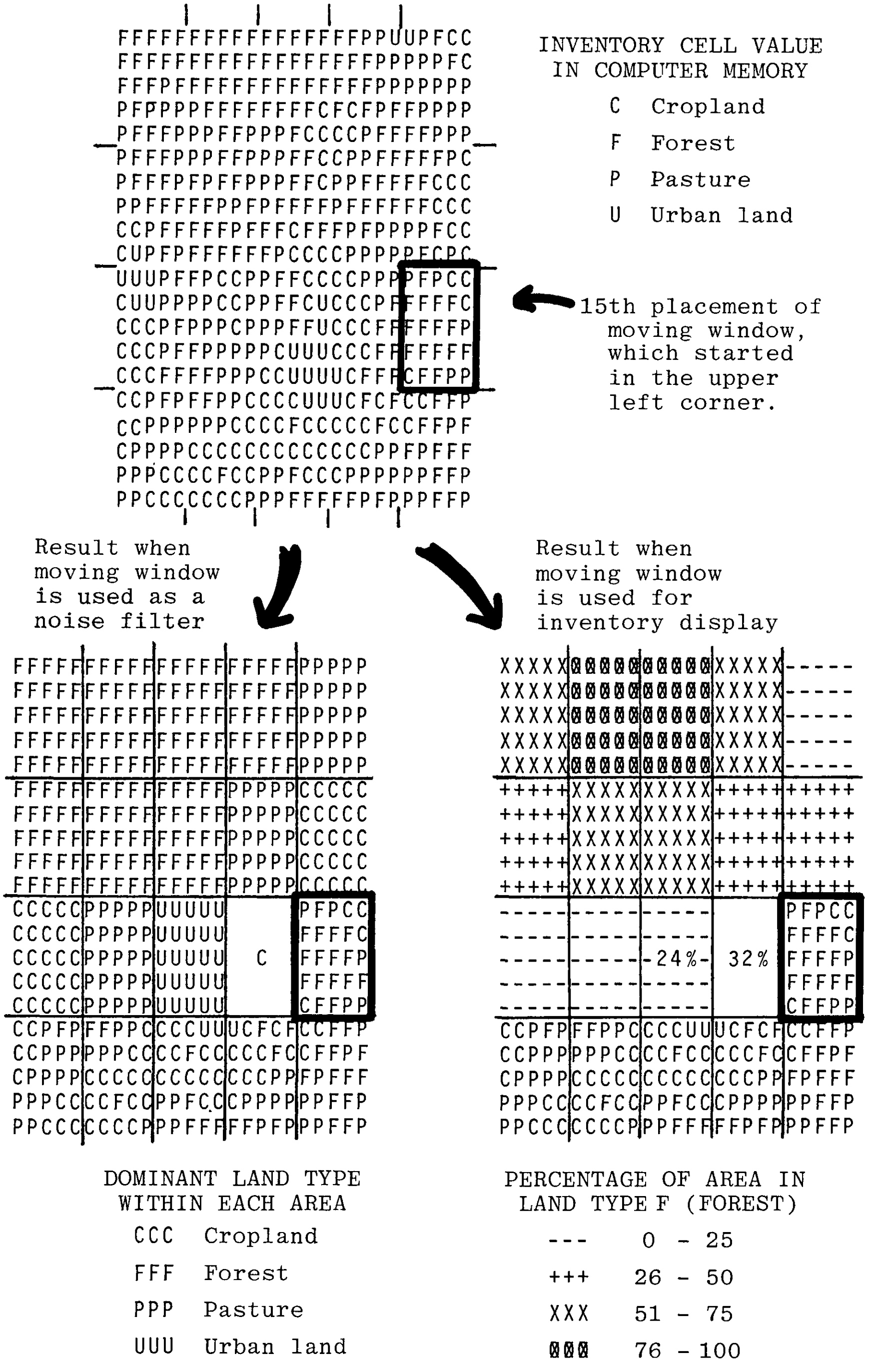


Figure 2. Two uses of a moving window, as a conventional noise filter (left) and as a tool to make probability blocks from inventory data (right). Figure 3 uses the same data.

These blocks, in turn, are keyed to a legend that talks about probabilities of occurrence of a particular trait somewhere within a large area, rather than about the actual value of a given cell. (Yes, we can hear the screams: "Why should we give up so much precision of location?" The answer is simple: if the data were gathered for an inventory (i.e., using point sampling), we didn't really have that degree of precision in the first place. Although we have a value stored for each small cell, the cell value is part of a statistical sample, not a place classification.)

<u>Probability isolines</u>. One of the drawbacks of the block approach is the coarse appearance of the window map, with its staircase boundaries and abrupt pattern changes. The maps are valid, in that they do not imply any more spatial precision than is warranted by the data, but readers may feel that they look too crude for their price tag. This is basically a PR problem, but it is a real one and must be addressed. Even so, we must emphatically refuse to condone a return to a fine-resolution graphic display for purely aesthetic reasons. Our third approach is just a way to "dress up" the probability blocks in order to produce a more impressive display. The method consists of three steps:

1) Calculate the percentage of a category (say forest) in each window placement as for a block map.
2) Assign this figure to the central cell of the window.
3) Use these figures as control points for an isoline map.

The isoline map, of course, could be shaded or colored in the conventional manner; the legend would say something like:

"Percentage of forest in an arbitrarily selected 10 square kilometers:"

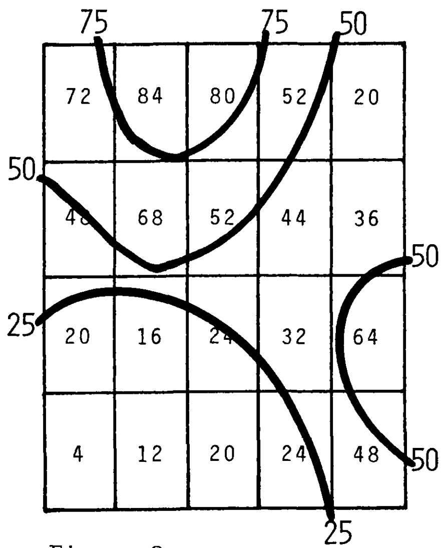


Figure 3

The moving-window isoline method, like any powerful cartographic transformation, should be used with caution. The size of the specified arbitrary area depends on the size of the window, which in turn should be chosen to screen out the minor kinds of repetitive variation in the data. The growing literature on spatial autocorrelation and image processing can provide the basis for a software algorithm to select an appropriate size and geometry for the moving window (Campbell, 1979). Suffice it here to note that the problem is much more complex than implied by standard sales department assertions ("It provides output at any scale you want").

CONCLUSION

Most planning-related information systems were justified as regional inventory tools. To be cost-effective as well as accurate, a resource inventory should generalize from some kind of point sampling method (Lund, 1981). Such a procedure is fundamentally incompatible with an alternative use of the data, namely the classification or evaluation of individual parcels of land. Data gathered for resource inventory will inevitably contain many land-parcel classification errors. Ideally, the primary output for an inventory should be a histogram; straightforward lineprinter maps of the stored cellular data are simply not appropriate for inventory logic. However, the pressures to produce some kind of "general" map are very strong, and indeed the data can yield a generalized picture of a resource distribution. The probability-based techniques described in this paper will allow the display of general patterns without revealing the possible inaccuracies at the scale of individual cells.

We are not offering a new technique -- indeed, it has been used for many years by a number of agencies (e.g. the U. S. Soil Conservation Service, for its county association maps, and the Canada Land Inventory, for its land suitability maps) What we are proposing is that the techniques used on these manually drawn maps should be made a mandatory part of most automated resource cartography. In other words, we think that the graphic display should be compatible with and should clearly communicate the underlying logic of the information system and the degree of generalization in the data.

It is fashionable among systems analysts to refer to scale and resolution as "data problems." To paraphrase several presentations on the subject: "Of course, we still have a data problem. Until the field people give us data at a sufficiently fine resolution, we will not be able to eliminate some of the errors you noted on our maps." We beg to differ. Scale incompatibilities will always plague a study that considers more than one environmental trait or variable. The inability to sense the logic, precision, and resolution of input data, and to adjust the output accordingly, is a <u>systems</u> problem (not a data deficiency). A method of displaying inventory data without inviting classificatory misuse should be built into any public resource information system.

POSTSCRIPT

We have demonstrated a fundamental incompatibility between two kinds of information systems, one designed for regional inventory and the other for parcel classification. It is logically and should be legally indefensible to use an inventory-based system for any classificatory purpose: tax assessment, prime-lands designation, zoning, or corridor selection. Although we did not address classificatory purposes directly in this paper, we do want to emphasize that they demand an information system whose sampling logic and scale of resolution are appropriate to that end use.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support of the Center for Urban and Regional Affairs of the University of Minnesota and the helpful advice and assistance of many people, especially Clark Archer, John Borchert, Greg Chu, Will Craig, Phil Muehrcke, and Darrell Napton.

REFERENCES

Brassel, K.E. & J.J. Utano 1979, Design Strategies for Continuous-Tone Area Mapping: The American Cartographer, Vol. 6, pp. 39-50

Campbell, J.B. 1979, Spatial Variability of Soils: Annals, Association of American Geographers, Vol. 69, pp. 544-556

Gersmehl, P.J. & D.A. Napton 1982, Interpretation of Resource Data: Problems of Scale and Spatial Transferability: Proceedings, URISA 20th Annual Conference: Practical Applications of Computers in Government, Minneapolis, Minnesota

Lillesand, T.M. & R.W. Kiefer 1979, Remote Sensing and Image Interpretation, John Wiley & Sons, New York, pp. 573-577

Lund, H.G. 1981, Point Sampling: The Role in In-Place Resource Inventories: Proceedings, In-Place Resource Inventories Workshop, Orono, Maine

Nichols, J.D. 1975, Characteristics of Computerized Soil Maps: Proceedings, Soil Science Society of America, Vol. 39, pp. 927-932

Tobler, W.R. 1973, Choropleth Maps Without Class Intervals?: Geographical Analysis, Vol. 5, pp. 262-265

Tomlinson, R.F., H.W. Calkins, & D.F. Marble 1976, Computer Handling of Geographic Data, The UNESCO Press, Paris

Wehde, M.E. 1979, Spatial Quantification of Maps or Images: Cell Size or Pixel Size Implication: Joint Proceedings, American Society of Photogrammetry and American Congress of Surveying and Mapping, pp. 45-65

Yanggen, D.A. 1979, Incorporating Soils Information into Land Use Controls: Planning the Uses and Management of Land, Soil Conservation Society of America, Madison, Wisconsin, pp. 957-981