Journal of the Minnesota Academy of Science

Volume 55 | Number 1 Article 3

1989

Geographic Information Systems, Data, and Water Resources

Dwight A. Brown
University of Minnesota

Philip J. Gersmehl
University of Minnesota

Recommended Citation

Brown, D. A., & Gersmehl, P. J. (1989). Geographic Information Systems, Data, and Water Resources. *Journal of the Minnesota Academy of Science, Vol. 55 No.1*, 14-17.
Retrieved from https://digitalcommons.morris.umn.edu/jmas/vol55/iss1/3

Geographic Information Systems, Data, and Water Resources

DWIGHT A. BROWN and PHILIP J. GERSMEHL

ABSTRACT—We evaluate three data handling methods for use in a GIS analysis of land-cover change impacts on runoff. A universe of 2560 point samples is analyzed to provide runoff calculations that would serve as a comparison base to evaluate different attribute logic systems. The attribute logics we evaluate are two variations of tag and one of count. We chose a two by five mile area of Dakota County, Minnesota as the test site, and prepared raster GIS maps of soil hydrologic groups and two plausible land covers. The count method for handling the generalization of data produced results that were substantially closer to the characteristics of the universe than either of the tag approaches. To minimize error in assessment of water resources with a GIS, analysts should start with primary data, control all phases of data manipulation, and use count methods to abstract large-area data.

Introduction

Analyses of water resources take many forms, examine many parts of the hydrologic cycle, and focus on a wide range of geographic scales. Broad types of studies include:

- drought severity, storm event rainfall, or stream flow;
- contamination potential from soil erosion;
- sensitivity of groundwater to contamination or loss of recharge potential;
- impact of climate change on productivity of forest or farmland;
- population pressures on surface water recreation;
- impact of planned land development on runoff or groundwater recharge.

All of these examples are now being examined at various locations with the assistance of Geographic Information Systems (GIS). The trend will likely increase because a growing amount of data are being archived in GIS files. One of the attractions of GIS is the ability to provide locational specificity to the products of these analyses. However, the user of these systems must recognize that data are of different spatial resolutions or sampling densities, of different time periods, and often reported in different units. Some data are recorded with different attribute logics. These logics can be described as tags or counts. In tag logic data, the attribute of a cell is tagged with a name that reflects the dominant trait of a mapped area. Count logic merely implies that we keep track of the frequency counts of various traits within each mapped area (1). Some data are primary, and some are second-hand files that have lost some of the original detail in processing.

Attribute logic presents problems for analysts because the theories that underlie much water resources simulation assume that data have many characteristics that may be lacking in the tag data that is often the most readily available to water resources planners and managers. The problematic characteristics of tag data can derive both from the design of the data collection process and from the subsequent handling of the data with GIS manipulation that is less than optimal. Some of the error induced in the analysis process can be avoided. The results of improved data collection, archival techniques, and manipulation procedures could improve the effectiveness of resource policy development and the adequacy of management decisions.

The objective of this paper is to illustrate the difference that data handling can make in the outcomes of analyses typically performed for water resources management and planning.

Research Design

We took three primary data sets and performed a series of analyses to demonstrate the effects that common differences in data handling can have on the outcomes of analysis. These disparities in data often result from differences in the purpose of data collection or from its subsequent manipulations with a GIS. The problem we used to illustrate the advantages of appropriate data collection and handling procedures is relatively simple — assessment of the impact of a development plan on surface water yield. The data used in these analyses came from the upper Vermillion River Basin in eastern Dakota County, Minnesota. The area was selected to include differences in complexity of landscape texture; it features a Late Wisconsinan terminal moraine in the west and nearly level outwash gravels in the east with thin loess cover in some areas. The site comprises a rectangular tract of land that is 2 by 5 miles. The 50-year, 24-hour storm (5.3 inches of rainfall) was selected for illustrative purposes because its intensity is sufficient to produce runoff on the most absorbing soil-landcover combination.

Data Handling

Runoff was calculated according to procedures described in USDA TR55 (2). The program was chosen for its convenience and simplicity. The primary data needed are soil hydrologic groups and landcover, for both a "natural" vegetation and for a planned development. Data were recorded as point samples on a 100 meter grid (2560 points, roughly equivalent to a sample at the corners of a 2.5 acre grid or 256 samples per square mile). The soil data were taken from the Dakota county soil survey and entered as series data

---

Dwight Brown and Phil Gersmehl are professors in the Department of Geography at the University of Minnesota, Minneapolis.

into files for the EPPL7 GIS program (3). Within the GIS the data were reclassed into soil hydrologic groups.

The two land-cover patterns are hypothetical. One map approximates the pre-settlement vegetation, a land cover of forest, grasslands, marsh, and water that might result from speculative ownership prior to development. The other is a planned development, primarily residential, in three density levels, with some retail commercial, institutional, parks, and water. These land-cover maps were designed to be as realistic as possible in both characteristics and geographic pattern. Both were encoded for the same grid-cell points as the soil data. The maps of the soil hydrologic groups and landcover are shown in Figures 1-3. Together these data sets provide the basis for comparison of data handling methods within a GIS.

We examined four data handling procedures to explore their effects on the analysis of the role of land-cover change on runoff. The first type used the soil hydrologic group file in combination with each of the landcover files to produce a two-variable reclass into a single map that represented the SCS curve numbers for the various combinations of the hydrologic group and land cover. The curve-number maps for both land covers were entered into TR55 to derive the water yield for a 5.3 inch rainfall event (the 50-year, 24-hour storm). Each of the resulting pair of runoff maps (2560 point samples) was extrapolated to acre-feet of runoff to form the baseline data set against which the other three data handling techniques were compared (Figures 4 and 5). For the remaining three techniques we abstracted the data to 10 geographic units of one square mile. Distributed element runoff models often have limits on the number of cells they can handle. This requires abstraction of the original data into fewer analytical units. Ten was an arbitrary choice, but does coincide with the maximum number of cells allowable in TR55 version 1.1 when computing a hydrograph.

The second data handling system is a pure tag approach, in which we "tagged" each square mile on each map according to its dominant soil hydrologic group or land cover (For a more complete discussion of the tag vs. count logic, see reference 1). The tagged square-mile cells were then used to produce curve numbers for use in determining runoff, which was converted to total acre feet for the entire study area (Figures 6 and 7).

The third method of data handling is a modified tag method in which the dominance tag for each square mile was applied to derived data rather than the primary data. We determined the SCS curve numbers for each of the 256 sample points within each square mile cell and then examined the data in order to determine the appropriate dominance tag for each square mile. The resulting yields for the two land cover schemes are shown in Figures 8 and 9.

The final method is a pure count method, in which we took a regular sample of 16 points, within each of the 10 sections and produced curve number and runoff maps based on that sample count of runoff values. The choice of 16 sample points per square mile resulted from our desire to use a simple arithmetic sampling procedure that was easily performed within the EPPL7 GIS software. Within cells of 256 points a regular sampling scheme is limited to populations of 1, 4, 16

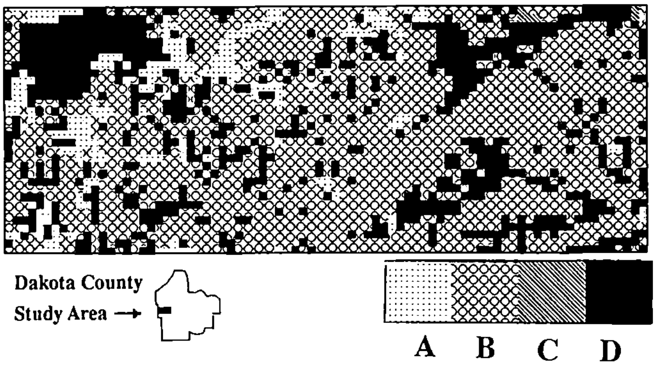


Figure 1. Hydrologic Soil Groups

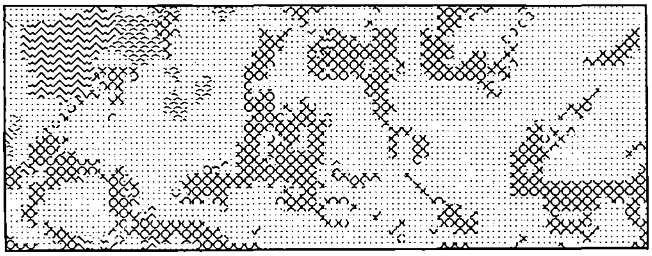


Figure 2. Natural Vegetation Land Cover

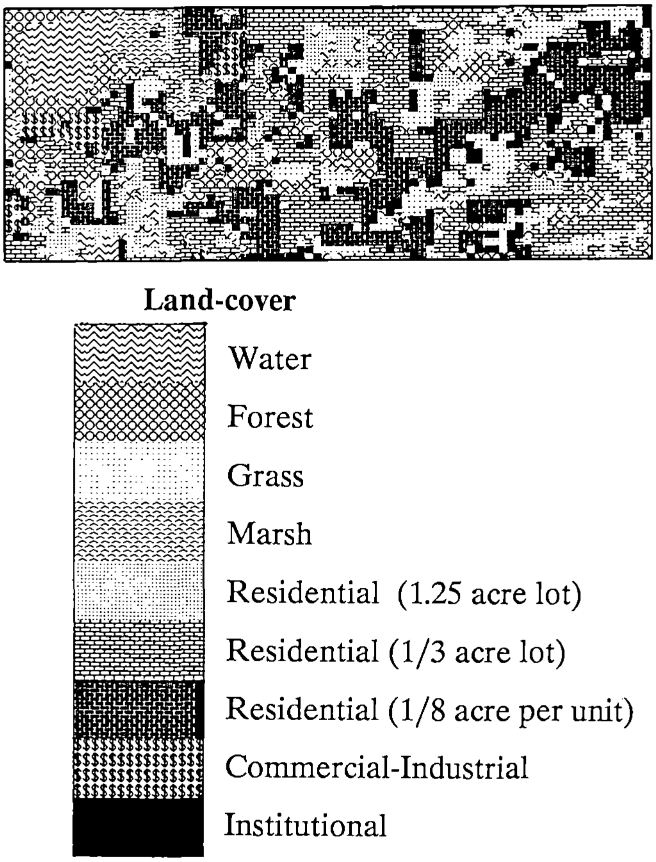
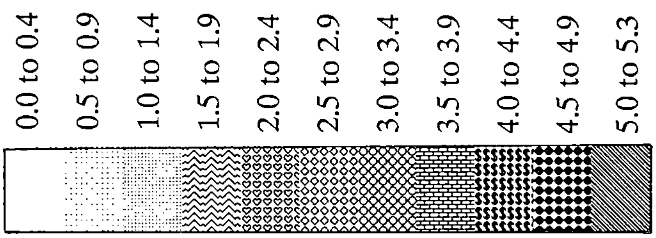


Figure 3. Planned Urban Land Cover

or 64 points, but can be accomplished in a two-step rescale process. A sample of 4 was insufficient to describe such a fine-textured landscape. A sample of 64 was not a sufficient abstraction to illustrate the power of a small sample in a count methodology. The values derived from this method are shown in Figures 10 and 11. Earlier work suggested that a sample of 16 points per square mile may not be quite sufficient to characterize a very complex landscape (4). We reasoned that if this method provides results that were significantly different from the tag approach, the magnitude of that difference should be detectable with a somewhat smaller than optimal sample size.

Runoff from 5.3-inch, 24-hour rainfall in inches



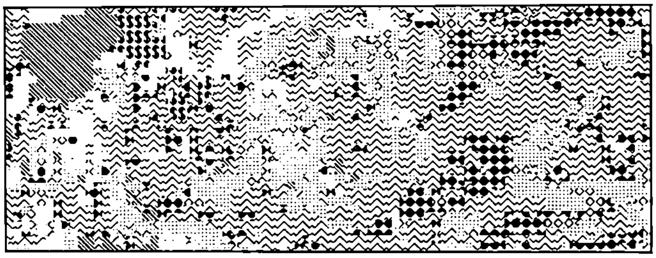


Figure 4. Calculated Runoff for Baseline Data Using Natural Vegetation Land Cover

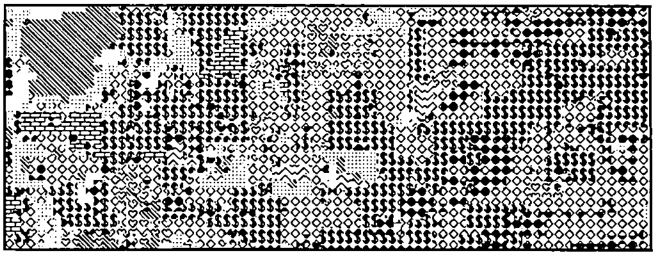


Figure 5. Calculated Runoff for Baseline Data Using Planned Urban Land Cover

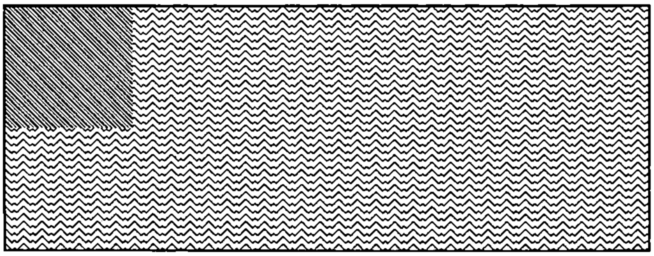


Figure 6. Calculated Runoff for Pure Tag Data Using Natural Vegetation Land Cover

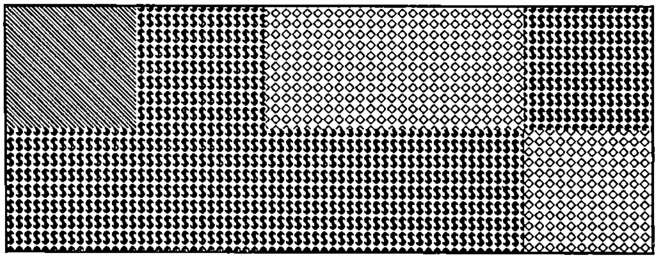


Figure 7. Calculated Runoff for Pure Tag Data Using Planned Urban Land Cover

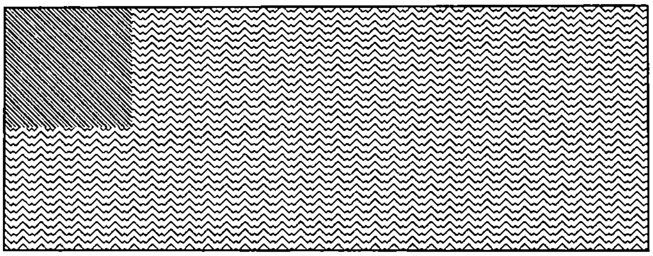


Figure 8. Calculated Runoff for Modified Tag Data Using Natural Vegetation Land Cover

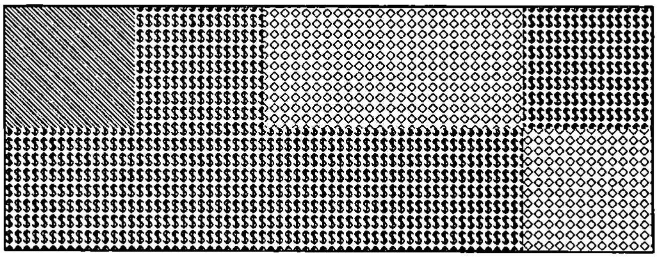


Figure 9. Calculated Runoff for Modified Tag Data Using Planned Urban Land Cover

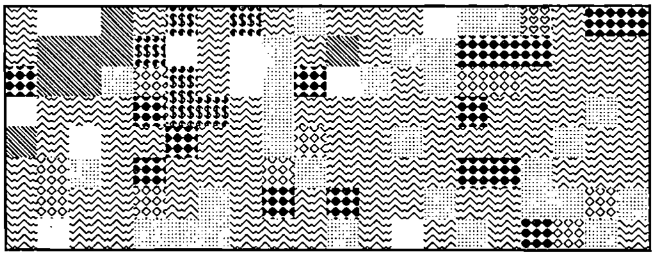


Figure 10. Calculated Runoff for Pure Count Data Using Natural Vegetation Land Cover

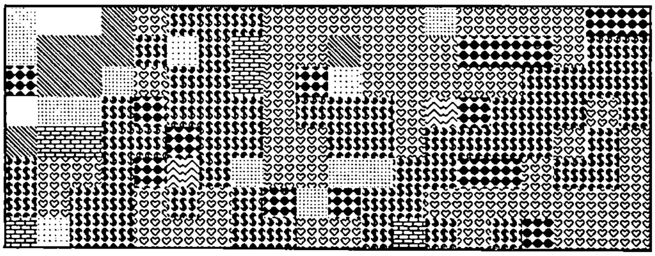


Figure 11. Calculated Runoff for Pure Count Data Using Planned Urban Land Cover

Analyses

Table 1 shows the differences in calculated runoff for the control and for each of the three data handling tests. For the tag methods there is little difference whether the primary data were converted to larger cells at the primary data stage or after the curve numbers were determined. There is a major difference between the results of the runoff analyses between the count and tag approaches. The count approach much more closely matched the product of the control maps.

If we focus on the type of analysis that we were attempting (i.e., how much the change in land cover will affect runoff from the 50-year storm), we conclude that the tag methods suggest increases of 99 and 101 percent (Table 1). The count method suggests an increase of 74 percent, very close to the 73 percent increase derived from analysis of the detailed control maps.

The reason for this discrepancy lies in the nature of the data in relation to the way the one square mile cells are characterized. In a fully developed area, urban features dominate each square mile. Minor areas that are highly absorptive are thus under-represented by the tag dominance approach. The count approach (with the sampling used here) gives features that comprise less than 1 percent of the total area a chance to influence the runoff calculations. The small difference between methods for the natural vegetation pattern is due to the complexity of that data set. The overlaying of four hydrologic groups and only four cover types produces a combination map with larger polygons, and hence few minor features to be misrepresented than are likely with nine or more land-cover classes.

The conclusion that we draw from this analysis is that tag maps, while useful in portraying geographic patterns, are of substantially diminished utility in a quantitative analysis process. This is especially true if the analytical tools are based on a theory that is built around point processes, where some minor features can have an impact on the analysis that is disproportionate to their area. Most water resources models view the hydrologic cycle as a "point process" and, in the instance of runoff, the very high runoff and very low runoff soils are generally of minor areal importance. Because these soils have so much leverage on the outcomes, their under-representation in a tag GIS imposes a hefty price in terms of error.

Table 1. Comparison of runoff Changes based on Three methods of Data Handling in GIS

Runoff			Data Handling Method
Total Acre Feet	Percent of base	Percent Increase Predicted	
Natural Vegetation			
1059	100		Base point, 100m data
939	89		Pure Tag, square-mile
1051	99		Modified Tag, square-mile
1038	98		Count, 6 percent sample
Urban Land Cover			
1628	100	73	Base point, 100m data
1867	115	99	Pure Tag, square-mile
1883	116	101	Modified Tag, square-mile
1635	100	74	Count, 6 percent sample

Recommendations

From the analysis presented here and from other relevant research we suggest several ways to minimize error when using a GIS and GIS data files in analysis of water resources:

- start with primary data to control all phases of data manipulation;
- avoid using tag data for large areas;
- avoid spurious high-resolution data (i.e. data that were rescaled from coarse resolution primary data to a high resolution grid) (1, 5);
- carry no decimal places greater than the crudest of the primary data nor carry any spatial resolution finer than the coarsest of the data sets (1, 6) and
- finally, make no specific site determinations from the analysis based on GIS (or other mapped data).

It is imperative that the targeting of locations be couched in terms like "there is a 30 percent increase in runoff in this square mile, based on a sample of 16 points." This in necessary in order to reflect properly the accuracy standards of primary sources such as topographic maps, soils maps, and climatic data. None of these sources is capable of providing data specific to any selected point (6).

GIS can be of tremendous service if supplied with proper data and if we don't ask it to answer site questions with data that only have regional-analysis capabilities. And those, quite frankly, are the capabilities of most very good maps.

Acknowledgement

Part of the work reported was produced under a grant from the Graduate School, the University of Minnesota to the senior author. We must also express appreciation to the Legislative Commission on Minnesota Resources which funded a Water Resources Research Center project that helped the authors develop the methods used in the research for this paper.

References

1. Gersmehl, P., D. Brown and K. Anderson, 1987. File Structure and Cell Size Considerations for a Water Resources GIS. Chapter 2 in D. Brown and P. Gersmehl, (eds.) *File Structure Design and Data Specifications for Water Resources Geographic Information Systems,* Water Resources Research Center Special Report No. 10, 404 pp.
2. Soil Conservation Service, 1986, Urban Hydrology for Small Watersheds, Technical Release 55, second edition Washington D.C., U. S. Department of Agriculture.
3. Land Management Information Center, 1988, EPPL7, version 1.1, Saint Paul, MN: Minnesota State Planning Agency.
4. Anderson K., J. Corbett, C. Gersmehl, D. Brown, and P. Gersmehl, 1986, Cell Size and Simulation Modeling of water resources at a regional scale, in Bernard J. Nieman, Jr. Editor, *Proceedings, Urban and Regional Information Systems Association,* Denver, 1:203-213.
5. Anderson, K., 1987, *Bear Creek Surface Water Simulation Modeling Demonstration,* Water Resources Research Center Special Report No. 13, 24 pp.
6. Gersmehl, P., J. Corbett, R. Greene, 1987, Soil Data for a Water Resources GIS, Chapter 10 in D. Brown and P. Gersmehl, (eds.) *File Structure Design and Data Specifications for Water Resources Geographic Information Systems,* Water Resources Research Center Special Report No. 10, 404 pp.