Reprinted from

Ninth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Natural Resources Evaluation

June 21-23, 1983

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

INFORMATION REQUIREMENTS FOR EVALUATION AND MANAGEMENT OF WATER RESOURCES

R.M. RAGAN

University of Maryland/Remote Sensing
Systems Laboratory/Department of Civil
Engineering
College Park, Maryland

ABSTRACT

Water resource decision making must reconcile a complex array of often conflicting issues. The quality of these decisions is contingent on the availability of appropriate data and its translation into meaningful information. Cell size and accuracy of a data element must be determined through careful evaluation of the role of the data on the information developed and, in turn, the real use of the information in the decision making process. The paper discusses these issues and presents examples illustrating the consequences of cell size selection and data sensitivity on the behavior of information required for proper evaluation and management of water resources.

I. INTRODUCTION

The evaluation and management of water resources must always reconcile an array of complex and frequently conflicting demands on the system. Decisions leading to the resolution of these demands are developed from a consensus among representatives of the agencies of government and private interests concerned. Obviously, these decisions must be based on comprehensive evaluations of the problems involved. It is fundamental that the data used to support the decision making process must be timely, of sufficient detail to define the problem, and in a form that can be readily used. Relative to the forms for use, there is increasing acceptance that the data must be put into digital formats to allow more efficient manipulation and analysis with computer aided techniques. The availability of computer aided techniques provides tremendous opportunities and temptations to break a problem into very small components and simulate its behavior over extended time series. One can then draw conclusions concerning the behavior of the system under an array of alternate operating strategies.

The availability of these opportunities require that we start giving more attention to what is meant by data quality of "sufficient detail to define the problem." We now have the power to use computer based systems having increasingly smaller grid cell sizes or time steps for "improved" problem definition or simulation. Our computer powers also allow us to run increasingly sophisticated mathematical models that can be used to provide information to aid the decision making process. However, the water resource analysts and, probably the analysts in many other disciplines, frequently do not spend the level of effort needed to fully understand the relationship between the needed level of data definition and the accuracy and use of results from a model simulation. Information requirements for the evaluation and management of water resources vary depending on the scope of the project and must be based on an understanding of the real use of the data-information, accuracy requirements, and sensitivity of the information in the decision making process.

This paper explores some of the issues on information requirements for water resources by centering on the role of data accuracy within grid cells and the impact of these accuracies on synthesized information used in the decision making process. The paper is general in scope and does not pretend to go into enough depth to really resolve the problems presented. Its intent is to point out some of the issues concerning the data-information interfaces that have not received enough attention in the water resource arena and, hopefully, stimulate discussion and more in-depth research that will lead to a better, more realistic, utilization of our machine aided processing capabilities.

II. BASIS FOR DECISION MAKING IN WATER RESOURCES

There have been numerous papers that define the steps to be followed in comprehensive water resource analyses. The steps outlined in Table 1 provide the reader with an overview of the sequence of events involved in the development of a water resource planning or management program.

As in any well conceived operation, the first steps are to clearly state the problem and develop a set of specific objectives. The translation of these objectives into a series of decisions that will control the program must develop through analyses involving sophisticated simulation models as well as subjective human interaction. These analyses cannot lead to sound decisions unless they are based on appropriately defined data and an effective means of translating that data into the information required for the decision making process.

Table 1. Steps in the Development of a Water Resource Planning or Managment Program

1. Definition of Problem and Objectives
   - Statement of Problem
   - Statement of Objectives
   - Definition of Scope
2. Definition of Data-Information Requirements
   - Evaluate Sensitivity of Decisions to Data
   - Real use of Data-Information
   - Define Accuracy Requirements
   - Define Modeling Requirements
   - Adopt Information Management-Modeling System
3. Assemble and Format Raw Data Base
   - Hydrologic
   - Land use/cover
   - Existing Systems
   - Socio-Economic
   - Geomorphic
   - Socio-Political Constraints
4. Finalize Information Management-Modeling System
   - Input and Check Data
   - Test System Flexibility and Performance
   - Refine
   - Develop User Understanding and Confidence
5. Evaluation and Forecast Modeling
   - Simulate Alternate Strategies
   - Regional vs. Critical Zone Analyses
   - Interpretation and Feedback
6. Program Adoption Through Consensus

The second step in Table 1 addresses the definition of data-information requirements. The importance of this step and its consistency with the conditions of step 1 cannot be overemphasized. Traditionally, we encounter situations where decisions have had to be made in an arena of inadequate data. With new technologies, however, we now have the capability to overdefine data or information relative to the ability of the analytical procedures to reflect the inputs. Thus, it is critical that rather detailed analyses be undertaken to evaluate the sensitivity of the anticipated decisions to the accuracy of the data. At this point, water resource analysts must decide what the real use of the data is going to be and what information is to be obtained from it. Then, the modeling requirements can be ascertained. A key element in this second step and the subsequent effectiveness of the steps leading toward a final set of decisions is the adoption of an information management-modeling system. Mathematical modeling has a key role in modern water resource analysis. Indeed, there are so many conflicting issues involving complex interrelationships that it is difficult to properly understand a system without mathematical modeling. The models are defined by massive quantities of data from an array of sources. Thus, an efficient, flexible, computer-based system must be adopted that can readily manipulate data, translate it into needed information, enter it directly into the models, and output the results in a format required by the decision makers.

The third step of Table 1 recognizes the importance of assembling and formating the raw data base. Data concerning the streamflow, rainfall, snow, groundwater and other hydrologic variables will come from a variety of sources. Land use and land cover and existing systems, such as utillities or transportation, may be in the form of digital data from orbiting satellites or in hard copy as maps or tabulated data. Socio-economic data may be from census tracts, tax maps, etc. The geomorphic requirements refer to the need to define slope, soil type, the channel network, etc. The term "socio-political constraints" recognizes the need to bring in information concerning decisions that have been made or are being made in related arenas. This would include master planning for land use, proposed developments, etc.

The next step is to develop the final form of the infomration management-modeling system. It is mandatory that the data be checked to assure accuracy. It is probable that the system adopted will have to be refined or restructured to meet the needs of the particular task involved. A key element is the development of user understanding and confidence. Water resource decisions are made through consensus. In most water resource arenas, there are an array of political and private institutions having needs and desires that must be reconciled. Thus, at this point, the parties involved in the decision making must understand and gain full confidence in the procedures that will support their efforts.

Steps 1 through 4 lead to step 5 which is the key element in translating data into usable information. It is at this point that the mathematical models and human interaction come together to evaluate the current conditions and develop forecasts of what might be in the future under different programs. The models, assuming that they are consistent and defined in terms of adequate data, provide a mechanism under which the consequences of alternate strategies can be better understood by the decision makers. One level of data may be adequate to define the behavior of the overall region. However, there will no doubt be critical areas such as heavy urban development in a flood plain that will require much more detailed data and significantly stronger analytical techniques. Thus, the evaluation and forecasting at this stage will involve a number of different levels of modeling strategies and information requirements.

As the results from step 5 become available, they move into the user arena where interpretations are made in terms of human perceptions of need and financial constraints. These final decisions can be effective only if each of the above steps has been completed with a realistic data quality-information accuracy interface.

III. DATA QUALITY-INFORMATION ACCURACY INTERFACES

In order to gain some insight into the problem of data quality-information accuracy interfaces, two examples will be presented. First, we will explore the idea of trying to simulate the spatial distribution of rainfall infiltration within a watershed by defining the initial soil moisture as an array of small grid cells. We will then examine the consequences of aggregating data into larger cells on the resulting watershed runoff coefficents and the estimated peak flood flows.

A. SPATIAL DISTRIBUTION OF INFILTRATION

There is a general perception that information on soil moisture would be invaluable for improving the day to day operations of water resource systems. Research indicates that the measurement of surface soil moisture with remote sensing techniques is an achievable goal in the not too distant future. The role of soil moisture in estimating the amount of rainfall that will infiltrate into the soil during a storm provides an excellent base for examing some of the subtleties between cell resolution, accuracy of a data estimate and its value as information related to a water resource decision.

Frequently, water resource professionals want to subdivide a watershed into relatively small units in order to reflect their perceived understanding of the role of spatial variability. Suppose, for example, the spatial distribution of parameters across the watershed is defined as an array of rectangular cells. The water resource analyst may want to use very small cells in order to better model the consequences of one particular process. He recognizes that he is going to have to pay a premium in computer running time because of the increased quantities of data, but he concludes that it is worth the effort because of the "improved definition" of the process. However, the interaction between the sensitivity of the process and the accuracy of the data to reflect the true conditions within the cell being considered is often neglected or misunderstood.

Figure 1 illustrates this problem using sensitivity of infiltration to the initial soil moisture content. The rainfall excess, Q, that portion of the rainfall that cannot enter soil and must be considered in flood flow computations, is expressed as a percent of the rainfall excess occurring when the initial moisture is .25 cm/cm. These curves show that the rainfall excess is sensitive to small errors in the estimate of the initial soil moisture, especially for relatively small storms. For example, an error of .03 cm/cm in the estimate of the initial soil moisture results in an error of about 50 percent in the rainfall excess predictions for a one-hour, 5.1 cm rainfall and about 10 percent for the 3 hour, 7.6 cm storm. Thus, if one is attempting to model the partitioning of water entering the soil versus water flowing overland toward the channel network, he must have a very good estimate of the soil moisture within the cell of interest if his results are to be meaningful for average rainfalls.

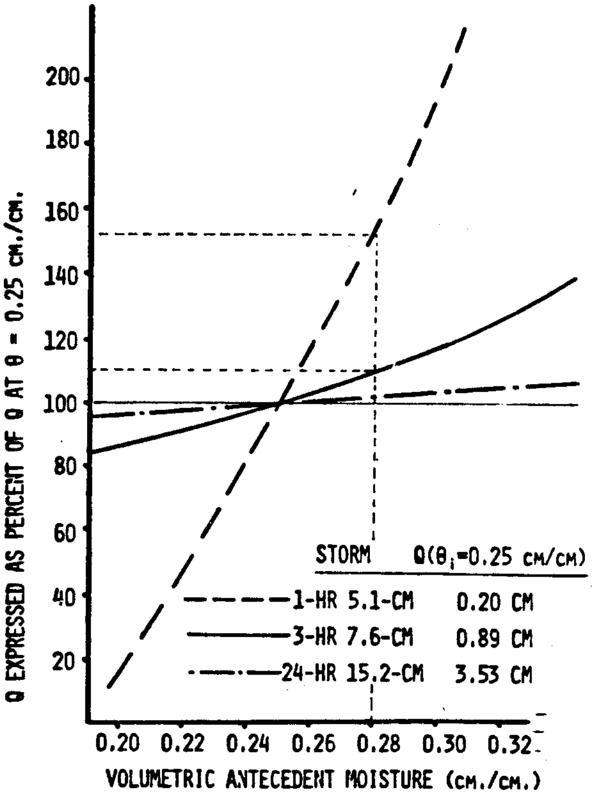

Figure 1. Percentage Variation in Excess Rainfall as Function of Error in Estimate of Antecedent Moisture

This problem of infiltration and soil moisture estimating is complicated by the spatial variability of the soil moisture within a small area. Figure 2 is an example of soil moisture variation within a 23 acre watershed administered by the USDA in Chickasha, Oklahoma. The mean soil moisture in this 23 acre area is 0.25 cm/cm with a coefficient of variation of 14 percent.

In order to illustrate the hazards of being overly ambitious in moving toward smaller and smaller cells, the following experiment was conducted. First, it was assumed that it was desired to represent the 23 acre watershed of

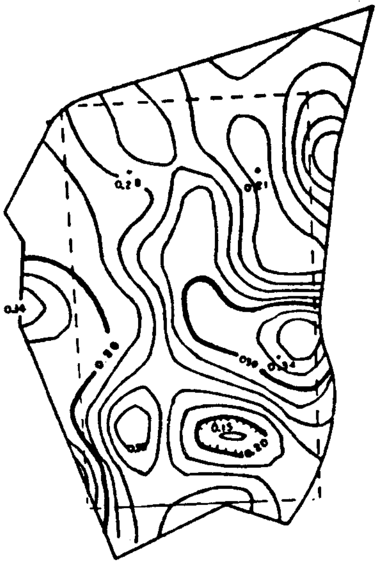

Figure 2. Soil Moisture Variation (top 5 cm)
Watershed R5 Chickasha, Olkahoma
(23 acres)
May 10, 1978

Figure 2 as an array of very small cells in order to "properly define" the influence of interconnections in the stream network on timing. At the same time, the initial soil moisture was assumed to be available only from some remote sensing device that had a footprint significantly larger than the watershed itself. In other words, the only real initial soil moisture data that would actually be available would be the 0.25 cm/cm mean value. Even though we could not measure it, the soil moisture actually varied in accordance with the distribution shown in Figure 2. The objective of this experiment was to pick a cell size such that one could assume a soil moisture of 0.25 cm/cm for each cell and not produce a rainfall excess for any cell that was more than 10 percent in error for the three hour 7.6 cm storm of Figure 1.

The soil moisture of Figure 2 was represented as an array of 2304 cells. When this cell size was used, 40 percent of the cells had actual soil moistures that resulted in rainfall excesses beyond the 10 percent accepted limit. The cells were then aggregated into 2 X 2 matrices and the mean moisture of the 4 cells considered to be the representative value. Each of these cells comprised 0.13 acres or 0.75 percent of the total watershed area. Thirty-three percent of these cells were still beyond the 10 percent threshold accepted as the allowable error in the rainfall excess. When the cells were aggregated into a 3 X 3 matrix of 0.51 acres or three percent of the total area, 20 percent of the cells exceeded the limit. Finally, when a matrix of 6 X 6 cells totaling 1.87 acres or 11 percent of the watershed area were used, all of the cells produced rainfall excesses that were within the 10 percent allowed value.

By aggregating the cells into larger units, the variability in soil moisture was smoothed to produce estimates that approached the mean of the entire area. However, trying to extend the limit limited soil moisture data into very small cells produced errors that would overshadow other processes if the 10 percent threshold was accepted.

B. DATA AGGREGATION INTO LARGER CELLS

While one set of goals may inspire the water resource analyst to strive toward progressively smaller cells, economics and difficulties in defining the data may lead him to aggregate data into progressively larger cells. Figures 3 and 4 are from the Ph.D. dissertation of Mr. Jack D. Fellows of the University of Maryland's Remote Sensing Systems Laboratory. These figures illustrate the consequences of attempting to aggregate the data into cells that are too large to meet the needs of an analytical procedure. Mr. Fellows conducted a series of experiments on the applications of geographic information systems to water resource problems that included an analysis of the consequences of cell size. His experiments centered on assessing the errors caused by increasing a cell size from 30 meters to 120, 220 or 300 meters. Two hundred and thirty-seven watersheds were involved in his analysis.

Figure 3 illustrates the change in a runoff coefficient, termed a curve number in the USDA Soil Conservation Service models, as the cell size changes. Curve number is a function of the land cover and the soil type. In Figure 3, the curve number defined by data tabulated in the array of 30 meter cells was accepted as the "correct" value. In essence, the results of the 30 meter cells would be zero error of the abscissa. The estimated curve numbers of each watershed for each cell size were then compared against the 30 meter data to determine the error created by the aggregated data. Figure 3 shows that as the cell size is increased there is increasing departure in the absolute error of the curve number in the smaller watershed sizes but that there is a tendency toward convergence as the watershed size increases. This trend is to be expected because there are very few cells in the small watersheds but there are large numbers of cells in the larger watersheds regardless of the resolution. At this point, one would conclude that the size of the cell does not produce serious departures in the mean land cover for the larger watersheds.

If one then wishes to use the SCS model to estimate the peak discharge that would result from a storm, elevation must also be considered along with land cover and soil type. As one aggregates cells and assigns a single elevation as being representative of the topography in each cell, significant smoothing takes place. One loses the terrain variability that is imbedded in the 30 meter cell. Because of the sensitivity of the SCS model to land slope, this elevation term is quite significant. Even though the errors

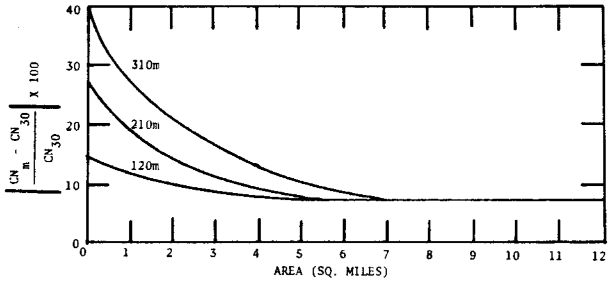

Figure 3. Changes in Estimated Curve Number Caused by Increasing Cell Size above 30 Meters

illustrated in Figure 3 are under 40 percent errors in the peak discharge rates caused by the smoothed elevation data are much larger as illustrated by Figure 4. In other words, if one had a three square mile watershed with this data, aggregating from a 30 meter cell to a 120 meter cell could produce changes in the estimate of the peak discharge up to approximately 18 percent. Going to the 210 or 300 meter cells, this figure would rise to approximately 25 percent. As the watersheds become smaller, the errors grow massively.

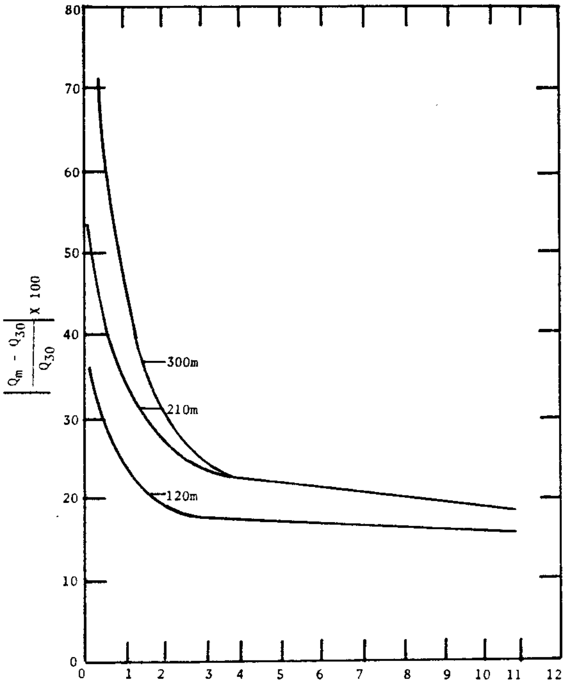

Figure 4. Changes in Peak Discharge Resulting from Increasing Cell Size Above 30 Meters.

IV. CONCLUSION

As our water resource projects grow in complexity and our recognition of the possible consequences of our decisions increases, consistancy between the quality of the data and the information it is to support must also increase. Fortunately, remote sensing techniques and computer-based geographic information systems have the potential to provide data that is both current and of high quality. Because we now have the technological power to define very accurate data and manage massive quantities, we must improve our understanding of what constitutes an adequate, yet economic, data base. In essence, data must be consistent with the capabilities of the analysis which, in turn, must be appropriate for the level of the decision to be made. Consistency is extremely important. We cannot simply strive to define data of higher resolution to fit into geographic information systems having smaller and smaller cells. There are going to be situations where the quality of data or sensitivity of a parameter relative to its role in process simulation does not justify moving toward a small cell size. However, in other situations the degradation of data caused by aggregating into larger cells may lead to results that have large errors that, in turn, would lead to erroneous conclusions concerning the behavior of the system.

This data quality-information accuracy interface issue has major implications with respect to the cost of a water resource analysis and the ultimate operational decision. Our remote sensing and geographic information systems capabilities are increasing dramatically. We are now at a point where research is needed to open the path to disciplined decisions on how much of this increasing capability to use in a given situation.