HYDROLOGIC MODELING THROUGH THE INTEGRATION OF REMOTELY SENSED DATA IN A GEOGRAPHIC INFORMATION SYSTEM

Mark S. Gregory
and
John D. Vitek
School of Geology
Oklahoma State University

A - 112

University Center for Water Research
Oklahoma State University
Stillwater, Oklahoma

August 1989
Acknowledgement:

The activities on which this report is based are financed in part by the Department of the Interior, U.S. Geological Survey, through the Oklahoma Water Resources Research Institute. The contents of this publication do not necessarily reflect the views and policies of the Department of the Interior, nor does mention of trade names or commercial products constitute their endorsement by the United States Government.
Runoff from non-urban land which enters portions of stream systems that are not controlled by reservoirs can have significant impacts on the timing of water released from flood control structures. Precipitation that occurs over portions of drainage basins not controlled by flood structures influences the manner in which downstream reservoirs are operated. Planners need this upstream information to effectively control the movement of water during extreme events whereby the impact of flooding can be minimized. With the ever increasing availability of computing power in the personal computer environment, more sophisticated and hopefully more accurate models are becoming available to local engineers and managers. The present research will demonstrate that diverse data can be integrated into a geographic information system (GIS) for the purpose of deriving information vital to managers of flood control structures.

With regard to improving runoff estimates, increased attention has recently been given to physically-based models of catchment hydrology (see Anderson and Rogers, 1987). Parameters for physically-based models are generally derived or measured as attributes of the land, such as land use, soil type, and topography. From these attributes a variety of parameters including infiltration rate, slope angle, and drainage pattern can then be derived. In many instances, however, common models lump parameters in order to describe catchment characteristics, even when modeling very small catchments. This procedure assumes a spatial homogeneity that rarely actually exists in real-world conditions. Admittedly, many micro-scale interactions or characteristics are poorly understood or are too complex to measure or accurately model. The opposite extreme from lumping parameters is to employ an ever-increasing number of parameters to model the runoff process. As the number of characteristics or parameters increases, however, the model often becomes extremely complex both conceptually and spatially.

Aside from conceptual problems involved with hydrologic modeling, problems often exist with regard to data acquisition and parameter processing and interaction. The development of remote sensing and GIS technologies offers great promise toward providing the tools necessary to include greater spatial detail and complex model interactions in hydrologic modeling. The present research employed these technologies in order to efficiently acquire and integrate diverse data into a distributed model for predicting hydrographs for runoff from non-urban land. Remote sensing and GIS technologies offer the means for production and integration of data characteristics critical to accurate runoff estimation. Some examples of characteristics utilized from these technologies include land use data, digital elevation data, and soil mapping unit information. Analysis of these data layers and their derived parameters were coupled with recorded rainfall data and mathematical equations for the amount and rate of runoff estimated for each cell in the analysis. The estimated runoff from each cell was accumulated and routed across a derived drainage network to simulate a runoff hydrograph. This research also investigated the scale of the analysis (size of the smallest unit or cell) and the effect on certain model outputs that results from a change in scale.
The result of this research is a model capable of estimating a runoff hydrograph for any location within the given area of study. The model is based upon characteristics of previously developed lumped models, now implemented in a more distributed manner. The distribution is carried out by incorporation of digital elevation data for development of a derived flow network of the drainage area. Model inputs are minimal and can easily be updated with remote sensing techniques or GIS manipulations. Finally, the model is highly automated to allow easy use by planners and managers concerned with runoff estimates.

BACKGROUND

The basis for this research is the ability of computer and remote sensing technology to efficiently and accurately generate detailed data layers significant to the estimation of runoff. These data and other ancillary data can then be manipulated through the use of a GIS for generation of runoff hydrographs.

Remote Sensing

Hydrologic modeling has become increasingly complex since Horton's overland flow theory (Horton, 1933). Hewlett's variable source area concept (U.S. Forest Service, 1961) represented the beginning of increasingly complicated models to account for the complex interactions of watershed runoff (Pearce et al., 1986 and Linsley, 1986). The increasing availability of personal computing power has allowed a move toward more physically-based, distributed models of increasing complexity (Beven, 1989). For almost all such models, however, land cover or land use is generally a significant parameter to determine the quantity or timing of runoff, as exhibited by the SCS runoff curve number model (S.C.S., 1972). As these and other models become more complex and data requirements more detailed, greater utilization has been made of automated data acquisition technologies such as remote sensing.

Remote sensing technology has continued to rapidly expand from the early days of aerial photography to satellite and airborne sensors with increasing detail and sensitivity of today. Numerous examples exist where remotely sensed data from satellites have provided much useful data for hydrologic modeling, including land cover information and information about soils, infiltration, and evaporation (Ottle et al., 1989). Consequently, much research has simply involved only a slight modification of existing hydrologic models to incorporate remotely sensed data (Engman, 1984).

Data from the Landsat satellite series was a logical choice as input to hydrologic models for several reasons. Since the satellite launch in 1972, data has generally been available for all parts of the world every 18 days. The spectral characteristics of the multispectral scanner (MSS) allowed for relatively high classification or mapping accuracy of land cover types found to significantly affect runoff (Alexander and Rao, 1985 and Harvey and Solomon, 1984). The 79 meter spatial resolution of the MSS data also resulted in a finer resolution than was previously available in an automated format. The digital characteristics of MSS data also allowed for easy computer analysis and incorporation into other
analysis schemes such as a GIS. The result was that many researchers attempted to improve runoff estimates by inputting Landsat MSS data directly into existing models such as the SCS runoff curve number model (Jackson and Bondelid, 1984 and Slack and Welch, 1980) as well as the U.S. Corps of Engineers STORM (Ragan et al., 1976). These and numerous other studies involving Landsat data identified a generalized set of relevant land cover classes which could be accurately mapped with satellite data, including highly impervious, residential, forest, agriculture, bare ground, and water (Troilier and Philipson, 1986). Whereas generalized land cover classes may satisfy certain hydrologic models, a more detailed classification or mapping scheme and an improved spatial resolution may be required to improve model accuracy or to extend the use of satellite data to more complicated models.

Similar in operation to the multispectral scanner but more detailed in spatial resolution is the next generation of U.S. civilian satellite sensors, the Thematic Mapper (TM). The TM scanner is now also on-board the Landsat satellite series but offers digital data with 30 meter (approximately 0.25 acre) spatial resolution compared to the 79 meter (1.1 acre) resolution of the MSS. In addition, greater spectral discrimination is available with TM data because the sensor has narrower and more spectral bands (seven versus four) than MSS data. The narrower spectral bands allow for less atmospheric attenuation of the recorded signal and result in a greater signal-to-noise ratio for the TM scanner. The additional spectral bands allow the TM scanner to sense and record information in portions of the electromagnetic spectrum previously not measured by the MSS (Sabins, 1987). These improvements in sensor number and resolution have resulted in improved spatial detail and accuracy of land cover analyses as they relate to hydrologic applications (Gervin et al, 1985 and Quattrochi, 1983). Thematic Mapper digital data was utilized in the present research to produce a digital data base of detailed land cover information as input into the hydrologic model.

Geographic Information Systems

All of the data for this research was managed through the use of a geographic information system (GIS). While Tomlinson (1972) may have begun the process of formal recognition of GIS, the technology was already evolving at that time as it still is today. Numerous discussions have been published concerning the origin and definition of a GIS (for example, see Cowen, 1988) as well as an international journal devoted specifically to the subject (Coppock and Anderson, 1986). For the present research, a GIS was defined as an information technology which can capture, store, analyze, and display spatial data (Parker, 1988).

GIS technology has been utilized for several years in a variety of hydrologic applications including sediment transport models (Beasley et al., 1982 and Gilliand and Baxter-Potter, 1987). Similarly, a variety of groundwater studies have also employed GIS technology to model contaminant flow or potential for groundwater contamination (see Baglio and Meade, 1986 and Broten et al., 1987). Conversely, fewer examples exist where GIS technology has been employed for
distributed watershed models concerned with surface runoff (Vieux, 1986). Most likely for rainfall-runoff applications is the situation where a GIS has been utilized to derive new data layers such as slope and erosion potential from original data sources such as topography and soil maps (Ventura et al., 1988). This latter case best describes the use of the GIS in the present research because model development will involve satellite derived land cover information integrated with soil mapping unit information from standard SCS county soil surveys, gaged precipitation data, and digital elevation data from U.S.G.S. topographic maps.

Digital Elevation Model

A key element of this research was the use of digital elevation data and subsequent development of a digital elevation model (DEM). DEMs have often been utilized in conjunction with Landsat data to improve land cover analyses, particularly in mountainous areas where elevation, slope angle, and slope aspect may affect the distribution and thus accurate mapping of Landsat derived land cover types (Guiden et al., 1981 and Straehler et al., 1978). Other areas where DEM data have been successfully combined with satellite derived land cover information have involved the estimation of non-point source pollution and sources of soil erosion in a GIS based analysis (Daniel et al., 1982 and Gesch and Naugle, 1984). More important to the present research are the efforts demonstrated by Band (1986) and Palacios-Velez and Cuevas-Renaud (1986) to derive topographic networks from digital elevation data.

Generation of topographic networks based on a grid have been accomplished with varying success in a number of studies dating back to the pioneering work of Leopold and Langbein (1962). It has also long been known that local topography can strongly influence the location and magnitude of surface runoff (Dunne and Black, 1970). Only recently, however, great attention has been given to the automatic derivation of drainage boundaries from digital elevation data (Marks et al., 1984 and Band, 1986). Similarly, the automatic derivation of drainage networks from digital elevation data is also a relatively recent phenomena (O'Callaghan and Mark, 1984 and Morris and Heerdegen, 1988). Following these efforts, researchers next attempted to derive topographic structure from DEMs such as the work by Jenson and Domingue (1988) and Band and Wood (1988). The final step in this developmental process is to incorporate DEM derived data into hydrologic models for runoff estimation (Band, 1989 and Beven et al., 1984). Whereas a great deal has been accomplished in this area of distributed hydrologic simulation, much work remains to accurately model surface runoff by utilizing digital elevation data, particularly in areas of minimal or reduced slope.

STUDY AREA

The study area selected for use in this project is the upstream portion of the Little Washita River drainage basin in south-central Oklahoma (see Figure 1). Specifically, only that portion of the watershed above gage 526 was utilized in this study. The area is a transition between the cross timbers and tall grass prairie ecoregions. The result is a variety of land cover types dominated by agriculture.
Figure 1. Little Washita River study area.
(cropland and pasture) and rangeland, with scattered forest and no major town or urbanized area. Topography varies from flat to broadly and steeply rolling. The climate is moist, sub-humid with a mean annual temperature of 17 degrees Celsius and mean annual precipitation of 95 cm.

The Little Washita basin has a long history of study by several government agencies. As a result, an extensive network of rain gages is located throughout the area and several years of detailed records exist for precipitation and gaged outflow from the basin. The precipitation data from the rain gages, indicated on Figure 1, were utilized as input into the developed model. Corresponding outflow data recorded at gage 526 was used to assess the accuracy of the runoff estimated by the model.

### DATA INPUTS

Thematic Mapper digital data from Landsat 5 was acquired for July 21, 1984, (scene identification number 50144216365). Data was reformatted onto the Center for Applications of Remote Sensing computer system and analyzed with the Earth Resource Laboratory Application Software (ELAS). The digital data were geometrically corrected to remove skew inherent in Landsat satellite data. During the geometric correction process, the data were resampled to a 30 meter by 30 meter resolution and referenced to a standard geometric grid system. The geometric reference system utilized was the Universal Transverse Mercator (UTM) system. An unsupervised classification procedure was employed to derive a map of land cover classes over the study area. After field checking and further analysis of the initial results, a final classified map of landcover was generated as input into the GIS.

Information about soil characteristics was acquired by digitizing standard SCS soil surveys for the study area. Because of the location of the watershed, portions of Caddo and Comanche counties were digitized. The soils from the two counties were correlated and a single soil legend of mapping units was formed. Soil data were entered into the GIS in a grid cell format in both 30 and 200 meter resolutions. Through software manipulation, these data were resampled to larger resolutions for analyses concerned with the effect of cell size on certain runoff models. From the basic soil data layer, information concerning soil texture, permeability, and infiltration capacity can also be derived within the GIS.

Precipitation and stream flow data were received from U.S.D.A. Agriculture Research Service for the entire year of 1984. Precipitation was allocated to cells across the study area by a Thiessen polygon method for the 14 rain gages in or around the study area. The amount of the study area covered by each polygon is listed in Table 1. Also included in the table are the recorded amounts of precipitation for two rainfall events of 1984 for which model results were computed.
TABLE 1 - PRECIPITATION BY THIESEN POLYGON

<table>
<thead>
<tr>
<th>Gage Number</th>
<th>Area (percent)</th>
<th>Amount (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>130</td>
<td>1.2</td>
<td>3/23 5.16</td>
</tr>
<tr>
<td>131</td>
<td>6.5</td>
<td>10/26 2.74</td>
</tr>
<tr>
<td>132</td>
<td>6.4</td>
<td>5.16 4.98</td>
</tr>
<tr>
<td>133</td>
<td>0.1</td>
<td>2.74 4.98</td>
</tr>
<tr>
<td>148</td>
<td>7.2</td>
<td>5.38 4.98</td>
</tr>
<tr>
<td>149</td>
<td>17.6</td>
<td>1.24 1.27</td>
</tr>
<tr>
<td>150</td>
<td>14.5</td>
<td>1.24 4.17</td>
</tr>
<tr>
<td>151</td>
<td>2.8</td>
<td>2.59 4.37</td>
</tr>
<tr>
<td>152</td>
<td>6.2</td>
<td>2.74 3.58</td>
</tr>
<tr>
<td>153</td>
<td>14.5</td>
<td>1.68 3.86</td>
</tr>
<tr>
<td>154</td>
<td>10.9</td>
<td>1.35 4.09</td>
</tr>
<tr>
<td>162</td>
<td>4.9</td>
<td>0.94 4.60</td>
</tr>
<tr>
<td>163</td>
<td>6.7</td>
<td>1.09 4.95</td>
</tr>
<tr>
<td>164</td>
<td>0.5</td>
<td>1.02 3.84</td>
</tr>
</tbody>
</table>

Digital elevation data were acquired from U.S.G.S. on magnetic computer tape at a scale of 1:250,000. A major goal of this past year's research effort was to acquire digital elevation data at 1:24,000 scale. DEM data at 1:24,000 scale would greatly enhance the present research as it is sampled at a 30 meter resolution, the same as the TM derived landcover data. Topographic data at this larger scale, however, was not available for the entire watershed. Likewise, the cost of manually acquiring such detailed data for the selected study area would have been beyond the scope of this research.

INVESTIGATION OF CELL SIZE

Prior to incorporation of the digital elevation data into the model, an attempt was made to determine the effect of input scale on model output. This limited investigation was performed by varying the cell size of input data for the SCS runoff curve number model. Once the above mentioned data layers were entered into the GIS, the data were manipulated and resampled into varying cell sizes. The GIS format allowed for rapid computation of weighted curve number (CN) values as standard input into the SCS runoff model (SCS, 1972). The amount of the drainage basin covered by the Landsat derived land cover categories in combination with corresponding hydrologic soil group is listed in Table 2. The terms in parentheses refer to CN definitions. The information in this table represents the results for a basic cell size of 30 meters. Similar data was computed for several cell sizes ranging up to 1600 meters.
TABLE 2 - PERCENT AREA BY LAND COVER/HYDROLOGIC SOIL GROUP

<table>
<thead>
<tr>
<th>Land Cover Type</th>
<th>Hydrologic Soil Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Cropland (type 1)</td>
<td>0.3</td>
</tr>
<tr>
<td>Cropland (type 2)</td>
<td>0.1</td>
</tr>
<tr>
<td>Pasture (good)</td>
<td>0.0</td>
</tr>
<tr>
<td>Rangeland (good)</td>
<td>0.7</td>
</tr>
<tr>
<td>Rangeland (fair)</td>
<td>0.1</td>
</tr>
<tr>
<td>Rangeland (poor)</td>
<td>0.1</td>
</tr>
<tr>
<td>Brush</td>
<td>0.2</td>
</tr>
<tr>
<td>Forest (fair)</td>
<td>0.3</td>
</tr>
<tr>
<td>Urban/Industry</td>
<td>0.0</td>
</tr>
<tr>
<td>Urban/Residential</td>
<td>0.0</td>
</tr>
</tbody>
</table>

The percent of the basin by land cover type changed markedly with changes in cell size. The resulting weighted curve number for the entire watershed, however, was relatively unchanged until computed for the largest cell sizes. See Table 3. An even smaller change in CN was observed for a Grant County watershed where a priori inspection revealed a homogeneous landscape dominated by cropland (wheat). A greater range in weighted CN values, however, was found for an apparent heterogeneous watershed located in Lincoln County. For this watershed a large change in CN did not occur until the resampled cell size reached 800 meters. The 800 meter cell size is quite large compared to the data resolution presently available from satellite sensors. For all three of the watersheds, the weighted curve number remained relatively unchanged until the cell size exceeded 400 meters. For any given watershed, such a small deviation in CN would result in an SCS Curve Number model estimate of approximately the same value of total runoff.

TABLE 3 - BASIN AVERAGE WEIGHTED CURVE NUMBER

<table>
<thead>
<tr>
<th>Resolution</th>
<th>L.Washita</th>
<th>Grant Co.</th>
<th>Lincoln Co.</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 m.</td>
<td>70.0</td>
<td>88.9</td>
<td>77.0</td>
</tr>
<tr>
<td>100 m.</td>
<td>69.3</td>
<td>88.9</td>
<td>77.1</td>
</tr>
<tr>
<td>200 m.</td>
<td>69.3</td>
<td>88.8</td>
<td>77.7</td>
</tr>
<tr>
<td>400 m.</td>
<td>69.8</td>
<td>88.2</td>
<td>77.4</td>
</tr>
<tr>
<td>800 m.</td>
<td>68.3</td>
<td>89.3</td>
<td>71.7</td>
</tr>
<tr>
<td>1600 m.</td>
<td>67.3</td>
<td>90.2</td>
<td>69.0</td>
</tr>
</tbody>
</table>

DRAINAGE NETWORK DERIVATION

Key to model development was the incorporation of digital elevation data. Digital elevation data can be represented in three different forms, uniform grid data, irregular triangular networks (TIN), or vectors of contour data. Each form of capture and representation of digital elevation data has advantages for computer processing and for topographic representation.
For the present research, the elevation data utilized was in a uniform grid format. The data were interpolated in a straight-line linear fashion between contour lines into a grid of points equally spaced in both x and y directions. The interpolated values were represented as the z value in a three-dimensional matrix. Organization of the elevation data in a uniform grid generally allows for greater flexibility and speed of computation compared to contour lines or TINs. Consequently, a uniform grid format was utilized as input into a computer algorithm that was written to automatically determine the drainage network for the study area. Throughout this research, each data point or matrix value of elevation was considered to represent the center of a uniformly shaped (square) grid cell.

To begin the procedure, a rectangular grid of elevation points was selected so that the study area (watershed) was within the matrix of data. The first step in the algorithm involved the determination of the direction of flow for all cells except those cells on the outside edge of the matrix. Direction of flow from cells on any outside edge of the matrix cannot be determined because of incomplete information concerning the elevation of the neighbors of those cells. In other words, no information is assumed about cells or elevations outside the given matrix. Cells on the outside edge of the matrix, however, can be recipients of flow from interior cells.

Flow direction for all interior cells is determined by sequentially processing each individual cell in the data matrix. Processing for each cell involves sequentially inspecting the elevation of the eight neighboring cells for the cell in question. A neighbor is considered to be any cell which touches the cell being processed, either horizontally, vertically, or on the diagonal. Thus, every cell has eight neighbors and, counting the cell itself, nine possible flow directions as indicated in Figure 2. Direction 5 indicates that flow for a cell occurs to itself. For orientation, direction eight represents north for the data set. It is important to note that by definition each cell can only flow to one of its eight neighbors or to itself if flow cannot occur to a neighbor. The numbers utilized to represent the direction of flow are patterned after the number key pad on a standard computer keyboard. The direction of flow numbers shown in Figure 2 are important as they are the basis for future processing by the algorithm and will be referred to several times in the following discussion.

Figure 2. Flow direction.
Equal Elevation Problems

For each cell in the matrix, direction of flow was determined by a gradient criteria. In other words, the flow direction for a given cell was defined as the steepest descent to one of the eight neighboring cells. If all of the neighbors of a cell were equal or greater in elevation than the cell in question, then additional criteria were used to determine flow direction. Figure 3 indicates examples of two common problems associated with the simple gradient criteria, namely two or more neighbors with the same minimum elevation. It is important to note that the path to a neighbor with the minimum elevation and the path of steepest descent are not always the same. For example, in Figure 3a, defined flow would occur from the center cell to the top, center neighbor (direction 8) and not to the upper right cell (direction 9) when descent or distance between the cells is considered.

```
45 43 43 B 45 46 48
46 44 45        43 44 46
48 46 45        45 43 46
```

Figure 3. Neighbors with equal minimum values.

The problem indicated in Figure 3b is similar to the previous example but more difficult to solve. In this situation two or more neighboring cells have elevations in which each represents both the minimum elevation and the cell of steepest descent. Whereas some researchers have employed complicated and computer intensive algorithms to deal with this problem, the present research utilizes a simple local average process. The same procedure is also followed when the elevation of the cell being processed and the minimum elevation of its neighbors are equal. For each neighbor cell that ties for the cell of steepest descent, an average elevation is computed of the neighbors of that (tied) cell. That is, a new, temporary elevation is computed for each cell which tied as the cell of steepest descent by averaging the elevation of a 3 by 3 matrix centered on the tied cell. The minimum value of these averages determines the flow direction for the original cell in question.

For the example in Figure 3b, averages are computed only for the two cells (in this case the cells with value 43) which originally tied as the cells of steepest descent for the cell with elevation value 44. Likewise, only these two cells can be considered as possible flow paths, even if the averaging produces increased values from their original elevation. Figure 4 illustrates the expanded neighborhoods for the two cells which tied for steepest descent. In this example, the local average for the left center cell is 43.89, while the bottom center cell has an average of 44.11. The result is that direction 4 would be selected as the flow direction for the original cell (indicated with value 43) and not direction 2.
Figure 4. Expanded neighborhood for local average.

If the 3 by 3 cell averages of the expanded neighborhoods are equal, the number of cells utilized to compute the average is expanded to a 5 by 5 matrix. Again the average matrix is centered on the original cells which tied for the cell of steepest descent. The procedure continues in this manner until the average for one of the cells which tied is less than the other averages considered. The direction of flow can then be properly assigned for the original cell and the algorithm proceeds to the next cell in the data set to determine direction of flow. All computed average values are temporary and not retained for the flow direction processing of the next original cell.

The procedure described is highly efficient and continues to work even when the algorithm encounters numerous contiguous cells of exactly the same elevation. Generally, such situations occur only on mesa-type environments and along some flood plain areas. No algorithm can accurately derive flow networks in a consistent manner for extremely flat or uniform conditions. Oftentimes flow patterns for these conditions in nature are random or are more detailed than the resolution of existing topographic data. Still, the present algorithm assumes the local trend of the slope regardless of the number of contiguous cells of the same elevation it encounters. In the absence of additional information, such an assumption would probably also be the case if the analysis were performed manually.

Topographic Sinks

The second major problem associated with the present methodology for determination of direction of flow from uniform grid data involves holes or sinks in the data. This situation occurs only when, after processing all cells in the topographic grid, a cell has a flow direction 5. In other words, the cell cannot flow to a neighboring cell, only to itself. Such sinks or holes may in fact be real, as in the case of playa lakes, karst environments, glacial terrain, or artifacts in the data caused by sampling point elevations from contour data. Figure 5 is an example of a single cell that the algorithm will define as a hole or sink. With any area of depression, at least one cell will, by definition, be defined as a hole or sink with flow direction of 5. For a large area of depression, the algorithm may require several iterations but will eventually result in flow out of the depression. In either
case, the present algorithm is designed to force flow to occur from every cell in the data matrix to another cell.

Figure 5. Sink as a result of point sampling.

The previously discussed areas deal with problems in determination of the flow direction for a cell and as such are handled during the sequential processing of each cell in the matrix. The removal of holes or sinks in the data set can only be dealt with after all cells in the matrix have been assigned an original flow direction. Thus, after the algorithm has processed all cells and assigned a flow direction to each, the algorithm again sequentially searches the data set of flow directions. The search continues until a flow direction 5 is found. Because of the decision criteria for determining flow direction, a cell with flow direction 5 cannot also have a neighbor with flow direction 5. A neighbor, however, can have the same elevation as a cell with flow direction 5. Whenever a cell is found with flow direction 5, the neighbors of that cell are examined for two conditions, equal elevation and flow into or away from the cell with flow direction 5. The combination of these two conditions results in the need for three possible procedures for eliminating defined sinks.

If a cell with flow direction 5 has a neighbor with the same elevation and the neighbor flows into the center cell or cell in question, then the elevation of the center cell is raised by one unit of elevation and the neighbor is now considered the center cell or cell in question. The process repeats itself by examining the eight neighbors of the new cell in question. Again the process looks for neighbors with the same elevation as the new center cell and flow directions into the cell. If the conditions are met, the elevation of the new center cell is raised one unit of elevation and a new center cell is established. This procedure continues until one of the two conditions are not satisfied by a neighbor of the cell in question.

For an original cell with flow direction 5, if a neighbor is equal in elevation but does not flow into the center cell, then the elevation of the original cell is raised by one unit of elevation. The algorithm then proceeds through the data set of flow directions in search of another cell with flow direction 5 to process.

The above two situations generally occur in extremely flat terrain. A more common situation for most topographic data was the case illustrated in Figure 5. In that example, all of the neighbors flow into the center cell and none of the neighbors have the same elevation as the cell with flow direction 5. In this
situation the elevation of the center cell is raised to the same elevation as the minimum elevation of its neighbors (in this case to value 39). Raising the elevation of the center cell only one elevation unit would served no purpose unless that value is also the minimum value of its neighbors. Without raising the elevation of the center cell to the minimum of its neighbors, the relationship and thus flow pattern between the center cell and its neighbors could not change. After the elevation has been adjusted, the algorithm continues the search through the data set of flow directions for another cell with flow direction 5.

For the three situations described above, flow direction for the cell in question is not re-determined at the time the elevations are changed. Even by only changing the elevations of sinks by one unit, it is possible that flow directions of neighboring cells may also be affected. Consequently, after all cells with flow direction 5 have been processed, flow directions are again determined by sequential examination of the entire data set utilizing the revised elevation values. After determination of the flow directions for all cells, the data set is again searched to find cells with flow direction 5 to again be processed in the above described manner. This iterative process continues until the search of up-dated flow directions does not find a sink or cell with flow direction 5. Although in extreme situations this process may require several iterations, the algorithm will insure that every cell is forced to flow into another cell within the data set.

Comparison of Derived and Actual Networks

Completed processing of flow direction for each cell results in a matrix of flow directions for the area defined by the input matrix of elevation values. An example of this completed data set is illustrated by Figure 6. In this printout the arrows indicate the direction of flow for all cells in the data set. It is important to note that no holes or sinks remain in the data set as all such features have been removed by the algorithm. Because true topographic sinks do not exist in this particular study area, the model conforms to reality in this respect.

Specifically, the particular data set shown in Figure 5 contains 3,500 cells or 70 rows with 50 elements per row. Five iterations of the algorithm were required to reach the situation where no sinks or holes remained in the flow direction data set. Most of the sinks were single cell holes and were removed on the second pass of the algorithm through the data. One area of depression was six cells in size and required the fifth iteration before the problem was eliminated. Because of a worst-case arrangement or ordering of the cells with respect to the sequential processing by the algorithm, the algorithm required an iteration for every cell in this particular defined area of depression. All of the problem areas or sinks found with the algorithm were, in fact, topographic artifacts caused by point sampling from vectors of contour data.

Several features are readily apparent upon examination of Figure 6. First, all cells are individually connected to another cell in the matrix. The connection is represented by an arrow indicating the direction of flow. This topological
Figure 6: Algorithm derived paths for direction of flow.
connection of cells is viewed as a flow path, and as an aggregate make up a 'connectivity matrix' for the area. This matrix of connections will be used in later analyses as a digital representation of the drainage network of the area. Eventually, all flow paths flow out of the matrix of elevations used as the data set. In reality, this is exactly what occurs for any arbitrary selection of a topographic surface within this particular geomorphic region.

Two general sorts of problems are present in the drainage network derived by the algorithm (Figure 6). Point A is located near two examples of flow lines which run parallel to each other for several cells. The number of flow paths from other cells into these parallel flows seem to indicate that one or both of these lines should represent small channels. In reality, however, rarely do channels run parallel this close together. Consequently, the algorithm should probably indicate that these parallel channels run together or connect earlier in the flow matrix.

The second problem type, shown at point B, is a more general version of the first problem. In this case several flow lines run parallel to one another, resulting in apparent fields or areas of sheet flow. For most situations, except possibly those at a micro scale, this flow pattern is probably unrealistic. These areas should probably contain a more tree-like pattern, gradually aggregating together to form more of a network pattern which results in increasingly larger channels as the drainage area expands.

If the connection matrix of flow direction is assumed to digitally represent the drainage network, the next step involves comparing this network to the actual drainage pattern of the area. From a purely qualitative viewpoint, this comparison was performed by manually overlaying the actual drainage network onto a printout of the derived network (Figure 7). The network utilized as reality was taken from U.S.D.A. Agricultural Research Service and U.S.G.S. sources which were mapped from large scale topographic data.

In general, good correlation appears to exist between the two separate data sets for both the drainage network and the resulting watershed boundaries. Selected areas of discrepancy between the two data sets are indicated as points A through D on Figure 7. Points A and B indicate the only two areas of significant deviation in the watershed boundary of the upper portion of the Little Washita River drainage. Examination of detailed topographic maps of these two areas indicates a series of small peaks and valleys interspersed along and on either side of the true topographic ridge at these locations. There is not a gentle or uniform gradient along the topographic divide as is generally common for this area. Instead, a mixture of slope angles and aspects, when combined with the sampling density of the topographic data, created incorrect flow directions at these two locations along the watershed boundary. Because of the scale of the digital elevation data utilized by this research, it is almost surprising that only these two small areas of disagreement in watershed boundary occurred for the study area.

Points C and D indicate the two major areas where the actual and derived drainage network patterns do not coincide well. Point C references an area where two moderate size drainages are separated by a smaller drainage. All three of
Figure 7. Drazilge network and channel compartment.
these drainages run almost parallel to one another into the Little Washita River. The derived network for this area indicates that most of the flow occurs toward the middle of the area into the smaller (in reality) drainage. Examination of large scale topographic data of this area shows a very gentle topographic trend from the watershed divide to the Little Washita River. The three tributary systems, while distinct and separate, are characterized by very narrow channel areas. Consequently, a more dense grid of elevation data would be required to accurately locate the two outside channel areas within the regional slope of this particular area.

Point D is located near the top center of the drainage pattern illustrated in Figure 7. This area appears to represent a large and probably the most significant discrepancy between the two data sets. Close examination of detailed elevation data of the area, however, appears to indicate that the derived network is not as far from reality as it would appear. In fact, except for a few cells on the top right side of this small, defined catchment, the topographic gradient from the detailed elevation data appears to indicate that the derived flow directions are correct. This evaluation results in spite of the fact that no perennial or intermittent stream is shown on the detailed topographic maps.

Another feature to note on the derived network in Figure 7 is that certain points along the drainage network are represented by small black squares. The software is designed to highlight channels by over-writing onto the flow pattern of arrows a small black square for any cell in the matrix which has been defined as a channel. Channels are defined within the software by requiring a minimum number of cells or area to flow into a given cell. For example, a minimum threshold of a 25-cell area was used to define the channels indicated in the derived network in Figure 7. Utilization of a smaller threshold value would result in a more detailed definition of channels within the defined drainage network. The analysis and display of any threshold computed channel designation is performed almost instantaneously with the proper software command. For this application, the software command only involves one keystroke on the computer keyboard followed by entering the threshold level of the number of cells used to define the drainage channel.

MODEL APPLICATIONS

Upon complete determination of the flow direction matrix, several applications and analysis results are available through the interactive software. For the software developed in this research, most present applications are initiated through the use of a mouse. The mouse moves a cursor (arrow) on the computer monitor (screen) to locate the position in the network for the analysis to occur. A click of the mouse results in the analysis referring to all cells upstream in the network from the point selected. With most application procedures of this software, the calculations necessary to perform the application were completed with the final processing of the direction of flow matrix. As with the above channel definition example, the analysis and the display of results occurs almost immediately after inputting the proper software command.
The basis for most aspects of the application software center on the definition of the area of interest as defined by the drainage network. Figure 8 indicates the selection of the upper portion of the Little Washita River watershed or the location of the area upstream from stream gage 526. This procedure was accomplished by locating the cursor at the proper location in the flow matrix displayed on the monitor and clicking the mouse. The area of drainage and likewise the drainage boundary is defined by the cells which eventually flow into the cell selected with the mouse. On the monitor (as illustrated on the resulting output product Figure 8) the cells within the defined drainage are almost instantaneously indicated as black squares overlaid onto the network of flow arrows. The area and the maximum path or flow length are also indicated on the screen. Another click of the mouse and the blackened squares disappear, leaving the original network or arrows. The cursor can then be moved to another location, such as a subwatershed or catchment, and clicked again for the definition of another drainage area. Because all calculations are performed during processing of the final flow direction matrix, response to application queries are almost instantaneous.

After selection of a given drainage area with the mouse, other statistics concerning runoff from the area are also immediately available. Figure 9 is a bar graph of the distribution of length of flow path of all cells in the drainage network defined by the area depicted in Figure 8. For this example, the length of flow is defined in terms of cell widths. Distance is computed as one unit for flow directions 2, 4, 6, and 8 and the square root of two units for the diagonal directions of 1, 3, 7, and 9 (see Figure 2). These non-dimensional units can easily be converted to actual distance by simply inputting the cell width in whatever units desired.

The next extension of this concept of distance or length of flow is to compute travel time. For this example, the velocity for each cell was estimated utilizing SCS's Upland Curves for velocity. The percent slope was automatically calculated for each cell upon completion of the drainage matrix. Landcover type was determined from the previously described Landsat TM analysis. All of these data layers were overlaid and registered together in the GIS. Travel time was computed for runoff across each cell in the data set. These cell travel times were then accumulated along the established line of flow or drainage. Again, upon selection of a drainage area with the mouse, one keystroke on the computer keyboard will immediately produce a bar graph of the arrival times of the cells in the selected drainage area (Figure 10).

The last element needed to construct a hydrograph is runoff volume. Input into this portion of the analysis was accomplished utilizing the Thiessen polygon approach previously discussed. Specifically, rainfall amounts were assigned to the cells within the GIS based upon the Thiessen polygon which coincided with each cell. Establishment of an accurate loss function to determine runoff, however, is a much more complicated process. Although quite simplistic in approach, this analysis utilized the soil infiltration rate as the loss function. This rate was assigned to each cell in the GIS based upon the soil type as defined by the previously digitized SCS maps of soil mapping units. In actuality, this procedure relies on several assumptions which may not be valid for this use, and is used in
max path len from cur square = 64.0122

total # square on screen = 1502

Figure 8. Area of analysis defined by drainage network.
Figure 10. Travel time of cells in selected watershed.
Figure 9. Length of flow paths in selected watershed.
this analysis only as an example.

The simulated hydrograph from the rainfall event dated March 23, 1984, is illustrated as the bar graph in Figure 11. The actual recorded hydrograph is indicated in Figure 11 as a solid line. For this example, the simulated and recorded hydrographs are quite dissimilar. The volume of the simulated hydrograph reflects an estimate over twice the volume recorded for the runoff event. The timing or distribution of the simulated hydrograph is skewed to the right while the recorded hydrograph is skewed to the left. One aspect between the two hydrographs which is similar is the length of time over which each hydrograph indicated an increase in stream flow. Although not shown in this report, similar comparisons of simulated and recorded hydrographs also resulted from the application of the software to the rainfall event dated October 26, 1984. The software produced equally accurate hydrographs in spite of the fact that this latter event produced a more uniform rainfall pattern than the unevenly distributed March 23 event.

The difference in volume between the simulated and recorded hydrographs is probably a function of use of an inappropriate loss function for the estimates. Whereas infiltration may be a major component of a rainfall loss function, by itself or applied in this manner, it does not accurately represent total abstractions for rainfall event which also can include vegetation interception and surface ponding.

The manner in which the loss function was applied and net runoff computed could also adversely affect both the volume and the timing depicted in the simulated hydrograph. For this simplified example, rainfall (input) and infiltration (loss) were both considered to occur for only a one hour period. The rainfall was assumed to occur uniformly over the one hour period. The infiltration rate was subtracted from the rainfall amount to estimate the excess or runoff from each cell. This net runoff per cell was accumulated through the drainage network to produce the simulated hydrograph. Only excess precipitation was accumulated through the network. Infiltration capacity not completely used in a cell was discarded and not considered when runoff from upstream cells was routed through the network. Finally, no consideration was given to antecedent moisture conditions which can significantly alter both the timing and volume of runoff from a rainfall event.

Another factor which could greatly affect the shape of the simulated hydrograph involved the empirical equations used to determine velocity of the flow. For this example, only overland flow characteristics, and not channel flow factors, were considered in computing velocity of runoff. As demonstrated in Figure 7, channels can be automatically defined within the drainage area. Runoff could then be routed through the network utilizing equations of channel flow for those cells designated as a channel. The result would be to skew the distribution to the left of the arrival times of cells reaching a given point on a defined channel. This particular aspect has not been added to the present generation of software. Incorporation of proper thresholds for channel definition and utilization of appropriate velocity equations should also skew the simulated hydrograph in
Figure 11. Simulated and recorded hydrographs.

Cubic Feet per Second

Within the Nth Hour

Arrival Within the Nth Hour
Figure 11 to more closely resemble the shape of the recorded hydrograph.

SUMMARY

Selection of the proper level of resolution or detail is a major consideration for any analysis or modeling effort. As shown by the brief example in this research, a large change in scale or cell size for some models (such as the SCS Curve Number Model) can have minimal effect on model outputs. Consequently, the increased spatial resolution of such media as the Landsat Thematic Mapper may not add any additional information or improve the results for some models. Certain model estimates have been improved, however, with utilization of remotely sensed data, and in particular Thematic Mapper data, by improving the accuracy of mapping landcover significant to hydrologic modeling (Wharton et al., 1984). At the same time, even Thematic Mapper data may still not be sufficiently fine in spatial resolution to accurately model runoff with landcover or other types of information in a grid cell format. Selection of a grid cell format, regardless of how small, assumes that the mapped phenomena is uniform across that cell. As addressed by Beven (1989), such an assumption is not valid for most data such as soil characteristics where micro interactions may significantly affect runoff. In effect, cell models, regardless of scale, are still conceptually lumped models, although possibly applied in a somewhat distributed manner. Research must therefore continue in an effort to determine the minimum scale or spatial resolution to adequately represent the surface and model runoff from that surface.

In spite of the simulated hydrograph produced in the present example, we still believe that the incorporation of remote sensing and digital drainage networks with geographic information system technology offers great promise for modeling of surface runoff. A more accurate loss function and consideration of antecedent moisture conditions and time factors for both rainfall and the abstractions should improve the model. Incorporation of channels and simple channel flow characteristics for velocity should also improve the simulated hydrograph.

More detailed topographic data as input into the analysis should produce a more accurate representation of the drainage pattern for an area. Although the 1:250,000 scale DEM utilized in this research produced a digital network similar to the actual drainage pattern, this scale elevation data is probably too generalized for all except near-regional analyses. DEM data at 1:24,000 scale or larger would provide digital elevation data with much greater topographic detail. Utilization of such detailed elevation data should improve the definition of watershed or catchment boundaries and help reduce the problem of parallel channels. Likewise, more detailed DEM data and research, like that reported by Roth et al. (1989) on micro level erosion patterns, may also aid in determining whether certain drainage areas should be described as apparent sheet flow or with a more tree-like drainage structure.

Even with these improvements in model inputs, the question still arises as to the determination of the minimum cell size or spatial resolution required to adequately or best describe the terrain and its characteristics. Wood et al. (1988)
suggests a "representative elementary area" exists for predicting catchment response. Establishment of a minimum area of analysis for model inputs such as rainfall, soil processes, and topographic fields remains an unresolved discussion. Even if a firm sampling resolution requirement could be established, the resolution of analysis is still a series of trade-offs concerning the theoretically required minimum resolution, cost of data acquisition, and computer processing capability and costs.

CONCLUSIONS

The analysis described and results depicted in this research resulted from the use of software developed and written specifically for this project. The software was written in C language and compiled with the Power C compiler. The software is presently operating on an IBM PS/2 Model 30 running DOS 3.3, with 512k memory, mouse, and VGA monitor. As a pseudo benchmark, processing time on the 8086 architecture machine required approximately four minutes for calculations for all flow diagrams, statistics, bar graphs, etc. for the 3,500 element data matrix containing elevation, soil type, and landcover data layers. Once the initial processing was completed, all output responses to application queries and location changes were near instantaneous.

Future utilization of the present research involves application of the software to additional rainfall-runoff events over a smaller gaged catchment. A smaller catchment or drainage area would allow for more detailed investigation. Compared to the large drainage area with several catchments used in the present study, a single catchment would also reduce the possibility that interactions between several catchments contribute to or cancel out total simulation errors (Bathurst, 1986). Other improvements and additions to the software could easily include the incorporation of unit hydrograph, time-area, and cascade of reservoir algorithms for runoff calculations. Variable source area concepts could also be added to the software.

In simplistic terms, runoff modeling involves only a loss function and a routing function. Development of a model capable of simulating runoff response for a variety of antecedent conditions, rainfall events, and varied topography, however, is not a simple matter. Continued efforts are still necessary to continue to improve runoff models that are accurate and presumably easy to understand and properly apply. The present research developed a highly interactive, user-friendly software package based on a personal computer. The software offers ease of use, flexibility of output products, and a unique approach for dealing with certain problems associated with the incorporation of digital elevation data in hydrologic models. This research and the developed software contributes to the effort seeking more accurate modeling of surface runoff.

ACKNOWLEDGEMENTS

The research on which this report is based was financed in part by the United State Department of the Interior as authorized by the Water Research Development Act of 1984. Contents of this publication do not necessarily reflect the
views and policies of the United States Department of Interior. Mention of trade names is for informational purposes and does not necessarily imply endorsement.

The authors wish to acknowledge and thank E. Teng Ong for his efforts on this project in the area of software development. We also wish to thank the Oklahoma State University Center for Water Research and its staff for their help in completing this project.
BIBLIOGRAPHY


ELAS. Earth Resources Laboratory Application Software. NASA. National Space Technology Laboratories, Earth Resources Laboratory. Bay St. Louis, MS. 1980.


