(NASA-TM-84984) AUTOMATED BASIN DELINEATION FROM DIGITAL TERRAIN DATA (NASA) 21 p HC A02/MF A01 CSCL 08B

N83-21487

Unclas
G3/43 09546

NASA

Technical Memorandum 84984

AUTOMATED BASIN DELINEATION FROM DIGITAL TERRAIN DATA

Danny Marks, Jeff Dozier, and James Frew

JANUARY 1983

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

AUTOMATED BASIN DELINEATION FROM DIGITAL TERRAIN DATA

by
Danny Marks
Hydrological Sciences Branch
NASA Goddard Space Flight Center
Code 924

Jeff Dozier, and James Frew
Computer Systems Laboratory
University of California
Santa Barbara, CA, 93106, U.S.A.

AUTOMATED BASIN DELINEATION FROM DIGITAL TERRAIN DATA

January, 1983

Danny Marks

Hydrological Sciences Branch
NASA Goddard Space Flight Center
Code 924

Jeff Dozier, and James Frew

Computer Systems Laboratory
University of California
Santa Barbara, CA, 93106, U.S.A.

While digital terrain grids are now in wide use, accurate delineation of drainage basins from these data is difficult to efficiently automate. We present a recursive 'order N' solution to this problem. The algorithm is fast because no point in the basin is checked more than once, and no points outside the basin are considered. Two applications for terrain analysis and one for remote sensing are given to illustrate the method, on a basin with high relief in the Sierra Nevada. This technique for automated basin delineation will enhance the utility of digital terrain analysis for hydrologic modeling and remote sensing.

INTRODUCTION

In order for digital terrain data to be useful in hydrologic modeling we must be able to delineate a drainage basin within the terrain grid. Usually the outline of a drainage basin is determined manually from a map, with or without the help of a coordinate digitizer. The coordinates of the digitized outline are then overlaid on the terrain grid. This is a slow process that must be redone for each drainage basin and sub-basin of interest, and there is no way to effectively check the accuracy of a basin outline. Consequently digital terrain data are not used as effectively as they could be in hydrologic models, in spite of their availability and low cost (at least in the U.S.). An efficient automated technique for drainage basin delineation would allow hydrologists to more easily include digital terrain information in their models.

BASIN DELINEATION

Most natural drainage basins are irregularly shaped, and the size of digital terrain grids (typically 10^4 to 10^6 grid points) requires that the delineation process be efficient. An effective delineation algorithm must be general enough to account for all terrain irregularities, must be fast, and must be able to cope with noise in the terrain data, edges of the grid, and flat regions. The solution we have developed satisfies these conditions.

By basin delineation, we mean that we first locate the grid point on a digital terrain grid that best corresponds to the outflow of the basin of interest, a gaging station, confluence, or other critical point. We then must locate all other points in the terrain grid that are "upstream" from the starting point, i.e. connected through some path down which water would flow. The basin delineation algorithm must be able to respond to irregular corners in the basin shape, account for edges in the terrain grid itself, handle flat regions, and not be fooled by ridges and other terrain boundaries.

PREVIOUS WORK

The literature on automated basin delineation is not extensive. Moreover much of it exists in ephemeral publications, so we recognize that we might have missed an efficient algorithm. We do know that at least three independent research groups in the U.S. (including ours) are currently working on an algorithm, until recently without knowledge of each other's efforts.

Collins (1975) outlined a method by which all drainage basins in a terrain grid could be identified. His solution requires that all elevations in the grid first be sorted into ascending order. The lowest point is obviously in a drainage basin. If the second lowest point is not a neighbor of the lowest point, then it must be in a separate drainage basin. By connecting neighbors in this sequence, all points on the grid are assigned to drainage basins. The algorithm is of order NlogN time complexity, because of the time required by the fastest sorting procedures. It also appears susceptible to errors in the terrain data.

The U.S. Geological Survey is currently developing an algorithm that draws drainage basins by moving "upstream. . . along inflections in the contour lines. The outline of the basin is drawn based on an algorithm that selects 'ridge lines' in the data" (C.J. Robinove, personal communication, 1982). Without more details we cannot evaluate the efficiency of the method. Apparently it has problems with highly irregular basins (S. Jensen, personal communication, 1982).

A third method has been developed at the University of Maryland, based on a "region-growing" algorithm that proceeds point-by-point through neighbors (J. Fellows, personal communication, 1982). Clearly the method can be used to select terrain features (pits, peaks, ridges, etc.) with order N efficiency (Peucker and Johnston, 1972), but without more details it is unclear how efficiently the algorithm evaluates upstream connectivity.

ORIGINAL PAGE IS
OF POOR QUALITY

THE NEW ALGORITHM

Our solution delineates the drainage basin from intermediate data sets, calculated from the terrain grid containing the drainage basin of interest in order N time: slope and exposure. Slope is the angle between the terrain facet and the horizontal; exposure is the direction that the slope faces. We arbitrarily use a right-handed coordinate system, so exposure is measured from south, with positive angles counter-clockwise (east). If the terrain grid of elevations z has spacing Δh in north-south and east-west directions, with distances x increasing eastward and y increasing northward, then slope S and exposure E are calculated by:

$$\tan S = \left[(\partial z/\partial x)^2 + (\partial z/\partial y)^2\right]^{1/2}$$

$$\tan E = -\frac{\partial z/\partial x}{\partial z/\partial y}$$

Partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$ are computed from the discrete elevation grid by a finite difference scheme, e.g.

$$\frac{\partial z}{\partial x} = \frac{z_{i,j+1} - z_{i,j-1}}{2\Delta h}$$

Figure 1 shows the kernel of the algorithm. If we are at a point in the center of a 3x3 sub-grid, we want to determine which, if any, of the surrounding 8 points are "upstream" of the center point, and therefore "in" the same basin. We define a point as upstream if its exposure faces toward the center point, within a quantization level of 8 divisions of the circle. The slope values are only used if the terrain is flat, because in that case exposure is undefined. An adjacent point that is flat is considered "upstream", and therefore "in".

"Flat" points are most often found when the terrain contains areas of lakes and meadows. Unfortunately, errors in the digital representation of the terrain force us to redefine the concept of "flat" in the algorithm. Small variations in the elevation data, caused by noise in the digitizing

i−1, J−1	i−1, J	i−1, J+1
i, J−1	i, J ⊗	i, J+1
i+1, J−1	i+1, J	i+1, J+1

Figure 1. Kernel of the basin
masking algorithm.

process, coupled with the coarse spatial resolution of the grid points, causes two significant problems. In digital terrain grids, lake surfaces do not have identical elevation values across their entire area, and areas of nearly flat terrain, such as meadows, tend to exhibit random slope features and consequent drainage patterns that are not realistic or physically possible. Because these variations are small (usually only a few meters elevation difference between adjacent grid points) the problem is overcome by assigning a threshold value for the definition of "flat". This threshold value can be set based on the quality and spatial resolution of the digital terrain data, and the terrain characteristics of the basin being delineated. A value of 3° is used for the examples presented in this paper.

Note that the exposure values are needed to a resolution of only 8 divisions of the circle. In our work on solar radiation in mountainous terrain, we usually calculate exposure to a resolution of 256 divisions of the circle, the maximum precision that can be stored in an 8-bit word. Conversion to a scale of 8 divisions of the circle is done rapidly by right-shifting the 8-bit word by 5 bits (equivalent to division by 2^5), reducing it to a 3-bit number with range 0 to 7. Figure 2 shows how this method establishes a simple range of exposure values that determine whether an adjacent point in the sub-grid is "upstream" of the center.

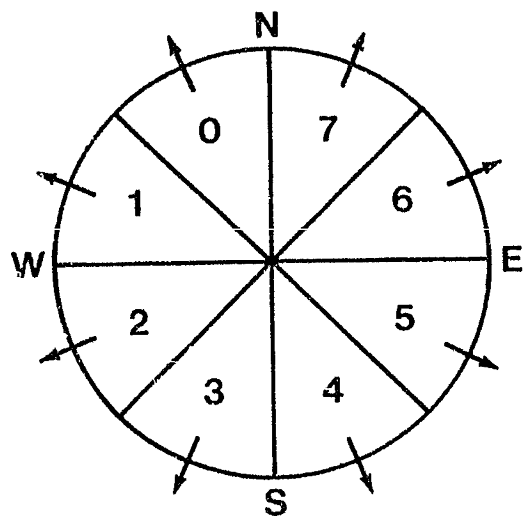
The full basin is marked by a recursive algorithm that first positions the center of the 3x3 sub-grid at the grid point designated as the basin outlet, and then does the following:

1) Mark the center point as "in".
2) Check the adjacent eight points, one by one:

   *<u>If</u> previously marked as "in", ignore;

   *<u>If</u> not marked, <u>and</u> the point is either "flat", <u>or</u> "upstream" of the center, set that point as the new center of the sub-grid and re-call the procedure.
3) Return to the last calling location and continue.

The algorithm travels recursively through the exposure grid until it encounters either a basin



Exposure reduction from 8 to 3 bits.

4 or 5	3 or 4	3 or 2
6 or 5	This point x always in	1 or 2
6 or 7	0 or 7	0 or 1

Exposure values considered "in" basin.

Figure 2. Scheme for exposure simplification used to determine whether or not an adjacent point is upstream of the center in the kernel of the basin masking algorithm.

boundary or the edge of the grid. It then back-tracks, repeating the procedure until the entire basin has been marked as "in". Once it returns to the initial starting grid point, the procedure is finished and the entire basin has been marked. The program's output is a grid of the same dimensions as the terrain grid with all bits off for points outside the basin, and all bits on for points in the basin. (Marking a point as "in" in the output file simply turns all bits on for that point.)

This method provides a simple, fast solution to automated basin delineation. Comparisons are made only once for grid points that are in or just adjacent to the drainage basin, while the rest of the terrain grid is ignored. The solution is order N; moreover the number of comparisons depends only on the number of grid points in the drainage area above the starting grid point. It is independent of the size of the full grid.

BASIN MASKING

Program output is a mask of the drainage basin that may be digitally overlaid on any geographic or image file that corresponds exactly to the terrain grid used to create it. The masking is done with a Boolean and of the mask file to any other file. A Boolean and of any value with a binary 0 (all bits off) results in a binary 0, whereas an and of any value with an "in" indication (all bits on) results in no change to the original value. This allows any grid value in a file anded to the mask file that is in the basin to 'fall through', while grid values outside the basin are set to binary 0.

The basin mask file can also be compressed, so that each point is represented by a single bit. This gives the basin delineation and masking procedures the flexibility for a variety of applications. Three direct applications of the technique are presented below.

APPLICATIONS

1. Drainage Basin Delineation

The most obvious application is the delineation of drainage basins and sub-basins on a digital terrain grid. Figure 3 shows terrain contour maps of four stages of this process. The full grid illustrates the terrain characteristics of the region near the Convict Lake basin, a small drainage basin on the eastern slope of the southern Sierra Nevada. This area of 158.3 km^2 is represented by a terrain grid of 15,827 grid points at 100m resolution.

The Convict Lake basin is shown above the stream gaging station on Convict Creek just below Convict Lake. The basin is 53 km^2, just less than one third of the area of the full grid. Because of its irregular shape, this basin requires a square grid much larger than itself to fully contain it. It is interesting that a manual delineation of the basin missed the small finger that extends to the top of McGee Mt.

The Genevieve and Box Canyon sub-basins are 9.9 km^2 and 7.3 km^2 respectively. These areas were selected because they have different terrain characteristics that can be compared once the sub-basin areas have been delineated.

2. Comparison of Basin Characteristics

A drainage basin can be described by terrain characteristics such as the distribution of slopes, exposures, and elevations over its area. While the manual determination of these features is tedious, once the basin has been delineated on a digital terrain grid, this process is straightforward.
Table 1 and Figure 4 compare terrain features from the four maps presented in Figure 3. Terrain features examined are: area (Δh^2x number of grid points), slope area ($\Delta h^2 \sum 1/\cos S$), elevation, slope (S), exposure (E), and "southness" (defined as cosEsinS). Elevation is presented both as a distribution and as a hypsometric curve.

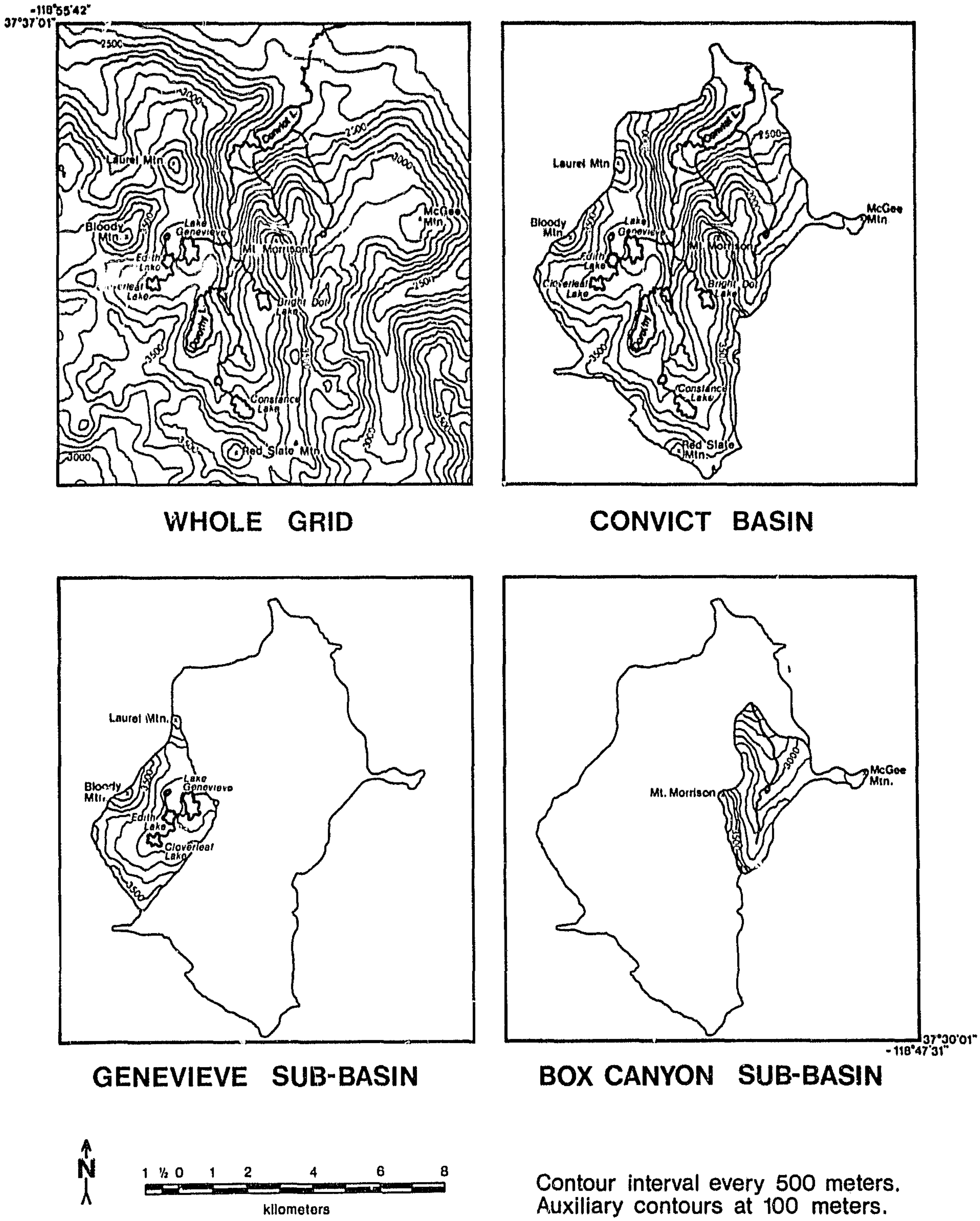


Figure 3. Terrain contour maps of square grid and four sub-basins. From NCIC digital terrain file Mariposa East, resampled to 100m grid spacing.

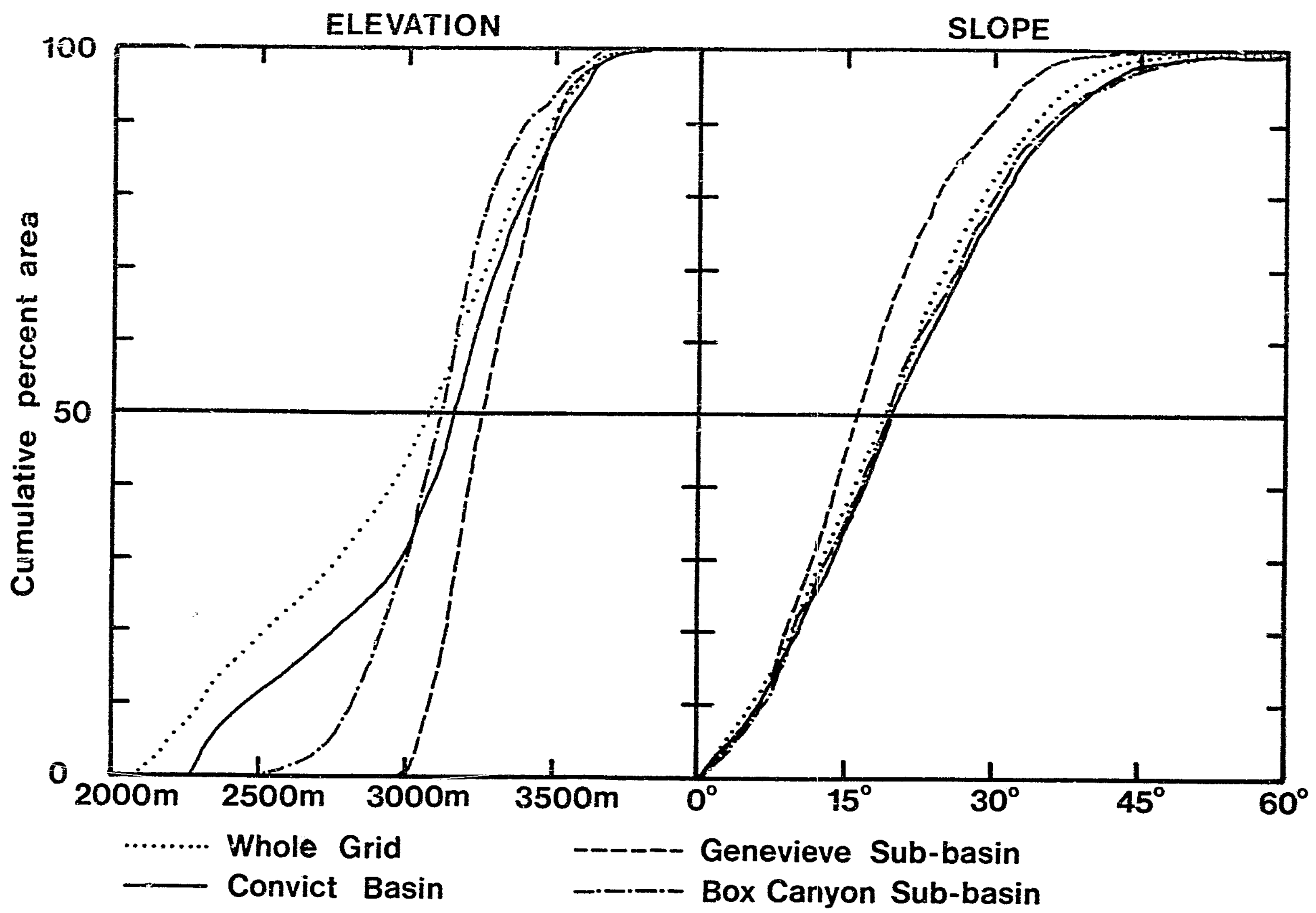


Figure 4. Comparison of elevation and slope features for square grid and three sub-basins shown in Figure 3.

Table 1

Terrain Characteristics of Four Drainage Areas

	Full Grid	Convict	Genevieve	Box Canyon
area (km^2)	158.3	53.0	9.9	7.3
slope area (km^2)	171.6	58.0	10.5	8.0
mean elevation (m)	2989	3098	3289	3119
median elevation (m)	3073	3158	3255	3120
mean slope	19°	20°	17°	20°
mean exposure	6°SW	8°SW	35°SW	Due S
mean "southness"	-0.0770	-0.0805	-0.0038	-0.1535

Both Table 1 and Figure 4 show the four areas are similar, because they are near one another, and morphological characteristics persist over the region. Differences between the median and mean elevation tend to increase with the area considered. The full grid and the Convict basin are nearly identical in all mean characteristics, while larger differences are evident between the two smaller sub-basins.

The slope, exposure, and 'southness' features show a sharp contrast between Genevieve and Box Canyon sub-basins. From the graphs in Figure 4 we see that the mean exposure values for these sub-basins result from different exposure distributions. The Genevieve sub-basin has a modal peak in the region of the mean, while the Box Canyon sub-basin is strongly bi-modal, with exposure peaks tending to the northwest and northeast. This is also the case for the full grid and the Convict basin. The mean exposure under these conditions is just an artifact of the averaging of the two modes. The combined slope and exposure differences between the sub-basins are

ORIGINAL PAGE IS
OF POOR QUALITY

shown by comparing the mean 'southness' measure. Box Canyon is significantly more 'north-facing' than Genevieve.

3. Masking Satellite Data

A third direct application is the masking of satellite data to delineate just that portion of the image that overlays a drainage basin. This is necessary if the satellite data are to provide spatial information on surface characteristics within the basin. Like basin outline delineation, this process can be done manually, but it is tedious and prone to errors. An automated technique has a multitude of potential applications such as automated snow cover mapping, surface temperature mapping, surface reflectance mapping, and net radiation mapping on a basin by basin basis. Figure 5 presents satellite images that correspond to the terrain contour maps in Figure 3. The data are from Landsat-2, April 21, 1978. They have been systematically corrected, registered to a UTM coordinate system, and resampled to the same spatial resolution as the terrain grid (100m spacing).

The full grid image contains few recognizable surface features because of the extent of the winter snow cover. The masked images provide satellite data on surface characteristics for each of the basins. These data can now be easily included as part of the hydrologic information system for each basin. Data from Landsat-2 are used here only for illustrative purposes. Because of severe saturation problems in all channels (more than 50% of the pixels in this image set are saturated) the Landsat MSS instrument provides useful information only on snow cover areal extent. This is the most common use of satellite data in hydrologic models.

CONCLUSION

Basin delineation for purposes of analyzing basin terrain morphology or isolating characteristics of hydrologic significance is a critical aspect of hydrologic modeling. The technique presented in

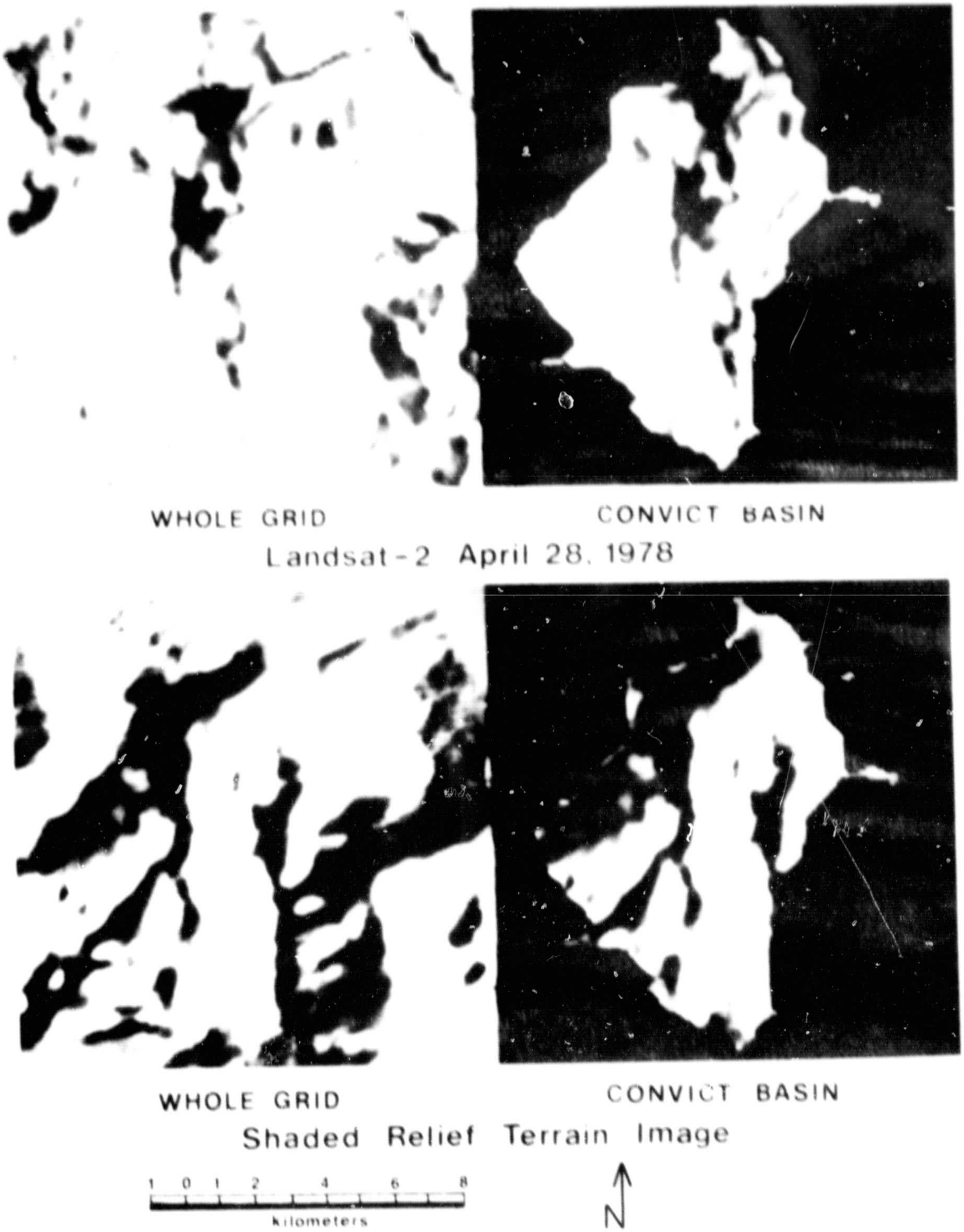


Figure 5. Masked and unmasked satellite and shaded relief images. Satellite image is from Landsat-2, band 6 (0.7–0.8μm), April 21, 1978 overpass. Shaded relief image is made from the same digital terrain file described in Figure 3.

this paper automates procedures that are necessary for a variety of hydrologic applications. This allows them to be done faster, at less cost, and more accurately than previously used manual procedures. It also allows digital terrain analysis and grid-based hydrologic modeling to be more generally applicable to any region for which digital terrain data are available.

The algorithm we have developed is reasonably portable. Its only hard requirement is that the computer environment allow recursion. It is helpful if the language used has provisions for bit manipulation, but if not, integer division or a more specific comparison scheme to determine whether a point in the sub-grid is "upstream" from the center can be substituted for the one we present. We show the algorithm using the C language (Kernighan and Ritchie, 1978). It could also be translated into Fortran, but this would be awkward if recursion were not allowed.

Basin delineation and terrain morphology analysis are simple procedures once a mask of the basin has been created. The method presented does not improve on these types of analysis, but automates the process, so that it is faster and more accurate than manual techniques. The method also provides a way that satellite and other grid based data can be easily included in the hydrologic modeling process. The technique is objective, and repeatable, and should give hydrologists a new tool for hydrologic modeling and terrain analysis.

ACKNOWLEDGEMENTS

This work was supported by NASA grant NGT-05-010-801, and NOAA grant NA81AA-D-00063.

REFERENCES

Collins, S.H., 1975. Terrain parameters directly from a digital terrain model. The Canadian Surveyor, 29:507-518.

Kernighan, B.W. and Ritchie, D.M., 1978. The C Programming Language. Prentice-Hall, Englewood Cliffs, NJ, 228 pp.

Peucker, T.K. and Johnston, E.G., 1972. Detection of surface-specific points by local parallel processing of discrete terrain elevation data. Technical Report TR-206, F-44620-72-C-0062, University of Maryland, Computer Science Center, College Park, MD.

ORIGINAL PAGE IS
OF POOR QUALITY

APPENDIX: THE ALGORITHM

We present the algorithm in the **C** *language (Kernighan and Ritchie, 1978). Translation into PASCAL would be easy, but translation into FORTRAN would be awkward if recursion were not allowed. It should compile in this form, requiring only a main program to handle data I/O. The program consists of two routines. Routine* **is_upstream** *performs the sub-grid procedure described in the text, and routine* **go_upstream** *performs the marking, checking, and recursion. The main program that calls these routines is described, but not presented. Because of the terseness of the* **C** *language, we have inserted numerous comments, delimited by '/*' and '*/'.*

1. Sketch of the Main Program

Main Program should do the following:

1. *Create the following external variables to minimize stacked information during recursive calls.*

   integer variables Maxrow *and* Maxcol *(maximum dimensions of terrain grid, starting at 0,0)* Kount *(to count points in the basin, initialized to zero), and* Mark *(to mark a point as in, usually a value corresponding to all bits on, i.e. 255 for an 8-bit word).*

2. *Read in slope and exposure files, and coordinates of starting grid point and set threshold value for* flat. *The variables* Basin, Slope, *and* Expo *are two-dimensional arrays, stored externally and known to all routines (to minimize stacked information).*

3. *Call basin delineation program.*

   go_upstream(start_row,start_col);

5. *Write output file.*

2. Basin Masking Routines

Recursively mark all points upstream from a given point. Input are slope and exposure matrices (stored as 1-byte binary fractions).

```
/*          create the exposure sub-grid as external          */
/*          (the following are octal constants)               */

#define  BIT0   1
#define  BIT1   2
#define  BIT2   4
#define  BIT3   010
#define  BIT4   020
#define  BIT5   040
#define  BIT6   0100
#define  BIT7   0200

unsigned char     upstream[3][3] = {
          { BIT4 | BIT5, BIT3 | BIT4, BIT2 | BIT3 } ,
          { BIT5 | BIT6,        0, BIT1 | BIT2 } ,
          { BIT6 | BIT7, BIT7 | BIT0, BIT0 | BIT1 }
};

/*          sub-grid procedure                          */
/*          returns 1 if point dr,dc is "upstream"      */
```

```
/*         returns 0 otherwise                        */

int
is_upstream(expo, dr, dc)                                       is_upstream
        unsigned char     expo;
        int               dr, dc;
{
        return((1 << (expo >> 5)) & upstream[dr + 1][dc + 1]);
}


/*         marking, checking, and recursion procedure          */

go_upstream(row, col)                                           go_upstream
        int       row, col;
{
        int       max_dr, max_dc, dr, r, dc, c;

        if (row < 0 || col < 0 || row > Maxrow || col > Maxcol)
                  return;

        /* are we at edge of grid? */

        max_dr = (row == Maxrow) ? 0 : 1;
        max_dc = (col == Maxcol) ? 0 : 1;

        /* point is upstream from itself, by definition */
        Basin[row][col] = Mark;
        Kount++;
        for (dr = (row == 0 ? 1 : -1); dr <= max_dr; dr++) {
                  r = row + dr;

                  for (dc = (col == 0 ? 1 : -1); dc <= max_dc; dc++) {
                            c = col + dc;

                            if (Basin[r][c] != Mark               /* not visited yet */
                            && (Slope[r][c] == Flat               /* flat always in */
                            || is_upstream(Expo[r][c], dr, dc)))/* or it's upstream */
                                      /* recursively call procedure again */
                                      go_upstream(r, c);
                  }
        }
}
```