4. Methods

4.1 Data Acquisition and Development

The success of any hydrologic analysis is dependent on the availability, quality, and application of relevant data. In highway drainage facility design, the hydrologic data, especially peak flood flow rates, are the primary factors that affect the size of a facility. Failure to use adequate and appropriate data can lead to damage of the facility and interruption of traffic from possible undersizing, or excessive construction costs due to oversizing.

The first steps in the design of a drainage facility are to:

- locate the site,
- identify the types of data required,
- determine sources of data, and
- acquire and manipulate the data.

This section outlines the data needs for typical hydrologic analyses for the design of highway drainage facilities, discusses potential data sources, and details the methods of establishing digital data for use in HDDS.

Data Needs

Section 3.3 outlines several hydrologic methods which have specific data requirements. That is, data requirements are dependent on the intended hydrologic method. In design, the availability and reliability of data affect the choice of hydrologic method. The most prevalent hydrologic parameters include:

- drainage area size and shape,
- topography,
- soil type,
- land use/land cover,
- rainfall characteristics,

- stream size, shape, and roughness, and
- storage volumes.

HDDS accommodates the first five items, and are discussed here. The last two items are by no means unimportant, but are beyond the current scope of this work.

Generally, the process of establishing parameters for use in hydrologic is manual: drainage area size and shape, watershed slopes and stream geometry still are established by interpretation of topographical maps, field survey data, and aerial photographs. Aerial photographs, land use maps, and vegetation maps are used to determine land use and land cover within a watershed. Soil information is acquired from county soil maps and soil reports. Design rainfall data is usually acquired by interpolation of rainfall depth-duration-frequency maps or tabular data.

The digital data requirements for hydrologic analyses are similar to those required using manual methods. **Table 4-1** presents a comparison of paper sources with digital data sources for hydrologic parameter development. An important consideration for spatial data is that of geospatial reference information including projection parameters, horizontal datum, and, if applicable, vertical datum. These were discussed in Section 3.2. All spatial data for a particular analysis must use the same geospatial reference parameters. The spatial data contained in HDDS have been set up in an Albers Equal Area projection, heretofore referred as the HDDS projection, with the parameters shown below.

Horizontal units: meters

Vertical units: meters

Spheroid: GRS 1980

Datum: NAD 83

First standard parallel: 29 30 00

Second standard parallel: 45 30 00

Longitude of origin: -96 0 00

Latitude of origin: 23 0 00

False Northing: 0.0

False Easting: 0.0

These parameters are used to represent data covering the contiguous United States, although most existing data uses NAD 27 as the horizontal datum. If a system such as HDDS was to be implemented for Texas, it would be preferable to set up all data using the Texas Statewide Mapping System projection parameters. This has not be done here for the following reasons:

1. Much of the source data were already in the nationwide projection.
2. By minimizing the amount of projection transformations, it was expected that projection errors would be minimized.
3. The main goal of this work is to demonstrate the potential of using a system such as HDDS for development of hydrologic data, not necessarily to have a version that is to be implemented.
4. If this type of system is to be used for nationwide analysis, the current projection parameters would be appropriate.
5. The possible application of scale factor adjustment techniques such as those discussed in Sections 3.2 and 4.2.2 could help negate or reduce the need for different projections for analysis at different locations. (Output generation would still employ the most appropriate projection).

Table 4-1: Comparison of Sources of Paper Data and Digital Data

Parameter	Paper Format	Digital Format
Location	Texas County maps (TxDOT), USGS topographical maps (1:250 K and 1:24 K scale)	Digitized highways, county boundaries, major cities.
Drainage area	USGS topographical maps, field survey data	Digital elevation models, digitized aerial photogrammetric data
Path Length and slope	As above	As above
Soil type	SCS County soil maps and reports	STATSGO (SCS)
Land use	Aerial photos, vegetation maps	GIRAS land use/cover
Rainfall depth	Rainfall depth-duration-frequency maps (e.g. TP 40)	Digitized spatial data by county

Data Sources

Table 4-2 presents a list of the existing data types, their sources, and original formats that have been used in HDDS.

Table 4-2: Data Sources for HDDS

Data	Source	Source Format	Georeferenced
15 arc second DEM	USGS - Oklahoma	ASCII	Yes
3 arc second DEM	USGS - Internet	ASCII	Yes
30 m DEM	USGS - Sioux Falls	ASCII	Yes
1:2 M highways	USGS - Internet	Digital Line Graph	Yes
1:2 M streams	USGS- Internet	Digital Line Graph	Yes
1:250 K land Use	USGS - Internet	Arc/Info - polygon	Yes
1:2 M drainage basins	USGS - Internet	Arc/Info - polygon	Yes
Hydrologic soil group	SCS STATSGO	Arc/Info - polygon	Yes
Stream gauge records	USGS-Austin	Tabular - ASCII	No
Stream gauge sites	USGS - Austin	Arc/Info - point	Yes
TP 40 24 hour rainfall	National Weather Service	Paper Contour Map	No
Texas county boundaries	USGS CDROM	Arc/Info - Polygon	Yes
1:250 K quadrangle index	USGS CDROM	Arc/Info - Polygon	Yes
1:24 K quadrangle index	USGS CDROM	Arc/Info - Polygon	Yes
Hydrologic regions	TxDOT	Paper map	No
Hydrologic Region boundaries	USGS - AUSTIN	Arc/Info - line	Yes
Rational Method rainfall coefficients	Texas Department of Transportation	Paper Tables	No
Texas state boundary	ArcUSA CDROM	Arc/Info - line	Yes

Data Development

Several data sets have been created which have been compiled into two packages, a large and a small set. For the large package, three data sets exist. One is based on 1:2000,000 (1:2 M) scale data, covers the whole of Texas and is entitled “tx”. The second is based on 1:250,000 (1:250 K) scale data covering an area of Northeast Texas near Paris and is entitled

”g”. The third set is named “s” and comprises processed 1:24 K data which covers two tributaries of the North Sulphur River. Short names are used to help minimize file name sizes. For the small package, the same named sets exist, but the data set “tx” covers the same geographic extent of Northeast Texas as set “g” at a 1:2 M scale. The set “g” is identical to the large package. The data was split up this way because the large package was difficult to transfer to other systems for testing and demonstration. Obviously, the larger package would be preferable for statewide implementation.

Table 4-3: Database Coverages

Data name	Feature type	Description	Primary attributes
raingrd	GRID	Design rainfall data	Rational method coefficients & 24-hour rainfall
txacc *	GRID	Flow accumulation	Cell value = no. of cells accumulated
txfil *	GRID	Filled DEM	Cell value = elevation
txstrms *	GRID	DEM-based streams	Cell value = 1
txrdgrd *	GRID	DEM-based roads	Cell value = txrds-id
txbas *	GRID	Drainage basins	Cell value = zone id
txslnk *	GRID	Stream links	Cell value = txsarc-id
txdir *	GRID	Flow direction	Cell value = direction
txslope *	GRID	Cell slope	Cell value = percent slope
arcbasns	POLYGON	Major Texas drainage basins	basin#, basinname
gsrgns	POLYGON	Hydrologic regions	Region#
statsgo	POLYGON	Soil data - soil names and percent hydrologic group	muid, muname, a-pct, b-pct, c-pct, d-pct
txcnty	POLYGON	Texas county boundaries	State_fips, state_name, cnty_fips, cnty_name
txlus	POLYGON	Texas land use	Anderson level II landuse code
tx24ndx	POLYGON	Index of 1:24 K quadrangles	Quadrangle name
tx250ndx	POLYGON	Index of 1:250 K quadrangles	Quadrangle name
txpoly	POLYGON	Texas state boundary	state_fips, state_name
txsarc *	LINE	Vector coverage of streams (txslnk)	txsarc-id = txslnk cell value
txrds *	LINE	major Texas (&OK) highways	txrds-id , hwy_name
txgages	POINT	Stream gauge sites	txgages_id, area, discharge

* also available at 1:250 K scale with prefix "g" instead of "tx" and 1:24 K scale with prefix "s".

The large package data set "tx" was established manually, using the procedures outlined below. The sets "g" and "s", and the small package set "tx" were created using the data preprocessor within HDDS (which is described in Section 4.2.2).

Table 4-3 presents a list of the digital data employed in this system and the relevant attributes. The following sections discuss the creation of the components of the aforementioned data sets.

Digital Elevation Model Data

Digital Elevation Models (DEM's) are files that contain a uniform grid of ground elevations. This type of data can be used for a wide variety of applications, including watershed delineation, which require digital description of the topography. The elevation assigned to a cell within the grid represents the ground elevation at the centroid of the cell. Digital elevation data are available from the United States Geological Survey at 1:2 M and 1:250 K scales for the whole US, and limited 1:24 K data are available.

The DEM data are crucial to the success of HDDS: a majority of the procedures rely on the digital elevation data. These are discussed in Section 4.2.

The 1:2 M (15 arc second) elevation data, which cover the contiguous United States, are experimental (unpublished) and were acquired from the USGS (Rea, 1994). The elevation sampling was at 500 m intervals. The original data are in Albers equal area projection, using the nationwide parameters discussed above, and the elevations are referenced to NGVD 1929.

For use in HDDS, the original 1:2 M scale data were converted to from an image format to an Arc/Info GRID format by Mizgalewicz (1994). Then, a window was set around the original data to create a data set which covered the whole of Texas and the approximate extent of all drainage basins that enter Texas.

The DEM data often contain artificial sinks and peaks (see discussion of problems) which may be eliminated using the GRID command **FILL**. This command allows the user to specify a limit within which peaks or sinks will be removed, however, it is unlikely that one threshold would be applicable to all instances, making an iterative approach necessary.

The 1:250 K DEM data are available from the USGS in 1 degree by 1 degree blocks (quadrants). The data are in geographic coordinates, horizontally referenced to the WGS 84 datum. Elevation data are in meters, vertically referenced to NGVD 1929 and sample spacing is 3 arc seconds (USGS, 1993, pp. 5). Individual quadrants are retrievable by binary file transfer protocol (ftp) from the following World Wide Web site:

http://edcww.cr.usgs.gov/glis/hyper/guide/l_dgr_demfig/ni14.html.

Two adjacent quadrants, Sherman East and Texarkana West, in Northeast Texas, were downloaded for use in HDDS. After downloading, it was necessary to delimit each file using the following UNIX command:

dd if=infile of=outfile ibs=4096 cbs=1024 conv=unblock

The resulting files were then imported into Arc/Info as a lattice using :

DEMLATTICE <infile> <outfile>

The above command must be issued from the ARC prompt, even though the resulting file is a grid coverage.

Since the original 1:250 K DEM data are in geographic coordinates, subsequent operations such as determining flow directions, cell slopes, and watershed areas, would be completely meaningless unless the data is projected. The subject quadrants were projected into the HDDS projection (Albers Equal Area). In addition to normal rounding errors, this process introduces sampling errors since the centroids of the cells in the projected file do not conform with the spacing of the original geographic data. The resulting spacing for the projected files was about 93 m. (This would vary with latitude for other data sets).

The two blocks of data (Sherman East and Texarkana West) were then merged to create a contiguous grid covering the extent of the North Sulphur River.

Once the data were in GRID format, projected and merged, the operations on the 1:250 K data and the procedures were the same as those described for the 1:2 M data.

As indicated earlier, only limited 1:24 K scale DEM coverage of Texas is available. The 1:24 K scale data consist of rectangular arrays of elevations horizontally referenced in the

UTM coordinate system using NAD 27. They are set up in 7.5-minute quadrangles. The sample spacing is 30 m, the elevations are in meters and are vertically referenced to NGVD 1929 (USGS, 1993, pp. 2&3).

Two quadrangle data files, Ladonia and Honey Grove, were obtained from the USGS (Dunn, 1995). These two adjacent quadrangles cover about one fifth of the geographic extent of the aforementioned 1:250 K data. The procedures for creating an Arc/Info GRID of the 1:24 K data are similar to those described for the 1:250 K.

An important issue for all scales of DEM data is whether the projecting of individual grids should precede the merging process. Without doubt, if the data are in UTM and extend across different UTM zones, the DEM's should be projected individually before merging into one grid. This is recommended here because the boundaries of adjacent blocks of data do not match. The process of transforming into a consistent projection should create matching boundaries (accuracy limitations notwithstanding). For DEM data in geographic coordinates, the order may not be so important, though it may be preferable to merge the individual blocks prior to projecting the data. This should reduce processing time and potential projection errors. In geographic coordinates, adjacent blocks of data should join without any erroneous gaps.

Potential Problems Related to DEMS

The most common problem results from errors in the sampling process which incur either false elevations, no data or artificial peaks and sinks. A sink is a topographical condition in which water collects to a point which has no outfall. For hydrologic analysis, artificial sinks are more worrisome than peaks because they could reduce the number of cells that should be contributing to the drainage area. In this project, all sinks, whether natural or artificial have been eliminated. It is recognized that for detailed analysis, it is necessary to differentiate between the natural sinks and artificial ones.

The GRID documentation (ESRI, 1991, pp. 1-14) leads the user through an involved process to identify sinks using the command **SINKS**. This process can be precluded by subtracting the original DEM from the filled DEM. All the cell locations that have not been filled will result in a value of zero, while those that have been filled will have values that represent the depth of the original sink. If the user has topographic maps or a knowledge of the terrain, the user may decide which, if any sinks are natural. There are means by which the user can then add the natural sinks back into the filled grid, or refill the original grid using an appropriate threshold.

Flat surfaces such as lakes and reservoirs pose potential problems similar to those associated with artificial sinks. The cells representing the water surface may be lower than any cell surrounding the water body, creating a sink. In reality, water may be released through gates at a much lower elevation. The elevation sampling accommodates only surface elevations.

Generally, sea level is represented by the value zero in the USGS DEM data. When issuing the **FLOWDIRECTION** command, the expanse of zero values associated with the sea results in meaningless flow directions which affect subsequent analysis. To avoid this problem, the **SETNULL** command may be used on the filled DEM prior to performing a flow direction analysis. This assigns all cells having a specified value (e.g. zero) as NODATA. The disadvantage is that any ground elevations of zero will also be eliminated.

Land Use Data

The USGS has land use/land cover data available mostly at the 1:250 K scale and some 1:100 K scale. The information contains georeferenced polygons in GIRAS format with land use codes assigned as attributes.

Originally, several 1:250 K quadrangles were downloaded from the Internet from the following FTP site:

http://sun1.cr.usgs.gov/glis/hyper/guide/1_250lulcfig/states/TX.html.

Table 4-4: Anderson Level I and Level II Land Use Codes

Level I		Level II	
Code	Description	Code	Description
1	Urban or Built up Land	11	Residential
		12	Commercial and Services
		13	Industrial
		14	Transportation, Communications and Utilities
		15	Industrial and Commercial Complexes
		16	Mixed Urban or Built-up land
		17	Other Urban or Built-up land
2	Agricultural Land	21	Cropland and Pasture
		22	Orchards, Groves, Nurseries, and Ornamental Horticultural Areas
		23	Confined Feeding operations
		24	Other Agricultural Land
3	Rangeland	31	Herbaceous Rangeland
		32	Shrub and Brush Rangeland
		33	Mixed Rangeland
4	Forest Land	41	Deciduous Forest Land
		42	Evergreen Forest Land
		43	Mixed Forest Land
5	Water	51	Streams and Canals
		52	Lakes
		53	Reservoirs
		54	Bays and Estuaries
6	Wetland	61	Forested Wetland
		62	Nonforested Wetland
7	Barren Land	71	Dry Salt Flats
		72	Beaches
		73	Sandy Areas Other than Beaches
		74	Bare Exposed Rock
		75	Strip Mines, Quarries, and Gravel Pits
		76	Transitional Areas
		77	Mixed Barren Land

Abstracted from USGS (1986, pp. 4)
Note: Codes for Tundra and snow or ice cover were omitted.

The data were imported into Arc/Info using the ARC command GIRASARC, however, on attempting to join adjacent quadrangles, it was noticed that an excessive non-conformal overlap was occurring. It is the author's belief that, although the original data are reported to be in UTM projections, the x and y shifts (origins) differ from quadrangle to quadrangle with no recorded information about these parameters. This problem could have been overcome by measuring the overlap and projecting each quadrangle using the measured overlaps as false eastings and false northings. However, existing land use coverages by drainage basin were acquired from the USGS (Tan, 1995). These coverages were then merged, projected into the HDDS projection, and cleaned to create coverage of the State of Texas. The original files contained only the Anderson Level II land use codes which appear in **Table 4-4**.

Hydrologic Soil Group Data

The Soil Conservation Service provides a spatial database of soil characteristics called the State Soil Geographic Data Base (STATSGO). STATSGO contains georeferenced polygon data of soil types which are available for the whole United States and are based on a generalization of detailed soil report and are mapped on U.S. Geological Survey 1:250 K scale topographic quadrangle maps. The data contain an extensive array of soil attributes contained in several cross-referenced tables. The data are available in Arc/Info format but some considerable processing was involved to link appropriate data and eliminate unwanted data. For this project, the soil names and percentages of occurrence of soils in each of four hydrologic soil groups (A, B, C, and D) were considered to be sufficient. A sample of the soils and hydrologic groups appears in **Table 4-5**: the complete data set consists of more than four thousand soil names and so is not shown here. Group A represents the best draining soils such as dry sands, while Group D represents soils of poor drainage quality such as heavy clays as discussed in Section 3.3.3.

Table 4-5: Sample of Hydrologic Soil Group Data

STATSGO ID	MUID	MUNAME	A PCT	B PCT	C PCT	D PCT
488	TX530	SPURLOCK-GRUVER-TEXLINE (TX530)	0	76	15	9
489	TX213	GRUVER-DUMAS-SUNRAY (TX213)	0	64	29	7
490	TX049	CONLEN-BERTHOUD-PASTURA (TX049)	2	62	0	36
491	TX213	GRUVER-DUMAS-SUNRAY (TX213)	0	64	29	7
492	TX108	CONLEN-SUNRAY-SPURLOCK (TX108)	0	88	4	8
493	TX124	DALLAM-DALHART-DUMAS (TX124)	1	98	1	0
494	TX049	CONLEN-BERTHOUD-PASTURA (TX049)	2	62	0	36
495	TX213	GRUVER-DUMAS-SUNRAY (TX213)	0	64	29	7

Hydrologic Regions

The hydrologic calculations employed in HDDS need a pre-defined hydrologic region as an input variable. For GIS analysis, a polygon coverage of regions is required. No polygon coverage was found, however, digitized line coverage of the boundaries between regions was acquired from the USGS (Ulery, 1995). A polygon coverage was created by placing a polygon of the Texas state boundary in the arc coverage of hydrologic regions. The hydrologic region arcs were then edited to ensure intersection with the Texas boundary using ARCEDIT and the polygons were built using **CLEAN**. (Note: CLEAN creates nodes at any intersection prior to attempting to build polygons. BUILD does not). The name (numerical id) of each region was then added to the polygon attribute table (PAT) using **ADDITEM** and **ADD** in the TABLES subsystem. The hydrologic regions are used to identify regression equations from which can be derived discharge versus frequency relationships as discussed in Section 3.3.1. **Figure 4-1** shows the coverage of hydrologic regions employed in HDDS.

Political

County boundaries and the Texas State boundary were abstracted from the ArcUSA 1:2 M compact disk using the **RESELECT** command within ARCPLOT. Items named FIPS_1, FIPS_2, FIPS_3 and FIPS_4 were used to isolate those arcs whose attributes contained the Texas code of 48. County boundaries were abstracted by reselecting arcs whose FIPS item contained the Texas code (48).

Design Rainfall Data

One of the hydrologic methods used here, the SCS runoff curve number method (discussed in Section 3.3.3), uses 24 hour duration design rainfall as appears in the National Weather Service Technical Paper 40 (1961). This publication includes maps with isohyetal lines for rainfall frequencies of 1 year, 2 years, 5 years, 10 years, 25 years, 50 years, and 100 years. For HDDS, ASCII format tables were created containing the TP 40 rainfall for the 2 through 100 year frequencies abstracted by county (**Table 3-5**). The rain tables were imported into Arc/Info and joined to the ArcUSA 1:2 M county polygon coverages using the county name as the relate item. The unwanted attributes were removed and the remaining coverage was converted into a grid using POLYGRID. The polygon identifier was used to establish the cell values. The 2 through 100 year frequency rainfall items of the vector coverage were then joined to the VAT of the grid coverage. This was accomplished by adding the polygon id name as an alternate name for the cell value in the grid using the TABLES command ALTER. Then the vector coverage PAT and the grid VAT were joined using the polygon id as the relate item. This resulted in a grid with a VAT containing the design TP 40 rainfall values. **Figure 4-2** shows the distribution across Texas of the design 24-hour, 100 year rainfall.

In anticipation of incorporating the Rational method into HDDS, the rainfall intensity coefficients for all Texas counties were added as attributes in the same manner as that described for the TP 40 data. HDDS has not been set up to use this data at the moment since recommended use of the Rational method is limited to about 200 acres or less (TxDOT, 1985, pp. 2-14). The author considers the resolution of the DEM data currently available for use in HDDS to be too coarse for determining such small areas.

Highways and Streams

The USGS publishes 1:2 M scale Digital Line Graphs (DLG) of hypsography and transportation features (USGS, 1990) and are available from the following Internet address:

http://sun1.cr.usgs.gov/glis/hyper/guide/2milfig/mapindex

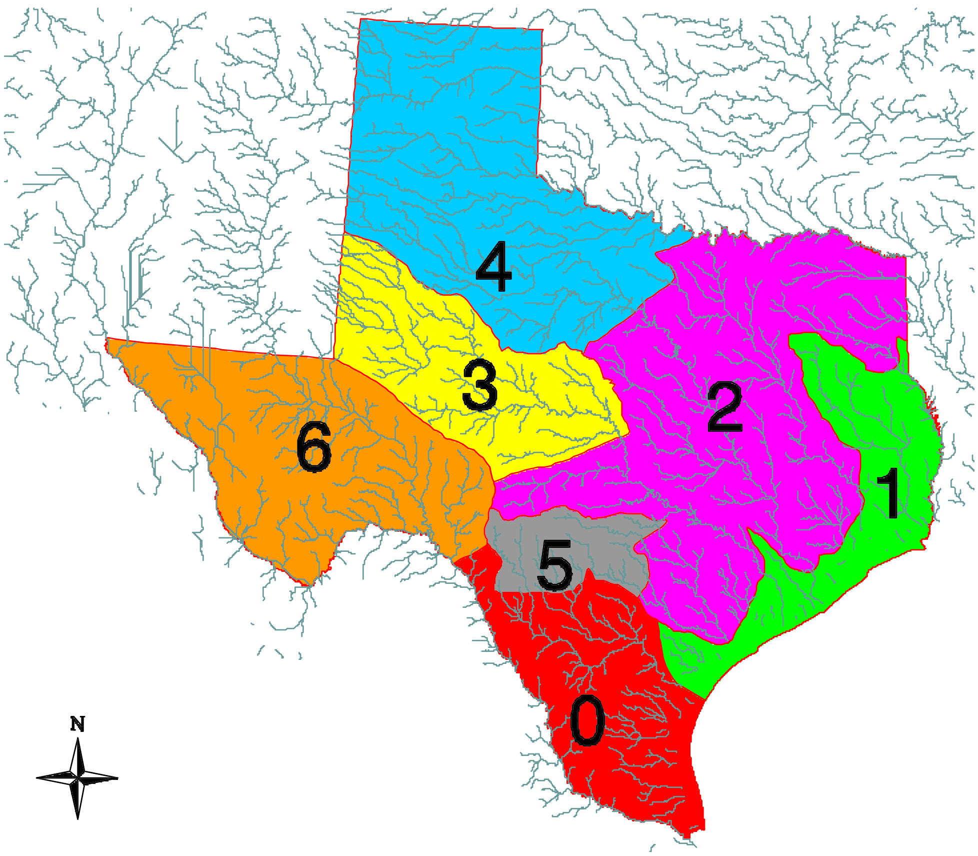

Figure 4-1:
HDDS coverage of hydrologic regions in Texas for regional regression equations
Number 0 represents undefined area.

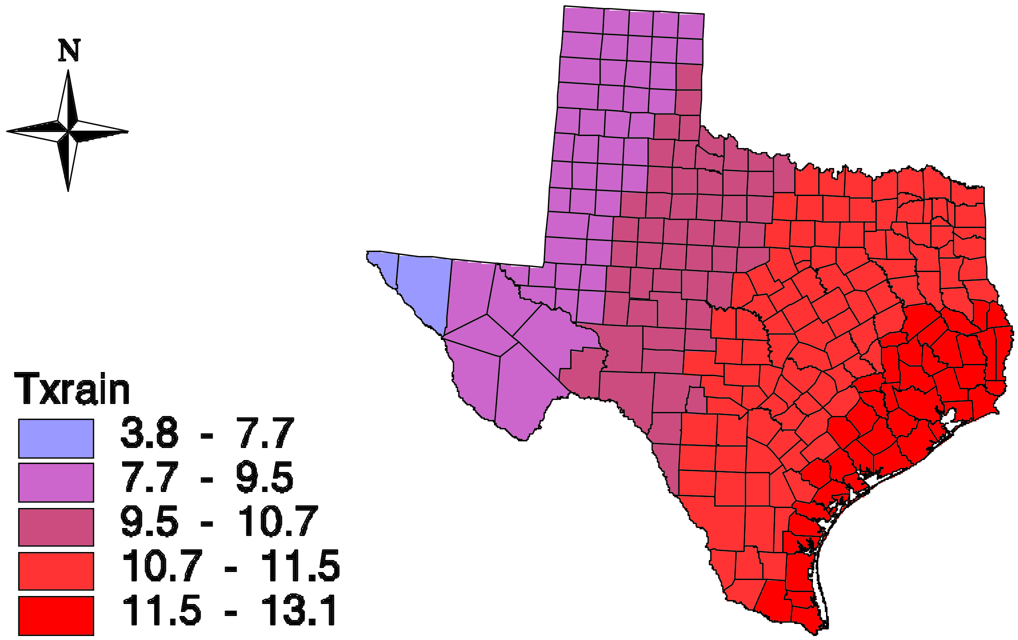

Figure 4-2:
Distribution of design 24-hour 100-year rainfall in Texas

Texas is covered by two portions (North and South). Two file types exist: standard and optional. It does not matter which is selected since Arc/Info seems to be able to use either. For highways, only major roads (e.g. Interstate and State highways) are presented. The attributes include items such as class but no highway names are included. Similarly, no names are available for the stream coverages. The following procedure was employed to convert the data into Arc/Info coverages. The optional file format was selected.

1. The downloaded file was delimited using the UNIX command

   dd if=<input> of=<output> ibs=8000 cbs=80 conv=unblock

2. The files were imported into Arc/Info using the ARC command

   DLGARC OPTIONAL <infile> <outcover> # ATTRIBUTED

3. Topology was established using the ARC command

   BUILD <incover> lines

4. Highway names were added as attributes of the highway vector coverage (txrds). Time constraints precluded attributing more than a few stretches of highway:

   In ARC Tables,

   ADDITEM txrds.aat hwy_name 7 7 C

   This specifies a data width of 7 characters, an output data width of 7 characters and the item type is alphanumeric (character)

5. Still in ARC Tables, the AAT for the highway coverage was opened: **SELECT txrds.aat**

6. The highway name was then added using **ADD** and following the prompts.

   Using the same approach other attribute data can be appended to the following:

   - Arc Attribute tables
   - Polygon attribute tables
   - Grid value attribute tables
   - Info lookup/relate tables.

No attributes were added to the streams at this stage, since the digital elevation model was used to delineate streams as discussed below. **Figure 4-3** shows the extent of highways and stream coverage at the 1:2 M scale.

Preprocessed Data

One objective of this project was to identify the procedures that are common to all hydrologic analyses and perform such procedures up front to minimize real-time processing. Several procedures were identified as being required for all sites, most of which involve sequential processing of the digital elevation data as follows:

1. Remove all sinks from the DEM using the GRID command **FILL**.
2. Compute flow directions using the GRID command **FLOWDIRECTION**.
3. Compute flow accumulation (number of cells contributing to a cell - not actual flow volume or rate) using the GRID command **FLOWACCUMULATION**.
4. Establish streams as being those cells with a flow accumulation in excess of some defined threshold using the GRID command **CON**.
5. Subdivide delineated streams into links with unique id's using the GRID command **STREAMLINK**.
6. Convert the links into a vector coverage using the GRID command **STREAMLINE**.
7. Determine general drainage basins (not individual drainage areas) using the GRID command **BASIN**.
8. Calculate the slope of each cell using the GRID command **SLOPE**.
9. Convert vector coverage of highways into grid using the GRID command **LINEGRID**.

The filling process was described earlier. **Figure 4-4** indicates the processing involved in steps (2) through (6). Since all cells in a grid, except boundary cells, have eight adjacent cells, there are eight possible directions in which flow may proceed from one cell to the next. Arc/Info uses a binary geometric series to define the directions: 1 = East, 2 = Southeast, 4 = South, 8 = Southwest, 16 = West, 32 = Northwest, 64 = North, and 128 = Northeast. The flow direction assigned is based on the steepest calculated slope between adjacent cells. Problems may arise where two or more directions yield the same "steepest" slope, in which case a unique direction would not exist. This problem is most likely to occur in flat terrain with coarse cell resolution. In such instances, the flow direction value in the cell reflects the sum of the individual flow direction codes. This allows identification of the problem cells.

The flow accumulation process counts the number of cells that contribute flow to a cell using the flow direction grid. At any given cell, the drainage area to the cell (but not including the cell) is the product of the flow accumulation value and the cell area.

The stream grid is determined subjectively as those cells whose flow accumulation value exceeds a threshold value. For HDDS, a threshold of 500 cells for the 1:2 M scale data and 1000 cells for the 1:250 K scale data were considered reasonable. Recent unpublished work suggests that a better threshold may be determined as a function of area and annual rainfall (Olivera, 1995, personal communication). However, for HDDS, and certainly for demonstration purposes, the threshold is not critical as long as it is low enough to include most highway stream crossings yet not so small as to create an inordinately excessive stream network and mislead a user into believing that a stream exists when in fact it does not. As an aside, it is worthwhile to note that, apparently, there has been no consistent method applied to the delineation of streams on the USGS 1:24,000 topographical maps. The use of drainage area, average annual rainfall, and possibly section geometry, in a GIS environment may be a suitable means of standardizing a method to designate streams.

Step (7) is similar to the watershed delineation process which is discussed in Section 4.2.2. However, instead of using user-specified pour points, the process considers any boundary cell or sink as a pour point.

The aforementioned procedures are time-consuming, yet perfunctory. They have been performed for each DEM data set and the resulting data have been added to the “permanent” database. Appendix A includes a preprocessing AML named preproc.aml which performs the above steps and may be applied to data for other areas. The preprocessor is discussed in Section 4.2.

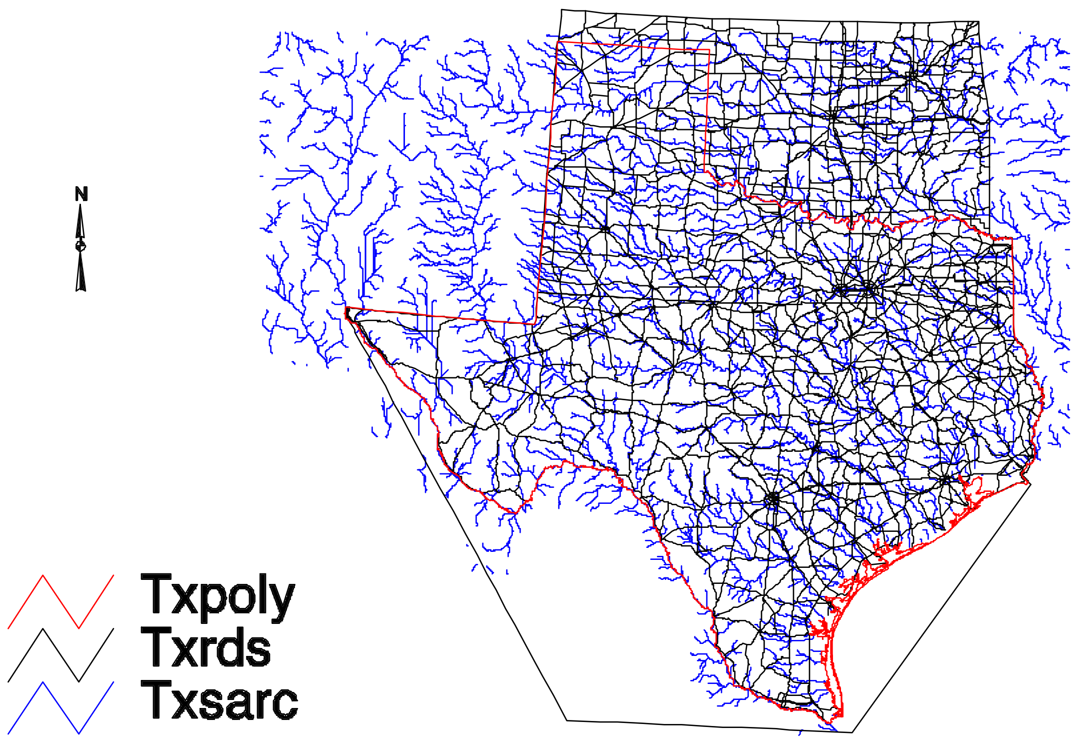

Figure 4-3: Extent of 1:2 M scale data

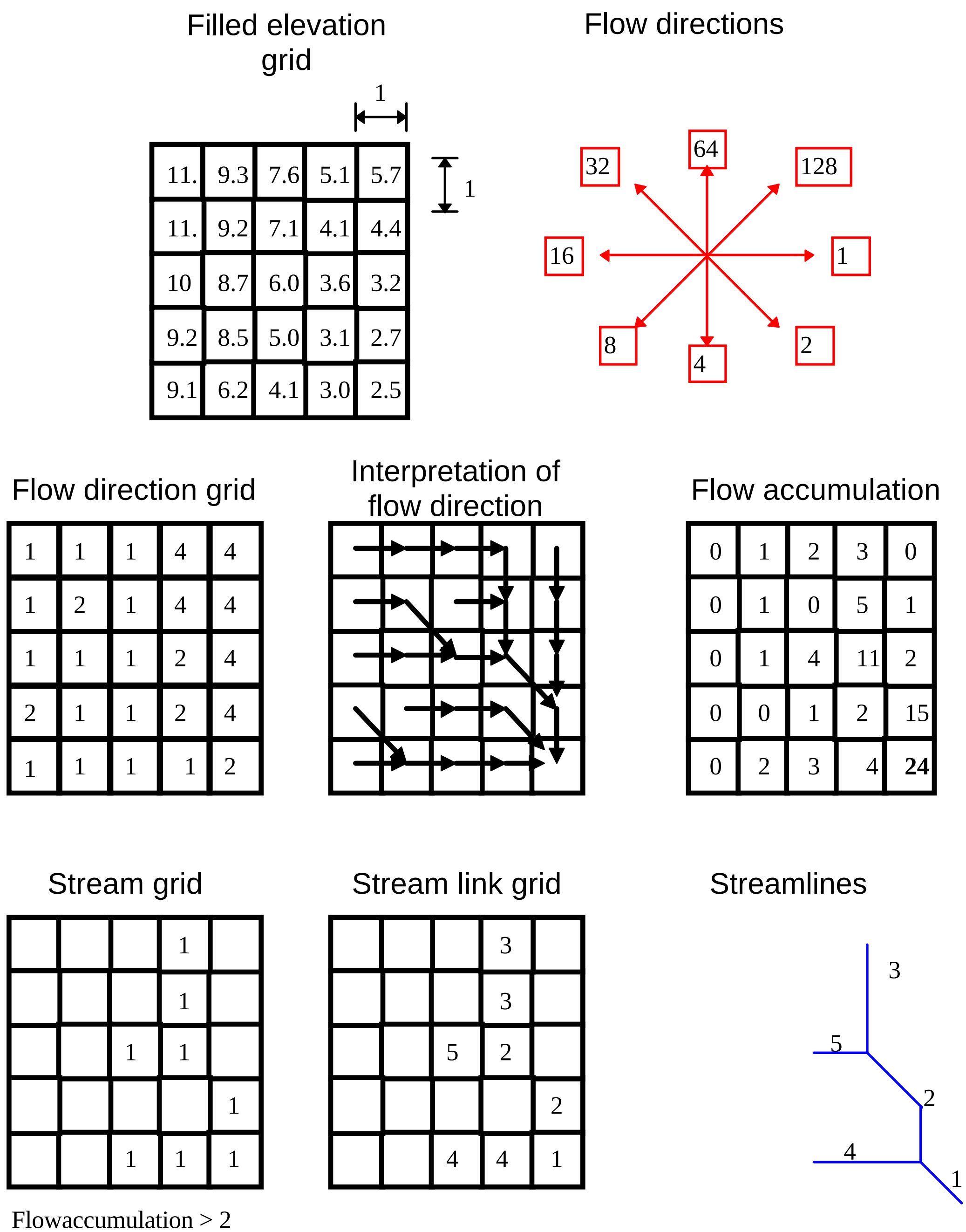

Figure 4-4: Processing the DEM grid

Relating Land Use and Soil Group to SCS Runoff Curve Number By Lookup Table

HDDS contains procedures which estimate Soil Conservation Service runoff curve numbers and times of concentration as discussed in Section 3.3.3. The curve numbers can be estimated by referencing tables which relate land use/land cover and hydrologic soil group to runoff curve numbers (**Tables 3-6, Table 3-7, Table 3-8, and Table 3-9**). A designer can establish a runoff curve number for a defined area by identifying land surface use/cover, determining which description best fits the surface cover and selecting the suggested RCN for the appropriate hydrologic soil group. A similar approach is employed in HDDS and is described in Section 4.3.3. Since the existing land use data (**Table 4-4**) are limited to sgeneral land use categories, and the SCS runoff curve number tables do not contain land use codes, a means of relating these with suggested runoff curve numbers was needed. **Table 4-6** was developed as an INFO table to relate hydrologic soil group and existing Anderson Level II codes and additional, more detailed codes to runoff curve number. As can be seen from **Table 4-4**, the Anderson Level II codes provide only general descriptions. These may be appropriate for small scale analysis or where more accurate description is not warranted. Surface descriptions appearing in **Tables 3-6 to Table 3-9** were grouped into appropriate Anderson Level II categories. The most conservative range of RCN's was taken to relate the Level II code to hydrologic soil group. For example, the Level II code 11 represents residential use. **Table 3-6** includes curve numbers for six different residential average lot sizes, for which the most conservative (highest values) is the range for 1/8 acre lots. The runoff curve numbers for soil groups A, B, C, and D are 77, 85, 90, and 92 respectively. These values were assigned to the Level II code 11. Since HDDS is intended for design, the author deems it appropriate to assign the most conservative values. However, on average, lower values may be reasonable. The detailed level codes were established by assigning numbers in which the first two digits relate to the assigned Level II group and the remaining digits are unique and sequential identifiers for the cover description. HDDS provides the capability of reassigning curve numbers to existing land use codes (see Section 5).

Some analyses will require a more refined description of land use than the Anderson Level II descriptions allow. A third level of description was established here by assigning land use codes (lucode) to **Tables 3-6**, **Table 3-7**, **Table 3-8**, **and Table 3-9**. These tables show the assigned codes for clarity, but the actual data used in HDDS appear in **Table 4-6**. This table is entitled "rcns.dat" in HDDS. The general steps employed to create this table follow:

1. An ASCII format table was prepared containing all the desired data without headers.
2. The INFO command **DEFINE** was used to establish an INFO data file frame in which the item names, field widths and data type were established. The items were named to conform with named items in appropriate coverage attribute tables. i.e. "lucode" is as appears in the polygon attribute table of the land use coverage "txlus", and "hyd-a", "hyd-b", "hyd-c", and "hyd-d" appear in the polygon attribute table of soil coverage "statsgo".
3. The ASCII file was incorporated into the defined INFO file using the command **ADD FROM**.

By creating cross-referenceable items between the INFO table and coverage attribute tables, a "look up" capability results, the use of which is discussed in Section 4.2.

Table 4-6 also contains velocity coefficients, the use of which are described later. The values were estimated on the assumption that surface roughness characteristics can be related to land use/ land cover: Section 3.3.3 describes the relationship between velocity coefficient and surface cover description. The curves appearing in **Figure 3-16** can be described by **Eq.(4-1)**.

$$v = b \cdot S^{0.5} \qquad \textbf{(4-1)}$$

where,

v is the surface flow velocity (m/s),

S is the average topographical slope (m/m), and

b is a coefficient dependent on surface cover herein named velocity coefficient.

Table 4-7 presents calculated values of velocity coefficient for each cover description.

The velocity coefficients assigned in **Table 4-6** are subjective interpolations of the aforementioned curves: they were determined by the author on the basis of his judgement and experience of estimating velocities for computing time of concentration. HDDS provides the capability of reassigning these values as discussed in Section 5.

Table 4-6: Estimated Runoff Curve Numbers (RCN) and Velocity Coefficients

	Description		RCN by Hydrologic Soil Group				
LUCODE	DES_A	DES_B	HYD_A	HYD_B	HYD_C	HYD_D	VCOEFF
0	Incase-of-zero data		100	100	100	100	5.00000
11	Residential	Level_2	77	85	90	92	4.62000
111	Residential	1/8_acre	77	85	90	92	4.62000
112	Residential	1/4_acre	61	75	83	87	3.66000
113	Residential	1/3_acre	57	72	81	86	3.42000
114	Residential	1/2_acre	54	70	80	85	3.24000
115	Residential	1_acre	51	68	79	84	3.06000
116	Residential	2_acre	46	65	77	82	2.76000
12	Urban_85%_imperv	Comm_&_business	89	92	94	95	5.34000
13	Urban_72%_imperv	Industrial	81	88	91	93	4.86000
14	Streets_&_roads	Level_2	98	98	98	98	5.88000
141	Paved	parking_lots-roofs	98	98	98	98	5.88000
142	Streets_&_roads	Paved-curbs/gutter	98	98	98	98	5.88000
143	Streets_&_roads	Paved-open-ditches(w	83	89	92	93	4.98000
144	Streets_&_roads	Gravel(w/ROW)	76	85	89	91	4.56000
145	Streets_&_roads	Dirt(w/ROW)	72	82	87	89	4.32000
16	Mixed_Urban		80	86	89	92	4.50000
17	Other_urban	Level_2	89	92	94	96	5.34000
171	Western_Desert_Urban	Natural_desert	63	77	85	88	3.78000
172	Western-Desert-Urban	landscaping	96	96	96	96	5.76000
173	DEVELOPING-URBAN	Newly-graded-area	77	86	91	94	4.62000
18	Urban-Open-space	General	68	79	86	89	4.08000
181	Urban-Open-space	grass<50%	68	79	86	89	4.08000
182	Urban-Open-space	grass50%-75%	49	69	79	84	2.94000
183	Urban-Open-space	grass>75%	39	61	74	80	2.34000
21	AGRICULTURAL	Level_2	77	86	91	94	4.62000
2111	Fallow	Bare-soil	77	86	91	94	4.62000
2112	Fallow	CR-poor	76	85	90	93	4.56000
2113	Fallow	CR-good	74	83	88	90	4.44000
2114	Row-crops	SR-poor	72	81	88	91	4.32000
2115	Row-crops	SR-good	67	78	85	89	4.02000
2116	Row-crops	SR+CRpoor	71	80	87	90	4.26000
2117	Row-crops	SR+CR-good	64	75	82	85	3.84000
2118	Row-crops	C-poor	70	79	84	88	4.20000
2119	Row-crops	C-good	65	75	82	86	3.90000
2120	Row-crops	C+CR-poor	69	78	83	87	4.14000
2121	Row-crops	C+CR-good	64	74	81	85	3.84000
2122	Row-crops	C&T-poor	66	74	80	82	3.96000
2123	Row-crops	C&T-good	62	71	78	81	3.72000
2124	Row-crops	C&T+CR	65	73	79	81	3.90000
2126	Small-grain	SR-poor	65	76	84	88	3.90000
2125	Row-crops	C&T+CR-good	61	70	77	80	3.66000
2128	Small-grain	SR-good	63	75	83	87	3.78000
2129	Small-grain	SR+CR-poor	64	75	83	86	3.84000
213 0	Small-grain	SR+CR-good	60	72	80	84	3.60000

Table 4-6(cont.)

LUCODE	DES_A	DES_B	HYD_A	HYD_B	HYD_C	HYD_D	VCOEFF
2131	Small-grain	C-poor	63	74	82	85	3.78000
2132	Small-grain	C-good	61	73	81	84	3.66000
2133	Small-grain	C+CR-poor	62	73	81	84	3.72000
2134	Small-grain	C+CR-good	60	72	80	83	3.60000
2135	Small-grain	C&T-poor	61	72	79	82	3.66000
2136	Small-grain	C&T-good	59	70	78	81	3.54000
2137	Small-grain	C&T+CR-poor	60	71	78	81	3.60000
2138	Small-grain	C&T+CR-good	58	69	77	80	3.48000
220	Close-seeded	SR	66	77	85	89	3.96000
222	legumes	SR	58	72	81	85	3.48000
23	Farmsteads	Level_2	59	74	82	86	3.54000
24	OTHER-AG	General	68	79	86	89	4.08000
241	Grass	poor	68	79	86	89	4.08000
242	Grass	fair	49	69	79	84	2.94000
243	Grass	good	39	61	74	80	2.34000
244	Rotation	C-poor	64	75	83	85	3.84000
245	Meadow	C-good	55	69	78	83	3.30000
246	Meadow	C&T-poor	63	73	80	83	3.78000
247	Meadow	C&T-good	51	67	76	80	3.06000
248	Meadow	Non-grazed	30	58	71	78	1.80000
31	Herbaceous	Level_2	70	80	87	93	4.20000
311	ARID-SEMIARID-RANGE	Herbaceous-poor	70	80	87	93	4.20000
312	ARID-SEMIARID-RANGE	Herbaceous-poor	60	71	81	89	3.60000
313	ARID-SEMIARID-RANGE	Herbaceous-poor	50	62	74	85	3.00000
32	Shrub-and-Brush	Level_2	55	67	80	85	3.30000
321	Sagebrush	poor	55	67	80	85	3.30000
322	Sagebrush	fair	40	51	63	70	2.40000
323	Sagebrush	good	25	35	47	55	1.50000
324	Desert-shrub	poor	63	77	85	88	3.78000
325	Desert-shrub	fair	55	72	81	86	3.30000
326	Desert-shrub	good	49	68	79	84	2.94000
33	Mixed-Rangeland	Level_2	48	67	77	83	2.88000
331	Brush_mix	poor	48	67	77	83	2.88000
332	Brush_mix	fair	35	56	70	77	2.10000
333	Brush_mix	good	30	48	65	73	1.80000
41	Deciduous_Forest	Level_2	55	66	74	79	3.30000
411	Oak-aspen	poor	55	66	74	79	3.30000
412	Oak-aspen	fair	37	48	57	63	2.22000
413	Oak-aspen	good	25	30	41	48	1.50000
42	Evergreen-Forest	Level_2	60	75	85	89	3.60000
421	Pinyon-juniper	poor	60	75	85	89	3.60000
422	Pinyon-juniper	fair	45	58	73	80	2.70000
423	Pinyon-juniper	good	25	41	61	71	1.50000
43	Mixed-Forest	Level_2	57	73	82	86	3.42000
431	Woods-grass	poor	57	73	82	86	3.42000
432	Woods-grass	fair	43	65	76	82	2.58000

Table 4-6 (cont.)

LUCODE	DES_A	DES_B	HYD_A	HYD_B	HYD_C	HYD_D	VCOEFF
433	Woods-grass	good	32	58	72	79	1.92000
434	Woods	poor	45	66	77	83	2.70000
435	Woods	fair	36	60	73	79	2.16000
436	Woods	good	30	55	70	77	1.80000
51	Streams-&-Channels		100	100	100	100	6.00000
52	Lakes		100	100	100	100	6.00000
53	Reservoirs		100	100	100	100	6.00000
54	Bays-&-Estuaries		100	100	100	100	6.00000
61	Forested-wetland		100	100	100	100	6.00000
62	Nonforested-wetland		100	100	100	100	6.00000
71	Dry_salt_flats		25	25	25	25	5.00000
72	Beaches		25	25	25	25	2.00000
73	Non_beach_sand		25	25	25	25	2.00000
74	Bare_rock		98	98	98	98	5.88000
75	Quarries-gravel_pits		0	0	0	0	0.01000
76	Transitional_areas		75	80	85	90	3.00000
77	Mixed_Barren_land		75	80	85	90	3.00000

Table 4-7: Calculated velocity coefficients

Cover Description	Velocity Coefficient (b)
Forest	0.7495
Fallow	1.4158
Short Grass	2.0821
Bare Ground	2.9147
Grassed waterway	4.5804
Paved	5.4138

4.2 HDDS Technical Procedures

4.2.1 Database Design

Figure 4-5 shows the directory layout for HDDS. The system consists of a main directory, and four sub-directories as follows:

1. The previous section outlined the digital data employed in this project. These data layers were established as the permanent database from which hydrologic analyses could be performed within the geographical extent of the data.
2. The macros, menus and message files were grouped together. For larger systems it may be desirable to provide separate sub-directories for each type, however, grouping these allowed easier programming.
3. Only one lookup table is currently present, rcns.dat, described earlier.
4. A user could choose to establish a workspace in any location. This example includes the workspace as a sub-directory of HDDS.

Once a database has been established an almost limitless number of possibilities exist for subsequent analysis. A system of menus, macros, and text files were established to perform the following "on-the-fly" analyses or functions:

- selection of database,
- creation of workspace,
- preprocessing of DEM data and highway coverage,
- viewing of basins, highways, streams coverages, and user-selected data,
- identification of watershed outlet (outfall) by intersecting stream and road, or user-identified cell or polygon,
- selection and relocation of stream gauges
- drainage area delineation and measurement,
- longest flow path delineation and measurement and ellipsoidal scale factor adjustment,

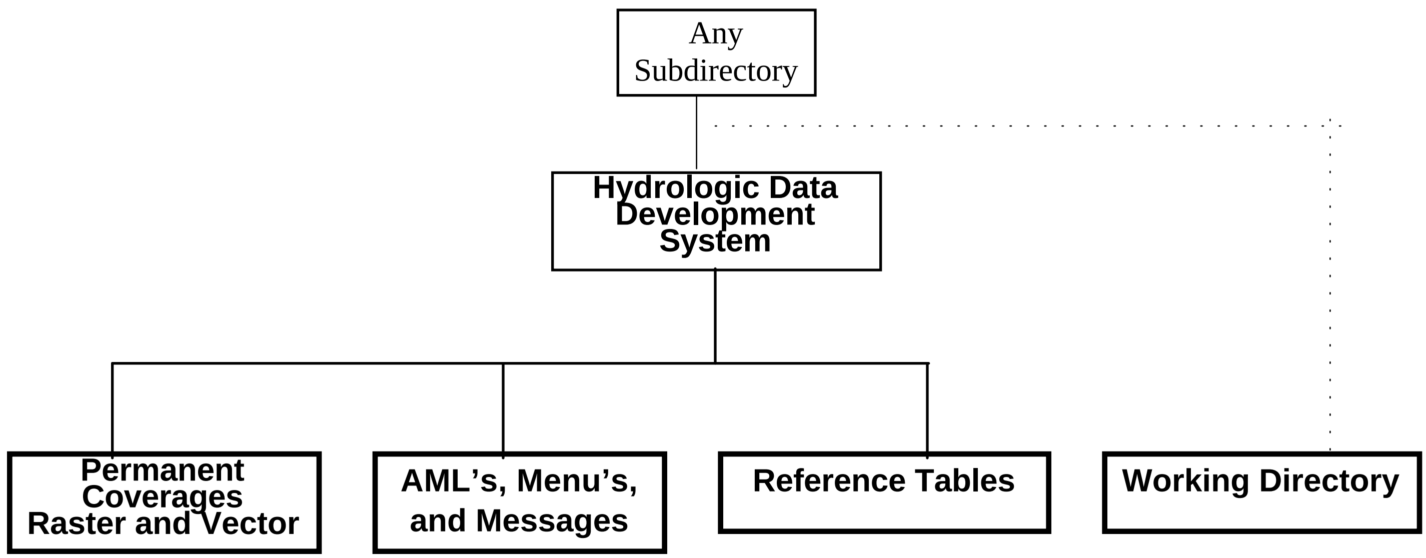

Figure 4-5:
Database layout for the Hydrologic Data Development System

- average watershed slope determination,
- watershed shape factor,
- subarea delineation and measurement,
- land use code designation/modification,
- weighted SCS runoff curve numbers by watershed or subarea,
- designation of overland and in-stream flow velocities,
- calculation of watershed and subarea times of concentration,
- hydrologic region designation,
- determination of required higher resolution quadrangles (1:250 K and 1:24 K DEM),
- execution of external hydrologic program (THYSYS),
- clipping of watershed hydrologic soil group data,
- clipping of watershed location data (political), and
- cleaning up of workspace.

Analyses may be performed in any user-created workspace.

4.2.2 Program Components

This section discusses the significant steps incorporated in HDDS. Flow diagrams of main menu routines and sub-menus are presented in **Figures 4-6**, **Figures 4-7**, **and Figures 4-8**. Annotated menus and routines are provided in Appendix A.

HDDS employs a combination of Arc/Info vector processing (ARC), raster or cell-based processing (GRID) and Arc Macro Language (AML) to determine hydrologic parameters. In general, vector coverages of features are overlaid upon a raster (GRID) image of major drainage basins. The drainage basin grid is used primarily as a visual aid. That is, a backdrop for other data such as streams and highways to be displayed. Vector coverages are used for visual aid, identification of features, clipping of data, and storage of hydrologic parameters determined within HDDS. The majority of hydrologic parameters are determined using cell-based processing. All hydrologic parameters determined by HDDS are

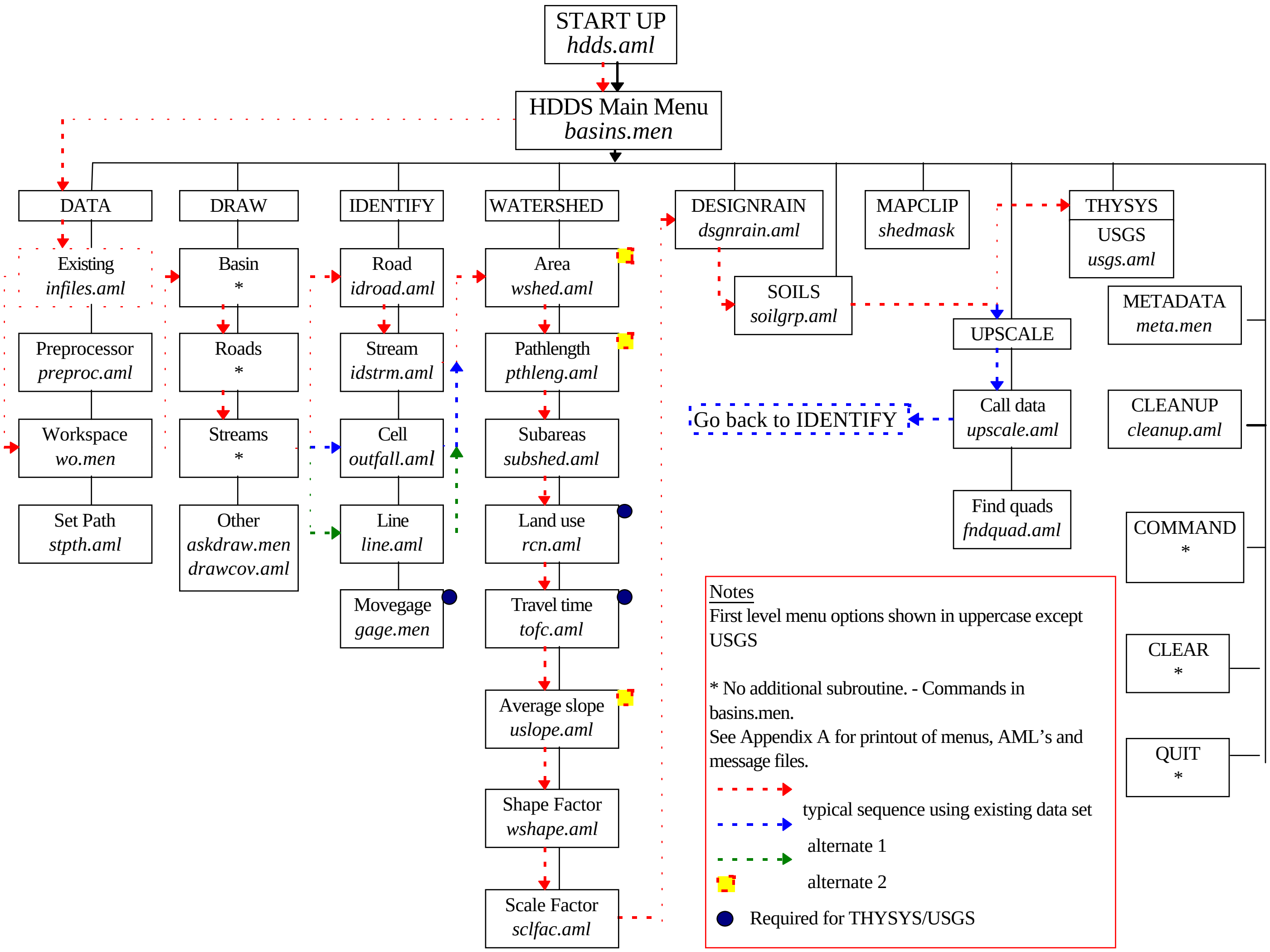

Figure 4-6: HDDS Main Menu Options

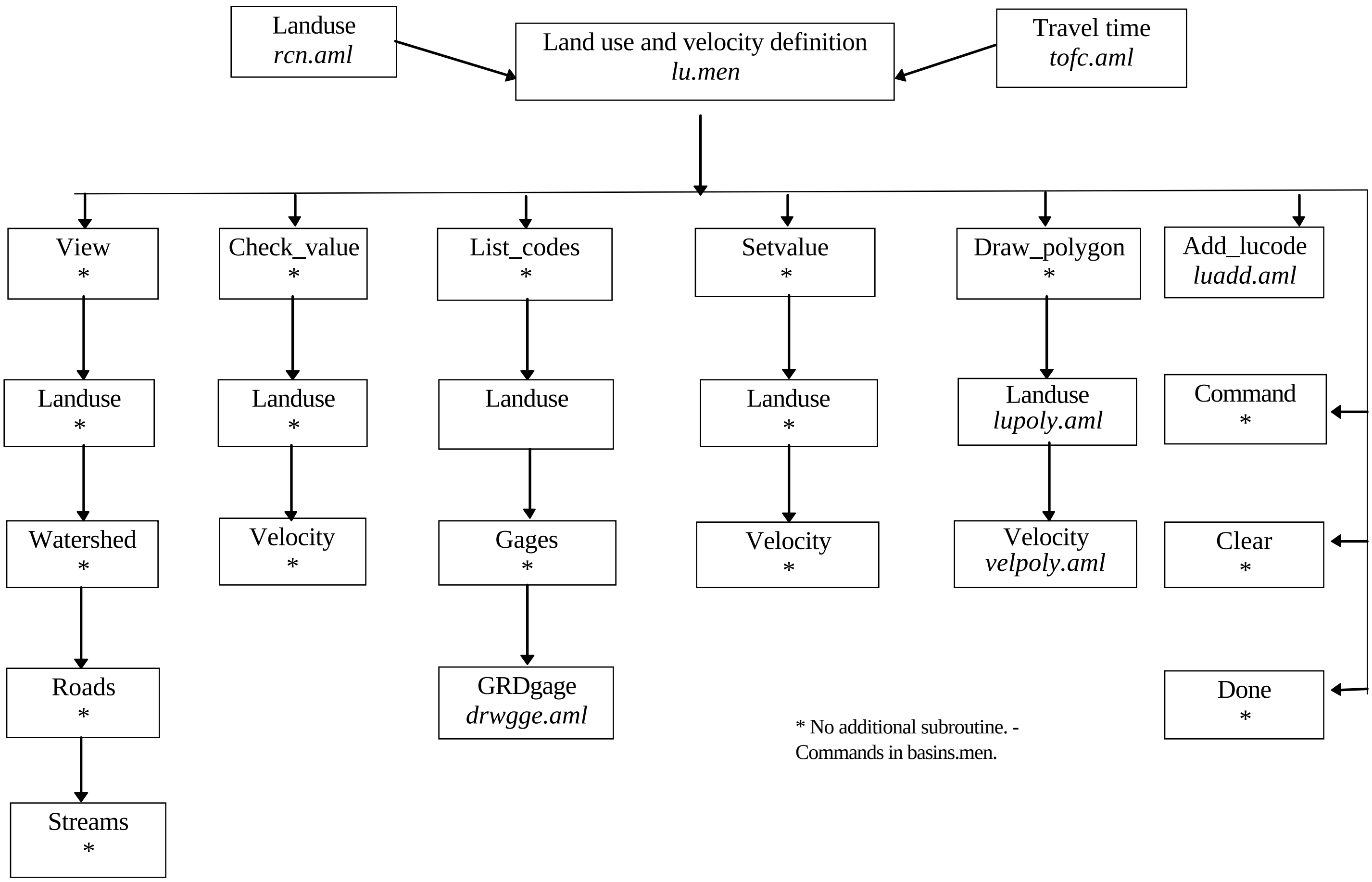

Figure 4-7
Menu layout for defining land use and velocity

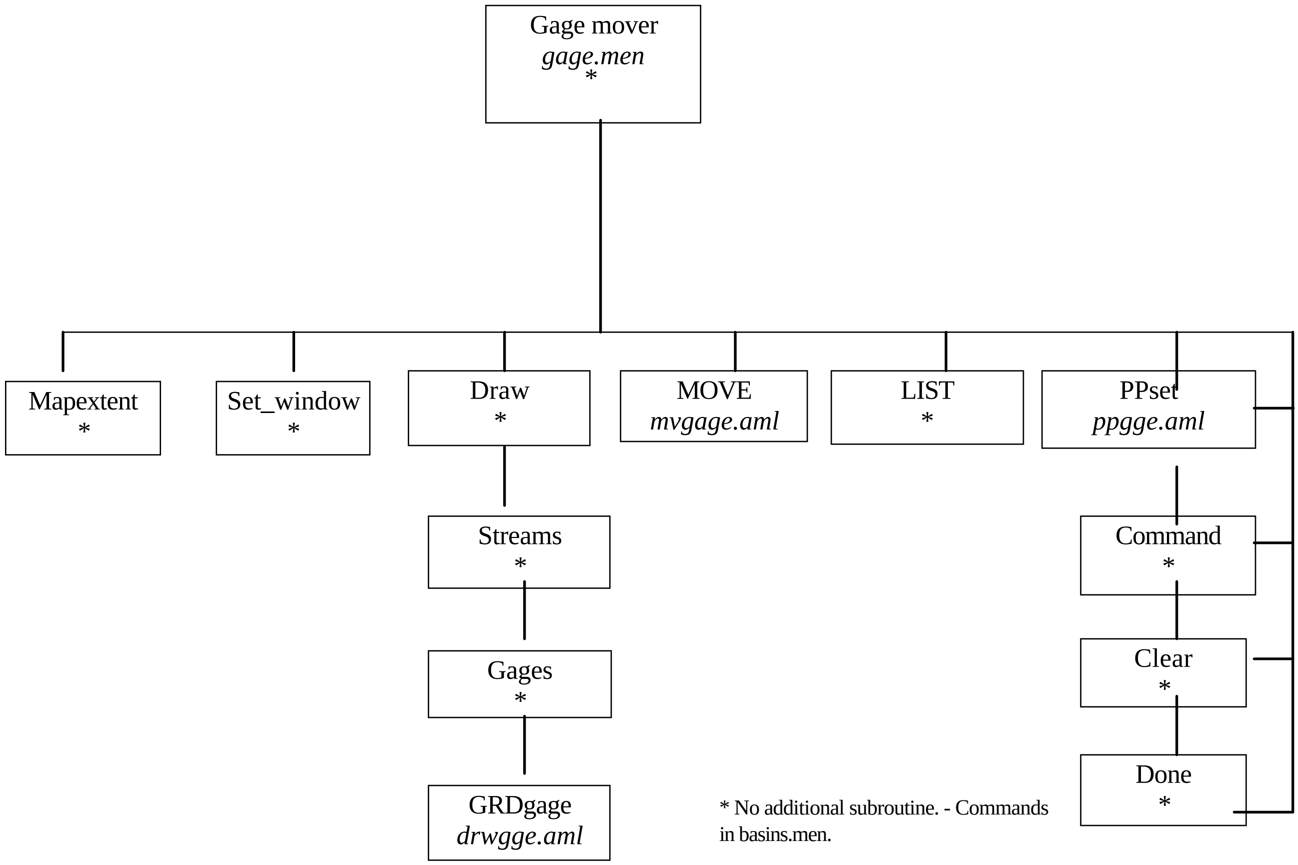

Figure 4-8: Menu layout for Gage Mover

added as attributes to either the delineated watershed or subarea coverage polygon attribute tables (PAT's). Calculated values that might be used in subsequent calculations within HDDS are stored as global variables.

Menus and Macros

HDDS comprises a main menu (basins.men) which is invoked from a variable initialization routine (hdds.aml). The initialization routine establishes default parameters and station settings such as the terminal and display environments. The main menu provides a user-interactive means of initiating functions or other macros which, in turn, may invoke sub-menus and macros.

File Naming Convention and Output Files

Rather than request user-defined file names, a simple convention is employed to minimize the need for user input: any grid or coverage that is created for other than temporary storage is named using three parts: (1) a prefix which represents the data set on which the process has operated, (2) an abbreviation representing the operation used to create the output, and (3) a user-defined suffix to make the name unique. The same suffix is applied to all created data during a particular series of analysis. For example, if the user uses the "tx" data set and specifies the suffix to be "BP", the grid of watershed area will be "txshedBP". If the user specifies a suffix for which files already exist for the current data set, the program will request a new suffix.

Data Preprocessor

The original data set "tx" was established manually as described in Section 4-1. A data preprocessor AML (preproc.aml) has been developed which emulates these steps. The preprocessor may be used on any Arc/Info format digital elevation grid to generate a new data set on which HDDS may operate. The routine also converts user-specified highway coverages in to grid coverages. Currently, this preprocessor is set up assuming the elevation data are in Texas so that the existing data such as design rainfall, land use, and soils data can be employed. However, it should not be difficult to modify the routines to accommodate data from other geographic regions.

Generally, the methods employed in the preprocessor are the same as those outlined in Section 4-1. However, automation of the map projection process required additional steps to overcome deficiencies in AML: the Arc/Info projection process allows a preset ASCII file containing input and output projection parameters. This is satisfactory only if projection files of the input coverage are always the same or if the user can create a specific ASCII projection file. In many instances, especially when the original data are in UTM, the input file projection parameters will vary. Arc/Info version 7.01 will not simultaneously determine the input projection parameters from the existing coverage and read an ASCII file of only output parameters.

A routine was developed to accommodate these shortcomings such that the user need not provide the appropriate parameters. Two ASCII files (utmalb.prj and geoalb.prj, Appendix A) were created in which the input projection parameters are defined as global variables and the output parameters are specified as those required for HDDS. The routine in preproc.aml (Appendix A, lines 173-179) uses the ARC command DESCRIBE to determine the input projection parameters and assigns the appropriate values to the same variable names as those established in the ASCII projection file. A conditional statement determines which projection file to use based on the selected scale of DEM data. When the projection process uses the projection parameter file, global variables are read for the input projection parameters and the fixed parameters are read for the output projection (Appendix A, preproc.aml, lines 180-194). This process could be modified further to avoid the need for two separate projection files.

Data Set Selection

The AML routine “infiles.aml” (Appendix A) allows a user to select a desired data set using standard AML procedures. As long as any new data sets are stored at the same directory level as the existing data sets (“tx” and “g”), they will be automatically available for selection. The name of the selected data set is stored as the prefix name (a global variable) for all subsequent data acquisition and file naming as discussed above.

Display Coverages

No analysis can be performed until data coverages have been displayed. The primary display coverages are major drainage basins (grid), major highways (lines), and streams (lines). Menu options in "basins.men" contain strings of standard ARCPLOT commands which draw these coverages on request. These coverages are the most basic display requirements for performing a hydrologic analysis at an existing highway crossing of a stream. If analysis at a location other than a highway is desired, other user-created vector coverages may be overlaid on the base data for visual aid.

Outfall Identification

A highway stream crossing constitutes the outfall to a watershed when the crossing is the subject of a hydrologic analysis. This would be the case for design of replacement bridges or culverts. No existing method is apparent in Arc/Info to easily and consistently select a specific location to be identified as an outfall. Several methods have been developed within HDDS for identification of an outfall location:

1. Highway/stream intersection
2. User-defined cell(s)
3. User-defined polygon(s)
4. User-adjusted stream gauge location(s).

The first method is designed for analyses at existing highway crossings of streams and relies on the presence of highway and stream coverages in both vector and grid format. The data were created such that the cell values of each gridded highway and stream link match the identifiers of the associated arc (vector coverage) of each highway and stream link respectively. The ARCPLOT RESELECT command is used to select interactively and store the identifiers of a desired highway and a stream. (Appendix A, idroad.aml, line 28 and idstrm.aml, line 27). In GRID, a conditional (CON) statement is employed to determine the cell at which the stream id and highway id are coincident (Appendix A, wshed.aml, line 53). The resulting cell is then stored as an outfall grid (or pour point) in which only one cell has a value of 1. **Figure 4-9** illustrates the process. It should be noted that, occasionally, a highway may cross the same segment of stream (arc) more than once, in which case more than one pour point could exist. The system will use the cell containing the lowest value, which may not be the desired location. One way to avoid this problem would be to modify the original highway or stream coverage by adding nodes between the highway/stream crossings to create intermediate arcs. This has not been performed for this project.

In some cases, an analysis may be desired at a stream location where no highway currently exists. Methods 2 and 3 were established for such instances. The cell method is the most simple in terms of programming: the GRID command SELECTPOINT is used to select interactively and store the value and location of the selected cell of a defined grid. HDDS incorporates a counted looping routine which allows the user to select as many points as desired (Appendix A, outfall.aml). A grid is created in which each selected point is represented by a cell containing a unique but sequential value set by the counter.

The polygon method is similar to the cell method except the GRID command SELECTPOLYGON is employed. This is nested within a zonal routine and conditional statement which determines the maximum flow accumulation cell contained within the defined polygon. Multiple polygons may be defined to create a grid of pour points (Appendix A, line.aml).

The fourth method creates a grid of pour points based on a grid coverage of stream gauges created using the gauge mover tool (see “Gauge Mover” later in this section).

The highway/stream intersection method of outfall identification is the most secure and consistent of the aforementioned methods: the repeated selection of a highway link and the stream link over which the highway crosses will always yield the same pour point. The polygon method will always select a pour point that falls within a stream as long as the user-defined polygon contains a stream cell. However, since the polygon method selects the cell containing the highest flow accumulation, the most downstream cell within the polygon will always be selected as the outfall. Repeatability then, is dependent on the user’s ability to

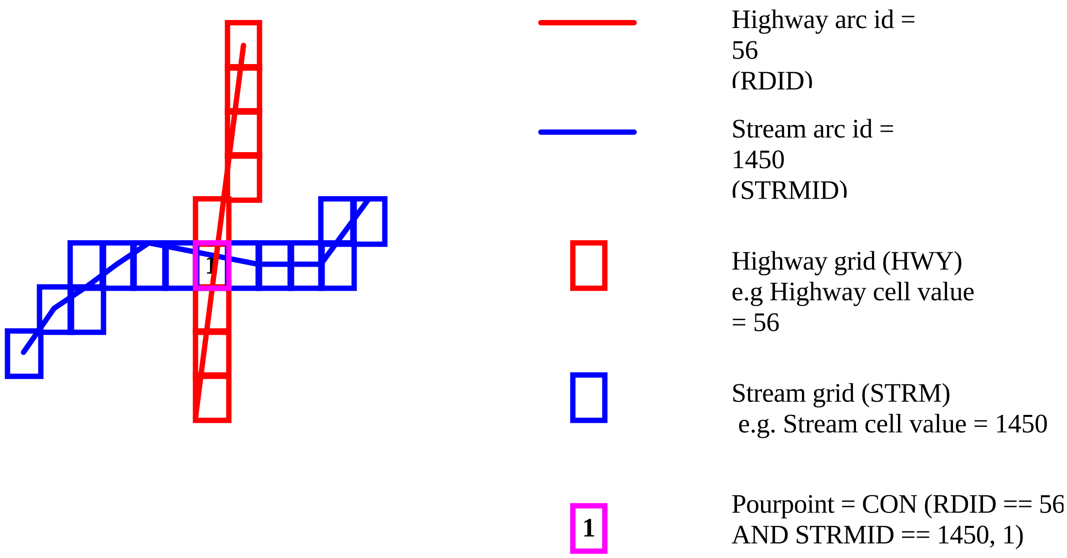

Figure 4-9: Locating the watershed outfall (pourpoint)

define a polygon in such a manner that the most downstream cell in the polygon is the desired location. Also, it is important that the defined polygon does not incorporate more than one stream. In many instances, the user may be satisfied with identifying the location of the outfall within an accuracy of one or two cells, in which case the polygon method will suffice.

The gauge location method is limited by the existence of stream gauges and the ability of the user to identify and select accurately the location of the gauge on a stream. Most stream gauges are sited on highway bridges, in which case it may be preferable to use the highway/stream intersection method.

The cell method is the least secure means of determining a pour point: the user must zoom in until the display resolution allows identification of individual cells. The cell selected may not always coincide with a stream (or highway) due accuracy limitations between the vector and grid coverages of streams and highways.

Watershed and Subarea Delineation

The outfall must be identified by one of the above methods. If only one outfall has been identified, a grid containing the outfall cell (pour point grid) is then used in conjunction with the flow direction grid to delineate the watershed boundary using the GRID command WATERSHED. This command uses the predetermined flow direction grid (described earlier) to identify all the cells that contribute flow to the defined pour point cell as indicated by **Figure 4-10**. The periphery of these cells defines the watershed boundary. The area is calculated by two methods:

1. The GRID command ZONALAREA is used to count all the cells containing the same cell value and then multiply the number of cells by the cell area to yield the total area of cells (Appendix A, wshed.aml, line 67).
2. The preprocessed data includes a flow accumulation grid for each data set (txacc and gacc). By accessing the cell in the accumulation grid that is coincident with the outfall cell (pour point), the area is calculated as the flow accumulation plus the outfall cell multiplied by the cell area (Appendix A, wshed.aml, line 65).

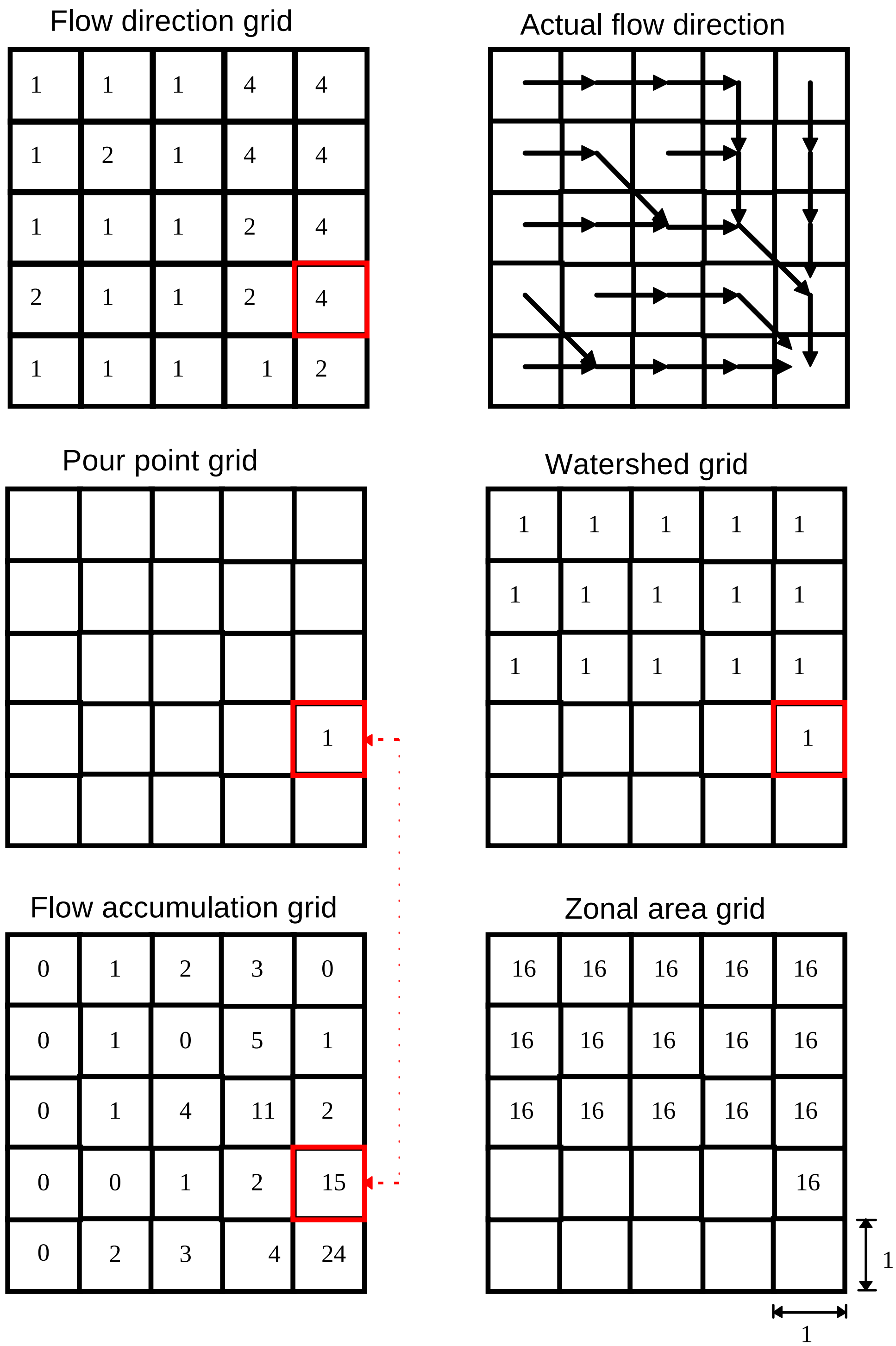

Figure 4-10: Determination of drainage area

The use of two methods allows for easier identification of errors. This is discussed in Section 5.

The resulting watershed grid is converted into a vector coverage and the calculated area (ZONALAREA) is added to the watershed polygon attribute table. Once the watershed has been delineated, the window area is reduced to just include the maximum extent of the watershed (Appendix A, winset.aml). This minimizes the amount of data needed for subsequent processing.

The subarea delineation process also uses the WATERSHED command. The difference is that subarea delineation uses multiple pour points as opposed to one. If a single pour point has been defined, the system combines the presence of stream tributaries and a user-defined area threshold to establish subarea pour points. Stream confluences with tributary sub-areas exceeding the threshold value will be defined as pour points. This step is bypassed if multiple pour points have been assigned using outfall identification methods (2), (3), or (4).

If multiple pour points were defined using the cell, polygon, or gauge mover methods, the subarea routine (subshed.aml) is used to determine drainage areas for the defined pour points rather than the watershed routine (wshed.aml). The pour points are assigned consecutive, unique values. These values are then assigned to the grid of delineated subareas such that each subarea can be uniquely identified. A vector coverage of the subareas is created using the subarea zone value as the polygon identifier. Calculated subarea sizes are added as attributes to the PAT of the subarea vector coverage and the values are written to an ASCII format file for possible use in external programs.

Longest Flow Path and Length

After the watershed has been delineated, the alignment and length of the longest flow path may be determined. GRID provides the command FLOWLENGTH which determines either the distance from the most remote cell to each cell in the grid (downstream option) or the distance from the outfall to every cell in the grid (upstream option). No provision is apparent in Arc/Info for providing directly the longest flow path.

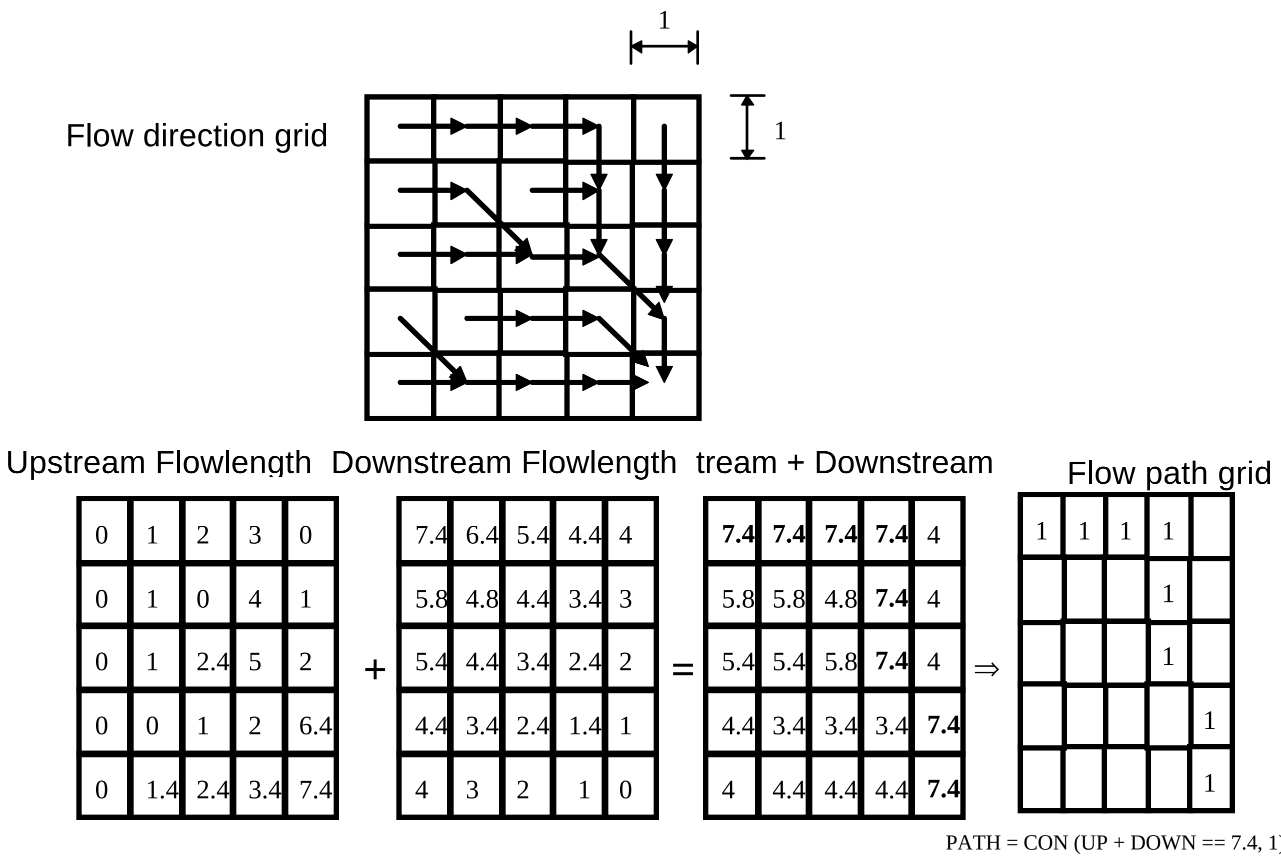

Figure 4-11: Method of flow path determination

The summation of an upstream flow length grid and downstream flow length grid yields a grid in which each cell in a unique string of cells has a value equal to the longest flow length. All other cells will have values that are lower than the maximum length. A conditional statement is then applied to isolate those cells whose value equals the maximum flow length. **Figure 4-11** demonstrates the process. Theoretically, the conditional statement could just check for each cell in the summation grid whose value is equal to the longest flow length. However, the occurrence of rounding errors precludes exact equality. Therefore, the conditional statement has been set to check for all cell values exceeding the sum of the value of flow length minus one (Appendix A, pthleng.aml, Line 48). This approach is reasonable as long as the cell resolution is greater than one meter! The resulting cells constitute the longest flow path.

The flow path grid is converted into a vector coverage so that it can be overlaid on the watershed display. The calculated path length is added to the watershed PAT.

Average Watershed Slope and Hydrologic Region

The average watershed slope was defined in Section 3.3.1 as the slope between the 10% and 85% points along the longest flow path. This system uses the delineated watershed and flow path, and calculated flow length to determine the cells at which the 10% and 85% points are located. Elevations are established at these points by accessing the elevation grid. The general procedure follows:

1. The cells in the upstream flow length grid that are coincident with the flow path are isolated using SELECTMASK with the gridded distance (not time) flow path as the mask grid operating on the upstream flow length grid (Appendix A, uslope.aml, line 38). This creates a grid containing a string of cells in which each cell value records the flow path distance of the cell from the watershed outfall.
2. Since no cell values in the string are likely to match the distances to the desired points, a conditional statement is employed to determine the cell in which the 85% point is located. A grid is created in which the desired point is represented by one

cell (Appendix A, uslope.aml, Lines 45 & 46). The same is done for the 10% point. The actual flow length values of the located cells are then stored for use in the slope calculation.

3. The grids containing the individual points are used as masks to determine the values of the coincident cells in the elevation grid using SELECTMASK (Appendix A, uslope.aml, Lines 53 & 56).
4. The slope is calculated as the quotient of the difference in the elevations determined in 3 and the difference of the flow length values determined in 2.

Generally, for highway design, the watershed average slope is computed for use in rural regression equations as described in Section 3.3.1. HDDS assumes that if the calculation of average slope is requested, the hydrologic region will also be required for use in the regression analysis. The hydrologic region number is determined by using the delineated watershed boundary coverage to clip the statewide hydrologic region coverage (gsrgns). The ARCPLOT commands RESELECT and SHOW SELECT are used to store the region number (Appendix A, uslope.aml, Lines 78 & 79).

The computed slope and identified hydrologic region number are added as attributes to the watershed PAT.

Estimation of Weighted Runoff Curve Number (RCN)

Section 3.3.3 discusses the use of the Soil Conservation Service runoff curve number method and indicates the derivation of curve numbers based on a description of land use/ land cover and hydrologic soil group. Typically, the manual process of determining weighted runoff curve numbers is tedious. HDDS mimics the manual process but on a much more refined scale than would normally be done manually. In order for the RCN routine to be performed, the watershed must have been delineated. The following steps outline the process:

1. The window extent of the watershed coverage is used to create a rectangular clip coverage (Appendix A, rcn.aml, Lines 47-78).

2. The land use (txlus) vector coverage is reduced to the watershed extent using the clip coverage created in 1 (Appendix A, rcn.aml, Line 82).
3. The reduced land use coverage is then converted in to a grid coverage using the same cell resolution as the elevation grid from which the watershed was delineated. The land use code number (lucode) is specified as the item on which the land use coverage is to be gridded (Appendix A, rcn.aml, Line 83).
4. The base soil data vector coverage (statsgo) is directly reduced and gridded to the window extent of the watershed coverage using the soil polygon identification number (statsgo-id) as the grid item. (Appendix A, rcn.aml, Line 88).

   The same one-step process could have been used for the land use data, however the processing speed would be much slower due to the inordinate number of polygons in the land use coverage.
5. The existing land use data for the current watershed may be modified interactively, the methods for which are described below.
6. The runoff curve number table (rcns.dat) is joined to the value attribute table (VAT) of the watershed land use grid by assigning an alternate name of "lucode" to the value item and using the item "lucode" as the relate item. Also, the attributes of the soil coverage (statsgo) are added to the watershed soil grid VAT in a similar manner using the "statsgo-id" as the relate item (Appendix A, rcn.aml, Lines 93-111).
7. Using either the watershed grid or the subarea grid as a mask grid, a runoff curve number is calculated for each cell in the watershed. This is accomplished using a DOCELL routine in which the fraction of each soil group (from the soil grid VAT) is multiplied by the runoff curve number associated with the land use code and soil group (from the land use grid VAT). The products are then summed to yield a weighted runoff curve number (Appendix A, rcn.aml, Lines 116-125).
8. Step 7 provides a grid in which every cell may have a unique value. The GRID command ZONALMEAN is used to determine the runoff curve numbers weighted by either watershed area or by subareas (Appendix A, rcn.aml, Line 128).

The resulting runoff curve numbers are appended to either the watershed PAT or the subareas PAT as appropriate.

This process is at least as accurate as current manual methods. In fact, it is likely that a user would describe varying land use conditions with more detail when using an automated, interactive system than when performing the task manually.

Modification of Land Use Codes

HDDS accommodates user-specified land-use codes which may be used to override existing land use data. This is achieved by allowing the user to specify a value of land use code to be established as the value for all cells within a user-defined polygon. A looping mechanism was developed to allow definition of as many polygons as desired to be drawn over the existing land use data. The current value of land use code is applied to each polygon. The land use code may then be changed and additional polygons defined (Appendix A, lupoly.aml , Lines 30-39).

Each defined polygon results in a separate grid. After each set of polygons has been defined, the polygon grids are merged with the existing land use grid, the override precedence being on the most recently defined polygon. That is, if any polygons overlap, the most recently defined values are used (Appendix A, lupoly.aml , Lines 40-48).

Since a merging process is used, wherever no polygons have been drawn within the watershed, the existing land use codes will remain. This allows the user to perform a before and after land development comparison and to accommodate changes that have occurred since the original data were developed.

Travel Time and Path

The time of travel of water over the land surface is estimated as a function of longitudinal slope and surface roughness as described in Section 3.3.3. The steeper the slope for a given surface roughness, the faster is the velocity of flow and the shorter the travel time. The path that determines the longest travel time (time of concentration) could be different from the longest distance path. The flowlength function is used in a similar manner to that described for flow path, however, a weighting factor is used to convert the length into time (Appendix A, tofc.aml). This is done by creating a grid of the reciprocal of the estimated velocity of flow in each cell which is used as the weight for the FLOWLENGTH function. Several means of establishing velocities are incorporated in HDDS (Appendix A, tcwt2.aml). These are:

1. The default method looks at the slope (S in m/m) of each cell, assigns a uniform velocity coefficient (b) to all cells, and calculates a velocity of flow (v in m/s) over each cell using **Eq. (4-1)** from Section 4.1 which is repeated here for clarity (Appendix A, tcwt2.aml, lines 67-74).

   $$v = b \cdot S^{0.5} \qquad (4\text{-}1)$$

   The default value of b is 4.58 (metric) and assumes short grass waterway (**Table 4-7**).
2. If selected by the user, the existing or user-modified land use coverage can be used to look up the velocity coefficient as assigned in **Table 4-6**. The velocity of flow in each cell is then calculated using **Eq. (4-1)** (Appendix A, tcwt2.aml, lines 47-53).
3. The user may draw polygons over the land use data to which are assigned user-defined velocities (Appendix A, luadj.aml). The user-defined velocities, the method for which is described below, will override those computed using method (2).

For methods (1) and (2), an implicit minimum velocity is applied by virtue of a threshold cell slope: if the cell slope is less than 0.3%, a cell slope of 0.3% is used to establish a lower velocity limit dependent on the assigned velocity coefficient. This is done instead of specifying a lower velocity threshold for two reasons: (1) If the slope is zero, no root exists and use of **Eq. (4-1)** would fail. (2) An explicit minimum velocity would not accommodate varying surface roughness characteristics. This approach seems reasonable considering that the calculated velocities can be overridden with user-defined velocities.

The longest time of travel is then computed using the inverse of the velocity as a weight factor in the flow length routine. (Time = length / velocity). The calculated watershed time of concentration is then added to the watershed PAT.

It should be recognized that methods (1) and (2) are only reasonable for overland sheet flow or shallow concentrated flow. Estimates of flow velocities in streams should be made using the average sectional properties of the stream and supplying the estimated velocities using method (3).

If the user has requested subarea delineation, HDDS will calculate subarea times of concentration. The process of computing the watershed time of concentration results in two time grids: one in which each cell value reflects the calculated travel time from its own location to the watershed outlet cell (upstream time), and the other in which each cell value reflects the travel time from the drainage divide to the cell (downstream time). The subarea time routine (Appendix A, tofc.aml, lines 86-126) uses these grids and the grid of subarea pour points to determine subarea path and times in the following manner:

1. The inverse of the watershed time path is created using the GRID command ISNULL. That is, a grid in which the flow path is reflected by NODATA and all other cells have a value of 1.
2. The upstream and downstream time of travel grids are summed such that every cell reflects the total time from the drainage divide to the outfall for the flow path that runs through the cell.
3. The cells constituting the main watershed path are removed from the total time grid using the GRID command SELECTMASK and using the grid created in (1) as the mask.
4. The remaining maximum times for each subarea are determined with the GRID command ZONALMAX using the subarea grid as the zone grid operating on the grid created in 3. That is, a grid is created in which all the cells within a subarea have the value of the maximum time of travel for the subarea to the watershed outfall (not the subarea pour point). If the main flow path cells had not been eliminated, the subareas through which the main flow path runs would result in cells containing the time of concentration of the whole watershed.
5. The flow paths for each subarea are determined in the same manner as the main path by recognizing that the sum of upstream and downstream time grids by

subarea should yield a string of cells in each subarea in which the sum equals the maximum time for the subarea.

6. Finally, the times of travel for each subarea are calculated as the difference between the time of travel between the most remote cell and the pour point in each subarea flow path. This is performed using the GRID function ZONALRANGE with the subarea path grid as the zone grid operating on the downstream time of travel grid.

Subarea times of concentration are added as attributes to the PAT of the delineated subarea vector coverage.

Specification of Flow Velocities

A grid of the reciprocal of cell flow velocities based on velocity coefficients is created in the time of concentration routine, as described above. Velocities may be defined to override use of the velocity coefficients in much the same manner as that described for land use codes. The existing reciprocal of velocity grid is merged with the polygons of the reciprocal of defined velocities (Appendix A, velpoly.aml).

This is especially useful for defining estimated flow velocities in streams: the stream segments may be subdivided as desired and a velocity assigned to each segment. Thus, the time of travel computation process can emulate the manual process with much less effort and much more detail, if desired.

Calculation of Ellipsoidal Scale Factors

Section 3.2 provided discussion on geodesy and map projections to familiarize the user with some important concepts relating to spatial accuracy and representation of data. All of the spatial data employed herein are projected on to a flat plane using the Albers Equal Area projection. This implies that relative area (size) is preserved but direction and distances are not. HDDS incorporates a scheme by which distances measured on the plane may be adjusted to account for the curvature of the Earth at the surface of the reference ellipsoid. It is recognized that the order of accuracy of measured lengths based on the 500m and 93m grids used here may result in greater uncertainty than the projection error. However, the process incorporated in HDDS demonstrates a potential capability that could allow analyses to be performed in one projection rather than defining a specific projection for each individual project.

A grid scheme was developed in which a longitudinal factor (h) and a latitudinal adjustment factor (k) are calculated for each attributed cell in the path length grid using the process described in Section 3.2.5 and the calculated centroidal coordinates of each cell. The scheme determines the direction of travel over each cell, the distance traveled over each cell, and creates a grid of adjusted travel length over each cell as follows:

- for East-West or West-East direction: adjusted length = cell size * k
- for North-South or South-North: adjusted length = cell size * h
- for NE-SW, SW-NE, NW-SE, or SE-NW: adjusted length = cell size * $(h^2 + k^2)^{0.5}$

The sum of the length values in each cell then yields the total adjusted length. This is performed using the GRID command ZONALSUM (Appendix A, sclfctr.aml). The adjusted length is appended to the watershed PAT.

Use of the Albers Equal Area projection negates the need for determining an area factor so HDDS does not attempt to do so. However, the scheme could be modified for use with other projections: the product of h and k can be determined for all cells in a zone (watershed or subarea) and summated using ZONALSUM. The resulting value can then be multiplied by the cell area to determine the adjusted area of the zone.

Watershed Shape Factor

The watershed and path length routines store the calculated area and maximum path length, respectively, as global variables. The watershed shape factor is a calculation of the watershed area divided by the square of the path length. Use of the global variables in this manner reduces the time-consuming need to access data stored in the attribute tables. The resulting shape factor is appended to the watershed PAT.

The current regression equations for determining flood frequency described in Section 3.3.1 do not call for a shape factor. However, on-going research indicates the need to consider such a variable (Slade, 1995, personal communication).

Design 24-Hour Rain

Section 3.3.3 described the use of design rainfall depths in the SCS runoff curve number method. HDDS uses the watershed grid or the subarea grid as a zone grid to weight the rainfall values of the coincident cells in the rainfall grid (raingrd). The rainfall grid contains the 2, 5, 10, 25, 50, and 100 year recurrence rainfall values (in inches) in the VAT. The GRID command ZONALMEAN is applied to each frequency item of the VAT to yield a grid of design rainfall amounts averaged over either the watershed or each subarea. The units are converted from inches into millimeters and the results are appended to the watershed PAT or subareas PAT, whichever is appropriate (Appendix A, dsgnrain.aml).

Digital Elevation Model Quadrangle Names

As indicated in Section 4.1, HDDS contains 1:2 M digital elevation data covering the whole of Texas and 1:250 K data covering a small portion of northeast Texas. It is anticipated that the 1:2 M data will suffice for basic hydrologic analysis of very large watersheds (of the order of thousands of square kilometers). Larger scale data will likely be needed for smaller watersheds, however, the 1:2 M data are still useful as a first level approximation: the drainage area can be delineated quickly using the 1:2 M data. The resulting watershed extent can be used to clip data from 1:250 K or 1:24 K scale quadrangle index coverages to determine the names of the higher resolution quadrangles required for more detailed analysis. The routine "fndquad.aml" (Appendix A) clips appropriate quadrangle index data and determines the number of polygons in the clipped coverage. The routine "slect.aml" (Appendix A) selects the name item of each polygon within the clipped index coverage and writes the name to an ASCII file. This file is subsequently displayed and can be used in the preprocessor (preproc.aml) to create a new data set, assuming the quadrangles of digital elevation data are available.

Gauge Mover

The presence and use of stream gauge data have been discussed in Sections 4.1 and 3.3.2 respectively. Often, the stream gauges are not coincident with the appropriate stream cell. This is a result of the combination of inaccuracies in the stream locations as determined from digital elevation data, and inaccuracies in the locations of the stream gauges as defined by coordinates of latitude and longitude. HDDS incorporates a routine that is designed to reduce the tedium of manually adjusting stream gauge locations. The relocated stream gauges may be used as pour points for subsequent drainage area delineation.

The following outlines the process, which assumes that the gauge coverage and gridded streams are displayed (a tool is provided to do so):

1. The stream gauge data are stored in a point coverage named txgages. The ARCPLOT command RESELECT is employed to select interactively a desired gauge (Appendix A, mvgage.aml, Line 34).
2. The identifier of the selected gauge is determined and stored using SHOW SELECT (Appendix A, mvgage.aml, Line 35).
3. The cell in the stream grid to which the gauge is to be relocated is identified using the GRID command SELECTPOINT (Appendix A, mvgage.aml, Line 40). This creates a grid with one cell containing the value of the gauge identifier.
4. Steps 1 to 3 are repeated for as many gauges as desired. A separate grid is created for each relocated gauge (Appendix A, mvgage.aml, Lines 32-45).
5. On completion of selecting gauges, the individual grids of relocated gauges are merged into one grid coverage (Appendix A, mvgage.aml, Lines 46-49).
6. The attributes of the original gauges are joined to the VAT of the relocated gauge grid (Appendix A, mvgage.aml, Lines 51-61).

This routine allows gauges to be placed only where a gridded stream exists.

The grid of relocated stream gauges may be used as a pour point grid. This is accomplished by renaming the grid as a pour point grid (e.g. txppa) and initializing a variable which indicates that a pour point grid has been determined for subsequent use. Also, the number of gauges in the grid is stored as a variable so that the program can prompt the user to use the watershed delineation tool when using a single pour point grid or the subarea tool for multiple pour points. The areas computed using HDDS may be compared with the documented drainage area which is stored as an attribute of the gauge grid VAT.

Hydrologic Computations

The system uses the Texas Hydraulics System computer program (THYSYS, 1977) to compute frequency versus discharge for frequencies of 2, 5, 10, 25, 50, and 100 years based on the regional regression equations discussed in Section 3.3.1.

The derivation of watershed area, average slope and hydrologic region were discussed above. The routine "usgs.aml" requests a job description and file name then writes this data and the global variables of area, slope and hydrologic region to an ASCII file in THYSYS input format. THYSYS can be invoked automatically from within HDDS using the newly created input file. THYSYS creates an ASCII output file from which HDDS abstracts the calculated peak discharge versus frequency data. These results are then appended to the watershed PAT (Appendix A, usgs.aml).

Currently, only the THYSYS regression analysis approach is run automatically from HDDS, however, the processes required for automatic analysis of the SCS runoff curve number method, or any other hydrologic method, are much the same.

Clip Coverages

Once the watershed is delineated and converted into vector format, it can be used to clip out any available data. This system contains the following data which are clipped by the watershed coverage:

- percentage of hydrologic soil group,
- county boundaries,
- highways,
- streams, and
- stream gauge sites.

Additionally, the watershed grid is used as a mask to create a grid of aspect of each cell in the watershed.

The resulting coverages and those already established by the methods previously described may then be imported into viewing and output generation software such as ARCVIEW.

Cleanup

Many coverages are created during the course of execution of this system. Those coverages that are not needed for documentation or possible subsequent analysis are removed automatically. The remaining coverages may be removed by the routine "cleanup.aml" (Appendix A) which checks that a coverage exists before eliminating it.

Summary of Data Created in HDDS

Table 4-8 provides a list of grids and vector coverages created by HDDS. This assumes a data set "tx" and a prefix "a" were used. Sample data are discussed in Section 6.

Documentation (Metadata)

Several tables describing geographic data set names and attributes are presented herein and constitute a data dictionary. These are useful for general reference. However, oftentimes specific detail of the data is required when data are used by someone other than the developer.s

Federal standards for documenting spatial data (metadata) were established by the Federal Geographic Data Committee (FGDC, 1994) with which all Federal government agencies should comply. These requirements are to ensure that the origin and intended use of spatial data are not lost during technology transfer or due to attrition within the agency developing the data.

The standards include the following items:

1. Identification information
2. Data quality information

3. Spatial data organization information
4. Spatial reference information
5. Entity and attribute information
6. Distribution information
7. Metadata reference information including citation, time period, and contact information.

The provision of metadata in accordance with the FGDC standards can be a daunting task. The standards currently apply to Federal agencies only, though it is likely that other agencies will adopt similar standards.

HDDS includes a menu-driven subsystem which is designed to conform with the intent of the FGDC standards (Appendix A, meta.men). Only two coverages, txrds (highways), and gsrgns (hydrologic regions) have metadata that can be accessed through HDDS. Additional coverage metadata can easily be added to the menu system once the information is written. Appendix B provides sample metadata for the highway coverage in the North Sulphur River (grds).

Table 4-8: Coverages created in HDDS

Data name	Feature type	Theme	Attributes
txsheda	GRID	watershed	No. of cells in the watershed
txtmpa	POLYGON	watershed	Area *
txsuba	POLYGON	subareas	Subarea id and size*
txptha	GRID	longest path	No. of cells in path
txptharca	LINE	longest path	path length
txtcwta	GRID	cell weight for t of c	reciprocal of velocity
txwtraina	GRID	rainfall grid	2 - 100 yr rainfall
txppa	GRID	watershed pour point	-
txsppa	GRID	subarea pour points	-
txaspcta	GRID	watershed aspect	direction of slope
txtptha	GRID	time of travel path	-
txtptharca	LINE	watershed time path	time of concentration
txstptha	GRID	subarea times of travel paths	-
txstptharca	LINE	subarea time paths	subarea t of c’s
txgagea	GRID	moved stream gauges	stream gauge data
txrcna	GRID	Runoff curve numbers	weighted RCN
txhydgrpa	POLYGON	hydrologic soil group	soil name, % hyd grp

* The path length, time of concentration, hydrologic region, average slope, weighted runoff curve number, shape factor, and calculated discharges are also appended if applicable.