Chapter 3

Data Sets

3.1 DATA FORMATS

For this project, numerous types of data were acquired in order to produce the model inputs and GIS coverages. Although each data set is discussed in more detail in later sections, Table 3-1 shows the general data sets used in the methodology.

Table 3-1 Digital Data used for GIS Database Development

Data	**Original Format**	**Source**	**Section(s) Discussed**
Digital Line Graphs	Exported Coverage	USGS, 1993	3.2.2
Digital Elevation Models	Grid	USGS, 1996	3.2.3
Land Use/Land Cover	Exported Coverage	Newell, *et al.*, 1992 and USEPA, 1996	3.2.4
Precipitation	Grid	Daly, *et al.*, 1994	3.2.5
Flow Data	ASCII	Texas USGS, 1996	3.2.6
Point Source Dischargers	Paper, ASCII	Armstrong and Ward, 1994, and Visnovsky, 1996	4.4
Channel Segmentation Information	Paper	Espey, *et al.*, 1971	4.2
Water Quality Information	ASCII Computer File	Ward and Armstrong, 1992	5.4

In addition, Table 3-1 also shows the particular section in which the data is further discussed, the original format of data, and the source from where the data was obtained. The data obtained was usually available in one of five different forms:

- Grids -- data which consisted of a spatial, uniform grid, with a measurement value applied to the center of each cell in the grid; imported into GIS using Arc/Info
- Coverages -- data obtained in an exported format (e00) and imported into GIS, using Arc/Info

- ASCII Files -- data, in a delimited format, obtained from the internet, an anonymous ftp site, or a personal acquisition from an organization
- Computer File -- data obtained in a software format (e.g., Microsoft Excel)
- Paper -- data which were obtained from a literature search or personal communication and keyed into the computer

Most of the data was imported into the computer and formatted for use in a PC or UNIX software environment. Subsequent sections provide further discussion on the manipulation of the data for use in a computer, specifically GIS, environment.

3.2 OBTAINING A DIGITAL DATABASE

To utilize GIS, characteristics of the study area needed to be digitally represented in a format compatible with the GIS software, Arc/Info 7.03 and ArcView 3. These characteristics included, but were not limited to, surface terrain, streamflow, areal precipitation, land use, and area hydrography. A majority of this data was obtainable via internet. Table 3-2 gives the internet addresses, current as of August 1996, for the data that was downloaded through the world wide web or file transfer protocol (ftp). The following sections describe how data for this digital representation were imported into GIS, using Arc/Info and ArcView. The result was a series of coverages, grids, and tables which contained information used in the digital representation of the study area. The attributes of many of the coverages, grids, and tables contained the parameters and input needed to run the water quality model. Appendix A contains a detailed list of all coverages, grids, and tables generated in GIS for this project.

Table 3-2 Internet Addresses for Data Sources

Data Source	Internet Address
Digital Line Graphs	http://sun1.cr.usgs.gov/eros-home.html
Digital Elevation Models	http://sun1.cr.usgs.gov/eros-home.html
Land Use/Land Cover Files	ftp earth1.epa.gov
Precipitation Data	ftp fsl.orst.edu
USGS Gauge Station Locations	http//txwww.cr.usgs.gov/cgi-bin/nwis1_server/
USGS Daily Discharge Values	http//txwww.cr.usgs.gov/cgi-bin/nwis1_server/

3.2.1 Map Projection

In order to represent the coverages and grids on a plane surface, a map projection was chosen. All of the coverages and grids for this were projected into USGS National Albers Equal Area Map Projection (USGS-Albers). USGS-Albers does not distort the area of the projected polygons. This quality is important in hydrologic modeling, since many characteristics of a watershed may be expressed per unit area. The parameters for this projection are found in Table 3-3, while the projection file, geoalb.prj referred to in various procedures is found in Appendix B. There are two datums listed in this table, World Geodetic System 1984 (WGS84) and North American Datum 1983 (NAD83). Since coverages initially existing in WGS72 cannot be projected into NAD83, WGS84 is used, instead. The errors involved with using NAD83 and WGS84 to spatially represent data in the same area are minimal.

Table 3-3 USGS Albers Equal Area Projection Parameters

Datum	Word Geodetic System 1984 (WGS84) or North American Datum 1983 (NAD83)
Units of Length	Meters
First Standard Parallel	29° 30′ 00″ N
Second Standard parallel	45° 30′ 00″ N
Central Meridian	96° 00′ 00″ W
False Easting	0.0
False Northing	0.0

3.2.2 Digital Line Graphs

Digital Line Graphs (DLG) which are available from the US Geological Survey (USGS), are a 1:100,000 scale representation of the hydrography in the study area and surrounding watershed (USGS, 1993). The maps, which are organized by USGS 1° x 2°

quad name, show streams, rivers, creeks, canals, lakes, and shorelines (Figure 3-1). For this project, the quads: Houston, Conroe, Anahuac, and Beaumont encompassed the entire watershed area. The zipped files, titled by each quad name, were downloaded from CD ROM (USGS, 1993). The unzipped files consisted of up to eight separate 15′ x 15′ maps which, when joined together, constituted the hydrography for an entire 1° x 2° quad. Over 30 of these smaller coverages were joined together to form one large map of the Houston Area (Figure 3-1). A short program, written in Arc Macro Language (aml), developed by Saunders (1996) and shown in Appendix C-1, explains how to obtain the files from CD ROM, unzip the files in UNIX, import them into Arc/Info, erase the map borders, and join the entire set of 15′ x 15′ maps to form one large map.

3.2.3 Digital Elevation Models

Digital Elevation Models (DEM) are available via the internet from the USGS web site (USGS, 1996). One DEM file is a grid which covers a 1° x 1° surface area and contains an elevation value every 3″. The result is a 1201 x 1201 cell grid with 1,440,000 data points describing the surface terrain. Four 1:250,000 scale DEM grids were merged to include the Houston Ship Channel area: Houston-West, Seguin-East, Beaumont-West, Austin-East (see Figure 3-2). These grids are originally in geographic coordinates (latitude and longitudes in decimal degrees). Once imported into Arc/Info, they were projected into USGS-Albers. Procedure 3-1 annotates the process used to import these grids from downloaded form to a merged and projected grid. The final product was a digital representation of the elevation over the entire potential study area, projected to 100 m x 100 m cells. With a process developed in Arc/Info (Section 4.3), this surface terrain model was used to delineate the watershed area draining into the Upper Houston Ship Channel (Figure 3-2).

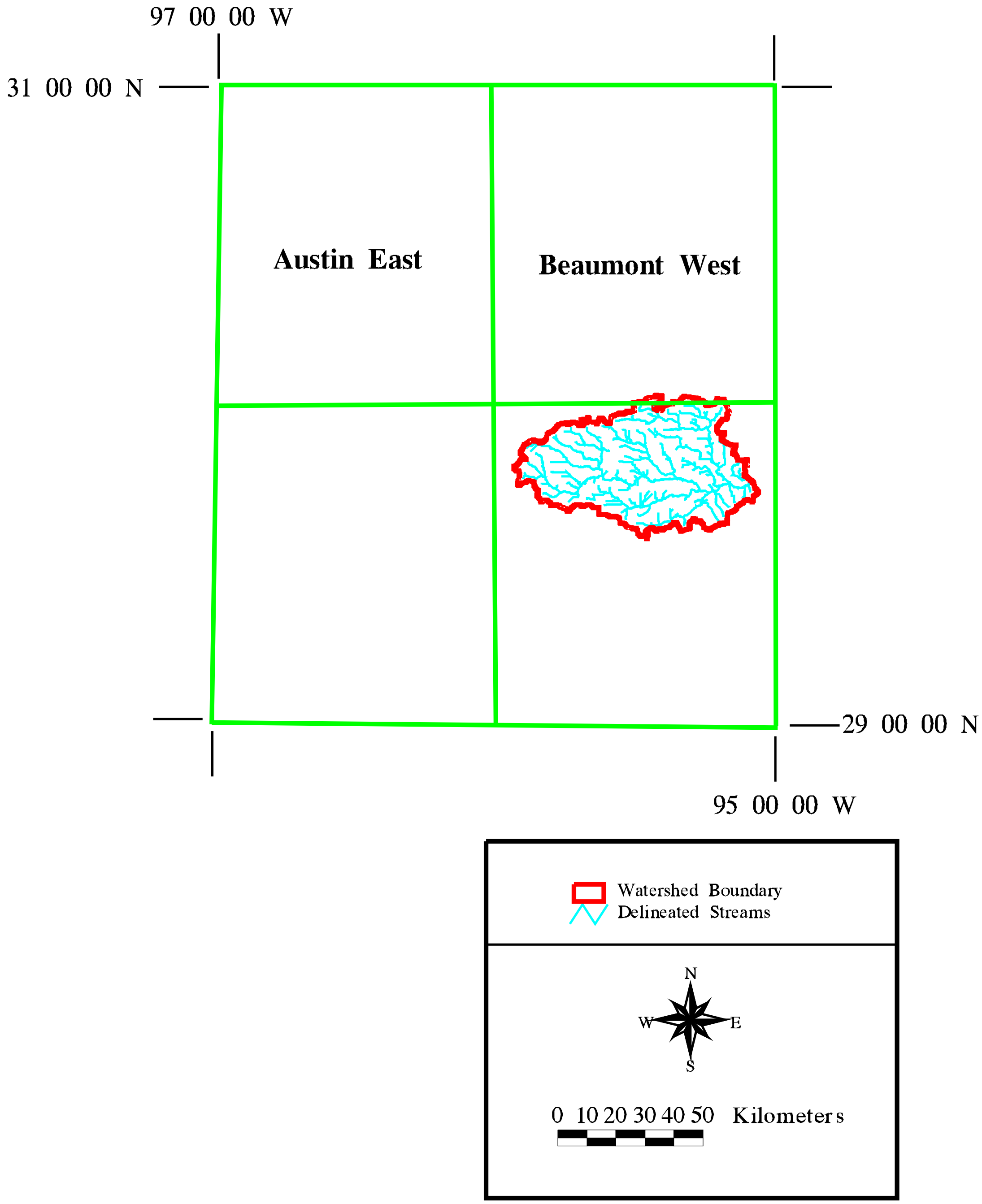

Figure 3-2 Digital Elevation Model (DEM) showing locations of USGS quads. Delineated subwatershed and streams are also shown (USGS, 1996).

$**uncompress filename**
uncompresses the file downloaded from USGS
$**dd if = <input> of = <output> cbs = 1024 conv = unblock**
typed at c-shell prompt to put the downloaded DEM obtained from the USGS website into arc format
$**arc**
invokes Arc/Info
arc: **demlattice <input file> <output grid> usgs**
converts the input file to a readable grid in geographic coordinates
arc: **grid**
invokes the Arc/Info subprogram, Grid
grid: **<totalgrid> = merge (<grid1>, <grid2>, ...)**
merges all of the imported grids into one grid
grid: **<projected grid> = project (<totalgrid>, geoalb.prj, #, 100)**
projects the geographic coordinate grid into a planer map projection using a cell size of 100m; the parameters of which are located in the projection file

Procedure 3-1 Commands used to import USGS Digital Elevation Models into Arc/Info.

3.2.4 Land Use/Land Cover Files

Since land use is used in this research to determine non-point source loading to the Upper HSC, it was important to obtain an accurate and recent digital representation of the land use in the Houston area. Land use/land cover for an area is usually determined via interpretation of aerial photographs or satellite imagery pictures. The interpretation typically puts the land use into one of eight categories: urban or built-up land, rangeland, agricultural land, barren land, water, forest, wetlands, or tundra. This division system, called Anderson Classification, has two levels of organization (Anderson, *et al.*, 1976). Level I consists of the categories described above. Each of these general categories is further subdivided into more descriptive categories for Level II. For example, the “urban” category in Level I is separated into residential, commercial services, industrial, transportation /communication, industrial/commercial, mixed urban, and other urban for Level II (Table 3-4).

In 1992, Newell, *et al.* (1992) developed GIS land use coverages from 1990 LANDSAT satellite imagery data. These coverages, which incorporated the entire Galveston Bay watershed area, were obtained from Rice University in exported format via ftp (LaWare, 1996). Each USGS subwatershed, as defined in the 1992 report, constituted a separate land use coverage. Table 3-5 and Figure 3-3 show the USGS subwatersheds

Table 3-4 Anderson Land Classification System and Corresponding Newell, *et al.* (1992) Classifications

Anderson Level I Class	**Anderson Level II Class**	**Newell, *et al.* (1992) Class**
1 Urban or Built-Up Land	11 Residential	6 Residential
	12 Commercial Services	1 High Intensity Urban
	13 Industrial	
	14 Transportation/Communication	
	15 Industrial and Commerical	
	16 Mixed Urban or Built-Up land	
	17 Other Urban or Built-Up Land	
2 Agricultural	21 Cropland and Pasture	3 Agriculture
	22 Orchards, Groves, Vineyards	2 Open/Pasture
	23 Confined feeding Operations	
3 Rangeland	31 Herbaceous Rangeland	N/A
	32 Shrub and Brush Rangeland	
	33 Mixed Rangeland	
4 Forest Land	41 Deciduous Forest Land	8 Forest
	42 Evergreen Forest land	
	43 Mixed Forest Land	
5 Water	51 Streams and Canals	7 Water
	52 Lakes	
	53 Reservoirs	
	54 Bays and Estuaries	
6 Wetland	61 Forested Wetlands	5 Wetlands
	62 Nonforested Wetlands	
7 Barren Land	71 Dry Salt Flats	4 Barren
	72 Beaches	
	73 Sandy Areas	
	74 Bare Rock	
	75 Strip Mines, Quarries, Gravel Pits	
	76 Transitional Area	
	77 Mixed Barren Land	
8 Tundra	81 Shrub and Brush Tundra	N/A
	82 Herbaceous Tundra	
	83 Bare ground	
	84 Wet Tundra	
	85 Mixed Tundra	
	91 Perennial Snowfields	
	92 Glaciers	

Source: Newell, *et al.*, 1992 and Anderson, *et al.* 1976.

Table 3-5 USGS Subwatersheds which Correspond to the Upper Houston Ship Channel Study Area

USGS Subwatershed	Newell, *et al.* (1992) Abbreviation	Area (km^2)
Addicks Reservoir	ad	347
Barker Reservoir	bk	316
Brays Bayou	br	329
Buffalo Bayou	bf	272
Greens Bayou	gr	539
Sims Bayou	sm	241
Ship Channel	sc	430
Whiteoak Bayou	wo	285
Total:		2759

Source: Newell, et al., 1992

from the Newell, *et al.* (1992) report which lie in the watershed used in this project. There are some discrepancies in the digitized USGS watershed obtained from Newell, *et al.* (1992) and the delineated boundary in GIS. These differences are discussed further in Chapter 5. To acquire a total coverage for the entire Upper HSC watershed, all of the subwatershed land use coverages were joined into one coverage with the *append* command so that the attributes of each coverage were preserved. Figure 3-4 shows the watershed boundary and the 1990 land use coverage that resulted from the joining of the 8 subwatershed land use coverages. It should be noted that the Newell *et al.* (1992) study somewhat reclassified the traditional Level I Anderson System in the following ways:

- The Anderson Level I "urban or built-up land" classification was divided into "residential" (single family homes) and "high density urban" (all other Level I urban) subclasses.
- The Anderson Level I "agricultural land" was divided into an "agricultural" subclass (i.e., cultivated land) and an "open/pasture" subclass.

The final classification system used by the 1992 study is shown in Table 3-4. Newell *et al.* (1992) considered the Level I classification sufficient, since no further subdivision enhances the non-point source loading calculations (Newell, *et al.*, 1992).

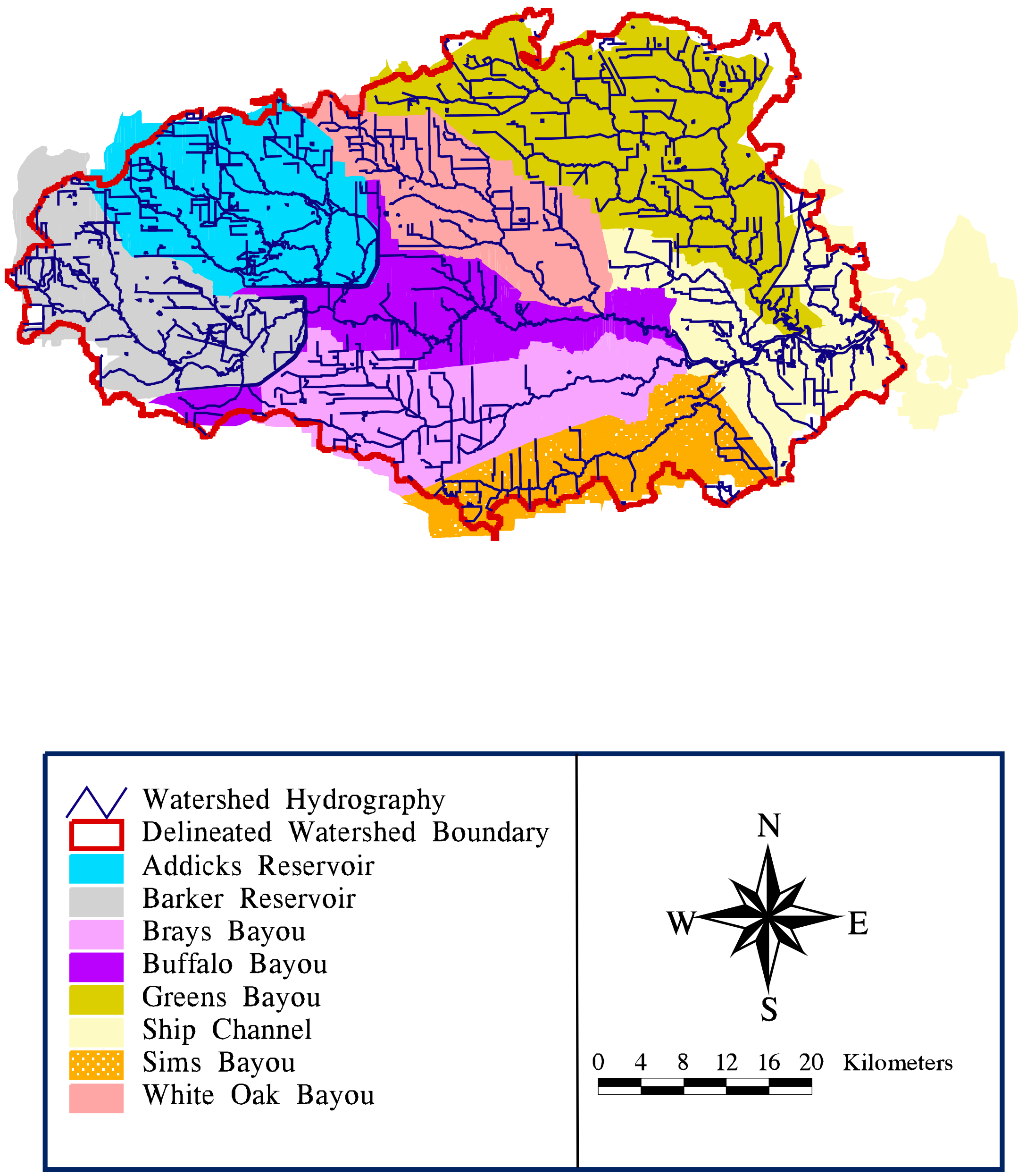

Figure 3-3 USGS subwatershed area and delineated watershed for the Upper Houston Ship Channel (Newell, et al., 1992).

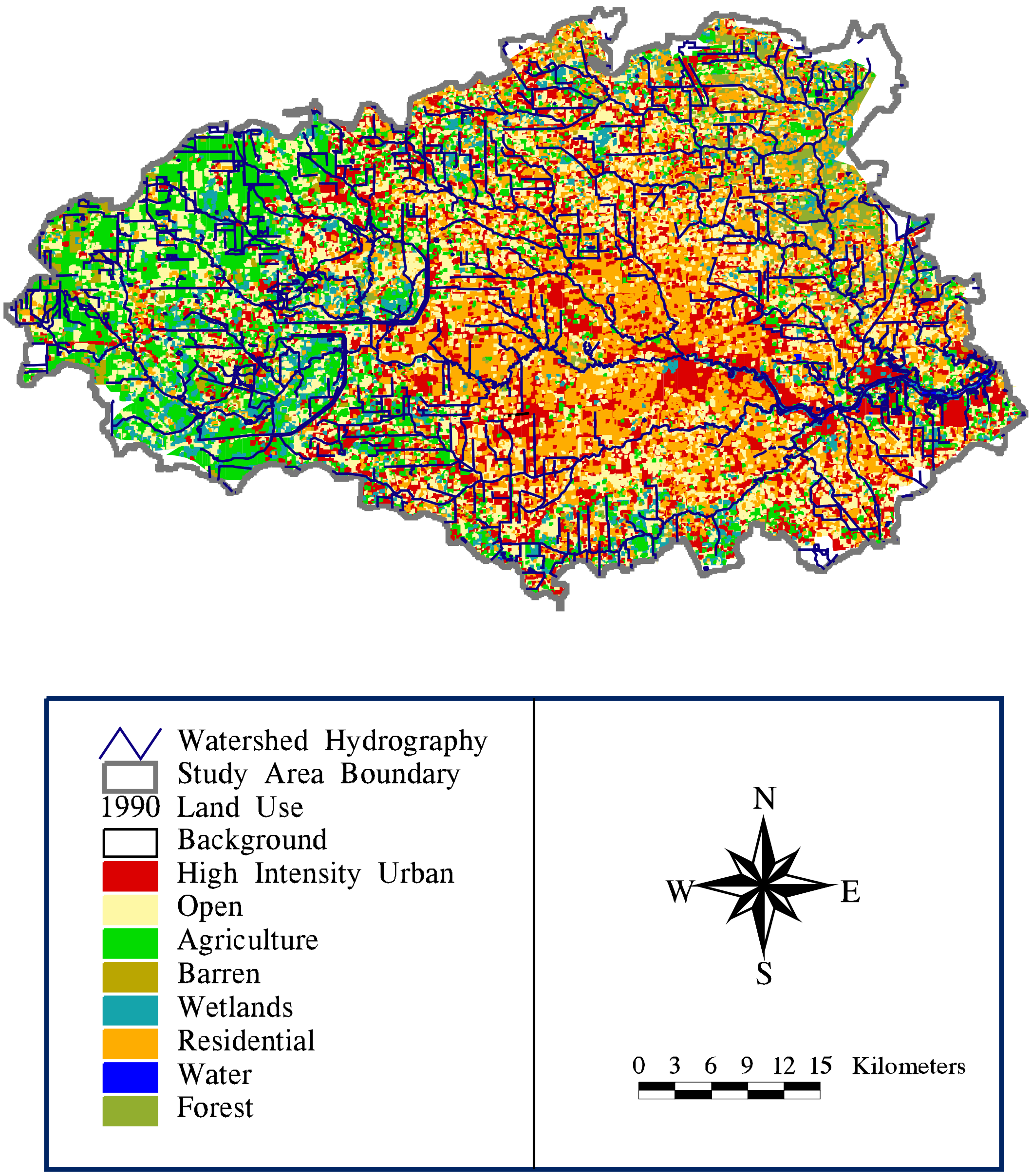

Figure 3-4 1990 land use coverage with delineated watershed boundary. Missing land use data is shown around edges of watershed boundary.

As shown in Figures 3-3 and 3-4, some outlying areas of the surface terrain (DEM) delineated watershed did not include the USGS subwatersheds. These gaps are probably due to minor errors that may have occurred in the digital delineation of the watershed or the digitization of the USGS subwatersheds by Newell, *et al.* (1992). These differences are further discussed in Section 5.2 Watershed Delineation. As a result, it was necessary to fill the gaps in the 1990 land use data, using 1980 data available from USGS. The 1:250,000 scale land use maps, organized by 1° x 2° quad name, were downloaded from the Environmental Protection Agency (USEPA) ftp site (USEPA, 1996). Appendix C-2 provides a detailed description of the process used to determine the names of the needed files, download the files from the USEPA ftp site, and import them into Arc/Info. The quad maps were joined into one large map and the delineated watershed area was clipped from the map. The 1980 coverage was then "reclassified" so that the land use categories for the older attributes matched those categories used in the Newell, *et al.* (1992) report. The gaps missing from the 1990 data were determined by using the *erase* command in Arc/Info. The *erase* command used the 1990 land use coverage as a "cookie cutter" and deleted the information of the 1980 coverage that occupied the same area as the 1990 coverage (Figure 3-5). The 1990 land use coverage was then joined to the 1980 coverage which contained just the small missing areas of the delineated watershed, to obtain a total land use/land cover map for the entire watershed area (Figure 3-6). Procedure 3-2 shows the process used in Arc/Info to create this coverage.

arc: **import cover <filename.e00> <coveragename>**
imports the 1980 land use coverages which are in exported format NOTE: If downloaded from USEPA these coverages are already projected in USGS-Albers
arc: **mapjoin lu_80**
will create a coverage, lu_80, which consists of one map, encompassing all coverages entered by the user, when prompted by Arc/Info
arc: **dissolve lu_80 lu_80dis grid-code**
arc: **dissolve lu_90 lu_90dis grid-code**
dissolves the coverage, lu_80 (or lu_90), by getting rid of any arcs that may exist between the adjacent polygons with the same value in the attribute, grid-code (the land use code); the final coverage is called lu_80dis (or lu_90dis)
arc: **clip lu_80dis covtotshd lu_80shd poly**
arc: **clip lu_90dis covtotshd lu_90shd poly**

Procedure 3-2 Process used in Arc/Info to obtain a full coverage of the land use/land cover**.**

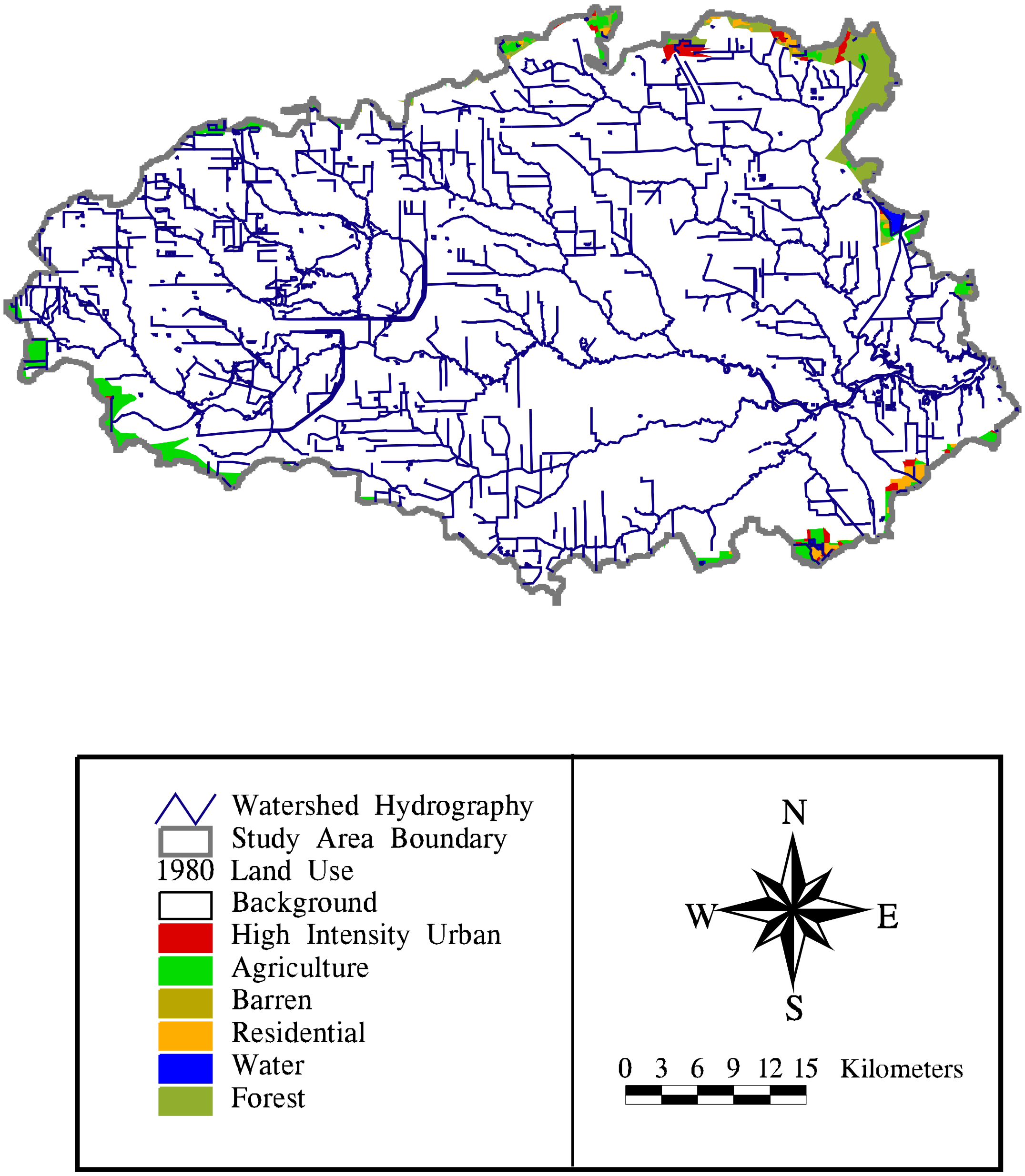

Figure 3-5 1980 land use coverage that was used to fill in the missing areas not covered by the 1990 land use data. Not that the "open" and "wetland" land uses did not exist in these areas (USGS, 1996).

clips the land use coverages with the watershed boundary so that only the land use within the delineated watershed are in the final coverage lu_80shd (or lu_90shd)
arc: **erase lu_80shd lu_90shd lu_80edge poly**
uses the 1990 land use coverage as a cookie cutter to delete the information in the 1980 land use coverage that occupies the same area; the result is a land use coverage of just the missing areas on the edge of delineated watershed from the 1990 land use coverage
arc: **mapjoin hsclu**
command used to join two or more polygon coverages into one, while maintaining the attributes of each coverage in the attribute table; in this case, the coverages lu_80edge and lu_90shd were joined to result in hsclu

Procedure 3-2 (cont.) Process used in Arc/Info to obtain a full coverage of the land use/land cover over the entire watershed area.

3.2.5 Precipitation Data

The Department of Agriculture, Natural Resources Conservation Service (formerly the Soil Conservation Service) has begun an effort to develop precipitation data for the entire United States. This study has produced a 2.5′ x 2.5′ cell grid of precipitation data over the US. There are 13 grids, consisting of precipitation values for each month (January through December) and one annual grid, averaged over the years of 1961 to 1990 (Daly, *et al.*, 1994). These grid were downloaded from the Oregon State University anonymous ftp site (see Table 3-2). From these grids, the Houston Ship Channel watershed was "clipped" from the national grids using a command called *set window* in Arc/Info's subprogram, Grid. This command allows one to chose a smaller area from a larger grid by manually drawing a box around the area of concern. These 13 smaller grids were projected into Albers and their resolution increased to 100 m x 100 m cells, in order to match the DEM cell size. Figure 3-7 illustrates an example of one of these grids. Sections 4.3 and 4.5.5 discuss the use of these grids in the rainfall/runoff and rainfall/flow relationships development.

3.2.6 USGS Gauge Stations and Flow Data

To obtain a representative record of the flow for this watershed area, data from the USGS was utilized. The USGS office in Texas (Texas USGS) maintains flow gauging

stations across the United States, some of which have continuous flow data in daily averages, as early as 1924 (Texas USGS, 1996). In the Upper HSC watershed area, there were a total of 37 possible gauging stations. Of those 37, ten were chosen for their continuous periods of record and locations (Figure 3-8). Table 3-6 shows the characteristics of each of those stations.

Table 3-6 USGS Gauge Stations used in Methodology

USGS Gauge Number	Name	USGS Drainage Area (km^2)	Location (DMS) Lat	Location (DMS) Long	Period of Record
8072730	Bear Creek near Barker Reservoir	56	29° 49′ 52″ N	95° 41′ 13″ W	07/01/77 - pr.
8073600	Buffalo Bayou at West Belt Drive	795	29° 45′ 43″ N	95° 33′ 29″ W	09/01/71 - pr.
8074500	White Oak Bayou at Houston	224	29° 46′ 30″ N	95° 23′ 49″ W	06/01/36 - pr.
8075000	Brays Bayou at Houston	246	29° 41′ 49″ N	95° 24′ 43″ W	06/01/36 - pr.
8075500	Sims Bayou at Houston	163	29° 40′ 26″ N	95° 17′ 20″ W	10/01/52 - pr.
8075730	Vince Bayou at Pasadena	21	29° 41′ 38″ N	95° 12′ 58″ W	10/01/71 - pr.
8075770	Hunting Bayou at IH 610	42	29° 47′ 35″ N	95° 16′ 05″ W	05/01/64 - pr.
8076000	Greens Bayou near Houston	178	29° 55′ 05″ N	95° 18′ 25″ W	10/01/52 - pr.
8076500	Halls Bayou at Houston	74	29° 51′ 43″ N	95° 20′ 06″ W	10/01/52 - pr.
8075900 *	Greens Bayou near Hwy 75	95	29° 34′ 30″ N	95° 25′ 05″ W	08/03/65 - pr.

* Used for rainfall/runoff relationship verification -- not used in actual relationship development
Source: Texas USGS, 1996

For each of these ten gauging stations, the flow data for 1961 to 1990, in mean daily averages, was downloaded from the website (Texas USGS, 1996). If the entire 30 years of data was unavailable, the period of record on file was obtained. This period of record was chosen to match the precipitation data's period of record (1961-1990), so an accurate rainfall to runoff relationship could be developed. In addition, a point coverage of the ten stations was developed from the location information given with each gauge (Figure 3-8). The watershed delineation process used for these stations is discussed in Section 4.3.

3.2.7 Tables and Coverages

Throughout the project, additional tables and coverages were needed to complete the representation of the study area in GIS. However, since the creation of these coverages and

tables varied in relation to their use, their development is best discussed in detail within the methodology chapter (Chapter 4). A brief summary of the different data is given below:

- *Point Source Dischargers* -- The primary industries discharging into each water quality segment were determined from locations provided by the Texas Natural Resource Conservation Commission (TNRCC) (Visnovsky., 1996). The result was a point coverage of this information linked to another table, containing the actual discharge values for each constituent of interest.
- *Channel Segment Information* -- The segmentation utilized in a 1971 Tracor report (Espey, *et al.*, 1971) was duplicated in GIS. A table of the parameters necessary for the modeling effort was then "attached" to the segmentation coverage.
- *Water Quality Information* -- The current water quality in the channel was spatially represented in GIS by creating a polygon coverage of the water quality results for each hydrographic segment described in Ward and Armstrong (1992).