Chapter 3: Data Sources and Description

The conclusions that this study presents are based on statistics calculated from 46,507 nitrate measurements taken from 29,485 wells throughout Texas. Following the methods outlined in Section 2.6, and described in detail in Chapter4, the spatial variation of the statistics is mapped to identify regions of high or low vulnerability to nitrate contamination. The spatial variation in the statistics is then compared to the spatial variation of potential water quality indicators, including soil parameters, average annual precipitation, and fertilizer sales, in order to assess the value of these data as indicators of water quality.

Because the structure and limitations of these data strongly influence the choice of the methods used, this chapter, which describes the data itself, is a necessary prelude to Chapters 4 and 5, which describe the methodology and procedures followed in the study. This chapter contains seven sections, one for each data set used in the study. These data sets can be divided into three groups:

1) Primary data, consisting of groundwater nitrate concentration measurements and descriptions of the wells where the groundwater was collected for testing. The nitrate data are described in Section 3.1 and the well data are described in Section 3.2.
2) Data to be considered as potential indicators of water quality. These include soil thickness and organic content described in Section 3.3; annual average precipitation, described in Section 3.4; and average annual nitrogen fertilizer sales, described in Section 3.5.

3) Independent measurements of nitrate and herbicides, used to test assumptions made in the study. These include measurements of nitrate in public water sources collected by the Water Utilities Division of the Texas Natural Resource Conservation Commission, described in Section 3.6, and the first year's results of the U.S. Geological Survey's reconnaissance of nitrate and herbicides in groundwater in the Midwest, described in Section 3.7.

3.1 NITRATE MEASUREMENT DATA

The nitrate measurements used in this study come from the Texas Water Development Board's (TWDB) Groundwater Data System (Nordstrom and Quincy, 1992). This statewide database contains physical descriptions of wells and their surroundings in Texas, and levels of chemical constituents measured by a variety of public agencies. The TWDB maintains the database to characterize the quantity and quality of groundwater available throughout the state, in support of the preparation of the Texas Water Plan (TWDB, 1994).

For every nitrate measurement listed in the Groundwater Data System as of October 1993—a total of 62,692 database records—the data fields listed in Table 3.1 were retrieved for use in this study. Of these data fields, the well ID, date, and nitrate level have values in all records. Many records have no values for the collecting agency or reliability remarks. The values in the flag field are discussed in section 3.1.1.

Table 3.1 Nitrate Measurement Data

Name of Data Field	Description
Well ID	Identification number of well where collected (see section 3.2.1)
Date	Date collected
Agency	Collecting agency (e.g. USGS, TWDB, etc.)
Reliability Remarks	Numeric code indicating handling and analysis reliability
Nitrate Level	Concentration (mg/l NO_3) of nitrate.
Nitrate Flag	Code ("<" or ">") indicating level is reporting limit rather than measured concentration

Table 3.2 Nitrate Measurements from Well 5740304

Year	Month	Reported Nitrate (mg/l NO_3)	Adjusted Nitrate (mg/l N)
1966	4	< 0.4	^{2} 0.10
1966	12	< 0.4	^{2} 0.10
1967	6	14.0	3.17
1968	6	12.0	2.71
1968	7	13.5	3.05
1971	6	8.0	1.81
1972	5	8.0	1.81
1974	3	5.9	1.33
1976	8	4.7	1.06
1980	3	3.9	0.88
1986	6	2.13	0.48
1991	8	0.44	^{2} 0.10

Nitrate concentrations in the TWDB database are listed as mg/l nitrate (nitrate-NO_3). However, unless otherwise noted the values used in this study's statistical analyses and reported here are in equivalent values of nitrate as nitrogen (nitrate-N), the units used in EPA regulations. 1 mg/l nitrate-N equals 4.42 mg/l nitrate-NO_3. Each nitrate-NO_3 value in the data set was converted to an equivalent nitrate-N value. To maintain a uniform reporting limit for all records used in the study, all values at or below a value of 0.1 mg/l nitrate-N will be treated as $\leq$ 0.1 mg/l. A nitrate concentration greater than 0.1 mg/l will be considered a "detection" and concentrations less than or equal to this value will be considered to be "below detection limit." As an illustration of this conversion and adjustment, Table 3.2 shows the nitrate measurements listed in the TWDB database for well 5740304 and the adjusted values used for analysis in this study. Of the twelve measurements shown, nine are considered detections of nitrate and three fall below the detection limit.

3.1.1 Nitrate Reporting Limits

The flag field in a nitrate measurement record may be blank or may contain a "<" or ">" character. A blank should indicate that the value listed for nitrate concentration in the nitrate level field is the actual value measured in the water; a "<" or ">" indicates that the value is a detection or reporting limit, rather than an actual value. The ">" character appeared 5 times in the retrieved data. The "<" character appears in 4047 (6.5%) of the records. A value of 0.40 mg/l nitrate-NO_3 (approximately equal to 0.1 mg/l nitrate-N) appears most frequently as a reporting limit, as the histogram in Figure 3.1 illustrates. (Not shown in the figure are 403 records with detection limits greater than 1 mg/l NO_3.)

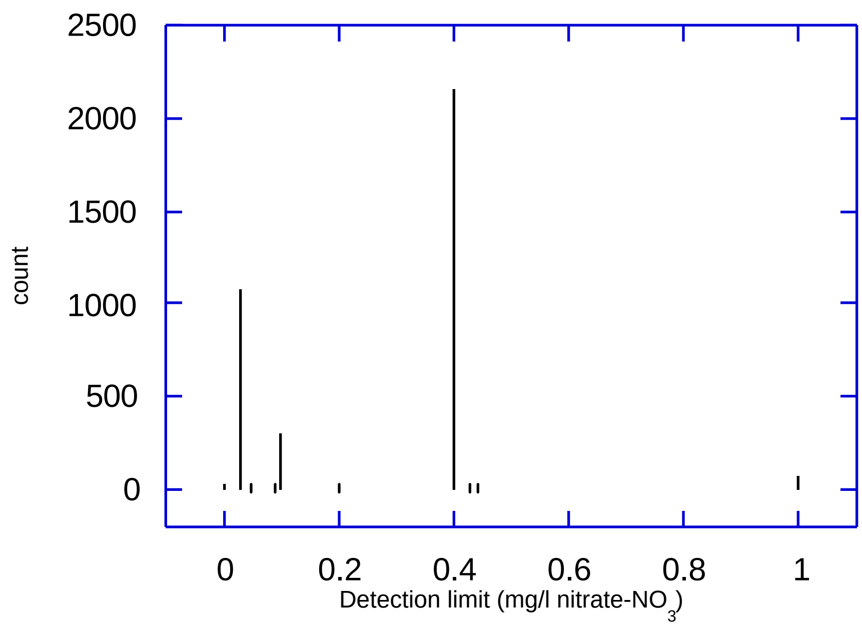

Figure 3.1 Reported Detection Limits for Nitrate

Although a blank in the flag field should indicate that the nitrate level in the record is a true measured concentration, the number of occurrences of some values suggests otherwise. Figure 3.2 shows a histogram of nitrate levels below 1 mg/l nitrate-NO_3 in records with blank flag fields. The value 0.4 appears 9,793 times in the 58,640 records with blank flag fields. It seems very unlikely that 17% of the water measurements reported in this database should have exactly this value. Since 0.4 is also the most common reporting limit value, a much more plausible explanation of this high incidence would be that the nitrate concentration in many of these cases was below 0.4 mg/l nitrate-NO_3, and that the "<" flag was omitted from the record. Because of the ambiguous meaning of

"0.4 mg/l nitrate-NO_3," this study will treat all occurrences of this value as meaning "less than or equal to 0.4 mg/l nitrate-NO_3."

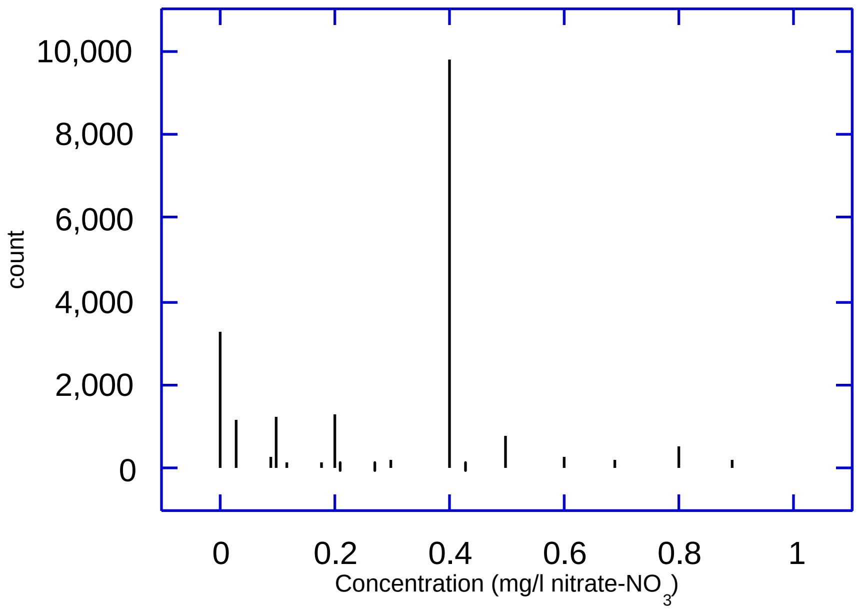

Figure 3.2 Reported Nitrate Concentrations

3.1.2 Sampling Period

The records retrieved from the TWDB database indicated sampling dates from 1896 to 1993. The histogram in Figure 3.3 shows the number of measurements taken in each year. As will be shown in the discussion of the results of this study in Chapter 6, there has been a slight increase over time in the amount of nitrate found in Texas groundwater. In order to reduce the effects of this increase on the data, the study was confined to measurements taken during the years 1962 to 1993. This period was chosen in part because of the sharp increase in the number of nitrate samples collected per year from 1962 onward.

Omitting nitrate measurements prior to this date retained a substantial majority of the database in the study while removing the measurements least likely to be representative of the present condition of Texas groundwater.

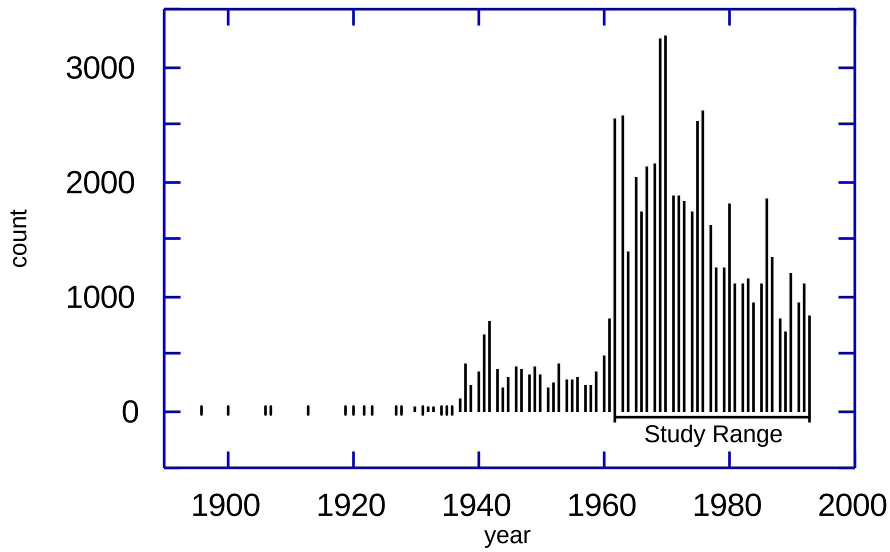

Figure 3.3 Nitrate Measurements Reported by Year

3.1.3 Measurement Record Accuracy

Because the nitrate measurements recorded in the TWDB database come from a variety of sources, they do not conform to a uniform set of quality control standards. In fact, there is evidence in the data to suggest that many values may be questionable. As the preceding section describes, it appears that a "<" flag was omitted from many records in the database. In addition, 140 records indicate nitrate concentrations over 500 mg/l NO_3, a suspiciously high level. (Concentrations of 500 mg/l have been found in waters in the unsaturated zone below irrigated crops, and levels over 1000 mg/l have been found in pools in the parts of Carlsbad Caverns where bats roost (Hem, 1989). It seems unlikely that concentrations this high are representative of natural groundwater.) 51,329 of the 62,692 nitrate records retrieved from the TWDB database had blank reliability remark fields; while this provides no grounds for excluding the records, it is not a ringing endorsement either.

In spite of these reservations, this study has taken an "innocent until proven guilty" approach to the measurement records. The data were included in the study "as is" unless substantial evidence indicated that they should be excluded. As shown in Table 3.3, records were excluded if reliability remarks indicated questionable collection or handling, if no record could be found of the well from which the water was collected, if the well had bad location data (see following section), if the reported value was "less than" a threshold greater than

Table 3.3 Excluded Measurement Records

Reason	Criteria	# Records Excluded
Reliability	Remarks = 01, 02, or 03	7,020
Well Data	No well record	11
Well Location	Well mis-located	418
Lower threshold	flag = "<" and nitrate > 0.45 mg/l NO_3	407
Upper threshold	flag = ">"	5
Collection Date	Year < 1962	9,087
Total Excluded		16,185

0.1 mg/l nitrate-N (0.45 mg/l NO_3), if the reported value was "greater than" any threshold, or if the measurement was taken before 1962 (see preceding section). These exclusions left 46,507 nitrate measurement records in the study. This set of nitrate measurement records will be called the "base data set" in the remainder of this document.

3.2 WELL DATA

The data providing physical descriptions of the wells included in the study comes from the same TWDB database as the nitrate measurement data. For each well for which a nitrate measurement was recorded—a total of 38,740 database records—the data fields listed in Table 3.4 were retrieved.

3.2.1 TWDB Well Numbers

TWDB has adopted a system of identification numbers for wells in Texas, based on the location of the wells expressed in latitude and longitude. The following description and Figure 3.4 explain the numbering system.

> [The numbering system] is based on division of the state into a grid of 1-degree quadrangles formed by degrees of latitude and longitude and the repeated division of these quadrangles into smaller ones as shown…
>
> Each 1-degree quadrangle is divided into sixty-four 7-1/2-minute quadrangles, each of which is further divided into nine 2-1/2-minute quadrangles. Each 1-degree quadrangle in the state has been assigned an identification number. The 7-1/2-minute quadrangles are numbered consecutively from left to right, beginning in the upper-left-hand corner of the 1-degree quadrangle, and the 2-1/2-minute quadrangles within each 7-1/2-minute quadrangle are similarly numbered. The first 2 digits of a well number identify the 1-degree quadrangle; the third and fourth digits, the 7-1/2-minute quadrangle; the fifth digit identifies the 2-1/2-minute quadrangle; and the last two digits identify the well

within the 2-1/2-minute quadrangle. (Nordstrom and Quincy, 1992)

Table 3.4 Well Description Data

Name of Data Field	Description
Well ID	Identification number of well (see section 3.2.1)
Aquifer Code	Alphanumeric code for aquifer or geologic unit associated with well
County	Numeric code for county where well is located (FIPS code)
Latitude	Latitude of wellhead location (DMS)
Longitude	Longitude of wellhead location (DMS)
Location Method	Numeric code indicating accuracy of latitude and longitude
Depth	Depth of completed well from land surface (feet)
Depth Method	Alphabetic code indicating source of depth measurement
Altitude	Elevation of land surface at wellhead (feet above mean sea level)
Altitude Method	Alphabetic code indicating source of altitude measurement
Primary Use	Alphabetic code indicating primary purpose served by well

The TWDB well-numbering system will be used throughout this report not only for wells and well locations, but also for numbering 1_, 7.5', and 2.5' quadrangles used to divide the state for analysis. Well number 5740304 is located in 1_ quad 57, 7.5' quad 5740, and 2.5' quad 57403.

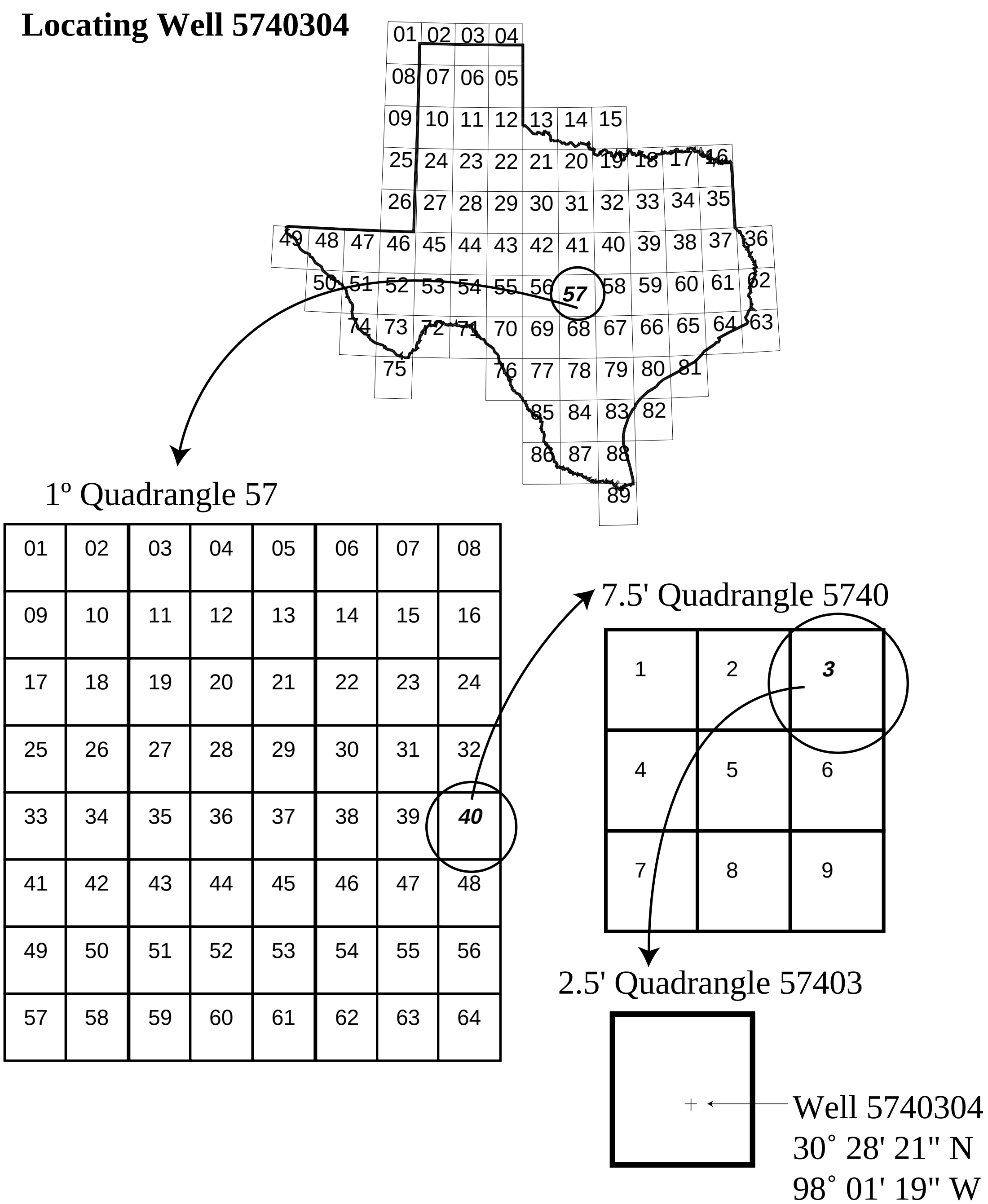

Figure 3.4 TWDB Well-Numbering System

3.2.2 Location Accuracy

The latitude and longitude of a well listed in the database do not perfectly represent the true location of that well. Different location methods have different degrees of precision and accuracy. The TWDB Ground-Water Data System assigns a numerical code to each well location, indicating the reliability of the given coordinates. The meanings of these codes are summarized in Table 3.5, which also lists the number of wells and associated measurements falling into each accuracy group.

Table 3.5 Location Accuracy Codes

Code	Accuracy	# wells	# measurements
1	± 1"	12,180	22,049
2	± 5"	2,832	4,801
3	± 10"	3,814	4,936
4	± 1'	12	17
5	*	5,628	7,412
none	unknown	4,779	7,260

*—latitude and longitude are given for center of 2.5' quadrangle

A location method code of 5 indicates that the given latitude and longitude are for the center of the 2.5' quadrangle, rather than the well itself. The TWDB states that this is a temporary measure, necessary to include wells listed in an older database that did not require latitude and longitude for well records. Nearly 20% of the wells included in the study (and 16% of the nitrate measurements) can be located only by 2.5' quadrangle.

3.2.3 Selected Aquifers

Wells and nitrate measurements were grouped for statistical and spatial analysis primarily by their location in the 7.5' quadrangles numbered according to the system described in Section 3.2.1. A subset of the wells and measurements selected for further examination were grouped by association with five aquifers, the Carrizo-Wilcox, the Balcones Fault Zone of the Edwards, the Hueco-Mesilla Bolson, the Ogallala, and the Seymour. The TWDB designates these as Major Aquifers, meaning that they supply "large quantities of water in large areas of the State" (Ashworth and Flores, 1991).

The field "Aquifer Code" in the Texas Groundwater Data System "is adopted from U.S. Geological Survey's WATSTORE Data File. The code consists of three digits designating the geologic Era, System, and Series followed by a four or five [character alphabetic] code designating the aquifer(s) or stratigraphic unit(s)" (Nordstrom and Quincy, 1992).

For example, the code "124WLCX" refers to the Wilcox Group, which belongs to the Cenozoic Era, the Tertiary System, and the Paleocene Series. The code has been modified to describe wells in ambiguous settings, or which draw water from more than one formation or aquifer (Nordstrom, 1994). For example, the code "110AVQW" refers to a combination of alluvium, Queen City Sands, and the Wilcox Group.

Based on the aquifer delineation criteria described by Ashworth and Flores (1991), and geologic descriptions from the Geologic Atlas of Texas (BEG, various years), wells were assigned to aquifer groups according to the TWDB aquifer codes listed in Table 3.6. Note that a well was assigned to an aquifer group only if the TWDB code associated it with a single formation or aquifer. A well with the code "110AVQW" was not assigned to the Carrizo-Wilcox, because it is associated with alluvium and the Queen City Sands as well as the Wilcox Group. The number of wells and measurements associated with these aquifers are summarized in Table 3.7.

Table 3.6 Aquifer Codes

Aquifer	TWDB Codes
Carrizo-Wilcox	124CRRZ 124WLCX 124CZWX 124CZWXA
Edwards (Balcones Fault Zone)	218EBFZA
Hueco-Mesilla Bolson	112HCBL 112MSBL
Ogallala	121OGLL
Seymour	112SYMR

Table 3.7 Wells and Measurements in Selected Aquifers

Aquifer	Wells	Measurements
Carrizo-Wilcox	2292	4597
Edwards (BFZ)	412	1691
Hueco-Mesilla Bolson	404	1908
Ogallala	3483	4430
Seymour	1993	2526

The five aquifers are shown in Figure 3.5. The map was created by combining the outlines of the aquifers from five GIS coverages prepared by TWDB, and represents that agency's estimate of the extent of the aquifers on surface and the limits of the unexposed (downdip) regions that provide usable water. Brief descriptions of the aquifers follow.

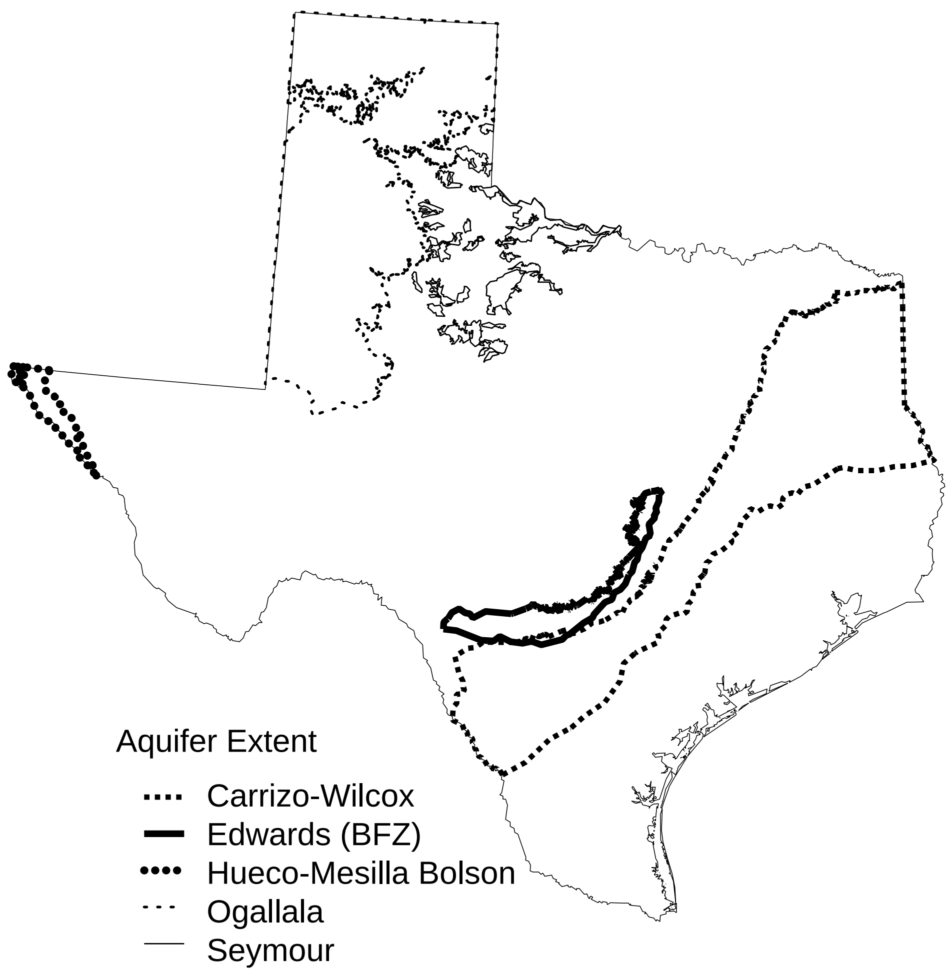

Figure 3.5 Boundaries of Study Aquifers as Identified by TWDB

Carrizo-Wilcox Aquifer. "The Carrizo-Wilcox aquifer includes the Carrizo Formation and the entire Wilcox Group. It extends across the State from Mexico to Louisiana" (Ashworth and Flores, 1991). The Carrizo Formation consists primarily of quartz sand, feldspar, and sandstone (BEG, 1974a and 1968). The Wilcox Group consists primarily of quartz sand, mudstone, clay, and silt (BEG, 1974 and 1968). The TWDB aquifer codes selected for this aquifer group are "124CRRZ" for Carrizo Sand, "124WLCX" for Wilcox Group, "124CZWX" for Carrizo Sand and Wilcox Group—Undifferentiated, and "124CZWXA" for Carrizo Wilcox Aquifer. (Norstrom and Quincy, 1992).

Edwards Aquifer (Balcones Fault Zone). "The Edwards (BFZ) aquifer consists of all the units formations and other members below the Del Rio Formation and above either the Glen Rose Limestone or, when it is present, the Walnut Formation." (Ashworth and Flores, 1991). The Balcones Fault Zone of the Edwards Aquifer is made up of a variety of limestone formations with some included dolomite and shale (BEG, 1974a and 1974b). The TWDB aquifer code selected for this aquifer group is "218EBFZA" for Edwards and Associated Limestones—Balcones Fault Zone.

Hueco-Mesilla Bolson Aquifer. "The Hueco-Mesilla Bolson aquifer consists of Cenozoic alluvial and bolson deposits that occur within the valleys that flank the Franklin Mountains; and extend north and west into New Mexico, and south into Mexico… Although hydrologically connected, the aquifer does not include the overlying Rio Grande alluvium." (Ashworth and Flores, 1991). The Hueco and Mesilla deposits include alluvium and "fluviatile deposits of clay, silt, sand and gypsum in bolsons" (BEG, 1993). The TWDB aquifer codes selected for this aquifer group are "112HCBL" for Hueco Bolson Deposits and "112MSBL" for Mesilla Bolson Aquifer.

Ogallala Aquifer. "The Ogallala aquifer consists primarily of the Ogallala Formation and extends north, west, and east into adjacent states. The boundary of the formation is mapped along the eastern High Plains escarpment and along the Canadian River Valley, where the formation outcrop is in contact with underlying formations of Cretaceous, Triassic, or Permian age. The southern extent is placed at the estimated formation pinchout" (Ashworth and Flores, 1991). The Ogallala Formation consists of "fluviatile sand, silt, clay, and gravel capped by caliche" (BEG, 1967). The TWDB aquifer code selected for this aquifer group is "121OGLL" for Ogallala Formation.

Seymour Aquifer. "The Seymour aquifer occurs in isolated, eroded alluvial remnants in north-central Texas. The areas delineated are based on surface extent, well development and usage. Consequently many smaller remnants that provide little water or are not developed, are not mapped" (Ashworth and Flores, 1991). The Seymour Formation consists of "Thick deposits… mostly sand, silty orange-brown to red, thick-bedded, massive, locally with large-scale cross-beds and gravel" (BEG, 1987). The TWDB aquifer code selected for this aquifer group is "112SYMR" for Seymour Formation.

3.2.4 Well Description Accuracy

In addition to the location of the well, the accuracy of a well's depth and aquifer code are of particular interest to this study. The histogram of well depths less than 200 feet shown in Figure 3.6 illustrates the overabundance of reported well depths equal to zero or integer multiples of 10 feet. Well depths are often reported by drillers or well owners, who may not always respond to data requests with scientific precision. Although the TWDB Ground-Water Data System Data Dictionary does not say so, the large number of zero depths suggests that zero may mean "no data" in many cases. The assignment of aquifer codes usually comes from a geologist's interpretation of driller's logs, or from data provided by an agency other than the TWDB, such as the U. S. Geological Survey or various state water districts, that provide well data to the TWDB. This process is not under a uniform quality-control program, and is certainly subject to some errors. However the number of erroneous classifications should be expected to be small in comparison to the database as a whole (Nordstrom, 1994).

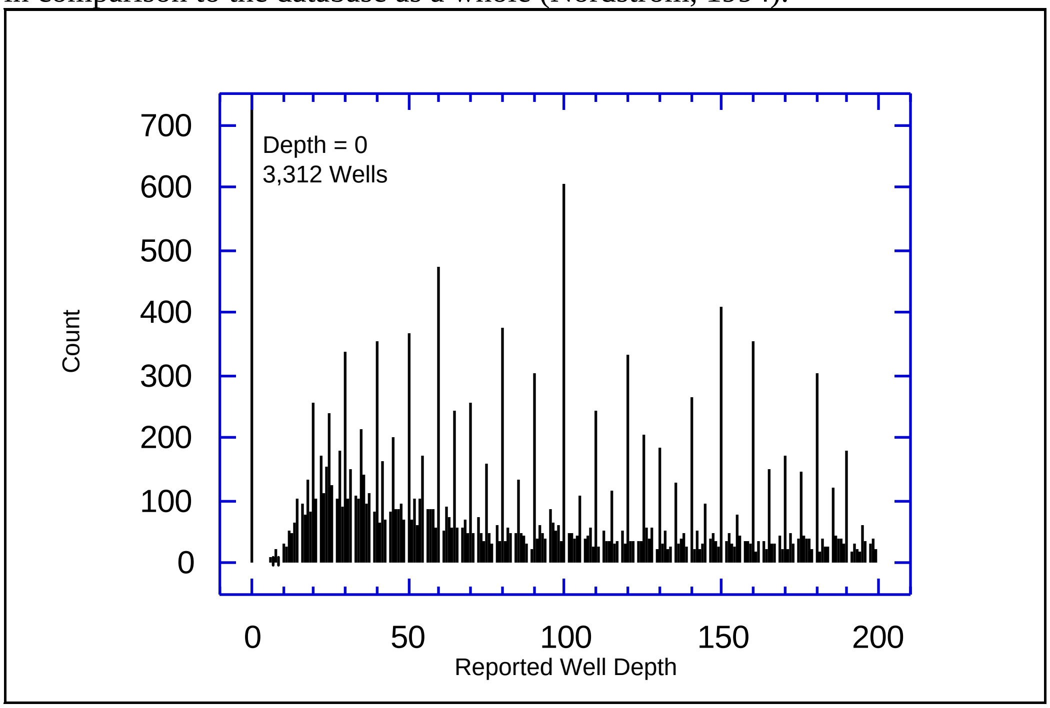

Figure 3.6 Well Depths (less than 200 feet)

The well description data included in the study, like the nitrate measurement data, were accepted "as is" without many exclusions. This does not mean that the data is considered error-free, but reflects the belief that the quantity of data is large enough that individual errors will not significantly effect the study's conclusions.

Well description records were excluded from the study if the well's latitude and longitude lay outside the quadrangle indicated by its ID number (290 records), or if no nitrate measurements from that well were left in the nitrate measurement table after the deletion of unsuitable records (9,485 records, including the mis-located wells). These deletions left 29,255 well description records in the study.

3.3 SOIL DATA

The soil data used in this study comes from the U. S. Department of Agriculture's State Soil Geographic Database (STATSGO) (USDA, 1993). This rather complex data set has two major components: maps—represented in a GIS—and several related database tables. This study draws data from the STATSGO map of Texas and three related database tables, the map unit, component, and layer tables. Both the map and the tables are stored and manipulated in Arc/Info. This section describes the organization of STATSGO data and the way that values for two soil parameters, soil thickness and average soil organic matter content, were extracted from the database for use in this study.

3.3.1 STATSGO Map and Data Structure

STATSGO maps are compiled from many sources, including soil survey maps, county and state general soil maps, state major land resource area (MLRA) maps, and LANDSAT images. The soil groups shown in these sources are transferred to USGS 1:250,000-scale base maps and digitized. The basic spatial unit of organization for STATSGO is the *map unit*, a combination of associated phases of soil series with a minimum size of approximately 6.25 km^2. A map unit is identified by a code (Map Unit ID or MUID) consisting of the two-character abbreviation of the state's name and a three-digit number (for example, TX071). Map units also have names reflecting the soil groups they contain (for example, TX071 is named "Brackett-Purves-Real"). The map units are not all contiguous; the map of Texas contains 4031 polygons classified into 632 map units, so on the average a Texas map unit is made up of six discontiguous polygons. Of the 632 map units in the STATSGO database for Texas, one (TX631) has no associated polygons, and one (TXW, the water group) has no associated soil parameter values. The remaining map units range in area from 10 km^2 to 21,500 km^2, with an average area of 1,082 km^2 and a median area of 570 km^2. The histogram in Figure 3.7 shows that a substantial majority of the map units cover areas of less than 1,000 km^2.

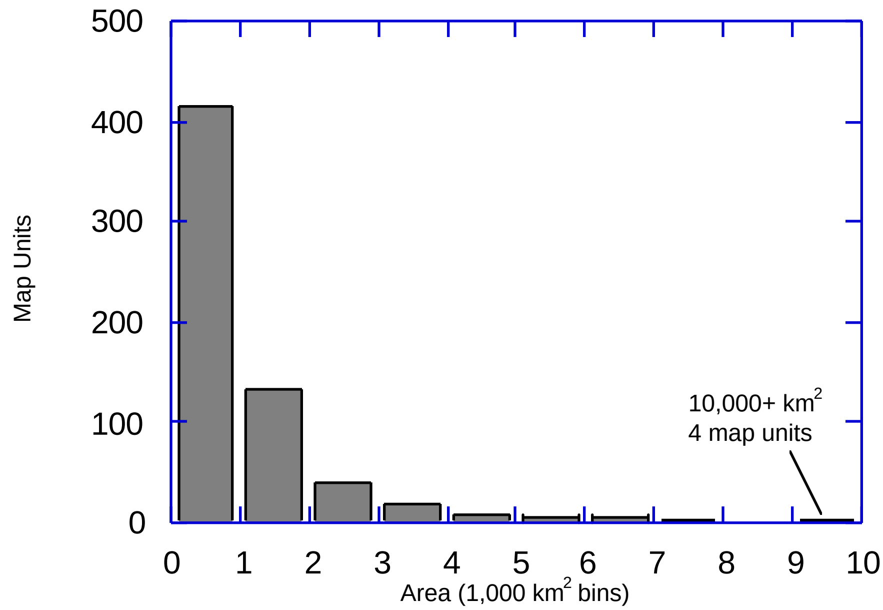

Figure 3.7 Map Unit Area Histogram

The relationship between the polygons, map units, and related tables is illustrated in Figure 3.8 and described in the following paragraphs. (The map units and data shown in Figure 3.8 are made up for purposes of illustration.)

The map units are made up of *components*, also called "soil sequences," or "soil series." Although the STATSGO map does not show components, they—like the map units—are horizontal divisions of the earth's surface, and the area of a map unit is the sum of the areas of the components it contains. Each map unit may contain from 1 to 21 components. In Texas, map units contain an average of 9 components. A component is uniquely identified by a map unit ID and a sequence number. STATSGO assigns 60 properties to the components, and stores

their values in the linked tables, including the component table. In the component table, the area of a component is expressed as a percentage of the map unit area.

The components, in turn, are made up of *layers*, which are vertical divisions of the soil. A component is a sequence of from 1 to 6 soil layers. In Texas, components contain an average of 3 layers. A layer is uniquely identified in the table by the map unit ID, the sequence number, and a layer number. STATSGO assigns 28 properties to each layer, and stores their values in linked tables, including the layer table.

The soil thickness, organic content, and bulk density values used in this study are stored in the layer table. All of these quantities are expressed as ranges, with maximum and minimum values listed in the table. For example, the minimum depth of the top layer in a component is zero, and the maximum depth of the bottom layer in a component is equal to the thickness of the component.

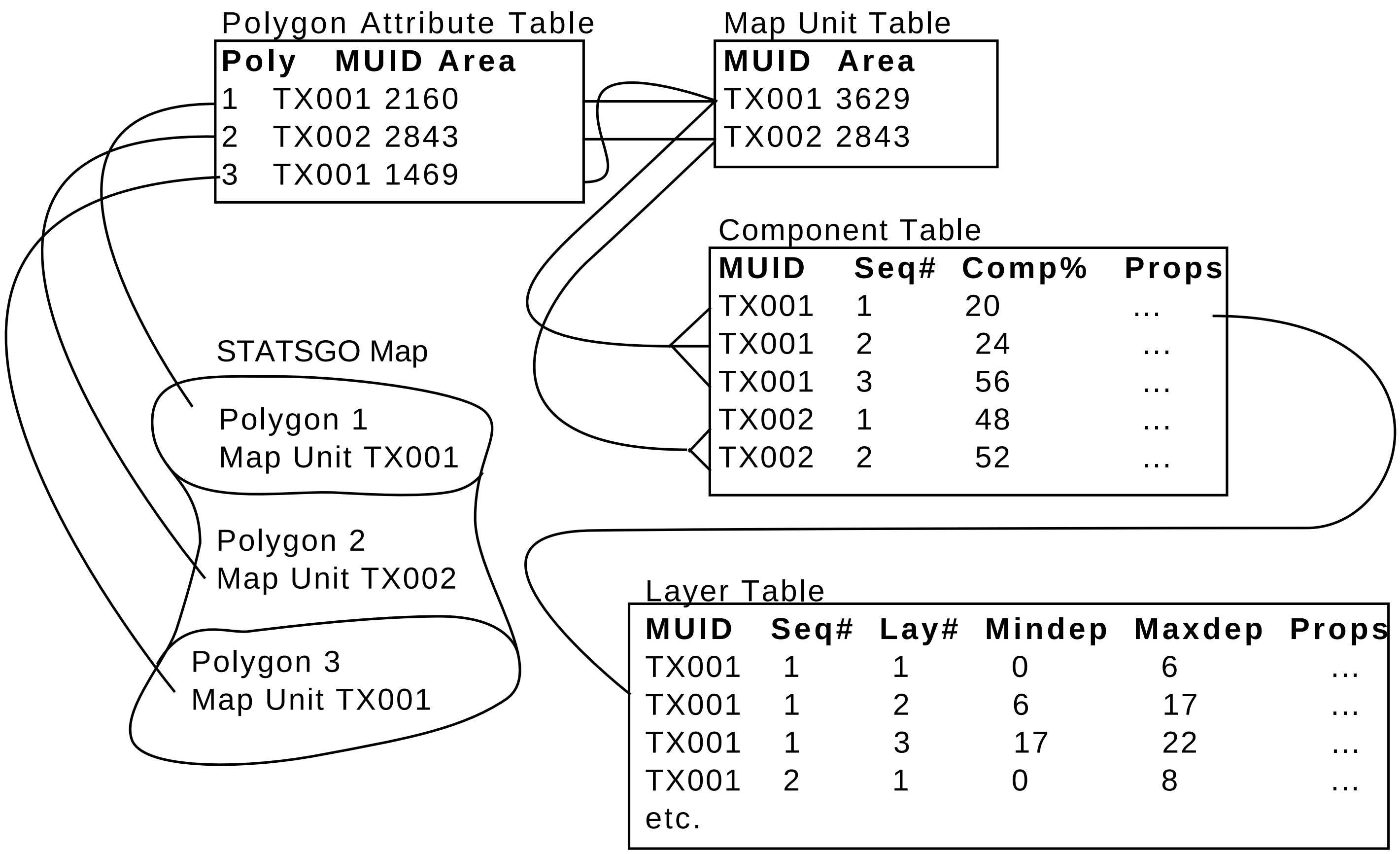

Figure 3.8 STATSGO Map and Data Organization

3.3.2 Using STATSGO Data

Figure 3.9 shows excerpts from the STATSGO map of Texas, giving some idea of the spatial structure of the map units. The area falling in the 1_ quadrangle between 30_ and 31_ N latitude and 98_ and 99_ W longitude (1_ quadrangle number 57 in the TWDB well-numbering system) is divided into roughly 140 polygons, which belong to 18 map units. The selected 7.5' quadrangle (number 5740) contains parts of two map units, which have identification codes "TXW" and "TX071." TXW is the code for all bodies of water in the state (in this case, part of Lake Travis), and TX071 is the "Brackett-Purves-Real," map unit . The soil series (also called "components") that make up TX071 are listed in Table 3.8.

Table 3.8, extracted from the component table, shows, for example, that the Purves soil series makes up 13% of map unit TX071. Table 3.9, extracted from the layer table, shows values for minimum and maximum layer depths in inches and minimum and maximum organic material content . The Purves series consists of three layers, which are 12, 2, and 6 inches thick, respectively. The total depth of the Purves series is thus 20 inches.

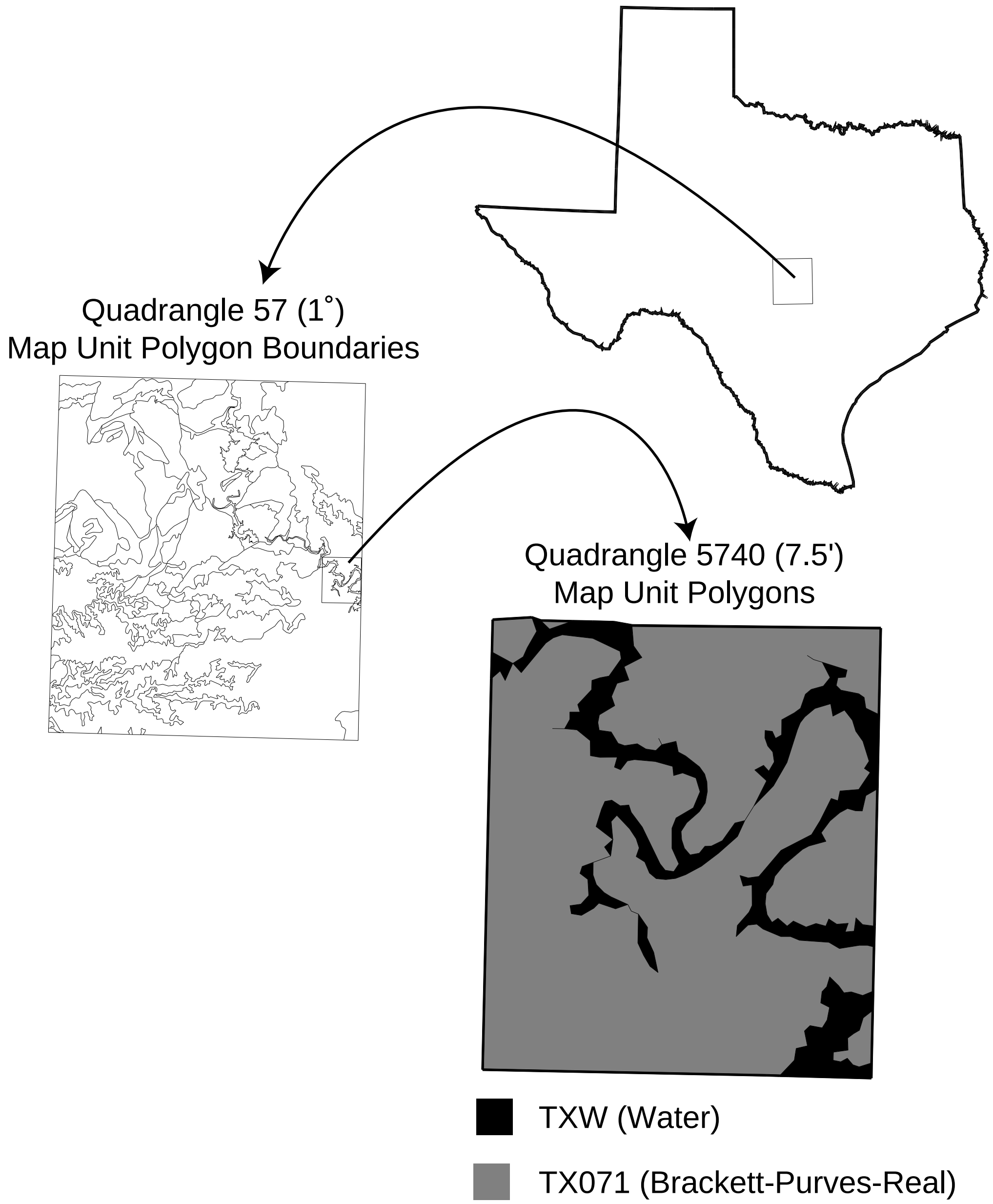

Figure 3.9 STATSGO Map Units

Table 3.8 Soil Series in Map Unit TX071 "Brackett-Purves-Real"

Seq. #	**Seq. Name**	**Comp %**
1	BRACKETT	28
2	BRACKETT	12
3	PURVES	13
4	REAL	7
5	REAL	3
6	ROCK OUTCROP	3
7	ROCK OUTCROP	3
8	COMFORT	6
9	BOLAR	4
10	DOSS	4
11	KRUM	4
12	ALEDO	5
13	OAKALLA	2
14	GRUENE	1
15	ECKRANT	2
16	BOLAR	1
17	SUNEV	1
18	TARPLEY	1
--	TOTAL	100

Table 3.9 Layers in Purves Component of Map Unit TX071

sequence number	**layer number**	**min. depth** (inches)	**max. depth** (inches)	**min. organic matter** (%)	**max. organic matter** (%)	**min. bulk density** (g/cm^3)	**max. bulk density** (g/cm^3)
3	1	0	12	1	4	1.25	1.45
3	2	12	14	1	2	1.25	1.45
3	3	14	20	0	0	0	0

Calculating the average organic material content for the layer requires more computation than the layer thickness. Organic matter is expressed as a percentage of soil mass, and must be multiplied by the bulk density of the soil to produce an organic mass density. For each layer, the average organic content and bulk density can be estimated as the midpoint between the minimum and maximum values (2.5%, 1.5%, and 0% organic matter, and 1.35, 1.35, and 0 g/cm^3, respectively). Multiplying these values by the layer thicknesses and summing over the layers produces an estimate of the organic material per unit area in the component.

$$M \approx \sum_{i=1}^{n} b_i \frac{(o_{min.} + o_{max.})_i}{2} \frac{(\rho_{min.} + \rho_{max.})_i}{2} \qquad (3\text{-}1)$$

where M is the density of organic matter (g/cm^2) for the component, b_i is the thickness (cm) of the layer, o is the weight percentage (by weight) of organic matter in the layer, ρ is the bulk density (g/cm^3) of the layer, and n is the number of layers in the component. A factor of 10 is used to convert g/cm^2 to kg/m^2. Table 3.10 shows how the organic content in the Purves series was calculated to be 11.32 kg/m^2. Note that the organic matter content for the component is expressed as a density by area, rather than volume because the organic content has been integrated over the depth of the soil.

Table 3.10 Derived Values for Soil Organic Content in Purves Series (Map Unit TX071)

sequence number	**layer number**	**thickness** (cm)	**mid bulk density** (g/cm^3)	**mid organic matter** (%)	**organic content** (kg/m^2)
3	1	30.5	1.35	2.5	10.29
3	2	5.1	1.35	1.5	1.03
3	3	15.2	0	0	0.00
3	all	50.8	--	--	11.32

Table 3.11 Soil Series Parameters for Map Unit TX071

Seq. #	**Seq. Name**	**Comp %**	**Thickness** (inches)	**Avg. om** (kg/m^2)
1	BRACKETT	28	60	23.26
2	BRACKETT	12	60	23.26
3	PURVES	13	20	11.32
4	REAL	7	36	9.07
5	REAL	3	36	9.07
6	ROCK OUTCROP	3	80	0
7	ROCK OUTCROP	3	80	0
8	COMFORT	6	20	4.44
9	BOLAR	4	44	18.86
10	DOSS	4	48	13.03
11	KRUM	4	72	28.61
12	ALEDO	5	20	5.83
13	OAKALLA	2	60	11.18
14	GRUENE	1	80	0
15	ECKRANT	2	30	17.86
16	BOLAR	1	44	18.86
17	SUNEV	1	72	22.69
18	TARPLEY	1	22	11.02
unit	TX071	100	48	15.77

Table 3.11 shows the calculated soil thicknesses and organic matter for the components of TX071. The map unit values shown on the last line of Table 3.11 are area-weighted averages, calculated by summing the products of the parameter values and the component percentages. Although values can be calculated for the soil parameters at both component and map unit levels, only the map unit averages can be located on the STATSGO map. For example, the Purves series makes up 13% of map unit TX071, but STATSGO provides no information about which 13% that is. For this reason, the STATSGO data cannot properly be applied to any areas but the STATSGO map units.

STATSGO's relatively poor spatial resolution presents a difficult problem for users of the data. In this study, the well and water quality data are organized on spatial units of 2.5' quadrangles, which are much smaller than STATSGO map units. Figure 3.10 shows the relative sizes of map unit TX071, a 1_ quadrangle, a 7.5' quadrangle, and a 2.5' quadrangle. Map unit TX071 covers about 6,700 square kilometers; in the same part of the state, a 1_ quadrangle covers about 10,000 square kilometers, a 7.5' quadrangle covers about 166 square kilometers, and a 2.5' quadrangle covers about 18.5 square kilometers. A 2.5' quadrangle is roughly the same size as the Oakalla component of map unit TX071.

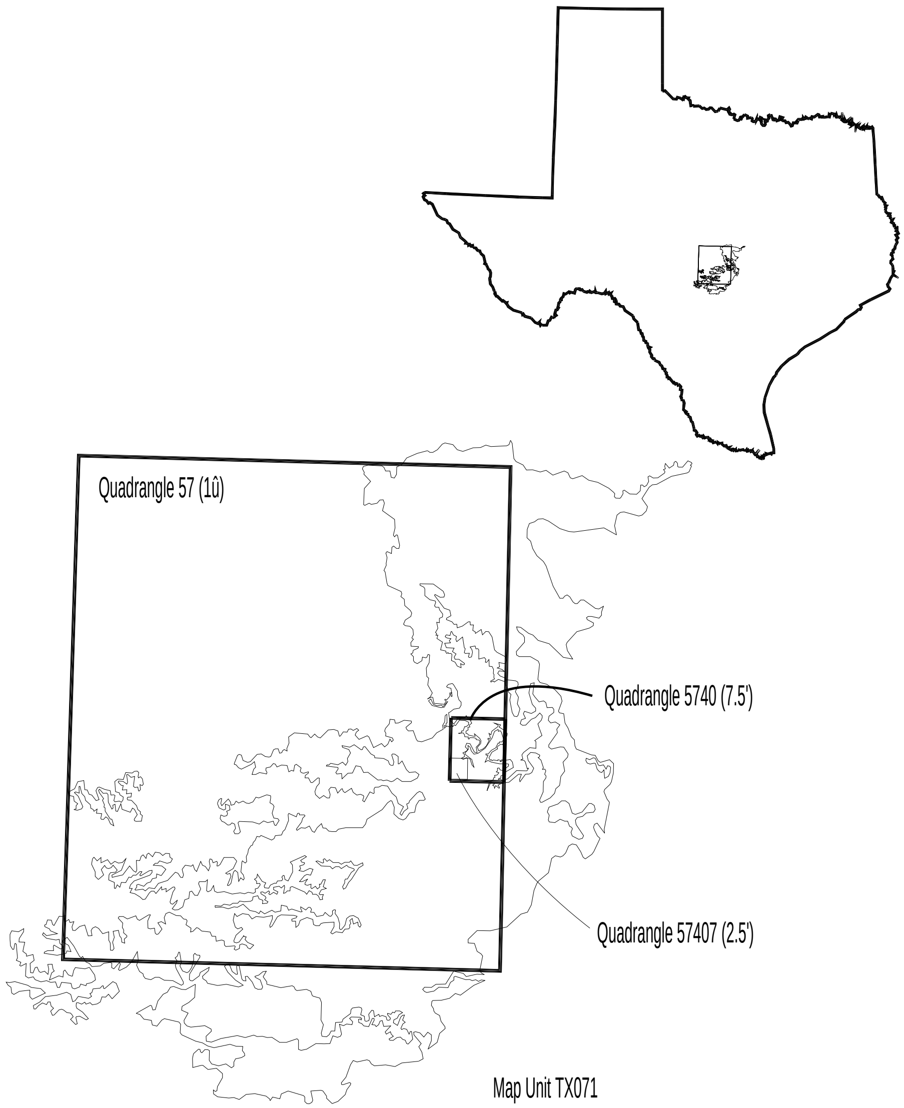

Figure 3.10 Map Unit TX071 with
Quadrangles for Size Comparison

Applying map unit values to areas other than the map units themselves—such as 2.5' quadrangles—requires the user to assume a spatial distribution of the soil series within the map units. The simplest assumption, and the best available without requiring supplementary data, is that the area-weighted averages of soil parameter values are uniformly distributed properties of the map units. This assumption contradicts fact, and the STATSGO user's guide specifically warns against it.

In spite of this warning, this study employs just this assumption. This use of the data can be justified on a variety of grounds. First, this study seeks to describe the variation of water quality through Texas using a database organized in 2.5' quadrangles. The STATSGO map units are organized in different divisions of the land surface and the two systems are irreconcilable; one must be compromised. Since the well data are primary, compromise of the STATSGO data must be tolerated. Secondly, the map units, by their nature, are groups of associated soils, so the variation in soil properties between map units ought to be greater than the variation within map units. Thirdly, since this is a statewide study, it is reasonable to assume that the errors introduced by mishandling the STATSGO data small enough that they will not significantly influence the conclusions drawn over so large a study area.

Using this compromise, soil parameters will be estimated by the following procedure. Any region (e.g., a 7.5' quadrangle) lying entirely within the boundaries of a STATSGO map unit will be assigned the average parameters for that map unit. Any region that crosses STATSGO map unit boundaries will be assigned soil parameter values equal to the area-weighted average of the values associated with the non-water map units that lie within the region. For example, since quadrangle 5740 is composed entirely of water (TXW) and portions of map unit TX071, it would be assigned values equal to the averages for TX071.

3.3.3 Range and Distribution of Soil Parameter Values

The average soil thickness in the non-water map units ranges from a minimum value of 22.4 inches to a maximum of 88 inches. The area-weighted average of the soil thickness is 65.2 inches, and the median values is 69.9 inches. The histogram-like chart in Figure 3.11 shows the map unit area associated with ranges of soil thickness in 5-inch bins. The distribution of soil thickness over the surface of Texas is illustrated in Figure 3.12.

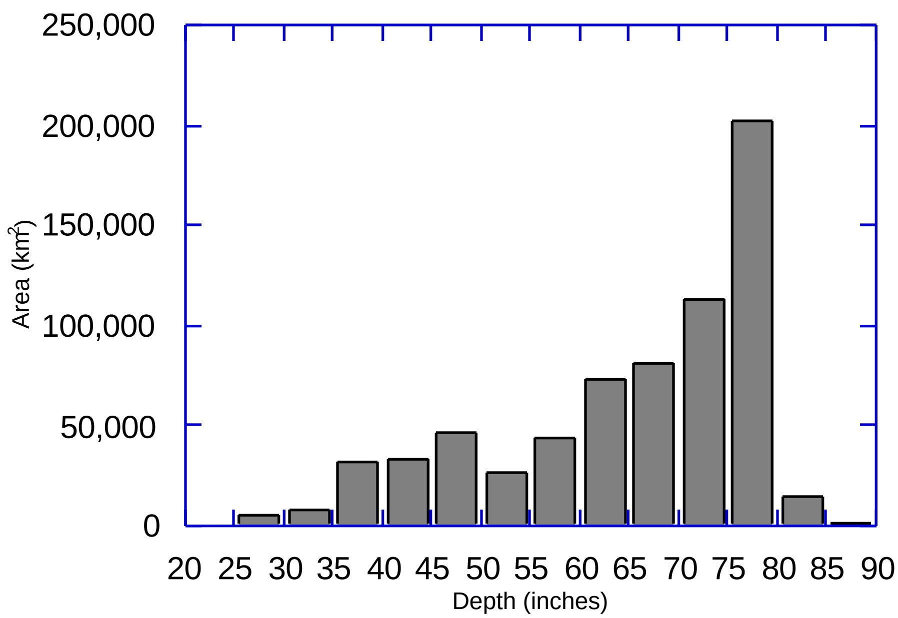

Figure 3.11 Soil Depth Histogram

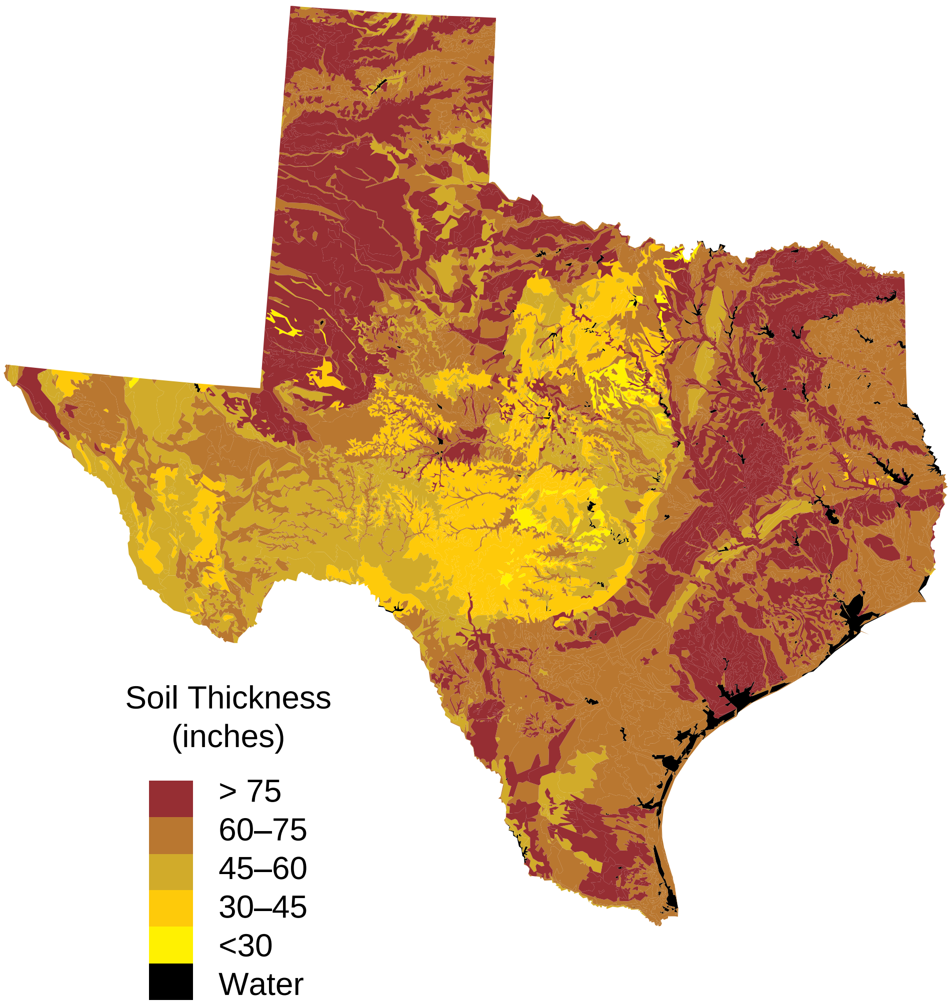

Figure 3.12 Spatial Distribution of Soil Thickness

The average soil organic content in the non-water map units ranges from a minimum value of 0.76 kg/m^2 to a maximum of 74.9 kg/m^2. The area-weighted average of the soil organic content is 16.2 kg/m^2, and the median values is 15.1 kg/m^2. The histogram-like chart in Figure 3.12 shows the map unit area associated with ranges of soil organic content in 5-kg/m^2 bins. The distribution of soil organic content over the surface of Texas is illustrated in Figure 3.13.

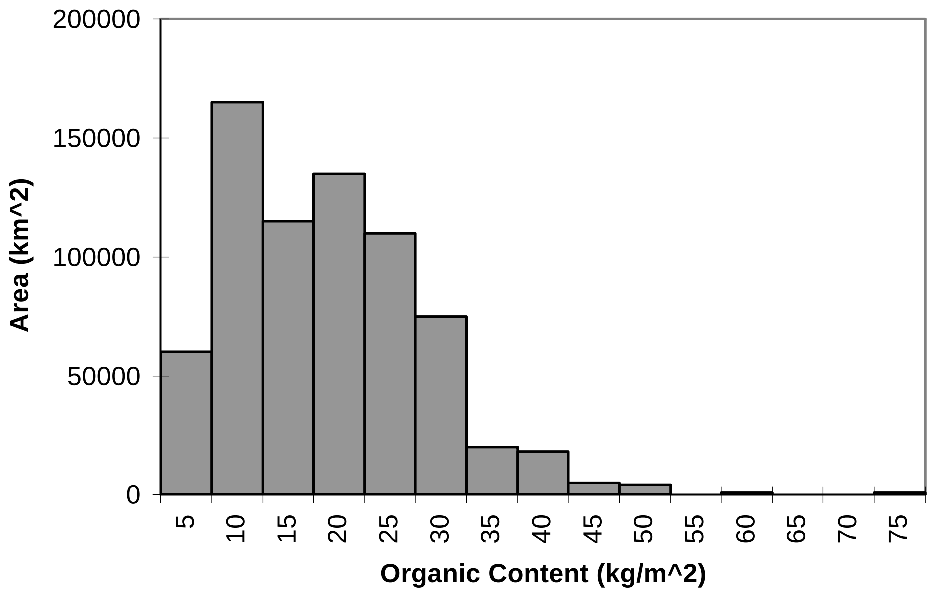

Figure 3.13 Soil Organic Matter Histogram

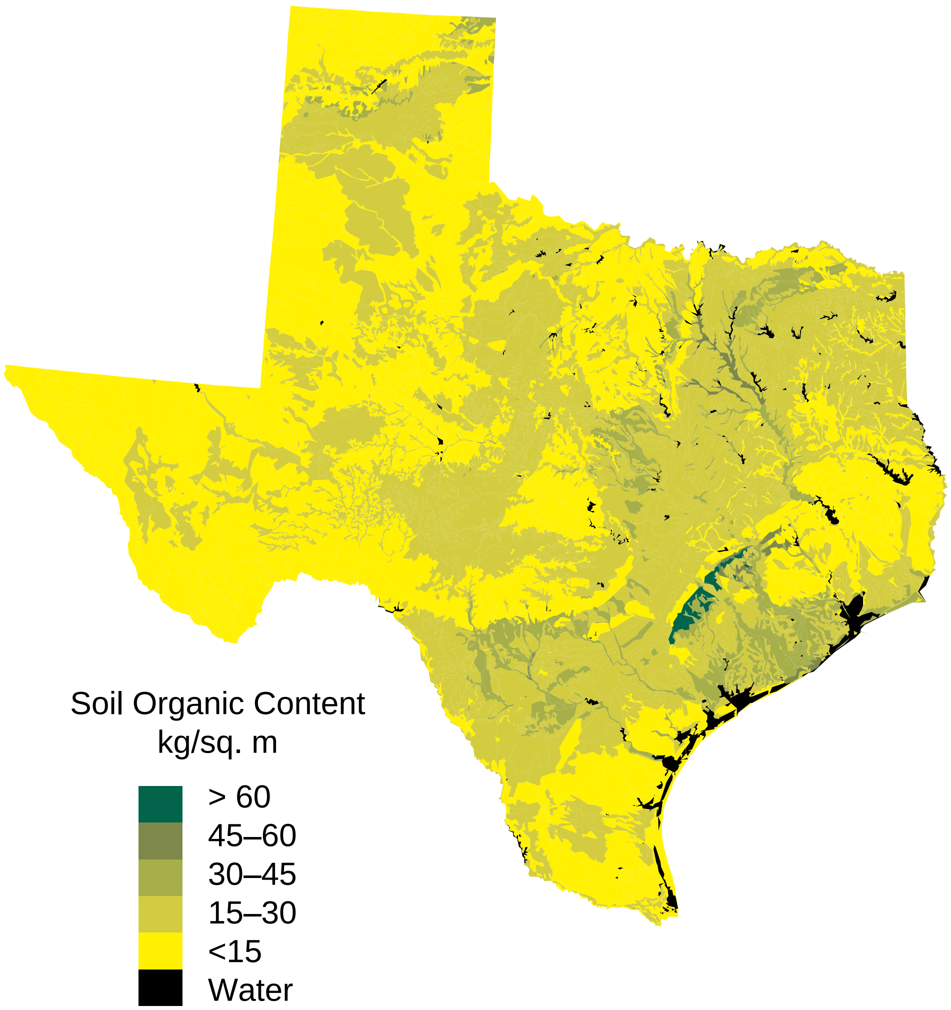

Figure 3.14 Spatial Distribution of Soil Organic Content

3.4 PRECIPITATION DATA

The precipitation data used in this study were copied from, or derived from data included in Hydrosphere Inc.'s *Climatedata* CD-ROMs (Hydrosphere Data Products, Inc., 1994). This data set consists of GIS coverages showing point locations of the observation stations, and database tables listing the daily observations of climatic data for the period of record of the TD-3200 Summary of the Day Cooperative Observer Network database of the National Climatic Data Center (NCDC).

3.4.1 Preparation of Annual Average Precipitation Map

The annual average precipitation map used in this study is intended to reflect the variation of expected rainfall across Texas. The objective in preparing the map was not to produce the best possible prediction of average annual precipitation at each station, which would require that the entire period of record be used for each station, but rather to produce the best estimate of the relative magnitudes of precipitation at different stations, which requires that the same period be reported for all stations.

This goal sets up an interesting set of conflicting requirements. For any map, including more points improves the spatial resolution, and for any time series, extending the period of record increases confidence in the calculated average values. Requiring that the period of record be the same for all stations means that stations operating for only a part of the period cannot be included in the map, so a longer period of record leads to fewer points, and vice versa.

After a trial-and-error exploration of the data, the following criteria were used to select the data for the map used in this study:

1. The period of record for the map extends from 1951 to 1980.

2. A station is deleted from the map if a sequence of than two years is missing from the station's records. (NCDC considers a year "missing" if it contains a missing month. A month is "missing" if more than nine days of data are absent.)

The selected period of record includes periods of both very low precipitation (the early-to-mid 1950s) and very high precipitation (the early 1970s), and can be considered a representative period for precipitation in Texas. Requiring a longer period of record (1951–1990) or tolerating only single-year gaps resulted in roughly 25% reductions in the number of stations included in the map.

The procedure used to generate the precipitation map is described in Section 5.2.3. The resulting map appears in Figure 3.15.

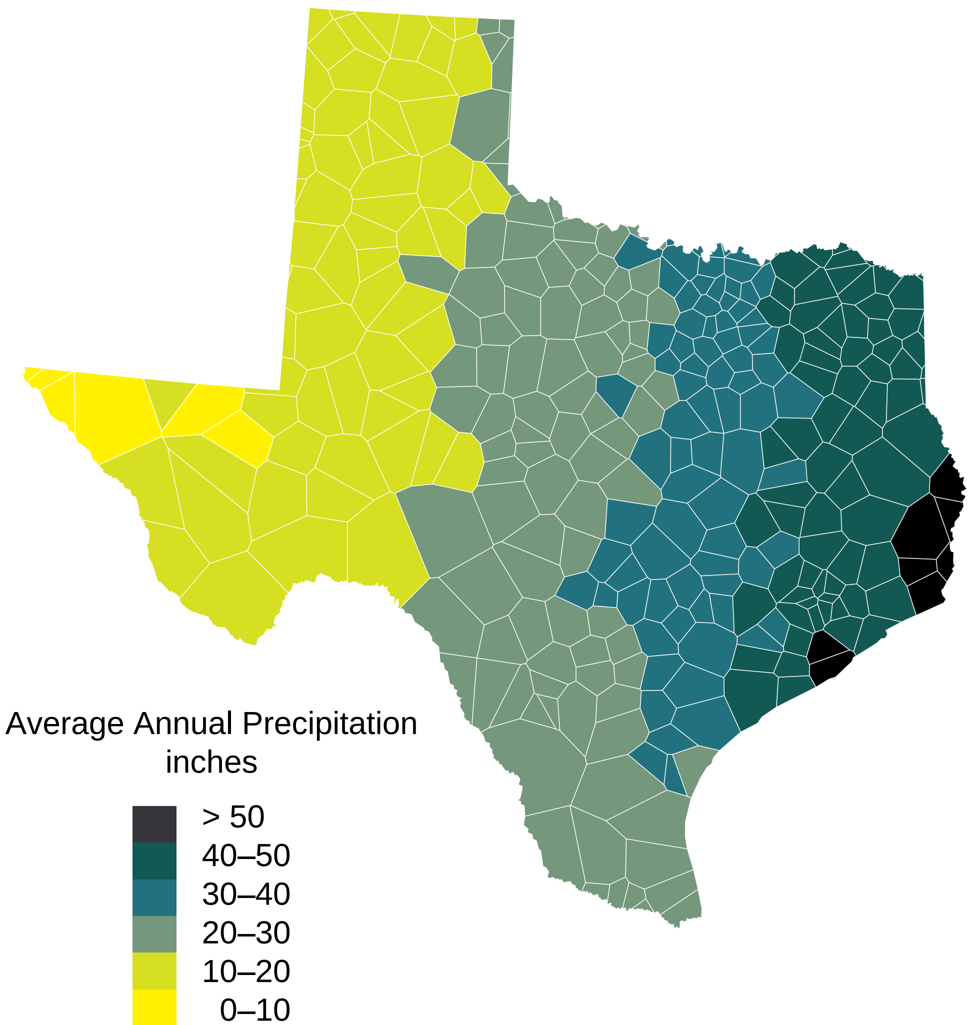

Figure 3.15 Spatial Distribution of Annual Average Precipitation (Thiessen Polygons)

3.4.2 Range and Distribution of Precipitation Data

The Thiessen polygons range in size from a minimum area of about 10 km^2 to a maximum area of about 10,600 km^2, with an average area of 2,130 km^2 and a median area of 1,690 km^2. The size of the polygons is inversely related to the density of gauges and hence to population. Polygons are small around cities and large in the unpopulated areas of west Texas. Figure 3.16 shows the frequency distribution of Thiessen polygon sizes for the study's precipitation gauging network.

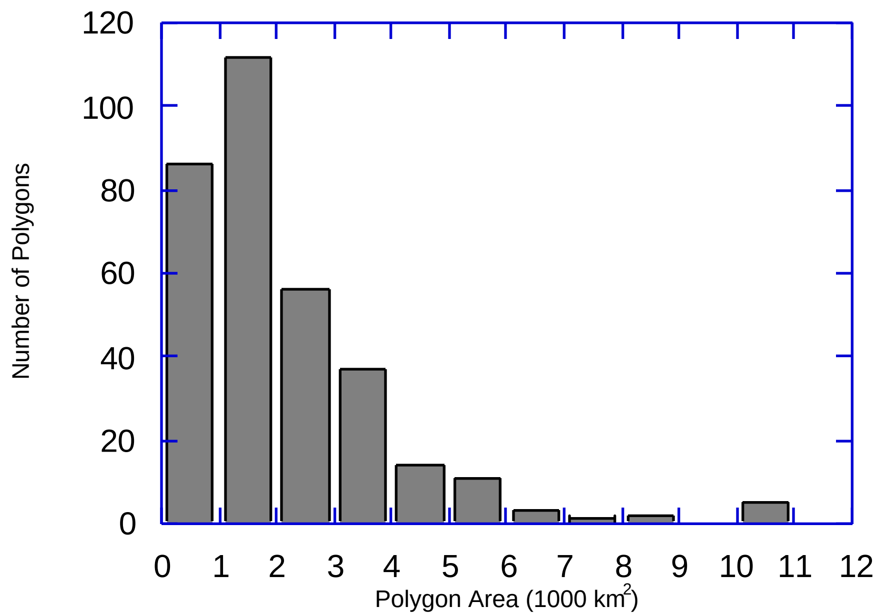

Figure 3.16 Thiessen Polygon Area Histogram

By the reckoning described in section 3.4.1, average annual precipitation ranges from a low of 7.8 inches in El Paso to a high of 59.1 inches in Orange. The area-weighted average precipitation for Texas as a whole is 26.8 inches and the area-based median is 24.5 inches (meaning that half the area of the state averages more than 24.5 inches of precipitation per year and the other half averages less). The histogram-like diagram in Figure 3.17 shows how the Thiessen polygon area associated with the various levels of precipitation.

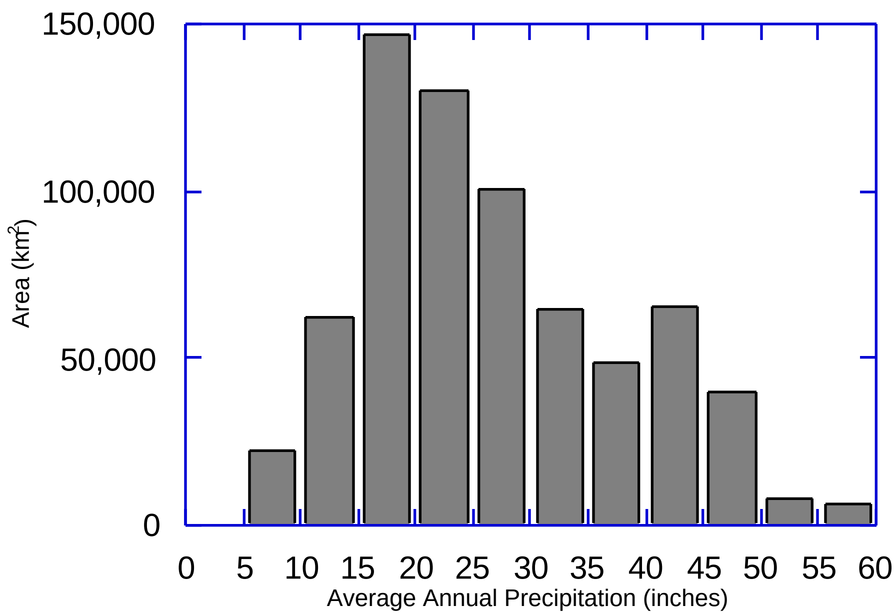

Figure 3.17 Precipitation Histogram

3.5 FERTILIZER SALES DATA

The nitrate fertilizer application data has the poorest spatial resolution of all the data used in this study. Figure 3.18 was generated from annual total fertilizer sales collected nation-wide on a county level by the EPA's office of Policy Planning and Evaluation. Battaglin and Goolsby (1995) related sales figures for the years 1986–1991 to county maps of the United States as part of a project to illustrate nationwide trends in agricultural chemical use with GIS (Mr.

Battaglin made the fertilizer data used in this study available to the author prior to the publication of the cited report). In addition to listing the total number of tons of fertilizer sold in each county, Battaglin and Goolsby divided the tons of fertilizer sold by the total area of the county to compensate somewhat for the range of variation in size of counties. The result is a number that they call "use" in tons per square mile. For the map in Figure 3.17, six years of use were averaged for each county. These averages range from a low of 0, meaning no recorded nitrate sales in the county for the six years, and a high of 18.9 tons per year of recorded nitrate fertilizer sales per square mile of county.

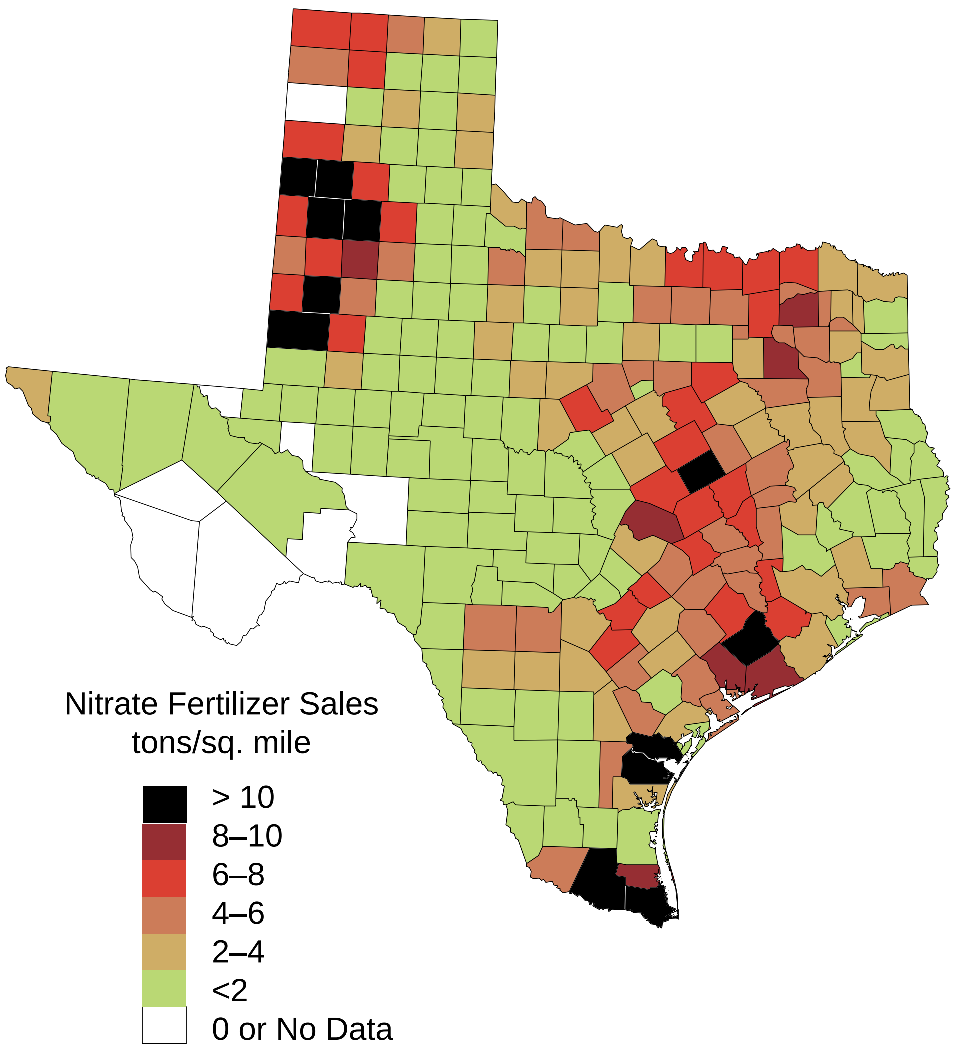

Figure 3.18 Sales of Nitrogen Fertilizers by County
(Annual Average 1986–1991)

3.6 WATER UTILITIES DIVISION NITRATE MONITORING DATA

Nitrate measurements collected by the Water Utilities Division (WUD) of the Texas Natural Resource Conservation Commission as part of their Primary Drinking Water Standards enforcement efforts, were used as an independent data set to test nitrate vulnerability predictions based on the TWDB data.

The nitrate measurements reported by the WUD are collected at points of entry to public water distribution systems, i.e., after water from multiple sources has been mixed and treated. A water system may have several points of entry and several wells or surface intakes supplying those points of entry. Water samples from points of entry do not represent individual wells unless the point of entry is tied to only one well.

The data provided by the WUD include nitrate concentrations measured at points of entry, identifications of those points of entry and the wells and surface intakes supplying them, and the locations of the wells. These were represented in two database tables and a GIS coverage. The nitrate measurement table includes the system and point of entry identification for each measurement, along with the date of sample collection and analysis results. The point of entry table contains one record for each well, listing the well ID, system ID, and point of entry ID. (WUD well numbers are not the same as TWDB well numbers. They are based on county and water supply identification, rather than geographic coordinates.) By linking nitrate concentration to points of entry, points of entry to wells, and wells to locations, it is possible to tie nitrate concentrations to quadrangles for comparison to the quad exceedence probabilities calculated from the TWDB data. The process and results of this comparison are described in Sections 5.8 and 6.4.

3.7 HERBICIDE AND NITRATE DATA FROM MIDWESTERN U.S.

Because of the lack of a sufficient quantity of measurements of herbicides and other man-made agricultural chemicals in Texas groundwater, it is not possible to determine whether vulnerability to nitrate is correlated to vulnerability to other agricultural chemicals in Texas. However, in order to generalize the results of a study of vulnerability to nitrate contamination to other agricultural chemicals, it is necessary to assume some relationship between nitrate and those other chemicals. The data presented by Kolpin et al (1993) is used to test the rather mild assumption that geologic conditions favorable to a high rate of detection of elevated nitrate levels will also be favorable to a high rate of herbicide detections.

The data were collected in 1991 from 300 wells in the Midwestern U.S. The nitrate and herbicide data were collected as part of an effort to characterize the spatial and seasonal distribution of agricultural chemicals in groundwater, and to provide data for an exploratory statistical analysis of the influence of anthropogenic, and geologic and other natural factors on the occurrence of herbicides (Kolpin and Burkart, 1991).

A full account of the reconnaissance can be found in the cited references. The data used here included the reported concentrations of nitrate and nine herbicides or herbicide metabolites (alachlor, atrazine, cyanazine, deethyl-atrazine, deisopropyl-atrazine, metolachlor, metribuzin, prometon, and simazine), and two geologic descriptors of well surroundings (depth to top of aquifer, and aquifer type—bedrock or unconsolidated). The use of the data is explained in Section 6.5.