Environ Geol (2007) 51:931–939
DOI 10.1007/s00254-006-0456-1

ORIGINAL ARTICLE

# Combining rough sets and GIS techniques to assess aquifer vulnerability characteristics in the semi-arid South Texas

**V. Uddameri · V. Honnungar**

Received: 30 August 2005 / Accepted: 12 July 2006 / Published online: 22 August 2006


**Abstract** The coastal semi-arid region of South Texas is undergoing significant growth causing an enormous burden on its limited water resources. Understanding regional-scale vulnerability of this resource is important for sustainable water resources management and land use development. In this study, DRASTIC methodology is integrated with an information-analytic technique called rough sets to understand groundwater vulnerability characteristics in 18 different counties of South Texas. The rough set theory provides three useful metrics: the strength factor which depicts how vulnerability characteristics occur over the area; the certainty factor computes the relative probabilities for various vulnerability states within a county and the coverage factor which elucidates the fraction of a specific vulnerability state present in each county. The coupling of rough sets with GIS is particularly advantageous to cluster counties exhibiting similar vulnerability characteristics and to obtain other related insights. The application of the approach indicates that the groundwater vulnerability exhibits greater variability along the coast than in the interior sections of the area. The shallow aquifer in Aransas, DeWitt, Goliad and Gonzales counties is the most vulnerable, while the aquifer in Duval, Jim Wells, Karnes, Live Oak, Nueces and San Patricio is less vulnerable. This approach should prove useful to regional planners and environmental managers entrusted with the protection of groundwater resources.

V. Uddameri (✉) · V. Honnungar
Department of Environmental Engineering,
MSC 213, Texas A&M University-Kingsville,
Kingsville, TX 78363, USA
e-mail: vuddameri@tamuk.edu

**Keywords** Geographic information systems · Multi-criteria decision making · Fuzzy logic · Soft computing · Data mining · South Texas

## Introduction

Establishing guidelines for aquifer protection is an important step towards sustainable management of groundwater resources in fast growing semi-arid regions such as South Texas. Often these studies integrate diverse set of hydrologic, geologic and environmental data with index-based techniques, such as DRASTIC (D: Depth to water table, R: aquifer Recharge, A: Aquifer media, S: Soil media, T: Topography, I: Impact of vadose zone, C: hydraulic Conductivity), to provide relative comparisons and screening level assessments (Aller et al. 1987). Given the ease of application of index methods, and the difficulties in carrying out detailed fate and transport modeling owing to limited resources and paucity of site-specific data (Fredrick et al. 2004), DRASTIC and other similar techniques are being used worldwide to identify areas that have a potential for groundwater contamination (Al-Zabiet 2002; Edet 2004; Fredrick et al. 2004).

Over the last decade, the utility of DRASTIC has been further enhanced by integrating it with informational technologies. For example, DRASTIC has been integrated with geographic information systems (GIS) and remote sensing tools for easy visualization (Evans and Myers 1990; Al-Adamat et al. 2003) and coupled with fuzzy logic to deal with uncertainties in the procedure (Cameron and Peloso 2001). An important application of DRASTIC is to evaluate the vulnerability of the aquifers in different administrative units (e.g., counties) within a region. Such an evaluation is clearly useful to guide future land use development as well as prioritize and allocate scarce fiscal and logistic resources for remedial activities. GIS provide tools to reclassify vulnerability maps to depict areas of high, medium and low vulnerability based on some pre-specified criteria. However, most administrative units (counties) will usually have a mixture of areas with varying degrees of vulnerability thus precluding an easy ranking of these units. As such, mathematical methodologies that provide insights related to the relative vulnerabilities of different administrative units within a region are of interest to regional planners and decision makers.

The concept of rough sets was first introduced by Pawlak (1982) to deal with classificatory analysis of data. It is a data mining technique aimed at synthesizing core knowledge embodied within a database by identifying and eliminating indiscernible information. The utility of rough sets in environmental applications is slowly being recognized and has been used in some environmental applications (An et al. 1996; Shen and Chouchoulas 2001; Tan 2005). However, to the best of the authors' knowledge the integration of rough sets with DRASTIC methodology to classify different administrative units within a region according to their aquifer vulnerability has not been demonstrated. The present study is a seminal application in this regard and it is hoped that this illustration will add to the toolkit of decision makers and regional planners entrusted with sustainable development of land and water resources.

## Materials and methods

### Study area

Eighteen counties in the coastal bend region of South Texas were selected for this illustrative application and are depicted in Fig. 1. The area is mainly underlain by the Gulf coast aquifer, which is comprised mostly of sedimentary deposits of sand and clays of quaternary era (Baker 1979). A small portion along the western section of the study area in Live Oak, McMullen, Karnes and Gonzales counties is underlain by the tertiary Carrizo-Wilcox formation. The topography of the region consists of gently rolling plains and the elevation ranges from approximately 76 m above mean sea level on the western side to near mean sea level along the eastern Gulf coast. The precipitation ranges from approximately 1.016 m/year along the northeast sections to roughly 0.635 m/year along the southwestern boundary.

### DRASTIC-based vulnerability mapping

The data required to develop DRASTIC aquifer vulnerability maps were obtained from various sources summarized in Table 1. The aquifer media was assigned a constant rating as only sand and gravel formations were of interest. Thus, while this parameter contributed to the overall DRASTIC score it did not have any relative impacts on aquifer vulnerability. The organic matter in the soil acts to retard and assist in the removal of hydrophobic organic contaminants and this parameter was used to characterize the impacts of vadose zone. As such, the DRASTIC assessment carried

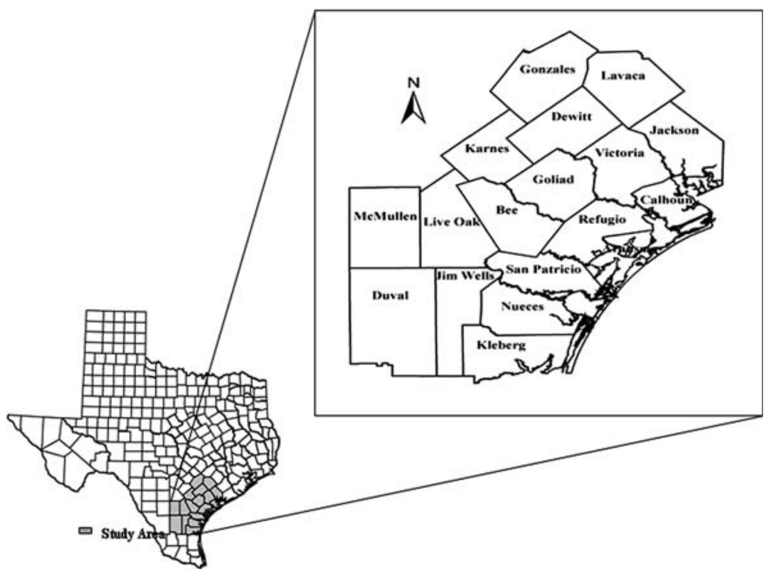


**Fig. 1** Counties in the study area and their locations in Texas

**Table 1** Data sources for DRASTIC parameters

| Data | Survey years | Data format | Scale | Source | Original projection |
|---|---|---|---|---|---|
| Adminstrative map of Texas | 2000 | GIS vector | 1:100 K | U.S. Census Bureau | GCS_NAD1927 |
| Groundwater database | 1966–2000 | MS access | – | TWDB | – |
| Topography | 1992 | DEM | 1:250 K | USGS | GCS_WGS1972 |
| Soil maps | 1994 | STATSGO | 1:250 K | USDA | GCS_NAD1983 |
| Precipitation | 1965–2002 | ASCII | – | NWS-NOAA | – |

*TWDB* Texas Water Development Board, *STATSGO* State Soil Geographic Database, *DEM* Digital Elevation Model, *USGS* United States Geological Survey, *USDA* United States Department of Agriculture, *NWS* National Weather Service, *NOAA* National Oceanographic and Atmospheric Administration

out in this study is related to hydrophobic organic contaminants. Point estimates for water table elevations for 2,886 wells in the shallow unconfined aquifers of the study area were obtained from the Texas Water Development Board (TWDB) groundwater database. The potentiometric map was contoured using the point estimates of average water levels. The Inverse Distance Weighting (IDW) scheme algorithm in ArcGIS spatial analyst extension was employed for this purpose. As measured recharge values were not readily available, the potential for aquifer recharge was estimated using the Williams-Kissel formula, which utilizes soil hydrologic group and precipitation measurements to provide the potential for recharge at any given location (Williams and Kissel 1991). All GIS data were re-projected into UTM N 14 projection with NAD 83 datum. All vector data were converted into raster format using the ArcGIS/Arc-Info version 9.0 (ESRI Inc., Redlands, CA) with a cell size of $30 \times 30$ m. The weights for different DRASTIC parameters are listed in Table 2 and are based on suggested values in the literature (Aller et al. 1987) and professional judgment of the authors. The regional scale variability of different DRASTIC parameters can be seen from maps presented in Figs. 2, 3, 4, 5, 6, and 7. All maps were reclassified using rating factors presented in Tables 3, 4, and 5. The DRASTIC vulnerability index was computed by multiplying the ratings at individual locations with corresponding weights using GIS overlay operations and map algebra. The final DRASTIC vulnerability map was reclassified into three vulnerability states namely: high, medium and low.

**Table 2** Weights assigned for DRASTIC factors

| Factor | Description | Assigned weights |
|---|---|---|
| D | Depth to the water table | 5 |
| R | Aquifer recharge | 4 |
| A | Aquifer media | 3 |
| S | Soil type | 2 |
| T | Topography | 1 |
| I | Impact of vadose zone | 2 |
| C | Conductivity | 3 |

## Data mining using rough sets

The rough set theory was developed by Zdzislaw Pawlak (Pawlak 1982) and is a complementary variant of the fuzzy set theory (Chanas and Kuchta 1992). It concerns itself with extracting generalized rules from databases. A comprehensive introduction to rough sets along with a thorough review of pertinent literature has been presented by Komorowski et al. (1999). Fundamental to the theory of rough sets is the idea of an information table, which is essentially a finite data table consisting of different columns (attributes) and rows (objects). Such an information table, $S$ consists of four parts: (1) a finite set of objects or records of the information table ($U$), (2) a finite set of attributes $A$, (3) $V$ is a set of attribute values, and (4) a function ($f$) that assigns particular values from the domains of the attributes to individual objects. Mathematically, an information table can be characterized as:

$$S = <U, A, V, f> \quad (1)$$

the attributes, $A$, are further defined into two disjoint subsets: a set of conditional attributes $C$, and a set of decision attributes $D$. Furthermore, the conditional and decision attributes are mutually exclusive. Mathematically these relationships can be expressed as:

$$A = C \cup D \quad \text{and} \quad C \cap D = \phi \quad (2)$$

typically, the first three characteristics of the decision table, namely $U$, $A$, $V$, are known and the function, $f$, that assigns particular values to the individual objects is obtained from available data and rough set theoretic algorithms. In this regard, another fundamental aspect of rough sets, namely the indiscernability relation is particularly useful. If two objects of an information system have the same sets of attributes then it is not possible to distinguish between them. For a given set of

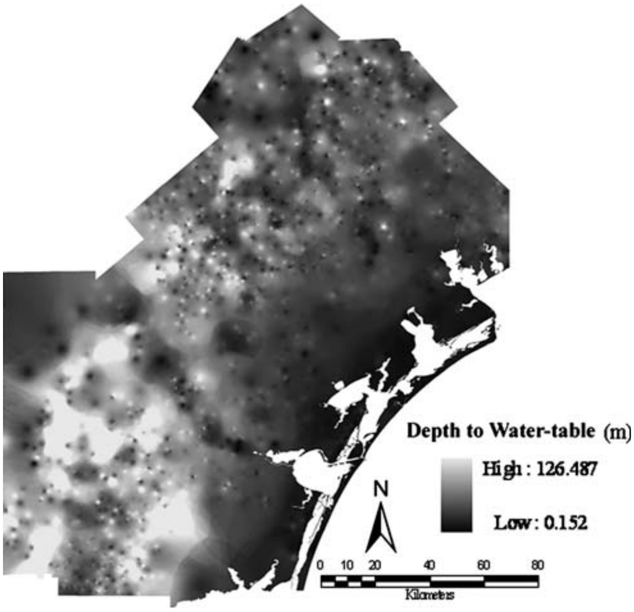


Fig. 2 Variability of the depth to the water table in the study area generated from point measurements at different wells using inverse distance weighting method

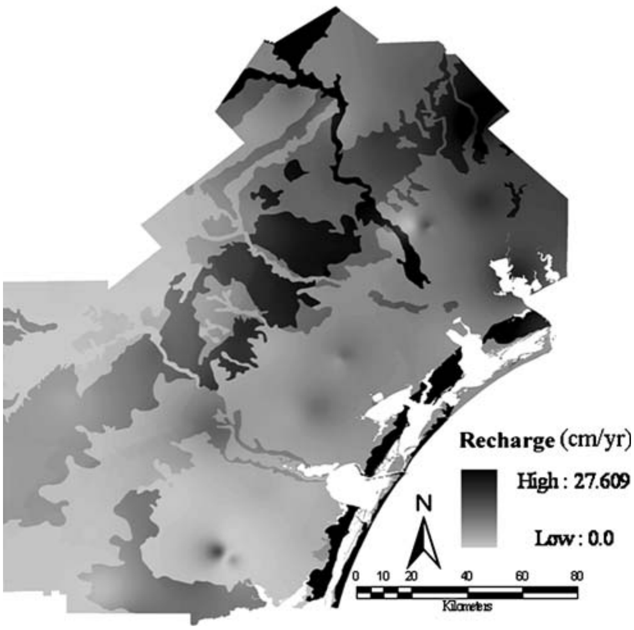


Fig. 3 Recharge rating map for the study area obtained using the Williams-Kissel method

attributes these objects contain the same information and as such are redundant. By eliminating such redundant information, the core knowledge embodied within the information system can be extracted. Rough set methodologies have also been proposed to remove redundancies in selected attributes and resolve inconsistencies that arise when two objects have the same set

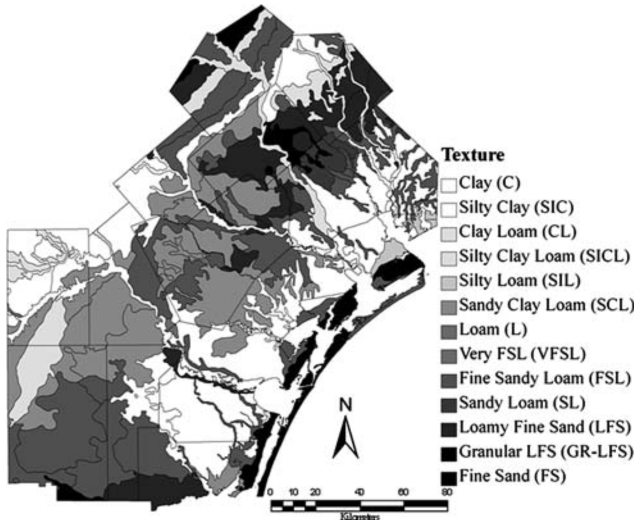


Fig. 4 Surface soil texture map for the study area

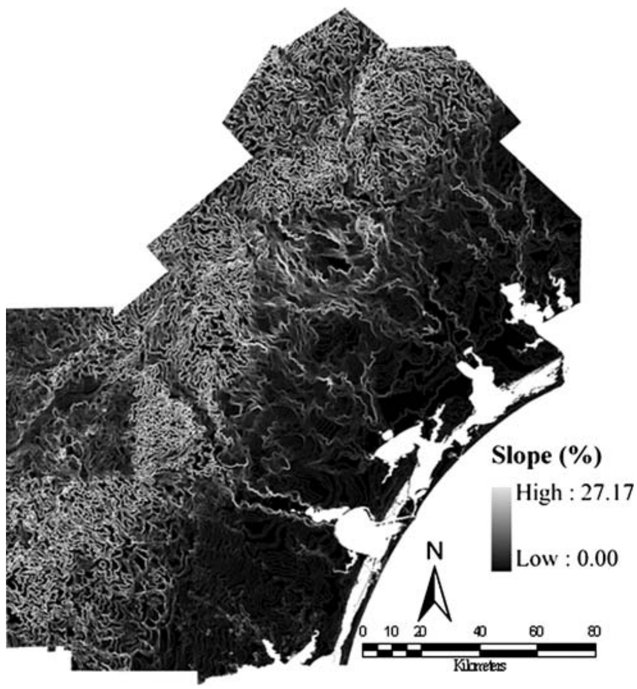


Fig. 5 Topographic map depicting slopes developed using 1:250 K DEM

of conditional attributes but different decision attributes (Komorowski et al. 1999). Recently Pawlak (2005) has extended rough sets to analyze conflicts and provided a set of mathematical constructs that can be used to obtain insights related to grouping of various parties based on their voting patterns. This mathematical framework can be used to cluster different geographic regions according to their aquifer vulnerability and is presented below:

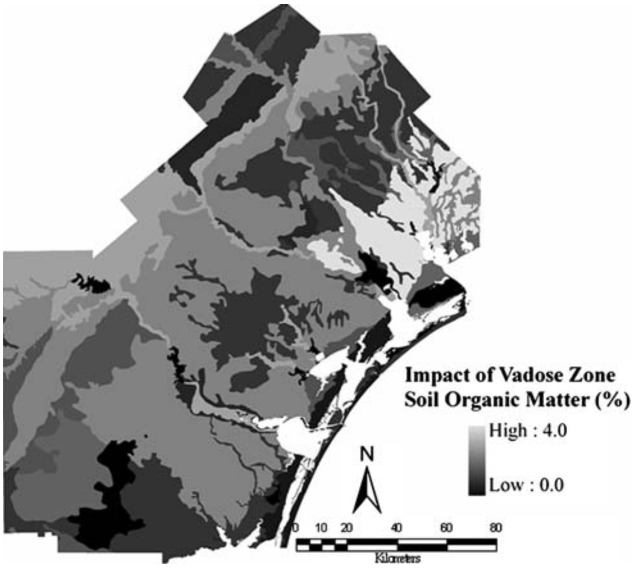


**Fig. 6** Soil organic matter map used to rate the impacts of the vadose zone

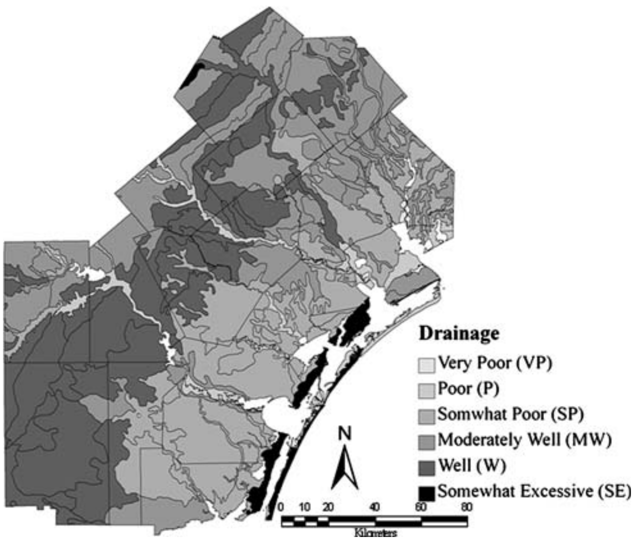


**Fig. 7** Soil drainage characteristics map used to rate conductivity factor

Following Pawlak (2005), consider an information table, $S = <U, C, D>$ where $U$ is a set of objects, $C$ is a set of conditions and $D$ is a set of decisions. Every object, $x$, in the set, $U$, will have a subset of m conditional attributes and a subset of n decision attributes. As such, a rule can be induced for each object $x$ in $U$ as follows:

$$\forall\, x \in U: \quad C = [c_1, \ldots, c_m] \quad \text{and} \quad D = [d_1, \ldots, d_n] \tag{3}$$

the sequence is called the decision rule induced by $x$ and is denoted as:

**Table 3** Ratings for depth to water table, aquifer recharge and aquifer media

| Depth to water (m) | Rating |
|---|---|
| Factor: depth to water table (D) | |
| 0–1.524 | 10 |
| 1.524–3.048 | 9 |
| 3.048–4.572 | 8 |
| 4.572–6.096 | 7 |
| 7.62–9.144 | 6 |
| 9.144–10.668 | 5 |
| 10.668–15.24 | 4 |
| 15.24–22.86 | 3 |
| 22.86–30.48 | 2 |
| >30.48 | 1 |
| Recharge | Rating |
| Factor: aquifer recharge (R) | |
| 0–2 | 1 |
| 2–4 | 3 |
| 4–6 | 6 |
| 6–8 | 8 |
| 8–10 | 9 |
| >10 | 10 |
| Aquifer media | Rating |
| Factor: aquifer media (A) | |
| Sand and gravel | 6 |

**Table 4** Ratings for soil media and topographic factors

| Factor: soil media | | | Factor: topography | |
|---|---|---|---|---|
| Soil type | Symbol | Rating | Slope (%) | Rating |
| Fine sand | FS | 10 | 0–1 | 10 |
| Gravelly loamy fine sand | GRLFS | 10 | 1–3 | 9 |
| Loamy fine sand | LFS | 9 | 3–6 | 7 |
| Sandy loam | SL | 8 | 6–10 | 5 |
| Fine sandy loam | FSL | 7 | 10–15 | 4 |
| Very fine sandy loam | VFSL | 6 | 15–20 | 2 |
| Loam | L | 5 | >20 | 1 |
| Sandy clay loam | SCL | 4 | | |
| Silty loam | SIL | 3 | | |
| Silty clay loam | SICL | 2 | | |
| Clay loam | CL | 2 | | |
| Silty clay | SIC | 1 | | |
| Clay | C | 1 | | |

$$C \rightarrow_x D \quad \text{or} \quad C \rightarrow D \tag{4}$$

the support of the decision rule $C \rightarrow$ D for $x$ is obtained as the absolute value of the intersection of the conditional set with the decision set and stated as:

$$\text{supp}_x(C, D) = |C(x) \cap D(x)| \tag{5}$$

where $C(x)$ and $D(x)$ are knowledge about the conditional and decision variables contained by a specific object, $x$, in the database and are termed as the granules induced by $x$ in rough sets terminology (Pawlak

**Table 5** Ratings for the impacts of vadose zone and conductivity

| Factor: impact of vadose zone | | Factor: conductivity | |
|---|---|---|---|
| Soil organic matter (%) | Rating | Conductivity | Rating |
| 0.00–0.25 | 10 | Somewhat excessive (SE) | 10 |
| 0.25–0.50 | 9 | Well (W) | 8 |
| 0.50–0.75 | 8 | Moderately well (MW) | 6 |
| 0.75–1.25 | 6 | Somewhat poor (SP) | 4 |
| 1.25–1.50 | 5 | Poor (P) | 2 |
| 1.50–2.00 | 4 | Very poor (VP) | 1 |
| 2.00–2.50 | 3 | | |
| >2.50 | 1 | | |

2005). The operation, $|x|$, implies cardinality of $x$. The strength of the decision rule $C \rightarrow$ D is denoted as $\sigma_x(C, D)$ and is given by:

$$\sigma_x(C, D) = \mathrm{supp}_x(C, D)/|U| \tag{6}$$

A certainty factor ($\mathrm{cer}_x(C, D)$) can be associated with each decision rule $C \rightarrow$ D and is defined as follows:

$$\mathrm{cer}_x(C, D) = \mathrm{supp}_x(C, D)/|C(x)| = \sigma_x(C, D)/\pi(C(x)) \tag{7}$$

where from Eq. 6

$$\pi(C(x)) = |C(x)|/|U| \tag{8}$$

The certainty factor can be interpreted as the probability of obtaining a decision, $D$, given a set of conditions, $C$, or the certainty with which $C$ implies $D$. If $\mathrm{cer}_x(C, D) = 1$ then rule $C \rightarrow$ D is called a certain decision rule. If $0 < \mathrm{cer}_x(C, D) < 1$ then the decision rule $C \rightarrow$ D is referred to as an uncertain decision rule. An approximate decision rule denoted by $C \Rightarrow D$ can be stated if $\mathrm{cer}_x(C, D) > 0.5$ and can be interpreted as $C$ mostly implies $D$.

In addition to support and certainty relations, a coverage factor of the decision rule can also be computed to give explanations for various decisions. Mathematically, the coverage factor, $\mathrm{cov}_x(C, D)$, can be expressed as:

$$\mathrm{cov_x(C, D) = supp_x(C, D)/|D(x)| = \sigma_x(C, D)/\pi(D(x))} \tag{9}$$

where,

$$\pi(D(x)) = |D(x)|/|U| \tag{10}$$

if $C \rightarrow$ D is the decision rule then $D \rightarrow$ C is called the inverse decision rule. The coverage factor can be interpreted as the probability of having a set of conditions, $C$, given a decision, $D$, and as such is the certainty factor of the inverse decision rule.

## Results and discussion

### DRASTIC map for the study area

The final DRASTIC vulnerability index map for the study area is depicted in Fig. 8. The values for the vulnerability index ranged from 46 to 188 with a mean value of roughly 90 and a standard deviation of about 20. The vulnerability index was re-classified into three categories namely high, medium and low. As DRASTIC is a relative ranking scheme, the medium vulnerability was defined as mean ± 0.5 standard deviation and ranged from 80 to 100. Areas having values lower than the lower bound of the medium range (i.e., <80) were classified as low vulnerability regions and those areas with values greater than the upper bound of the medium range (i.e., >100) were classified as high vulnerability regions. As can be seen from Fig. 8, Aransas County along the coast is comprised of medium and high vulnerability areas and McMullen County in the southwestern section of the study area is mainly comprised of low and moderately vulnerable regions. All other counties have a combination of low, medium and high vulnerability zones and it is rather difficult to cluster counties exhibiting similar vulnerability characteristics via simple visual inspection alone.

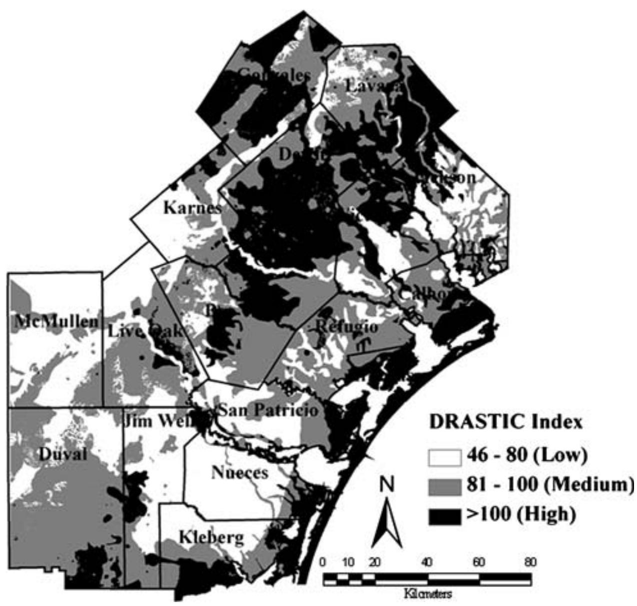


**Fig. 8** DRASTIC map for the study area

## Application of rough set methodologies

A data table containing the number of equal-sized pixels having low, medium and high vulnerability rankings in each county was generated by sequentially masking the generated vulnerability map (Fig. 8) with the administrative boundary maps of the individual counties in the study area. As per Eq. 5, these values represent the support for a particular decision (i.e., high, medium or low vulnerability) for different counties in the study area. The strength, certainty and coverage factors for low, medium and high rankings in each county were computed using Eqs. 6–9 (Pawlak 2005) and are presented in Table 6. Filtering and other Boolean operations in MS-EXCEL® were then used to ascertain approximate decision rules.

The sum of all the strength factors over all rankings and counties is equal to unity and this factor can be used to identify the relative proportions of low, medium and high permeability regions in the study area, which were noted to be 31.6, 40.6 and 27.8%, respectively. Figure 9 presents a more detailed breakup of different vulnerabilities for coastal and hinterland counties. As can be seen from Fig. 8, the sum of high and low vulnerability regions in hinterland counties account for 36% of the total area while moderate vulnerability regions in the hinterland counties account for nearly 30% of the total study area. By the same token high and low vulnerability regions in the coastal counties account for 23% of the study area and medium vulnerability regions in the coastal counties account for 12% of the study area. Therefore, the aquifer vulnerability is more heterogeneous in coastal counties than in hinterland counties. The coastal counties in the study area are where the majority of the population resides and many of these counties are experiencing significant population growth and contain relatively limited amounts of groundwater for various

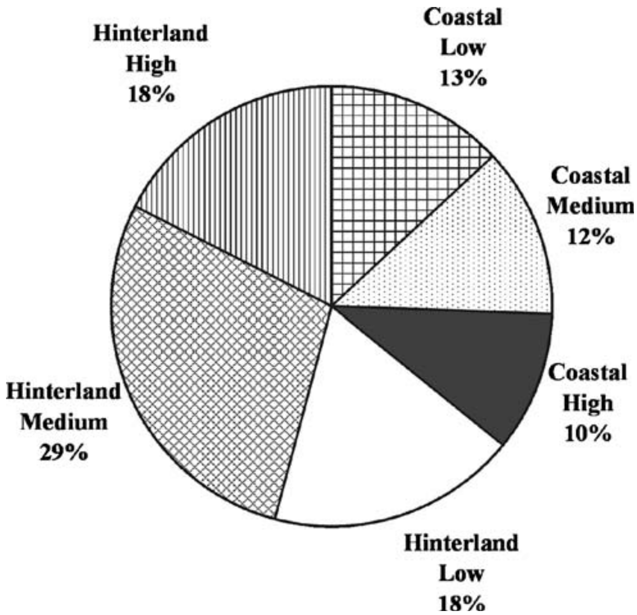


**Fig. 9** Fraction of total area for different geographic regions and their vulnerability

**Table 6** Computed strength, certainty and coverage factors for counties in the study area

| County | Support[a] | | | Strength | | | Certainty | | | Coverage | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Low | Medium | High | Low | Medium | High | Low | Medium | High | Low | Medium | High |
| Aransas[b] | 0 | 108,430 | 674,795 | 0.000 | 0.002 | 0.015 | 0.000 | 0.138 | 0.862 | 0.000 | 0.006 | 0.053 |
| Bee | 523,464 | 1,581,353 | 427,495 | 0.011 | 0.035 | 0.009 | 0.207 | 0.624 | 0.169 | 0.036 | 0.085 | 0.034 |
| Calhoun[b] | 85,899 | 680,945 | 718,388 | 0.002 | 0.015 | 0.016 | 0.058 | 0.458 | 0.484 | 0.006 | 0.037 | 0.056 |
| DeWitt | 131,002 | 786,631 | 1,702,121 | 0.003 | 0.017 | 0.037 | 0.050 | 0.300 | 0.650 | 0.009 | 0.042 | 0.134 |
| Duval | 1,024,572 | 3,594,966 | 459,853 | 0.022 | 0.079 | 0.010 | 0.202 | 0.708 | 0.091 | 0.071 | 0.193 | 0.036 |
| Goliad | 156,922 | 712,058 | 1,603,555 | 0.003 | 0.016 | 0.035 | 0.063 | 0.288 | 0.649 | 0.011 | 0.038 | 0.126 |
| Gonzales | 307,372 | 1,095,975 | 1,671,666 | 0.007 | 0.024 | 0.037 | 0.100 | 0.356 | 0.544 | 0.021 | 0.059 | 0.131 |
| Jackson[b] | 1,003,153 | 771,105 | 607,032 | 0.022 | 0.017 | 0.013 | 0.421 | 0.324 | 0.255 | 0.069 | 0.041 | 0.048 |
| Jim Wells | 1,412,859 | 689,274 | 385,114 | 0.031 | 0.015 | 0.008 | 0.568 | 0.277 | 0.155 | 0.098 | 0.037 | 0.030 |
| Karnes | 1,120,093 | 769,698 | 269,740 | 0.024 | 0.017 | 0.006 | 0.519 | 0.356 | 0.125 | 0.078 | 0.041 | 0.021 |
| Kleberg[b] | 894,222 | 725,695 | 911,764 | 0.020 | 0.016 | 0.020 | 0.353 | 0.287 | 0.360 | 0.062 | 0.039 | 0.072 |
| Lavaca | 418,852 | 1,110,270 | 1,262,562 | 0.009 | 0.024 | 0.028 | 0.150 | 0.398 | 0.452 | 0.029 | 0.060 | 0.099 |
| Live Oak | 1,161,152 | 1,593,703 | 245,132 | 0.025 | 0.035 | 0.005 | 0.387 | 0.531 | 0.082 | 0.080 | 0.086 | 0.019 |
| McMullen | 2,181,447 | 974,247 | 2,888 | 0.048 | 0.021 | 0.000 | 0.691 | 0.308 | 0.001 | 0.151 | 0.052 | 0.000 |
| Nueces[b] | 1,794,447 | 256,772 | 381,061 | 0.039 | 0.006 | 0.008 | 0.738 | 0.106 | 0.157 | 0.124 | 0.014 | 0.030 |
| Refugio[b] | 495,979 | 1,519,675 | 219,649 | 0.011 | 0.033 | 0.005 | 0.222 | 0.680 | 0.098 | 0.034 | 0.082 | 0.017 |
| SanPatricio[b] | 1,031,213 | 758,208 | 195,494 | 0.023 | 0.017 | 0.004 | 0.520 | 0.382 | 0.098 | 0.071 | 0.041 | 0.015 |
| Victoria[b] | 702,280 | 872,347 | 980,818 | 0.015 | 0.019 | 0.021 | 0.275 | 0.341 | 0.384 | 0.049 | 0.047 | 0.077 |

[a] Number of pixels

[b] Coastal counties

uses. Therefore, understanding vulnerability characteristics is of prime importance for future land development in the coastal areas of South Texas.

The sum of the coverage factors for all counties for any given vulnerability classification is equal to unity. Therefore, coverage factors are useful to identify the relative fractions of a specific vulnerability classification found within a particular county. For example, about 15% of all the low vulnerability regions can be found in McMullen County. Similarly, nearly 20% of all the medium vulnerability regions are present in Duval County and a little over 13% of all the regions classified as high vulnerability lie in DeWitt County. The eight coastal counties in the study area account for 41.6, 30.6 and 36.9% of all the low, medium and high vulnerability regions, respectively.

The coverage factor was also used to discover other insights present in the generated dataset. For examples, the seven counties—Duval, Jim Wells, Karnes, Live Oak, McMullen, Nueces and San Patricio account for nearly 90% of all the low vulnerability regions in the 18 county study area. On the other hand, the seven counties—Aransas, Calhoun, DeWitt, Goliad, Gonzales, Lavaca and Victoria account for nearly 67% of all the high vulnerability regions in the 18 county area indicating that regions of low vulnerability are more closely clustered along the southern and western sections of the study area and high vulnerability regions mostly occur along the northern and eastern sections of the study area.

The certainty factors for a specific county sum to unity. Thus, certainty factors indicate the relative fractions of high, medium and low vulnerability regions within a given county of interest. For example, 86.2% of the Aransas County is covered by high vulnerability regions and 13.8% is covered by moderately vulnerable regions. Also, as certainty factors are also the probability of obtaining a decision given a set of conditions, it can be interpreted as the likelihood of observing a particular vulnerability state and can be used to cluster counties in accordance with their most likely vulnerability characteristics. As certainty factors are seldom absolute, the approximate decision rule has been espoused by Pawlak (2005) for ranking purposes. Using Pawlak's approximate decision rule, the 18 counties can be clustered into following sets: high vulnerability = [Gonzales, Goliad, DeWitt, Aransas], medium vulnerability = [Live Oak, Bee, Refugio, Duval], low vulnerability = [Karnes, San Patricio, McMullen, Jim Wells, Nueces], and indiscernible = [Kleberg, Calhoun, Lavaca, Victoria, Jackson].

For counties in the indiscernible set, a single dominant vulnerability state cannot be isolated with a fair amount of certainty (probability >0.5 in this case). In some planning situations, indiscernible sets can pose challenges to decision making and a crisp classification of vulnerability states may be preferred. The most likely vulnerability states can be identified for counties categorized into the indiscernible set, albeit, with a certainty less than 50%. Therefore, when a crisp classification is preferred, the counties in the discernible sets can be classified as per their maximal vulnerability state. The addition of a fuzzy prefix to the vulnerability state as depicted in Fig. 10 can be useful to communicate the vagueness associated with the elements in the indiscernible set.

## Summary and conclusions

The assessment and classification of aquifer vulnerability in different administrative units (e.g., counties) of a region is often carried out to guide land use and water resources planning endeavors. The use of the DRASTIC approach with GIS for visualizing aquifer vulnerability characteristics has become a standard practice with the easy availability of the required data in digital format. The present study is based on the premise that the utility of digital vulnerability mapping is further enhanced when the generated data is effectively mined for insights and information. Rough set theoretic approaches provide classificatory algorithms to extract knowledge embedded in databases. The rough-set based mathematical formalisms recently put

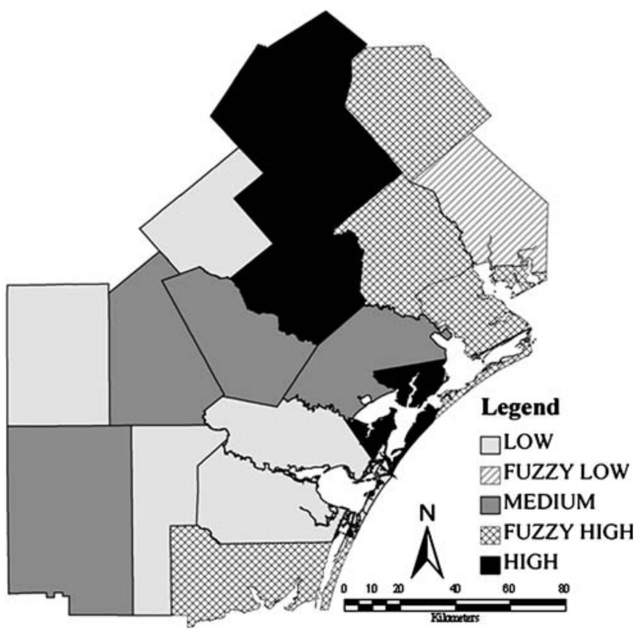


**Fig. 10** Rough-sets based aquifer vulnerability classification of counties

forth by Pawlak (2005) to perform conflict analysis were adapted in this study to understand regional-scale aquifer vulnerability and cluster counties in a region according to the perceived risks to groundwater resources based on hydrogeologic characteristics. The rough set theory provides three useful metrics: the strength factor depicts how vulnerability characteristics occur over the entire study area; the certainty factor computes the relative probabilities for various vulnerability states within a county and can be used to classify counties into different clusters based on the most likely vulnerability state; the coverage factor elucidates the variability of the vulnerability among different counties across a specific vulnerability state.

The integrated GIS–rough sets approach was used to rank 18 different counties in South Texas according to their vulnerability characteristics and obtain an understanding of potential areas of concern. A variety of insights were extracted from the database using rough sets. The results indicate that there is greater variability associated with aquifer vulnerability estimates along the coast than in the interior sections of the study area. The low vulnerability regions are mostly centered in the southwestern sections of the domain while high vulnerability areas are likely to be found in the northern and eastern sections. The certainty factors and approximate decision rules were used to classify counties according to their likely vulnerability characteristics. However, the approximate decision rule did not lead to a crisp classification and could not discern the dominant vulnerability state in five counties. An approach to overcome such indiscerniblity by re-grouping these counties according to their certainty factors, albeit, in an approximate sense has also been presented. Rough sets were noted to provide useful mathematical constructs to mine GIS derived datasets for aquifer vulnerability and as such are deemed useful for regional land and water resources planning.

**Acknowledgments** Financial support from the National Science Foundation—Combined Research and Curriculum Development (CRCD) program (Award No: 0203482) is gratefully acknowledged.

## References


Administrative maps for Texas: TIGER maps by US Census Bureau (accessed on June 2004) http://www.arcdata.esri.com/data/tiger2000/tiger_download.cfm

Al-Admat RAN, Foster IDL, Baban SMJ (2003) Groundwater vulnerability and risk mapping for the Basaltic Aquifer of Azraq Basin of Jordan using GIS, remote sensing and DRASTIC. Appl Geogr 23:303–324

Aller L, Bennett T, Lehr JH, Petty RJ, Hackett G (1987) DRASTIC: a standardized system for evaluating ground water pollution potential using hydrogeologic settings. EPA/600/2–87/035, US Environmental Protection Agency, Washington, p 455

Al-Zabiet T (2002) Evaluation of aquifer vulnerability to contamination potential using the DRASTIC method. Environ Geol 43:203–208

An A, Shan N, Chan C, Cercone N, Ziarko W (1996) Discovering rules for water demand prediction: an enhanced rough sets approach. Eng Appl Artif Intel 9:645–653

Baker ET (1979) Stratigraphic and hydrogeological framework of part of the coastal plain of Texas. Texas department of water resources report 236, Austin, Texas, p 43

Cameron E, Pelosa GF (2001) An application of fuzzy logic to the assessment of aquifers' pollution potential. Environ Geol 40:1305–1315

Chanas S, Kutcha D (1992) Further remarks on the relation between fuzzy and rough sets. Fuzzy Set Syst 47:391–394

Earth Resources Observation and Science (EROS) Data center, US Geological Survey (USGS). Digital Elevation Model (DEM) for topography (accessed on June 2004) http://www.edc.usgs.gov/

Edet AE (2004) Vulnerability evaluation of a coastal plain sand aquifer with a case example from Calabar, Southeastern Nigeria. Environ Geol 45:1062–1070

Evans BM, Myers WL (1990) A GIS-based approach to evaluating regional groundwater pollution potential using DRASTIC. J Soil Water Conserv 45:242–245

Fredrick KC, Becker MW, Flewlling DM, Silavisesrith W, Hart ER (2004) Enhancement of aquifer vulnerability indexing using the analytic-element method. Environ Geol 45:1054–1061

Komorowski J, Pawlak Z, Polkkowski L, Skowron A (1999) Rough sets: a tutorial Pal SK and Skowron A (eds) Springer, Singapore

National Climatic Data Center (NCDC), National Oceanic and Atmospheric Administration (NOAA), Historical precipitation records (accessed on June 2004) http://www.ncdc.noaa.gov/oa/climate/stationlocator.html

Natural Resources Conservation Service (NRCS), US Department of Agriculture (USDA). State Soil Geographic (STATSGO) database (accessed on June 2004) http://www.ncgc.nrcs.usda.gov/products/datasets/statsgo/

Pawlak Z (1982) Rough sets. Int J Inform Comput Sci 11:341–356

Pawlak Z (2005) Some remarks on conflict analysis. Eur J Oper Res 166:649–654

Shen Q, Chouchoulas A (2001) FuREAP: A fuzzy-rough estimator of algae populations. Artif Intell Eng 15:13–24

Tan RR (2005) Rule-based life cycle impact assessment using modified rough set induction methodology. Environ Model Softw 20:509–513

Texas Groundwater Database 2000 retrieved from Texas Water Development Board (TWDB) (Accessed on June 2004) http://www.twdb.state.tx.us/

Williams JR, Kissel DE (1991) Water percolation: An indicator of nitrogen-leaching potential Follett RF. Keeney DR, Cruse RM (eds) Soil Science Society of America, WI