Fuzzy Expert System for Site Characterization

Zhiying Hu[1], Christine W. Chan[2], Gordon H. Huang[3]
[1] Faculty of Engineering, University of Regina.
Regina, SK, Canada
e-mail : huzhiy11@uregina.ca
[2] Department of Computer Science / Energy Informatics Laboratory,
University of Regina.
Regina, SK, Canada
e-mail : Christine.Chan@uregina.ca
[3] Faculty of Engineering, University of Regina.
Regina, SK, Canada
e-mail : Gordon.huang@uregina.ca

Abstract: Remediation Selection Expert System (RSES) is a rule-based expert system which is used for the selection of remediation techniques for petroleum contaminated sites. In this paper, we describe a fuzzy logic–based sub-system: Site Characterization Sub-System (SCSS). It is an enhancement of the RSES, which is used to analyze the hydraulic properties of contaminated sites. This paper focuses on an explanation on how to apply fuzzy set theory for identification of soil types and hydraulic properties of a contaminated site. To illustrate application of fuzzy set theory to the problem, two sample cases are presented in detail.

1. Introduction

Contamination of soil and groundwater due to subsurface leakage of petroleum products from storage tanks has become an important environmental concern during the past three decades. Numerous site remediation techniques for eliminating petroleum contamination have been developed. However the techniques vary greatly in their characteristics in terms of the kind of contamination they address, hydrological conditions under which they work, and a variety of associated physical, chemical, and biological process. The decision on the choice of the remediation methods given different site characteristics is different and decision-making presumes expertise on both remediation technologies and site hydrological conditions. The expertise is scarce; and automated support for this decision making process by means of an expert system called the Remediation Selection Expert System (RSES) has been developed [Geng et al., 2001].

This paper presents a sub-system of RSES: the Site Characterization Sub-System (SCSS). The SCSS is an enhancement to the RSES. The task of the SCSS is to analyze and determine the hydraulic properties of a contaminated site and provide them as input into RSES, which then selects the appropriate remediation techniques.

2. Site Characterization Sub-System (SCSS)

The SCSS is necessary because site characterization is a difficult process. While site hydraulic characteristics are necessary input to RSES for the selection of remediation techniques, it is often not easy for the user to specify that information on the entire contaminated site. To provide the crucial input, the user needs to deduce the soil texture, permeability and other site hydraulic properties from the sampling data using a complex but comprehensive procedure. In this procedure, many subjective and uncertain conceptions or preferences are involved and different users will characterize a given site differently. To assist the user in decision-making, the SCSS has been built to deal with the uncertainty inherent in the procedure of analysis of the available sampling data. The user can input some basic soil sampling data and the SCSS will analyze the data based on a rule base provided by the domain expert and output the site characteristics to RSES. The approach taken in building the knowledge base is to use fuzzy set theory.

2.1 Structure of SCSS

According to opinions of domain experts and references on soil toxicology, definition of hydraulic characteristic of a contaminated site can proceed in two steps. The first involves determining hydraulic characteristics of each sample taken from the site in question. These characteristics include permeability and isotropism. The second step involves drawing conclusions on the hydraulic characteristics of the entire site derived empirically based on results from step 1.

The input data include information on the location, moisture and particle size distribution of each soil sample, and the physical characteristics of all the vertical soil layers above each sample. The site hydraulic characteristics produced by SCSS include the following:

(a) The degree of permeability of the entire site defined as high, medium, or low;
(b) The heterogeneity of the site defined as homogeneous or heterogeneous;
(c) The direction property of the site defined as isotropic or anisotropic; and
(d) The overall hydraulic property of the site defined as simple or complex.

2.2 Methodology

Fuzzy Soil Classification: Soil classification is an important step in defining hydraulic conditions of a site. Soil type determines the site's porosity, which, together with its moisture content, determines permeability of a soil sample.

(a) Theory of soil type classification

The method adopted in SCSS for classifying soil type was developed by the U.S. Department of Agriculture (USDA). The sizes of individual particles were used as criteria for soil classification. There are four basic soil types:

Gravel: >2.0 mm in diameter

Sand: 0.05 – 2.0 mm in diameter

Silt: 0.002 – 0.05 mm in diameter

Clay: < 0.002 mm in diameter

Distribution of particle size determines soil type; specifically, different volume percentages of gravel, sand, silt and clay determines the texture of a particular soil sample. A diagrammatic representation of the classification concept is shown in Figure.1. When the specimen includes a large amount of gravel, a correction will be necessary (Das, 1985).

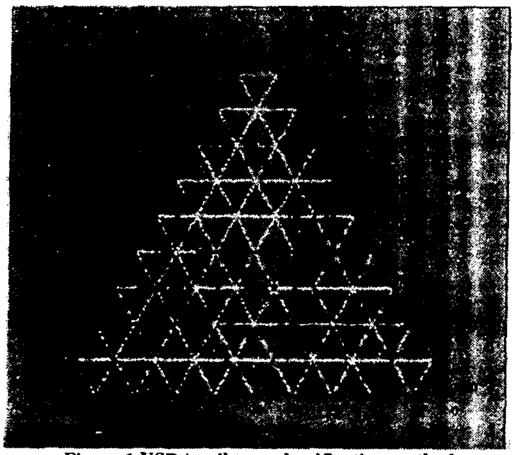

Figure.1 USDA soil type classification method

(b) Fuzzy Soil Type Classification

In this study, fuzzy set theory is adopted to convert the triangular chart of Figure.1 into continuous classes. An assumption is also made that each soil type has a representative set of values for percentages of the three components (sand, silt and clay). Thus, for soil type j, assume that its representative percentage values are D_j % of sand, T_j % of silt and C_j % of clay.

The membership values of a sample depends on their proximity to the strictly defined class centroids or means. They also depend on the membership functions used and the interval between the individuals and the centroids of the classes. If the data points in a sample has the same volume percentages of each component D_j, T_j and C_j, then this sample can be classified as a typical soil j, with a membership value of $\mu_j = 1$. The data points can also have one or more than one of its composition different from that of the typical soil j but located within an acceptable range R_j, in this case, the soil sample type can still be classified as type j, with a membership value lower than1. Here, R_j is three dimensional, with the three axis representing volume percentages of sand, silt and clay.

Each range is defined as a triangular fuzzy interval. The middle numeral of each section is the mean of the fuzzy interval. The tolerance interval of each fuzzy set R_{ij} (i = {1, 2, 3} = {sand, silt, clay}) is calculated based on the upper and lower bounds of each particle's soil type section. Table.1 lists all twelve soil types and their corresponding mean values and tolerance intervals in SCSS.

Table.1 Representative value and tolerance interval of each soil type

Soil type j	Sand(D_j)	R_{dj}	Silt(T_j)	R_{tj}	Clay(C_j)	R_{cj}
sand	93	7	8	8	5	5
loamy sand	80	10	15	15	8	8
Sandy loam	61	19	25	25	10	10
loam	38	14	39	11	18	10
silt	10	10	90	10	6	6
silt loam	25	25	25	25	14	14
sandy clay loam	62	18	19	19	29	9
clay loam	33	13	34	18	35	7
silty clay loam	10	10	56	16	35	7
sandy clay	54	10	10	10	46	10
silty clay	10	10	50	10	50	10
clay	22	22	20	20	77	23

The membership functions are defined as follows:

$$\mu_{1j} = \begin{cases} 1 & D_s = D_j \\ 1 - \left|\dfrac{D_s - D_j}{R_{Dj}}\right| & D_j - R_{Dj} \le D_s \le D_j + R_{Dj} \\ 0 & D_s \le D_j - R_{Dj} \text{ OR } D_s \ge D_j + R_{Dj} \end{cases}$$

$$\mu_{2j} = \begin{cases} 1 & T_s = T_j \\ 1 - \left|\dfrac{T_s - T_j}{R_{Tj}}\right| & T_j - R_{Tj} \le T_s \le T_j + R_{Tj} \\ 0 & T_s \le T_j - R_{Tj} \text{ OR } T_s \ge T_j + R_{Tj} \end{cases}$$

$$\mu_{j} = \begin{cases} 1 & C_s = C_j \\ 1 - \left|\dfrac{C_s - C_j}{R_{Cj}}\right| & C_j - R_{Cj} \le C_s \le C_j + R_{Cj} \\ 0 & C_s \le C_j - R_{Cj} \text{ OR } C_s \ge C_j + R_{Cj} \end{cases}$$

Thus, for a soil sample containing three kinds of particles, the soil type can be analyzed through a ternary fuzzy relation, according to the above binary method.

Let D, T, C $\subseteq \underset{\sim}{R}$ be universal sets, we have:

$$\underset{\sim}{R} = \left\{ \left((d,t,c), \mu_R(d,t,c) \right) \middle| (d,t,c) \subseteq D \times T \times C \right\}$$

Let $\underset{\sim}{D} = \{ \mu_{1j} | j = 1,2,\ldots 12 \}$, $\underset{\sim}{T} = \{ \mu_{2j} | j = 1,2,\ldots 12 \}$ and $\underset{\sim}{C} = \{ \mu_{3j} | j = 1,2,\ldots 12 \}$ be three 12-dimensional fuzzy vectors that represent membership functions of three particles belonging to each soil type. Then fuzzy relation b_j, which represents the membership grade for soil type j, can be obtained as follows:

$$\underset{\sim}{B} = \underset{\sim}{D} \circ \underset{\sim}{T} \circ \underset{\sim}{C} = \{b_1, b_2, \ldots, b_{12}\}$$

By using max-min composition, we have:

$$b_j = \sum_{i=1}^{m} \left(\mu_{1j} \cap \mu_{2j} \cap \mu_{3j} \right) =$$

$$\max\{\min(\mu_{11}, \mu_{21}, \mu_{31}), \min(\mu_{12}, \mu_{22}, \mu_{32}), \ldots, \min(\mu_{112}, \mu_{212}, \mu_{312})\}$$

where μ_{ij} is the grade of membership for the fuzzy relation between volume percentage of particle i and soil type j.

Site Hydraulic Properties: The second step in SCSS is to derive the texture and hydraulic property of the entire site based on characteristics of the random samples obtained.

(a) Site hydraulic conductivity

The average values for intrinsic hydraulic conductivity of each type of saturated soil are listed in Table.2.

Table.2 Hydraulic conductivity of saturated soil

Soil type	sand	loamy sand	Sandy loam	loam	silt	silt loam
Ks (cm/day)	713	350.2	106	25	6	11
Soil type	sandy clay loam	clay loam	silty clay loam	sandy clay	silty clay	clay
Ks	31.44	10.8	31.44	1.68	0.5	4.8

In SCSS, an experimental function is used to simplify the relationship between the hydraulic conductivity and moisture of unsaturated soil and those of saturated soil. This helps provide hydraulic conductivity of each sample.

The second step of deriving hydraulic property of the entire site from the random samples involves use of statistical and weighting methods. Letting n_j be defined as number of samples that belong to soil texture j (j=1, 2, ...12), the total number of random samples is then $\sum_{j=1,2,...12} n_j$. Considering the fraction level of each type j ($n_j/\sum n_j$) as the weight or influence of soil type on the entire site, the following function can then be used to evaluate the corresponding characteristic of the entire site:

$$f_{whole} = \sum_{j=1}^{12} f_j \times W_j$$

where:

f_{whole} = characteristic of entire site,

f_j = characteristic of soil type j,

W_j = weight of type j, $W_j = n_j \Big/ \sum_{i=1}^{12} n_i$.

Thus, the horizontal hydraulic conductivity of the entire site can be calculated as follows:

$$K_{x-whole} = \sum_{j=1}^{12} K_j \times W_j$$

(b) Site heterogeneity

Another site hydraulic condition considered in SCSS is site heterogeneity which is denoted as x and can be classified into three classes: homogeneous, heterogeneous, and extremely heterogeneous. Correspondingly, a fuzzy set of this hydraulic property is defined as follows:

μ_x = {μ_1/ homogeneous, μ_2/ heterogeneous, μ_3/ extremely heterogeneous}

The membership functions of μ_x are defined as the function of max (W_j) where j=1, 2, ...12, (Figure.2). If the weight of any soil type j is larger or equal to 0.8, the entire site is considered homogeneous, with $\mu_1 = 1$, and $\mu_2 = \mu_3 = 0$.

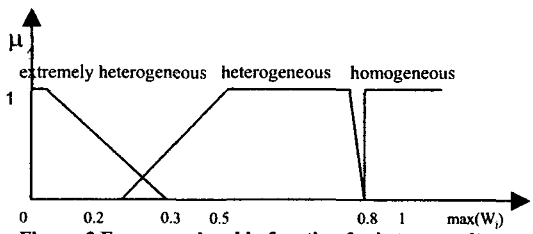

Figure. 2 Fuzzy membership function for heterogeneity

(c) Isotropism

SCSS only considers the two dimensions of horizontal-x and vertical-z directions. The vertical hydraulic conductivity of a sample can be calculated as follows (David, 1983):

$$K_{zi} = H_i \Big/ \sum (h_l / k_l)$$

where

H_i = the depth of the sampling point i,

h_l = the thickness of each layer of soil above the sample,

K_l = the estimated hydraulic conductivity in layer l above a sample.

According to the domain expert, the thresholds to categorize istropism is defined as follows:

When $K_{xi} / K_{zi} < 2$, sample *i* is isotropic. Otherwise it is anisotropic. If more than 20% of all samples are anisotropic then the entire site is considered anisotropic.

(d) Permeability

The horizontal hydraulic conductivity of the entire site described in section (b) is used here as the indicator of site permeability. Permeability denoted as z is classified into four fuzzy linguistic levels of high, medium, low, and very low permeability.

Correspondingly, the fuzzy set is defined as follows:

μ_z = {μ_1/ high, μ_2/ medium, μ_3/ low, μ_4/ very low}

The domain expert has provided the fuzzy membership function of each linguistic term.

(e) Site complexity

Hydraulic complexity of a site is determined by properties of heterogeneity, isotropism and permeability as described in sections (5.b) to (5.d). The fuzzy rules related to these variables are represented in the following forms:

If heterogeneity of a site is denoted as X_x with μ_x, isotropism denoted as Y_y with μ_y, and permeability denoted as Zz with μ_z, then the site complexity is denoted as V_v with $\mu_v = \min(\mu_x, \mu_y, \mu_z)$.

The complexity of this site is given by V_v with the largest membership value being μ_v, where v = {1, 2} = {simple, complex}.

The thirteen decision rules in this fuzzy rule base obtained from domain experts are detailed in Table.3. This table forms the basis of the SCSS. Based on these rules, SCSS can determine hydraulic property of a given site, which then becomes inputs for the expert system for remediation selection.

Table. 3 Decision rules

X	Y	Z(permeability)	V
Homogeneous	isotropic	high	simple
Homogeneous	isotropic	medium	simple
Homogeneous	isotropic	low	simple
Heterogeneous	isotropic	high	simple
Heterogeneous	isotropic	medium	simple
Heterogeneous	isotropic	low	simple
Extremely heterogeneous	anisotropic	high	simple
Homogeneous	isotropic	very low	complex
Extremely heterogeneous	isotropic	high	complex
Extremely heterogeneous	isotropic	medium	complex
Heterogeneous	anisotropic	high	complex
Heterogeneous	anisotropic	medium	complex
Heterogeneous	anisotropic	low	complex

4. Case Study

Two cases were examined for validation of SCSS. Here we provide a trace of the data as they are processed through the system in each case.

4.1 Case one

The data of the first case were obtained from the Site Remediation Laboratory at University of Regina. Different soil particles were mixed and put into a 3 x 1 x 1.5 m^3 cuboids container to simulate a unitary clayey subsurface soil structure. This set of data was used for validating SCSS

The particle size distribution of all sampling points are provided in Table.4.

Table. 4 Sampling data of case one

Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Gravel %	17	12	13	15	16	11	14	12	17	11	16	15	13	12	7	13
Clay %	45	56	66	48	72	47	42	58	55	47	66	29	59	44	58	49
Sand %	20	23	18	36	12	38	40	4	21	31	12	39	21	18	17	12
Silt %	18	9	3	1	0	4	4	26	7	11	6	17	7	26	18	26
Depth m	7	2	3	5	6	1	4	2	7	1	6	5	3	2	7	3
Above soil layer structure:																
Number of layers	2	1	2	2	2	1	1	2	2	1	2	2	2	2	1	2
Texture of layer 1	Sandy clay	Clay	Silty clay	Clay	Silty clay	clay	Silty clay	Sandy clay	Silty clay	Sandy clay	Silty clay	Silty clay	Silty clay	Sandy clay	Sandy clay	Silty clay
Thickness	4	2	0.5	3	3	1	4	1	5	1	2	1	1	1	5	2
Texture of layer 2	Silty clay		Clay	Si lt	clay			Clay	Clay loa m		clay	clay	Clay	Silt	Clay loam	silt
Thickness	2		2.5	2	3			3	2		4	3.5	1	1	2	1

Based on the input, SCSS concludes that the hydraulic characteristic of this site is "Simple", with properties of "homogeneous, isotropic, and low permeability". This conclusion coincides with the site characteristics that were designed at the laboratory.

4.2 Case two

The second case is a real world site located in southern Saskatchewan of Canada. The site was contaminated by spilled gasoline from a perforated underground storage tank in the past 5 years. The seepage leaked into the soil and has not reached groundwater. Thus, the contaminant-concentrated area was formed in the unsaturated zone. The area of the plume is between 2600 and 2900 m^2, and the volume is between 29000 and 30000 m^3.

Benzene, Toluene, E-benzene and Xylenes are the main components of gasoline known to be potentially toxic. The monitored mean concentrations are: benzene at 24 mg/kg, toluene at 14 mg/kg, e-benzene at 12 mg/kg, and xylenes at 78 mg/kg. The existing contaminants are in residual and free phases.

The Saskatchewan provincial standards for the four contaminants are: Benzene: 5 mg/kg, Toluene: 30 mg/kg, E-benzene: 30 mg/kg, and Xylenes: 30 mg/kg

Twenty samples were randomly collected from the site. Based on the input, SCSS decided that the overall site hydraulic property is complex with the detailed properties being heterogeneous, anisotropic, and high permeability. These inputs were provided to RSES, which provided the following recommended remediation alternatives:

(a) In-situ enhanced soil extraction
(b) In-situ bioremediation

The concentrations of contaminants at this site are all higher than acceptable levels. The contaminated volume is large. Therefore, ex-situ techniques are not applicable. The outputs of SCSS indicate that the site is heterogeneous, anisotropic, and highly permeable, and the hydraulic property is relatively complex. These coincide with the recommendation of domain experts.

One of the remediation alternatives from RSES is in-situ enhanced soil extraction. This integrated technique consists of shallow soil mixing, thermally enhanced vapor extraction, and soil vapor extraction. In general, the in-situ enhanced soil extraction technology can remove more than 70% of pollutants in the soil. This technology is especially suitable for this kind of contaminated sites with large volumes and high depths.

In-situ soil bioremediation has also been recommended by RSES. The four kinds of contaminants are biodegradable under aerobic conditions. In-situ soil bioremediation involves a mechanism for stimulating and maintaining intrinsic degradation processes where indigenous microbes convert contaminants to innocuous end products via electron acceptor and/or inorganic nutrient amendment. This is a cost-effective method. Among the recommendations suggested by the RSES, the site owner selected in-situ soil bioremediation for practical implementation.

5. Conclusions

A rule based fuzzy expert system named Site Characterization Sub-System (SCSS) has been developed as an enhancement for the existing Remediation Selection Expert System (RSES). Fuzzy set theory provides an improved extension of Boolean Logic for supporting definition of uncertainty in the SCSS. Methods of fuzzy knowledge representation and fuzzy reasoning were employed to convert uncertain system inputs to fuzzy linguistic information.

The results from two case studies indicate that SCSS can effectively process the input data and clarify the hydraulic characteristics needed for RSES. The system provides decision support to users in definition of site hydraulic characteristics, which are crucial inputs for identification of suitable site remediation alternatives.

References

[1] B. M. Alex. and O.A.O. Inakwu "Application of fuzzy sets in soil science; fuzzy logic, fuzzy measurements and fuzzy decisions", *Geoderma*, Vol. 77,1997.

[2] B. M. Das. *Principles of geotechnical engineering*, PWS publishing company, 1997.

[3] J. A. Soesilo and S. R. Wilson, *Site remediation planning and management*, Lewis publishe, Inc,1997.

[4] L. Geng, Z. Chen, C. W. Chan and G. H. Huang, "An intelligent decision support system for management of petroleum contaminated sites", *Expert Systems with Applications*, Vol. 20, pp. 251-260, 2001.

[5] D. K. Todd, *Groundwater hydrology*, John Wiley & Sons, 1983.