USE OF GEOLOGIC INFORMATION IN SITE CHARACTERIZATION

Tien H. Wu[1], Mohamed A. Abdel-Latif[2], Michael A. Nuhfer[3], and B. Brandon Curry[4]

INTRODUCTION

Adequate site characterization should provide the engineer with the necessary information to carry out the design. For many geotechnical problems, geologic details that are important to the safety of a structure cannot be identified with certainty. This was described by Terzaghi (1929) and the fundamental uncertainty in the characterization of such details remain valid today, although advances in site exploration techniques offer advantages that were not available in 1929. Probability methods may be used to assess the uncertainties associated with site characterization and to draw inferences from observed data (Baecher, 1972, 1978). Site characterization should start with the choice of an appropriate geologic model, based on geologic observations. Given a geologic model, probability methods can be used to evaluate available information and estimate the uncertainties associated with various elements of site characterization. This paper describes such a problem.

Geologic information is difficult to quantify because observational data are often inprecise and incomplete. One approach is to reconstruct the sedimentary environment and use the geometry of sedimentary units observed in modern environments or in exposures as a basis for interpretation. We have chosen the continental glacial environment and developed a probabilistic site characterization model, whereby geologic information including professional judgement can be combined with data from site investigation. The site characterization model consists of the following components: (1) a geologic model that summarizes observations on relevant geologic features, (2) a probability model that estimates the likelihood of certain events or design scenarios, given the geologic observations, (3) an encoding procedure that derives a subjective probability, which reflects the professional judgement of engineers and geologists, and (4) an updating model that combines subjective and objective probabilities and that revises the probabilities as information from site investigation becomes available. The

1, 2 Professor Emeritus and Research Associate, respectively. the Ohio State University. Columbus, OH.

3 Project Engineer, Metcalf and Eddy, Inc., Columbus, OH.

4 Associate Geologist, Illinois State geological Survey, Champaign, IL.

site characterization model can be used to estimate the most likely value of a design parameter or event and the uncertainty about that estimate. Such estimates could be used with decision models in planning site investigations, evaluating the potential benefits of additional investigations, and in comparing design alternatives.

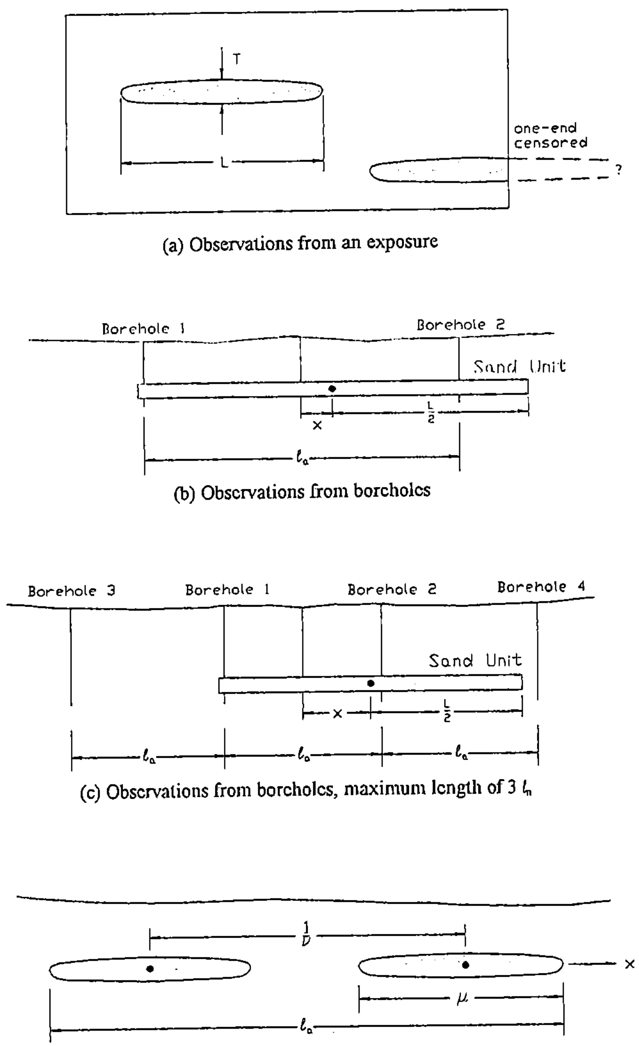
The geologic setting used here consists of continental glacial sediments of Illinois. We applied the conceptual models, (1) through (4) above, to a site for projected waste disposal. The issue of concern is leachate migration from the disposal site via groundwater flow, primarily along layers of granular sediments within relatively impermeable, fine-grained tills. An important issue is the continuity of the sand and gravel units. A case history of the site investigation and characterization of hydrogeology for such a problem was described by Curry et al (1994) for a site at Martinsville, Illinois. We applied our site characterization model to the site of the Fermi National Accelerator Laboratory (Fermilab.), near Fermilab, IL, as an example. This site was chosen because of: (1) the similarity between the geology at Fermi Lab. and at Martinsville, and (2) the large amount of data produced by the extensive site investigation at Fermilab.

The importance of heterogeneity in geologic materials on groundwater flow is well known. The problem addressed here is the continuity or interconnectedness between stratigraphic units. Studies by Wu et al, 1973; Fogg, 1986; Johnson and Dreiss, 1987; Jussel, 1994; etc. have shown the importance of continuity on the flow pattern. In this paper we describe a site characterization model that addresses the question of continuity.

GEOLOGIC MODEL

A geologic model should provide the properties or characteristics of geologic features pertinent to the problem. For the sites under consideration, the most important feature is the pervious sand and gravel units, hereafter called sand units, within the fine-grained diamicton, which consists of poorly sorted clay, silt, sand and rock fragments. The geology of the Fermilab site was described in detail by Landon and Kempton (1971). The properties of the sand units are the dimensions: length (L) or width, thickness (T), and occurrence rate (ν). The geologic information we gathered consists of the following items: (1) length and thickness of sand units observed in exposures in glacial drift, (2) facies or members to which the sand units are related, and (3) occurrence of sand units in boreholes made at Martinsville and Fermilab.

For Item (1) we studied published diagrams and photographs of exposures of glacial deposits at sites in northeastern United States and southeastern Canada (Table 1). The lengths (L) and thicknesses (T) of sand units were measured. Where a unit extends beyond the exposure, it is identified as having one end or both ends censored. The observed length, Fig. 1a, of a censored unit represents the minimum possible length for that unit. We note that L is merely the length observed in an exposure, which represents a random section through a sand unit, whose dimensions and shape are unknown. For the present problem, we consider L as maximum distance within which a sand unit has been



(a) Observations from an exposure

(b) Observations from boreholes

(c) Observations from boreholes, maximum length of 3 l_n

(d) Occurence and length of sand units

Fig. 1. Length and Thickness as Observed in Exposures and Boreholes

observed to be continuous. Since the data are collected from many different sites, a wide range of sedimentary environments are represented. We do not have sufficient information to differentiate between the various environments; and the data set is taken to represent the population of coarse sediments associated with glacial environments.

Table 1. Sites of Exposures Used in this Study.

Site	Reference
Wayne Co., Indiana	Gooding, GSA* (1961), 3:99-130
Southeastern Ohio	Goldthwaite et al, GSA (1981), 3: 409-432
Sundance, Manitoba	Dredge and Nielsen, GSA (1987), 3: 43-46.
Wedron, Illinois	Johnson et al, GSA (1987), 3:212-217.
Great Bend, Indiana	Bleuer and Shaver, GSA (1987), 3:343-348.
Hueston Woods, Ohio	Stewart and Miller, GSA (1987), 3:423-426.
Gowanda, New York	Calkin, GSA (1987), 5:107-112.
Nantucket, Mass.	Oldale, GSA (1987) 5:221-224.
Toronto, Ontario	Sharpe, GSA (1987) 5:339-344.
Yarmouth, Nova Scotia	Grant, GSA (1987) 5:427-432.

* GSA denotes Geological Society of America in References

Item (2) was derived from geologic evaluation of the materials. For the Martinsville and Fermilab sites, color, texture, and mineralogy were used to correlate the different till layers (Landon and Kempton, 1971; Curry et al, 1994). An example of the geologic profile is shown in Fig. 2. The sand and gravel layers have been interpreted as deposits in various glacial (ie. subglacial, englacial, ice-margin, and proglacial) and nonglacial environments. One may detect some general characteristics of some types of deposits. The thickest and most extensive sand and gravel deposits occur along boundaries between different till units, or above the bedrock. This is because such sands and gravels were deposited during the waning and waxing of glacial ice across the site. There are also many thinner sand lenses scattered within till layers, deposited during minor changes at the base of the glacier.

Information for Item (3) is illustrated in Fig. 1b, which shows a sand unit encountered in two boreholes. Another case is shown in Fig. 1c; a sand unit was observed in boreholes 1 and 2 but not in boreholes 3 and 4. Whether the sand unit is continuous between the boreholes is unknown. However, the identification of till layers and the depositional environment of the sand are known from (2), and may help the engineer or geologist to form an opinion about the continuity or length of the sand units.

PROBABILITY MODELS

Probability models are used to draw inference on the state of nature, given some observations. The probability that a certain state exist, given some data, is

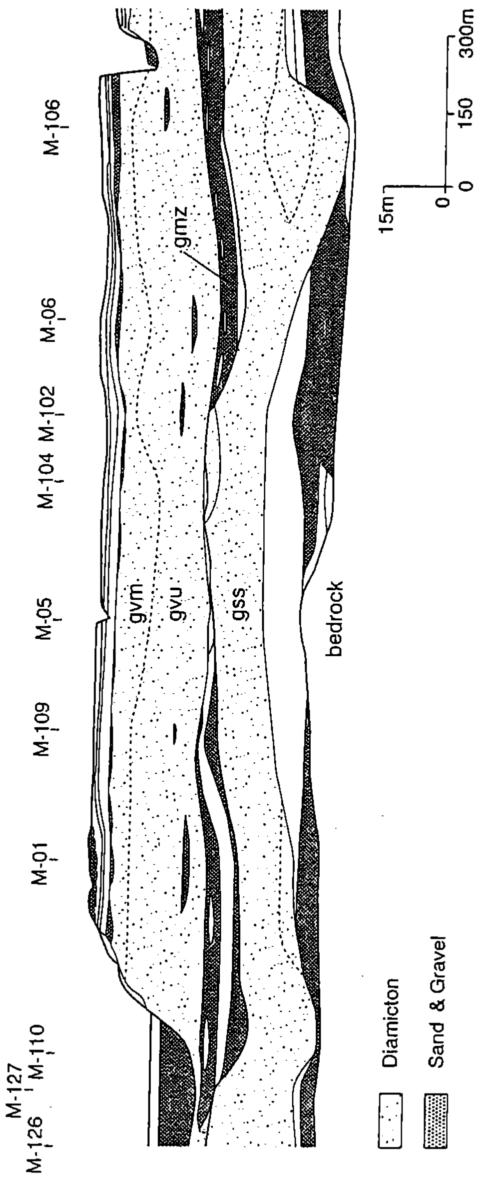


Fig. 2. Constructed profile for Martinsville site (from Curry et al, 1994).

$$P(\text{state} \mid \text{data}) = P[\theta \mid Z] \qquad [1]$$

where θ = state of nature or event and Z = observation or data. In probability, an event denotes occurrence of certain features, such as continuity of a sand unit between 2 points and data consist of observed lengths of sand units. Probability mass functions (pmf), $p(X)$, and probability density functions (pdf), $f(X)$, can be constructed for properties of features. Here, X = property, which may be frequency, number, dimension, etc., of a feature (e.g. rate of occurrence, ν, length, L, and thickness, T, of sand units). Data from exposures, borehole logs that show pervious layers, etc., may be used to construct pdf's of T and L which are $f(T)$ and $f(L)$, respectively. While the log-normal distribution has been used for $f(T)$ and $f(L)$ (Knutson et al, 1971; Wu et al, 1973), the exponential distribution has been favored from the view point of sedimentary processes (Krumbein and Dacey, 1969; Schwarzacher, 1972). Nuhfer (1996) has shown that the log-normal distribution gives a better fit to the data in Item (1).

Item (1) of the geologic model was used to estimate the distribution of mean (μ) and standard deviation (σ) of L, given an observed T, and the correlation between L and T. The data from exposures were separated into two groups: those from exposures with no censored data and those from exposures with censored data. For data from exposures with no censored data, values of L and T constitute the data points and are plotted in Fig.3. For the second group, Pahl (1981) gives equations of the estimated mean length, L^*, and $Var(L^*)$, of "traces" measured in an exposure:

$$L^* = w(1+c)/(1-c) \qquad [2a]$$

$$Var(L^*) = d\, n(n+1)^2 / [1+(1-p_2-p_0)n]^4 \qquad [2b]$$

$$c = (n_2-n_0)/(n+1) \qquad [2c]$$

$$d = 4w^2 [\, p_2+p_0-(p_2-p_0)^2\,] \qquad [2d]$$

where w = width of exposure, n = no. of traces observed, n_0 = no. of traces with both ends visible, n_2 = no. of traces with both ends extending beyond the exposure, $p_0 = n_0/n$, $p_2 = n_2/n$. However, a large number of observations from an exposure are required to obtain the ratios p_0 and p_2. In the present case, we have only a few measurements from each exposure. An approximate estimate was made by combing data from exposures within a range of widths and combining measured lengths for a range of thickness T. These groups were given in Table 2.There is an error from using the average w. The two sources of error are combined to give a coefficient of variation (COV) of L^*. The estimated mean and COV of L^* for each group are given in Table 2. The estimated mean L^* and the standard deviation are also shown in Fig.3.

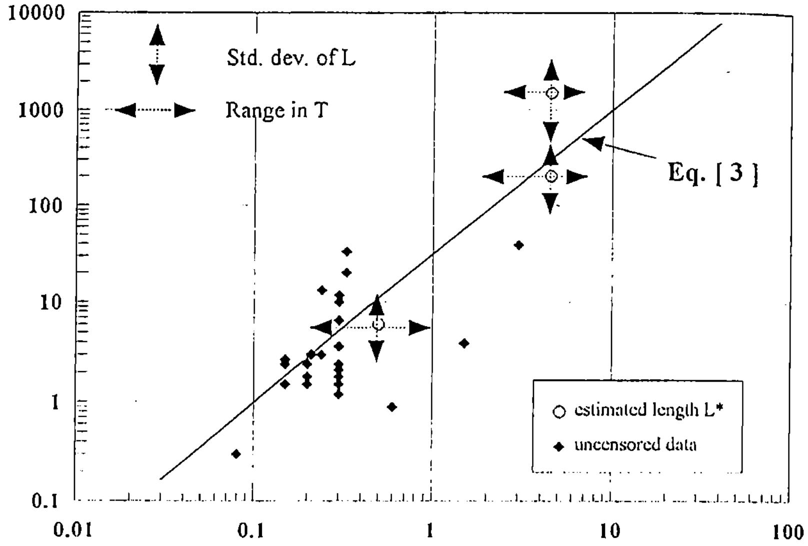


Fig. 3. Relation between Length and Thickness.

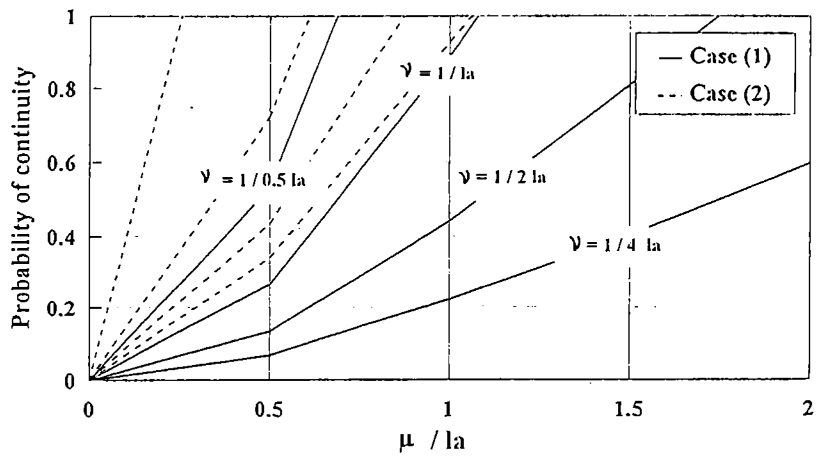


Fig. 4. Probability that a sand layer is continuous between two boreholes.

Table 2. Estimation of mean length from observations in exposures.

Sites	w (m)	n	n_0	n_2	T (m)	L*(m)	C.O.V
Sundance, S.E. Ohio, Fayette	30-50	18	14	0	0.3-1	6	0.43
Calkin, Nantucket, Wayne Co.	150-300	11	2	2	2 - 7	200	0.39
Yarmouth, Shelbyville, Wedron	500-700	6	1	4	2 - 6	1500	0.66

The censored data were combined with the uncensored data to obtain a regression relation. However, it is necessary to account for the difference in variance between the two data sets. Weights proportional to the variance of the data points were assigned to data points. For the censored data, the variance is equal to that given in Table 2. For the uncensored data, the variance is the error in the measurements of L and T. This is unknown and we assume the errors result in a COV of 0.1. The weighted least square procedure (Wetherill, 1986) was used and the regression coefficients β_0 and β_1 were obtained by minimizing the weighted variance. The correlation is

$$\log L = \beta_0 + \beta_1 (\log T) \qquad [3]$$

where β_0, β_1 =1.5 and 1.2, respectively. Eq.[3] is shown in Fig.3 and may be used to estimate L, given T, or L|T. The standard error of estimate of (log L) = 1. The large scatter in data is obvious, especially at T = 1 m and larger. The ratio L/T is between 30 and 100 for T between 1 and 10 m.

For the data in Item (3), illustrated in Fig. 1b and c, consider the occurrence of a sand unit along the X axis, at any given elevation, as a Poison process with mean rate of occurrence = ν, Fig. 1d. With a mean length μ, and the average distance between the midpoints as 1/ν, the fraction p_s of the length l_a occupied by the sand units is ν μ. Then,

$$\nu = p_s / (\mu) \qquad [4]$$

For estimation, p_s is taken to be the fraction of boreholes where the sand is observed.

The above information allows us to address the question: is the sand unit continuous between 2 boreholes in which the unit was found? The probability of a unit being continuous is P[θ|Z], Eq.[1], in which Z is the observation of a sand unit in boreholes, θ is the state of nature which is the presence of one continuous unit between two adjacent boreholes. Consider the case shown in Fig.1b. For one sand unit, the condition that makes it appear in boreholes 1 and 2 is shown in Fig.1b. The probability that this occurs is that a unit occurs with its center at X and its length is greater than

(l_a+2x). The probability of continuity is

$$P(C)=2\int_0^{\infty}P[L > (l_a+2X)]\ \nu dx=2\int_0^{\infty}\int_{L=l_a+2X}^{\infty} f(L)dL\ \ \nu dx \qquad [5]$$

For the case shown in Fig.1c, the length of the sand unit with its center at X, must be at least $(l_a + 2x)$ and can not exceed $(3l_a - 2x)$. The probability of this is

$$P(C)=\frac{P(CI)}{\int_0^{3l_a} f(L)dL \quad \nu l_a e^{-\nu l_a}} \qquad [6a]$$

$$\text{where} \quad P(CI)=2\int_0^{l_a/2}P[(l_a+2X)<L<(3l_a-2X)]\nu dx=2\int_0^{l_a/2}\int_{L=l_a+2X}^{(3l_a-2X)} f(L)dL\nu dx \qquad [6b]$$

It is assumed that f(L) has a Log-normal distribution. Computed probabilities for continuity of the sand units between adjacent boreholes are shown in Fig.4 for the two cases shown in Fig. 1b and c. Other cases have also been studied and the probabilities fall between the two shown, but are closer to those for Case 1.

Example: Consider the formation gm-z in Fig. 2. The thickness observed in boreholes M-06 and M-106 is 5.5 m. If one assumes that μ has a ratio of 100, the estimated mean length μ is 550 m. The sand is observed in 7 out 11 boreholes, so p_s = 0.67 and ν = 0.67/μ from Eq.[4]. The distance between the boreholes is 360 m, so μ/l_s = 1.5. From Fig. 4, P[C] = 1.0 and 1.0 for Cases 1 and 2 respectively.

SUBJECTIVE PROBABILITY

The subjective element has always played an important role in engineering decisions because it provides the incorporation of experience and judgement. Incorporation of expert opinion in site exploration strategy and site characterization have been done by Rehak et al (1985), Halim et al (1991) and Toll (1995). Subjective probability may be taken as a measure of the engineer or geologist's (expert's) best estimate, represented by the mean, and uncertainty, represented by the variance, about a design parameter. Experts' subjective probabilities were obtained for the following questions: The experts were given the geologic profile of the area west of Fermilab., and Fig. 5 and 6, which show the borehole logs along two profiles at the Fermilab site. Borehole pairs that encounter sand units and information about the sand layers are listed in Table 3. The experts were asked to express their opinion on the length of each sand

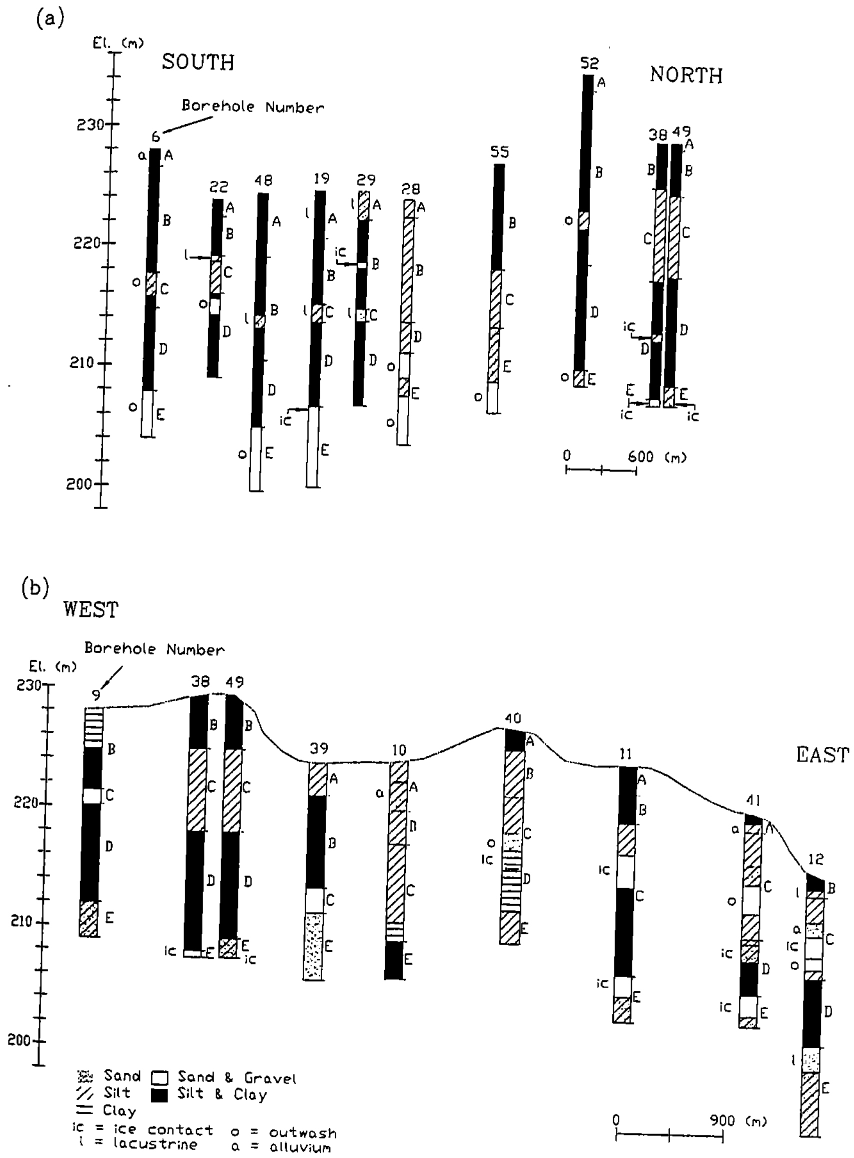


Fig. 5. Profile and borehole logs, Batavia Site (from Landon et al, 1971), (a) profile along A-A', (b) profile along D-D'.

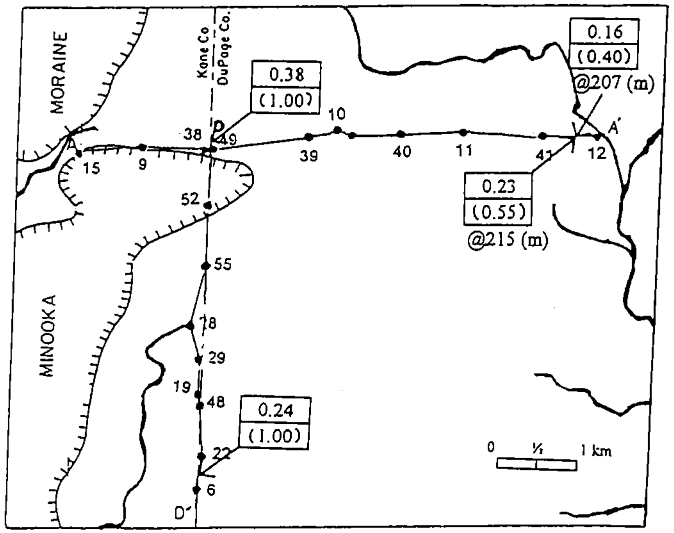


Fig. 6. Location of Boreholes (from Illinois State Geological Survey Circular 456)

Table 3. Answers to Questionaire.

Borehole Pair: Elevation: Thickness:	6 and 22 near 210 m 1.5, 1.4		38 and 49 near 207 m 0.6,1.7		11 and 41 near 210 m msl 2.4, 2.3		11 and 41 near 204 m msl 1.8, 0.9	
Expert No.	μ(m)	σ(m)	μ(m)	σ(m)	μ(m)	σ(m)	μ(m)	σ(m)
1	1033	649	1678	1592	2136	1733	1374	864
2	46	40	1676	821	1010	569	800	312
3	522	305	476	281	617	392	515	327
4	457	88	953	134	1600	711	1772	293
5	5	8	157	137	497	475	497	475
6	15	15	236	154	633	423	542	347
7	69	38	400	164	2134	449	876	309
8	686	220	-	-	1715	355	1562	355
9	525	29	210	25	446	196	446	196
Average	373	362	723	639	1199	701	931	509

unit. They were asked to estimate the range, the median and the quartiles of the length. Details of the encoding procedures are given in Nuhfer (1996). Note that Layers A - E in Fig.5 denote formations of different stages in glacial history (Landon and Kempton, 1971). This information comes from Item (3) of the geologic model, an item which is not considered in the probability model.

Table 3 summarizes the experts' answers to Question 2. Also shown is the mean of the experts' estimates and its standard deviation. We note that there are substantial differences between the lengths as estimated by the different experts. In addition, there are substantial differences between the lengths of the different sand units, as estimated by each expert, although the thicknesses of the sand units are not very different. The ratio L/T in the experts' estimates is larger than that shown in Fig. 3, except for the estimate by experts 6 and 7 on the length of the sand unit in boreholes 6 and 22. However, the experts' estimates fall within the range of data scatter shown in Fig. 3. The departure of the experts' estimate from the average in Fig. 3 may reflect their judgement based on the geology of the site. The experts were also asked to give the reasons for their estimates. An example is Expert No.7's reason for his large estimated lengths of sand units in boreholes 11 and 41 (Col. 5, Table 3). He considered the direction of drainage, which appears to be toward the west and his opinion is that sand units tend to have larger dimension in the direction of drainage. Expert 9's estimate of the lengths of units in boreholes 6-22, 38-49, and 11-41 elev. 200 was based on the fact that he identified these as proglacial deposits, which he believed would have greater lengths. On the other hand, his estimate of the length of the unit in boreholes 11-41, elev. 214 was based on the fact that he identified this as an ice marginal deposit. Other experts did not give definite reasons for their estimates.

COMBINING AND UPDATING INFORMATION

As shown in the preceeding sections, the pdf $f(L|T)$ may be obtained from statistics of observed lengths or from experts' opinion. The user is free to choose the source of information. Combining experts' opinions usually uses a weighted average

Bayesian updating may be used to combine objective and subjective probabilities. Updating gives a posterior probability

$$f''(\theta) = K\,\Gamma(Z|\theta)\,f'(\theta) \qquad [7]$$

where θ = parameter that represents the state of nature, such as L/T, Z = observed data, $\Gamma(Z|\theta)$.= likelihood function = probability of observing Z, given θ, and $f'(\theta)$ = prior probability = probability before data Z was obtained, K = normalization factor. One rationale is to use the statistical relationship in Fig. 3 as the prior $f'(\theta)$. This data apply to all glacial deposits in the region sampled. At a particular site, observations Z consists of the results of the site exploration, which are borehole logs and depositional environment. The experts' opinions after seeing the results of the site exploration constitute $\Gamma(Z|\theta)$. Further discussions with Expert 9 showed that he also considered T in his estimates of L; he expects the proglacial deposits to have a larger ratio (L/T) than ice-marginal deposits. From this we constructed f(L/T) which is log-normal with (λ_1 = 5.8, $\zeta_1 = 0.32$; λ_2 = 5.1, $\zeta_2 = 0.32$), where λ, ζ = mean and variance and subscript 1,2 denote proglacial and ice-marginal, respectively. Conversely, if given only (L/T), Expert No.9 would say the deposit is more likely proglacial than ice-marginal if (L/T) is large. This was used to establish $\Gamma(Z|\theta)$. The probability of Z = proglacial or ice marginal was assumed to be proportional to f(L/T). The prior of (L/T) was determined from Fig. 3. The posterior was found to have a log-normal distribution with means and variances (λ''_1 = 5.3, $\zeta''_1 = 0.37$; λ_2 = 4.9, $\zeta''_2 = 0.41$).

APPLICATIONS

We describe a possible applications to the site characterization problem described in the Introduction. The probability of continuity, P(C), was computed for several sand units at Martinsville and at Fermilab. A value of L/T = 100 was used and the results are given in Table 3. For Martinsville, the probability of 0.95 for Formation gm-z may be taken as high. Pumping tests from wells conducted at the site showed a drawdown radius of at least 600 m (Battelle and Hanson, 1990) which supports the high probability that the sand is continuous between the boreholes.

For the hypothetical problem of projected waste disposal at the Fermilab, consider the sand units found in Formation C as shown in the borehole logs (Fig. 5a and b). The probability of continuity between each pair of boreholes was first computed with the mean and variance of f(L) taken from the data in Fig.3. This represents the state where only limited information is available from boreholes. The values are indicated on the map, Fig.6. With detailed information on geology, expert's opinion could be incorporated into a posterior distribution of L/T as described in the preceding section. Next we used the posterior given in the preceding section to compute the probability of continuity. the results are shown in ( ) in Fig.6. These probabilities could be used to estimate the probabilities of seepage velocity and leachate migration. If the uncertainty is considered too great, additional site investigations should be made. Attention should then be directed to locations where the probabilities are closest to 0.5, which indicates the highest degree of uncertainty.

SUMMARY AND CONCLUSIONS

We have presented a site characterization model that accounts for geologic information, uses probability to draw inference from data, and can incorporate professional judgement. We have shown that statistical data on sand and gravel units in glacial sediments can be used for a first estimate in a site characterization problem. As more geologic information becomes available for a specific site, experts' opinions can be used to revise the estimate. The model contains assumptions that are admittedly simplifications of nature. However, we believe the model provides a viable starting point for planning site investigation. The simplifications can be improved or removed with more detailed studies.

Acknowledgment We wish to thank the following engineers/geologists for contributing their expertise to this study: M.R. McCord, D.W. Harrelson, G. D. McKenzie, R.O. Utgard, P.A. Beal, A.M. Woods, T. Aebie, R. Fernandez, J.A. Peeples, J. Montello, and J. Bates.

REFERENCES

Baecher, G.B. (1972) Site exploration: a probabilistic approach. Ph. D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.

Baecher, G.B. (1978) Analyzing exploration strategies. In Site characterization and exploration, C.H. Dowding, ed., American Soc. Civil Engrs., New York, NY.

Battelle Memorial Institute and Hanson Engineers (1990) Aquifer tests reports, Martensville alternative site. Clark Co., IL., Report to Illinois Dept. of Nuclear Safety.

Curry, B.B., Troost, K.G., and Berg, R.C. (1994) Quaternary geology of the Martinsville alternative site, Clark Co.,IL. Illinois State Geological Survey Circular 556, 85p.

Fogg, G. E. (1986) Groundwater flow and sand body interconnectedness in a thick multiple-aquifer system. Water Resources Research, 22 (5):679-694.

Geological Society of America (1961) Guidebook for Field Trips, Cincinnati meeting. New York.

Geological Society of America (1981) Guidebook for Field Trips, Cincinnati meeting. New York.

Geologic Society of America (1987) Centennial Field Guide, Vol. 3 and 5, Boulder, CO.

Halim, I.S., Tang, W.H., and Garrett, J.H. (1991) Knowledge-assisted interactive probabilistic site characterization. Geotechnical Congress, Geotechnical Spec. Publication 27, ASCE, 1:264-275.

Johnson, N.M., and Dreoss, S.J. (1989) Hydrostratigraphic interpretation using indicator geostatics. Water Resources Research, 25 (12):2501-2510.

Jussel, P., Stauffer, F., and Dracos, T. (1994) Transport modeling in heterogeneous aquifers: 1. Statistical description and numerical generation. Water Resources Research, 30 (6):1803-1817.

Kempton, J.P., Vaiden, R.C., Kolata, D.R., DuMontelle, P.B., Killey, M.M., and Bauer, R.A. (1985) Geological-geotechnical studies for siting the superconducting super