January 1993

# Interpretation of Drill Log Data: DLOG3D - A Probabilistic Tool for Analyzing Subsurface Soil Variability

Alaa El-Din Ali

Upmanu Lall

Utah Water Research Laboratory

# Interpretation of Drill Log Data: DLOG3D - A Probabilistic Tool for Analyzing Subsurface Soil Variability

Alaa El-Din Ali and Upmanu Lall

Technical Report TR-93-HWR-UL/001
Submitted to the Utah Division of Water Rights
November 1993

Utah Water Research Laboratory, Utah State University, Logan UT 84322-8200, USA
(801)-750-3155, FAX : (801)-750-3663

TABLE OF CONTENETS

LIST OF TABLES

LIST OF FIGURES

Figure Page

# Interpretation of Drill Log Data:

## DLOG3D - A Probabilistic Tool for Analyzing Subsurface Soil Variability

by

Alaa El-Din Ali and Upmanu Lall
Utah Water Research Laboratory, Utah State University, Logan UT 84322-8200

**Abstract**

Groundwater contamination potential is directly related to the location of zones of high hydraulic conductivity. Sedimentary depositional environments can be quite complex. Pockets of sand of varying thickness and orientation may be scattered over the aquifer. A contaminant pathway can exist where these pockets or lenses are interconnected . Aquifer pump tests do not provide useful information on the occurrence of such lenses. Drill log data directly sample the soil, but may represent very local sections of the aquifer, and are of relatively poor quality. Here we present a methodology (DLOG3D) that allows a probabilistic interpretation of drill log data that accounts for these uncertainties. DLOG3D output can be used to assess the likelihood that a zone of high conductivity exists in a location, and/or to map such zones in the vertical and the horizontal.

---

**Partial Support of this work was provided by the Utah Division of Water Rights under Contract no. 5-43346**

## 1) INTRODUCTION

Drinking water supplies in many areas are developed from deep aquifers since shallow aquifers are often contaminated. As local water demand increases, hydraulic gradients that were historically upward are reversed. Concerns about the likely contamination of the deep aquifers consequently arise. In alluvial sedimentary environments, shallow and deep aquifer systems are typically separated by discontinuous lenses or layers of markedly different hydraulic conductivity, geometry, and size (see Figure 1). Preferential pathways of varying size and orientation thus exist for the potential movement of contaminants through the heterogeneous aquifer system. Evaluation of such heterogeneities is a serious challenge.

The primary sources of information on subsurface hydraulic properties are pump tests and drill logs. Existing pump test methodologies are inappropriate for the identification of subsurface heterogeneities since they focus on the estimation of average hydraulic parameters, and the aquifer response to pumping is damped and smoothed over the discontinuities in hydraulic conductivity. Hence pumping tests can provide only limited information about the spatial variability of hydraulic conductivity. On the other hand, drill logs provide only local qualitative stratigraphic information. The traditional geologist's stratigraphic sections based on drill logs provide only a smoothed subjective interpretation of the subsurface. Extracting useful information from such data is clearly a challenge.

## 2) OBJECTIVE

The objective of this study is to use drill log information to develop a stochastic model of the aquifers' heterogeneity with emphasis on the following quantitative assessments:

1)The likelihood that a high hydraulic conductivity zone may exist at a certain depth in a general area.

2)The likely depth of the interface between high and low conductivity zones in a general area.

A product of this work is a contour map that describes the rate of occurrence of high hydraulic conductivity zones. Such a map could be used for selecting monitoring locations, designing pump tests, and providing an improved solution of the aquifer parameter identification problem.

## 3) BACKGROUND

The history of an alluvial sedimentary environment dates back hundreds, or may be thousands of years. The sedimentary depositional process is a function of the time during which it had formed. Some information can be revealed about such an environment by bore hole data. Each location reflects the conditions of the depositional and climatic environment at the time of deposition.

Meriam (1976) introduced two concepts to model the evolution of sedimentary deposits as a function of time. He first proposed a model in which a constant deposition rate (amount of sediment deposited/duration of deposition) is considered at any point in the basin during the history of deposition. The deposition amount during a period of time is assumed to be independent of the location of this point. For a large time scale (global scale), the sedimentation process rate changes due to climatic and topographic changes. Each change is called an event and is recorded by marker horizons. The thickness of sediment between two markers is called a bed. Clearly, the bed thickness is proportional to the recorded time and the rate of deposition. Each individual thickness is considered constant over the entire basin (Figure 2 a). Such a model is called a deterministic sedimentation model.

The second model extends these ideas. The sedimentary depositional process depends also on shorter time scale variations (i.e. local or seasonal variations related to the characteristics of the flow, sediment grain size distributions, topography,..etc). The flow that carries the sediment is affected by the seasonal climatic variability. Coarse grained sediments are likely to deposit more rapidly than finer ones. For a small amount of flow, most of the loaded coarse sediments will deposit before the flow reaches the basin. Topographic changes due to prior deposition also affect the posterior depositional process. The sequence of all of these changes is unpredictable. The situation described in the first model is now disturbed by unpredictable interruptions due to the short time scale variations in which a random change of the deposition (or erosion) rate occurs locally with time.(Figure 2b). The resulting depositional process is thus marked by random changes in space and time superimposed on the overall longer term trend.(Figure 2 a&b).

The bed depositional process occurs at the long time scale. Random deposits occur locally and form lenses with materials usually different from that of the bed. A coarse grained deposition

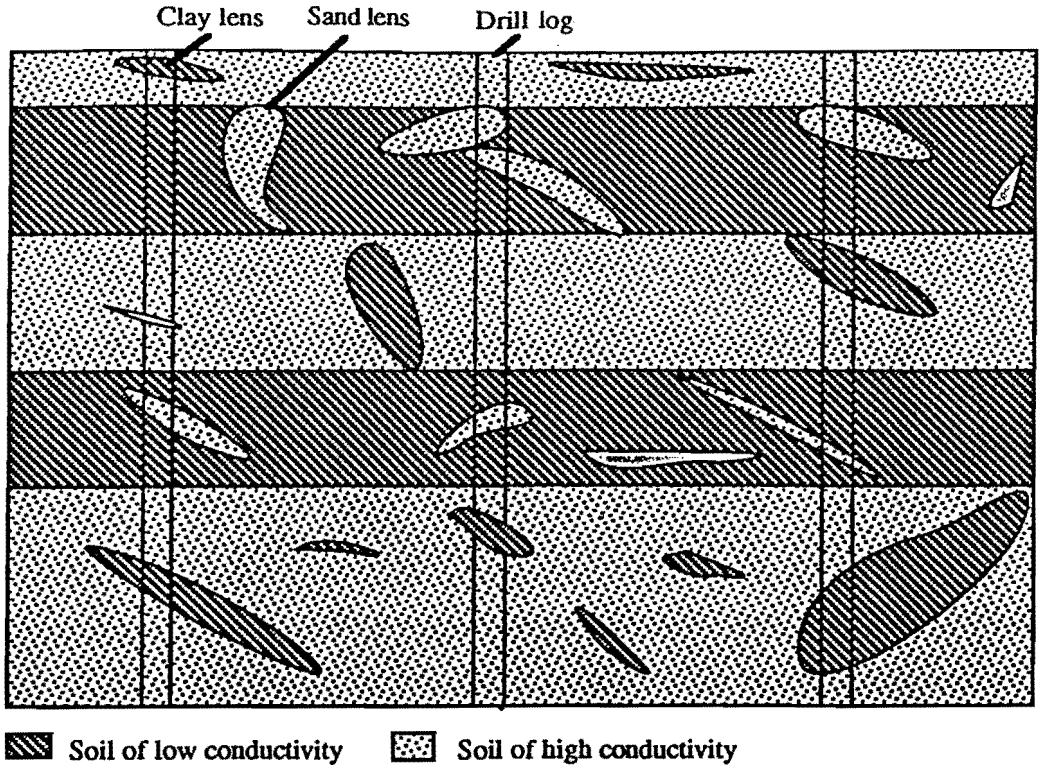



Figure 1. Soil heterogeneity in alluvial sedimentary environments.

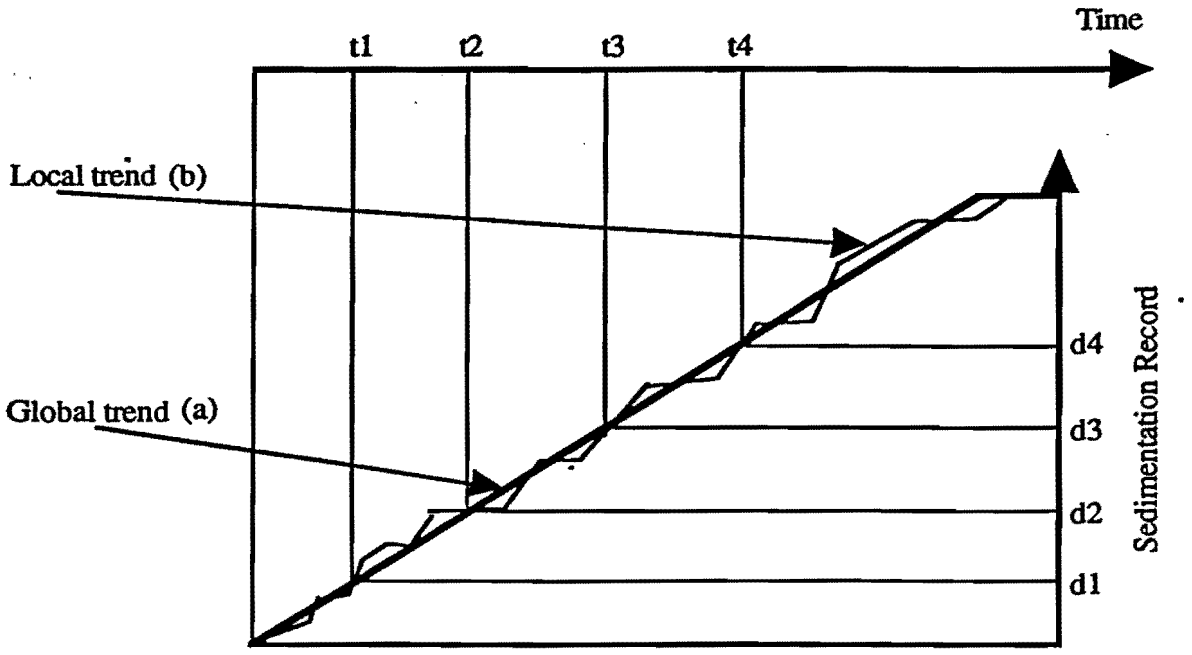



Figure 2. Long scale (a) and short scale (b) sedimentation models. (Following Meriam 1976).

may randomly occur during a clay depositional process forming coarse grain lenses and vice versa. The alluvial aquifer system is consequently characterized by lenses and layers of media of rather disparate hydraulic conductivity, and of variable size and geometry. The effective degree of hydraulic interconnection between the upper unconfined aquifer and the deeper confined aquifer depends directly on the lateral extent, location, and thickness of the high and low conductivity lenses in the strata loosely referred to as the aquitard.

The situation of interest is shown in Figure 1. The near surface unconfined aquifer, and the deep confined aquifer have hydraulic conductivities $K_u$ and $K_c$ respectively that are an order of magnitude higher than the hydraulic conductivity K' of the intervening aquitard. A binary classification of hydraulic conductivity (1 for $K_u$ or $K_c$ and 0 for K') can thus be adopted to reflect the potential for water flow and contaminant transport. The relevance of such a classification to the groundwater contamination problem of interest is seen through the following example. Figure 3 shows a vertical cross-section of the situation during the operation of a pumping well. High and low conductivity zones are distinguished by white and black nodes of a rectangular grid. Under the downward gradient, a contaminant particle migrates from the bottom of the unconfined aquifer downwards through the aquitard. It travels to the next downward plane of nodes by the shortest path. This means it travels to the nearest node of high conductivity at this plane. A possible migration or flow path is shown in Figure 3. Thus, identification of interconnected zones of high conductivity has direct bearing on the identification of contaminant migration paths and on the risk of contamination in an area.

The binary classification mentioned above is a quantitative description of the white/black node classification. Drill log data are usually rather imprecise. The driller may or may not adhere to the Unified Soil Classification System (USCS), and may not be diligent in recording the information. Even if the soil types are properly documented, direct inference of hydraulic conductivities from the drill log is infeasible. Johnson and Dreiss (1989) discuss two possible schemes based on interpretations of the USCS for the binary classification of drill log data into high and low conductivity regions (see Table 1). Interpretation (a) highlights layers of low conductivity and emphasizes the separation between the high conductivity layers, while interpretation (b) highlights the spatial structure of zones of high conductivity. Thus the choice of the binary classification scheme of material with hydraulic conductivity ranging on a continuum may emphasize different

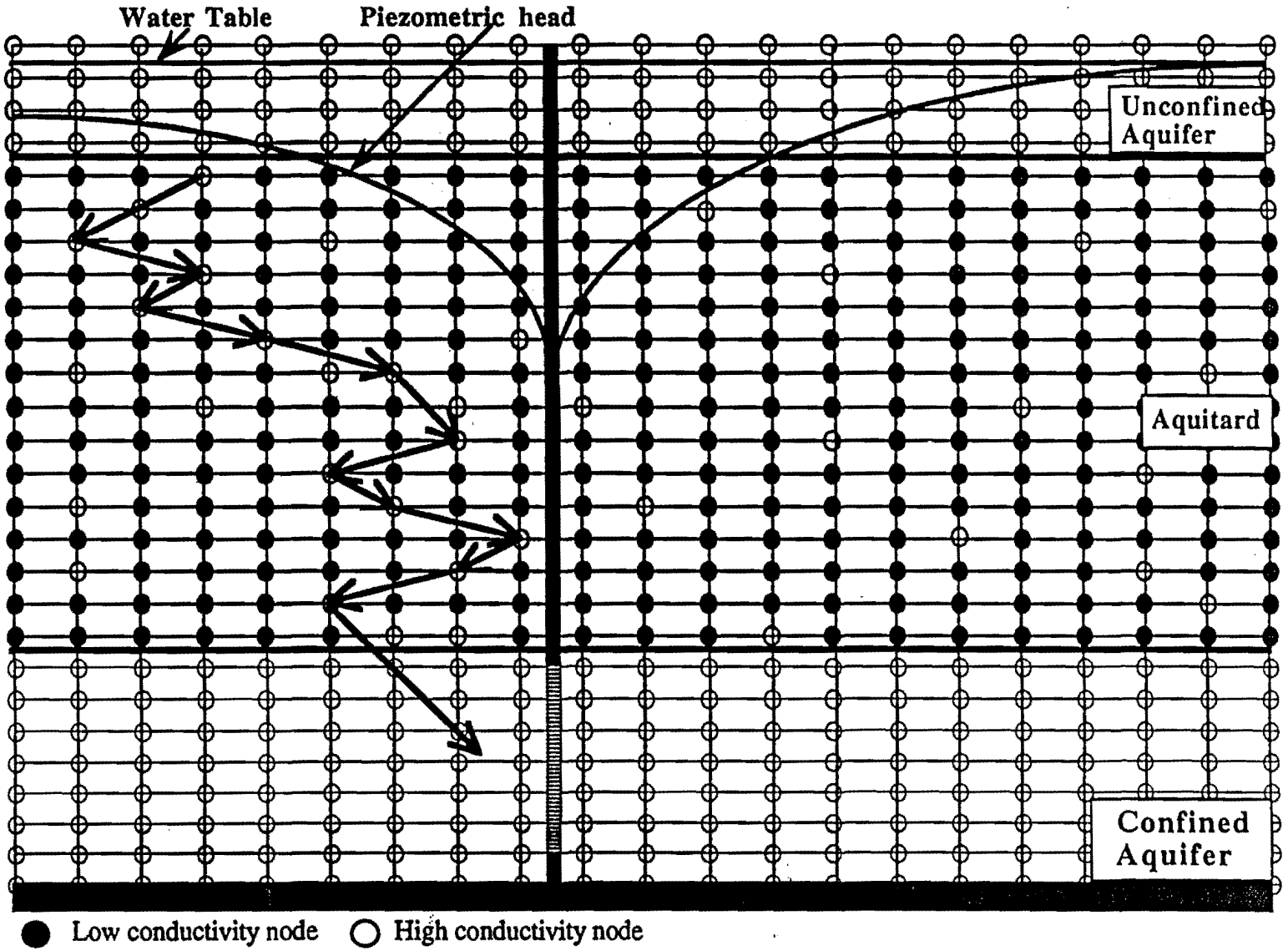



Figure 3. Hypothetical flow from node to node under the effect of the hydraulic gradient in the vicinity of a pumping well in a sedimentary deposites area.

attributes of the data. Johnson and Dreiss (1989) show that a subsurface with soil types as shown in Figure 4a may be interpreted as shown in Figure 4b according to these schemes.

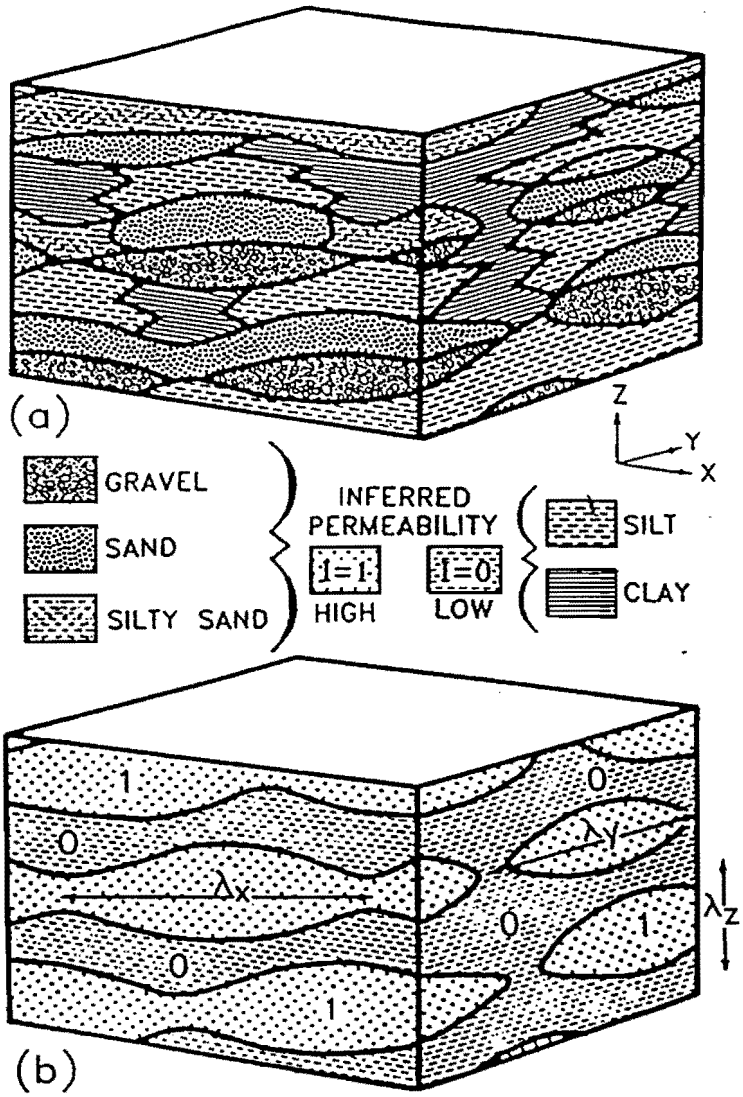
| High permeability<br>Indicator=1 | | Low permeability<br>Indicator=0 |
|---|---|---|
| | Interpretation (a) | |
| GW, SW, GM, SM<br>GP, SP, GC, SC | | ML, CL, OL<br>MH, CH, OH |
| | Interpretation (b) | |
| GW, SW<br>GP, SP | | GM, SM, ML, CL, OL<br>GC, SC, MH, CH, OH |

G, gravel; S, sand; M, silt; C, clay; O, organics; W, well graded, (i.e., poorly sorted); P, poorly graded; L/H, Low/High plasticity.

Table 1. Indicator Values for Alternative Interpretations of the United Soil Classification System (from Johnson and Dreiss, 1989).

The lesson from these observations is that even though a binary classification is adopted here, one has to be cognizant of the fact that the classification is somewhat subjective, transitions between the two states occur smoothly rather than sharply, and that statistically we should expect the lens occurrence process to be nonstationarity in the vertical and perhaps also in the horizontal. The drill hole typically samples the aquifer only to a few inches (cm) in lateral extent. The horizontal information provided is consequently very local and subject to a high degree of variability. Given that the spacing between drill logs may range from a few meters to hundreds of meters and the drill log data have abrupt rather than smooth transitions, one has to be very cautious in essentially extrapolating aquifer attributes between drill logs. Rather than interpolating the sand/clay horizons, it is more meaningful to interpolate the probability of sand at a particular location. This probability may be estimated by smoothing (locally averaging) the indicator function in the vertical over each drill log before we think about interpolation between the drill logs. Such smoothing can be accomplished using a moving weighted average of the indicator function over the vertical. This can be followed by an interpolation of these vertical averages between drill logs at desired horizons. A complete description of the method is provided in the methodology section.

Some basic definitions that are useful for understanding the model formulation are first presented.




and (b) the structure of these sediments when interpreted with indicator function. The length and width of a lens are given by $\lambda_x$ and $\lambda_y$ and the distance between the midpoints in low-permeability layers is $\lambda_z$ (from Johnson and Dreiss, 1989).

## 4) DEFINITIONS

**4.1)Random Variable (RV):** Assume we have drill log observations. Also assume that sand length observed within any arbitrary space in the drill log is called an outcome with value *y*. We call *y* a random variable since its values are determined to some extent by chance occurrences during sedimentation.

**4.2)Probability:** It is defined as relative frequency. If an experiment is conducted n different times and if event S occurs on $n_s$ of these trials, then the probability of event S is approximately:

$$P(S) \cong \frac{n_s}{n} \tag{4.1}$$

**4.3)Statistics of RV:** Let *ds* be a RV with n possible outcomes, then:

$$\text{Average or Mean} = \overline{ds} = \frac{1}{n}\sum_{i=1}^{n} ds_i \tag{4.2}$$

$$\text{Variance} = \hat{\sigma}^2_{ds} = \frac{1}{n-1}\sum_{i=1}^{n} (ds_i - \overline{ds})^2 \tag{4.3}$$

**4.4) Moving Weighted Average MWA:** Given data $y_i$ at location $x_i$

$$Wav_k = \sum_{i=1}^{n} \omega_{ik} y_i \tag{4.4}$$

Where $\omega_{ik}$ is the weight of the observation $y_i$ recorded at point i, for an estimate centered at point k.

Usually the weights sum to 1, i.e., $\sum_{i=1}^{n} \omega_{ik} = 1$.

**4.5)Kernel Estimator as a MWA:** To compute a local or MWA, we need to specify a weight sequence ("Kernel function"), and an averaging span ("bandwidth"). A kernel function prescribes the weights $\omega_{ik}$ used to estimate the MWA at point k using every data $y_i$ at point i that lies within the averaging span or "bandwidth". This is explained with the help of Figure 5. Consider a bandwidth $=h_L$ (e.g., 6) over which the MWA is to be computed, centered at point k with elevation $z_k$. Define a scaled distance to an observation located at i with elevation $z_i$ as $t_i=(z_i-z_k)/h_L$. Now consider a Kernel function defined as $K(t) = .75*(1-t^2)$. The weights $\omega_{ik}$ for the MWA at $z_k$ are calculated as follows: (Assume unit length is 1 foot and $z_k=10$)

1)Assume points $z_i$ (e.g., 6, 7, 9, 11, 12, 14) fall within the averaging span $h_L$ =6 (Note $z_k$=10).

2)Compute the corresponding scaled distances t (i.e., -.667, -.5, -.167, .167, .333, .667).

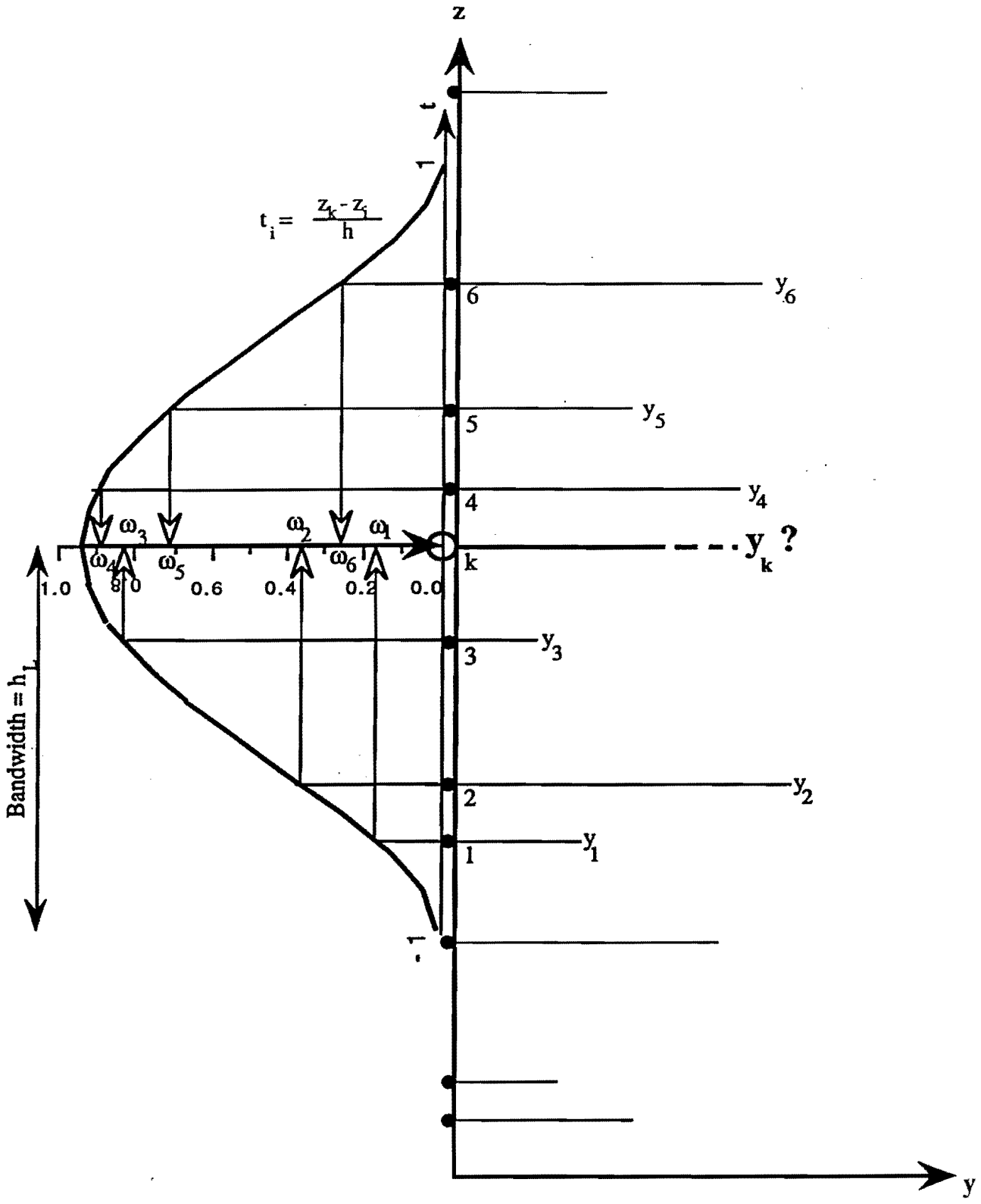



Figure 5. Estimating the weighted average at point k using Kernel Estimator.

3)Evaluate the kernel function $K(t_i)$ for each value of $t_i$ (i.e., .412, .563, .73, .73, .667, .412)

4)Compute the total sum of the values obtained in 3.($\Sigma K(t_i) = 3.514$).

4)The weights $\omega_{ik}$ are the normalized values of K(i) to ($\Sigma K(t_i) = 3.514$).(i.e., .117, .16, .208, .208, .19, .117). We notice that $\Sigma \omega_{ik}=1$.

say the values of $y_i$ corresponding to each point $z_i$ are (4, 6.5, 3.3, 2.1, 8.3, 6.8). The MWA at point $z_k$ would be (.117*4 + .16*6.5 + .208*3.3 + .208*2.1 + .19*8.3, .117*6.8 = 5). Note that if a Uniform Kernel was used, (i.e., all weights are equal to 1/6), the MWA would be ((4+6.5+3.3+2.1+8.3+6.8)/6 = 5.17. The above computations assume that Kernel function is symmetric about point $z_k$, however, this assumption is not valid if point $z_k$ is in a boundary region.

Kernel estimators usually suffer from boundary effects. Such effects arise because at the boundaries (near end points) we do not have a full complement of observations on both sides of the point of estimate. This leads to a biased estimate unless a special treatment is used.

The Kernel estimator concept can be generalized to higher dimensions.

**4.6)Types of Kernel functions:** In this paper, two types of Kernel functions are mentioned:

1)Epanechnikov Kernel: $K(t) = .75*(1-t^2)$ (4.5)

2)Bisquare Kernel: $K(t) = .9375*(1-t^2)^2$. (4.6)

For any of the two Kernels, the following properties should be satisfied:

$$\int_{-\infty}^{\infty} K(t)\,dt = 1 \qquad \int_{-\infty}^{\infty} t*K(t)\,dt = 0 \qquad \int_{-\infty}^{\infty} t^2*K(t)\,dt = \text{constant} \tag{4.7}$$

i.e., Area=1 Symmetry Finite variance

**4.7)Point Process:** Say the drill log length is divided into 1-foot intervals, and each interval is considered to be a point. The sand points are randomly located over the entire drill log. The random occurrence of sand points on the line is called a point process.

**4.8)Stationary Process:** If the values of the statistics (e.g., mean and variance) of a process are independent of location, the process is called stationary.

**4.9)Process Rate:** Assume a drill log divided into intervals of fixed length $dz$, and the observed length of sand within each interval is random variable $ds$. The ratio $\frac{ds}{dz}$ ($dz \to 0$) is also a random variable and called the process rate or the average rate of sand occurrence.

**4.10)Homogenous/Heterogeneous Process:** If the average process rate is constant over the drill log, the process is called homogenous. Conversely, if the average process rate varies systematically, the process is called heterogeneous.

**4.11)Poisson Process:** Assume that this drill log is divided into intervals of fixed length dz. Let N(z) the number of unit lengths of sand observed between (0,z). Let the event S denote that interval (z, z+dz) has sand and let the event C denote that it has clay. Let the probability of S be P(S)=p(dz), and that of C be P(C)=q(dz). Trivedi (1982, p 283-4) states that N(z) forms a Poisson Process if the following four conditions are satisfied:

i)N(0)=0. i.e. no sand can be observed in 0 length.

ii)Sand occurrences in nonoverlapping intervals of depth are mutually independent.

iii)Probabilities p(dz) and q(dz) depend only on the interval length dz and not on the location z for the Homogenous Poisson Process. They depend on both dz and z for the Nonhomogenous Poisson Process case. The rate $\lambda$ is constant for a Homogenous Poisson Process, *HPP*, and varies with location for the Nonhomogenous Poisson Process, *NPP*.

iv)For sufficiently small values of dz, i.e., dz is small enough to contain only one event of sand (e.g., one event of sand may be represented by 1-foot of sand occurrence). We can write:

$$P(s)=p(dz)=p[\text{one event of sand in the interval } (z,z+dz)] = \lambda dz+O(dz) \qquad (4.8)$$

and

$$P(s)=q(dz)=p(\text{no event of sand in the interval } (z,z+dz)] = 1-\lambda dz+O(dz) \qquad (4.9)$$

where: O(dz) represents terms that are an order of magnitude smaller than dz. (e.g., if dz=1 foot, O(dz)=.1 or .01 or may be less)

Note the direct relation between the probability of observing sand in an interval dz, and the rate $\lambda$ for that interval from Equation (4.8). To assure that only one event of sand occurs in z-z+dz interval, the sand length should be as close as possible to dz. If sand length $\cong$ dz =1, the function O(dz) $\cong$ 0. and the probability of sand occurrence P(S) $\cong \lambda$.

**4.12)Likelihood:** To illustrate the idea of the likelihood, assume that we flip a fair coin. The possible outcomes are a head (H) or a tail (T). The likelihood of getting a head is then 1/2. If we flip two fair coins, the possible outcomes (equally likely) are: 1)HH, 2)HT, 3)TH and 4)TT. This situation has 4 possible events, and getting two heads is one possible event. The likelihood of

getting two heads is then 1/4. Since the likelihood of the first head is 1/2 and the likelihood of the second head is also 1/2, we can directly calculate the likelihood of getting both in the same time by multiplying 1/2*1/2=1/4. Generally, the joint likelihood of a number of independent events occurring jointly is the product of their probabilities of occurrence.

**4.13)Cross Validation:** This is the process of judging the performance of an estimation scheme, where an observation is dropped and the function of interest (e.g. the rate) is estimated at that point using the remaining data. The estimation performance is usually judged by considering the sum of squares of differences between the observed (but deleted) value and its estimate using the remaining data.

**4.14)Mean Square Error MSE:** The difference between the true and the estimated function at a point is called the error at this point. The average of the errors squared for all data points is called the Mean Square Error (MSE)

# 5) METHODOLOGY

## 5.1) Introduction

As noted earlier, the sedimentary depositional process can be viewed as a series with rate changes in time. Figure 1 shows a picture of a hypothetical alluvial sedimentary environment. Note the following features:

1)Sand lenses are randomly scattered over the subsurface profile.

2)The drill logs do not capture all such lenses. This implies that we need to estimate the location and thickness of the uncaptured lenses.

3)The environment is typically heterogeneous. This means that the rate of sand occurrence varies as a function of location. Consequently, such a process is nonstationary.

A goal of the model presented is to quantitatively assess the likelihood that a sand lens exists in a certain neighborhood. The model considers the sediment deposition process as Nonhomogenous Poisson Process (NPP) in the vertical. The *NPP* is taken to be heterogeneous in the horizontal as well, to allow for localized variation in the deposition process. This model structure and its utility are illustrated by example below. Consider the following two situations (Figures 6a and b). Figure 6a shows two hypothetical drill logs, d1 and d2, of the same depth, several hundreds of meters apart. Note that each drill log has two sand layers separated by a clay layer. The layer thicknesses vary between d1 and d2. For an averaging interval completely contained within a sand layer, the average process rate $\lambda$ is clearly 1, and similarly it is 0 if one is strictly in a clay layer. An averaging interval, (that may be arbitrary small) centered at a clay or sand interface leads to an average rate of sand occurrence of .5 (half sand, half clay). Now we see that the rate $\lambda$ varies with depth for each drill log (from 1 to 0 to 1 again). Given the different layer thickness at the two drill logs, it also varies between drill logs at a given depth. The problem here is to interpolate the rate $\lambda(z)$ in the horizontal, i.e.between drill logs. One way to visualize this is to draw isolines of $\lambda(z)$ between drill logs. This assumes that the probability of seeing sand at a certain depth varies smoothly between the drill logs. This is equivalent to connecting each sand/clay interface at d1 to the corresponding interface at d2, and is consistent with a traditional Geologist's Stratigraphic Interpolation, GSI, where layers are directly connected in the description of a stratigraphic section.

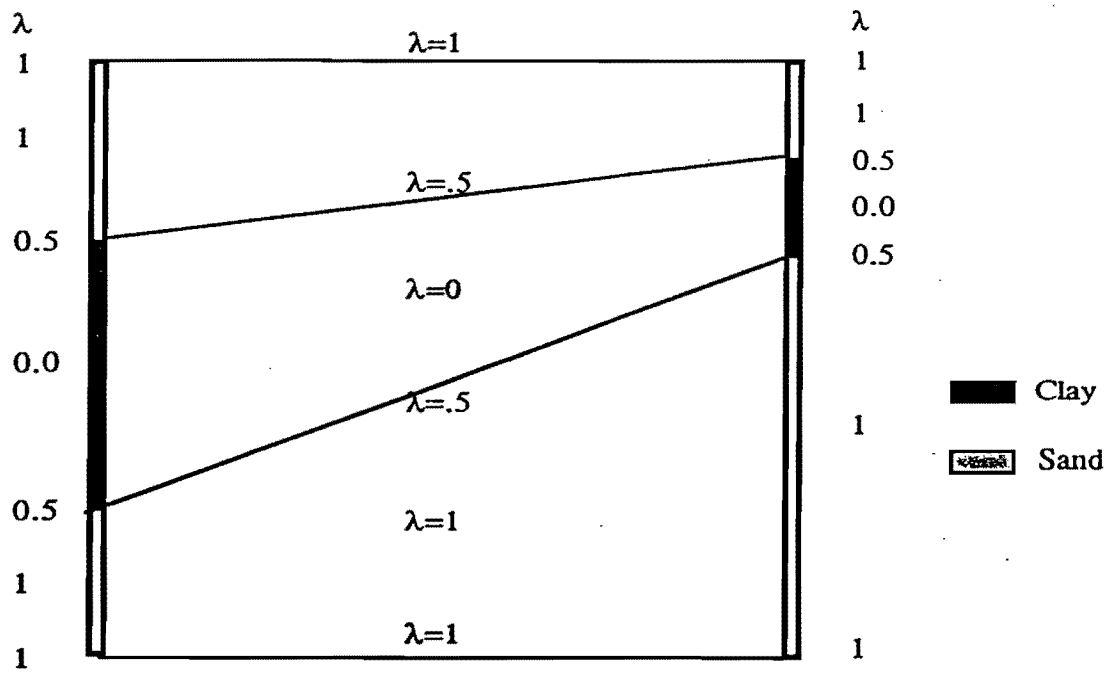



Figure 6a. Nearly constant beds with varying thickness

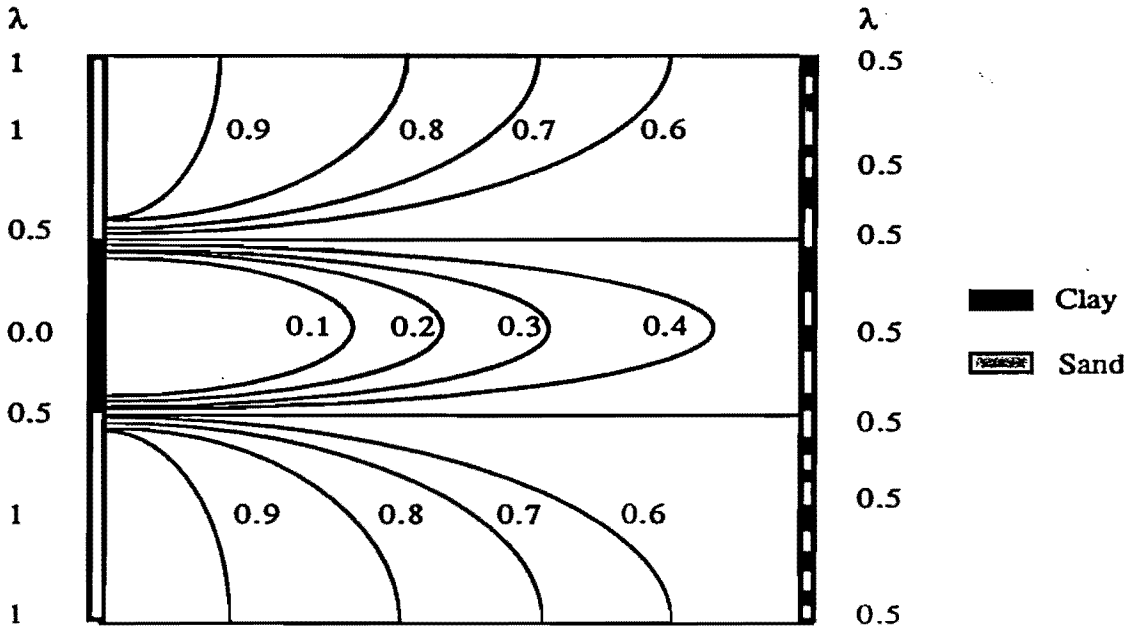



Figure 6b. Highly spatially variable depositional process.

Figure 6b shows the same d1 as in Figure 6a, while d2 has rapidly alternating thin clay/sand layers. Consider that the average rate of sand occurrence for d2 is 0.5 over the whole depth. We clearly notice that the GSI approach can not work in this situation. On the other hand, one can meaningfully construct contours of the rate of sand occurrence $\lambda$ between the two drill logs, under the assumption that there is a smooth variation in the probability of observing sand as one moves from d1 to d2.

In either case, the probabilistic approach of interpolating the rate $\lambda$ is useful, while GSI may or may not be useful. The strategy used in our work follows from the above examples.

1)Layers observed from each drill log i are classified according to a discrete binary classification,e.g., interpretation (a) or (b) described in Table 1.

2)Referring to definition 4.11, we can see that the probability of sand occurrence is related to the poisson process rate $\lambda$. For each drill log i, the NPP rate $\lambda_i(z)$, hence, the probability of sand occurrence is first represented as a function of depth using a Kernel method (vertical smoothing).

3)For each depth $z_j$ of interest, rate $\lambda_{(x,y)}(z_j)$, (for notation simplicity we will call it $\lambda(x,y,z_j)$), at a specified location $(x,y,z_j)$ is also represented by a Kernel method in terms of the $\lambda_i(z_j)$ values estimated at drill log $(x_i,y_i)$ (horizontal interpolation).

To estimate the probability of sand occurrence at any location (x,y,z), steps 2 and 3 are simultaneously carried out using the observed data in step 1 and a three-dimensional MWA using a kernel method. All neighboring observations, within specified averaging spans in the x,y, and z directions, are used for this estimate. A three-dimensional Kernel estimator was used to estimate the probability of sand occurrence at specified location $(x,y,z_j)$. The following is a formal presentation for this estimator. A brief description of the modelling process is shown in Figure 7.

**5.2) Three-Dimensional Kernel Estimator of sand occurrence rate.**

To solve the estimation problem, a three-dimensional MWA is needed. As was mentioned earlier (section 4.5), the concept of the Kernel estimator can be extended to multiple dimensions. A three-dimensional Kernel as a product of two Kernels was used. The first Kernel is one dimensional, with bandwidth $h_{v,i}$, applied to the vertical smoothing at each drill log i (see Figure 8a). The second Kernel is two dimensional, whose bandwidth is a radial distance $h_{r,z_j}$, that interpolates the rate between the drill logs at any horizontal level $z_j$ (see Figure 8b). The parameters $h_{v,i}$ and $h_{r,z}$ are computed in two sequential steps and are varied by location, i.e., each

drill log has a variable $h_{v,i}$ (section 5.3) and each depth has different $h_{r,z}$ (section 5.4). Assume that the values $h_{v,i}$ and $h_{r,z}$ are given. If the probability estimate is desired at point $(x,y,z_j)$, the estimation proceeds as follows:

i)Consider only the drill logs that lie within a horizontal distance $h_{r,z_j}$ from the point $(x,y,z_j)$ and the observations that lie within $A_i$ and $B_i$ drifting in the vertical (see Figure 8b). Note that $A_i B_i = 2h_{v,i}$.

ii)At each drill log i, the NPP rate $\hat{\lambda}_i(z_j)$ is computed using the observations that lie within the vertical distance $h_{v,i}$ from the horizontal level $z_j$.(i.e., within $A_i B_i$) as follows (see Figure 8a):

$$\hat{\lambda}_i(z_j) = \int_{z=z_j-h_{v,i}}^{z=z_j+h_{v,i}} \alpha_{i,z} * I_i(z)$$

or

$$\hat{\lambda}_i(z_j) = \frac{1}{h_{v,i}} * \int_{z=z_j-h_{v,i}}^{z=z_j+h_{v,i}} I_i(z) * K_v\left(\frac{z_j - z}{h_{v,i}}\right) dz \qquad (5.1)$$

where:

$\hat{\lambda}_i(z_j)$ = the smoothed rate parameter at drill log i and horizontal level $z_j$

$K_v(.)$ = One-dimensional Kernel (MWA) in the vertical defined as: $K_v(.) = \frac{15}{16}*(1-(.)^2)^2$.

$h_{v,i}$ = The optimal bandwidth in the vertical at drill site i.

$I_i(z)$ = Indicator function of drill log at site i; it takes the value 1 if the value of z lies within a high conductivity zone and 0 otherwise.

The weight $\alpha_{i,z}$ given to a sand ($I_i(z)$ =1) is determined as the sum (really integration) of the weights assigned by the Kernel function centered at the estimation point to points in the sand layer that fall within a bandwidth of the estimation point. In Figure 8a, the shaded areas under the Kernel are the weights corresponding to the layers for which the indicator function equals 1. The unshaded areas under the Kernel are the weights corresponding to layers where the indicator function is 0. Table 2 gives a numerical example for estimating $\lambda_i(z)$ using the data shown in Figure 8a, for two choices of $h_{v,i}$.

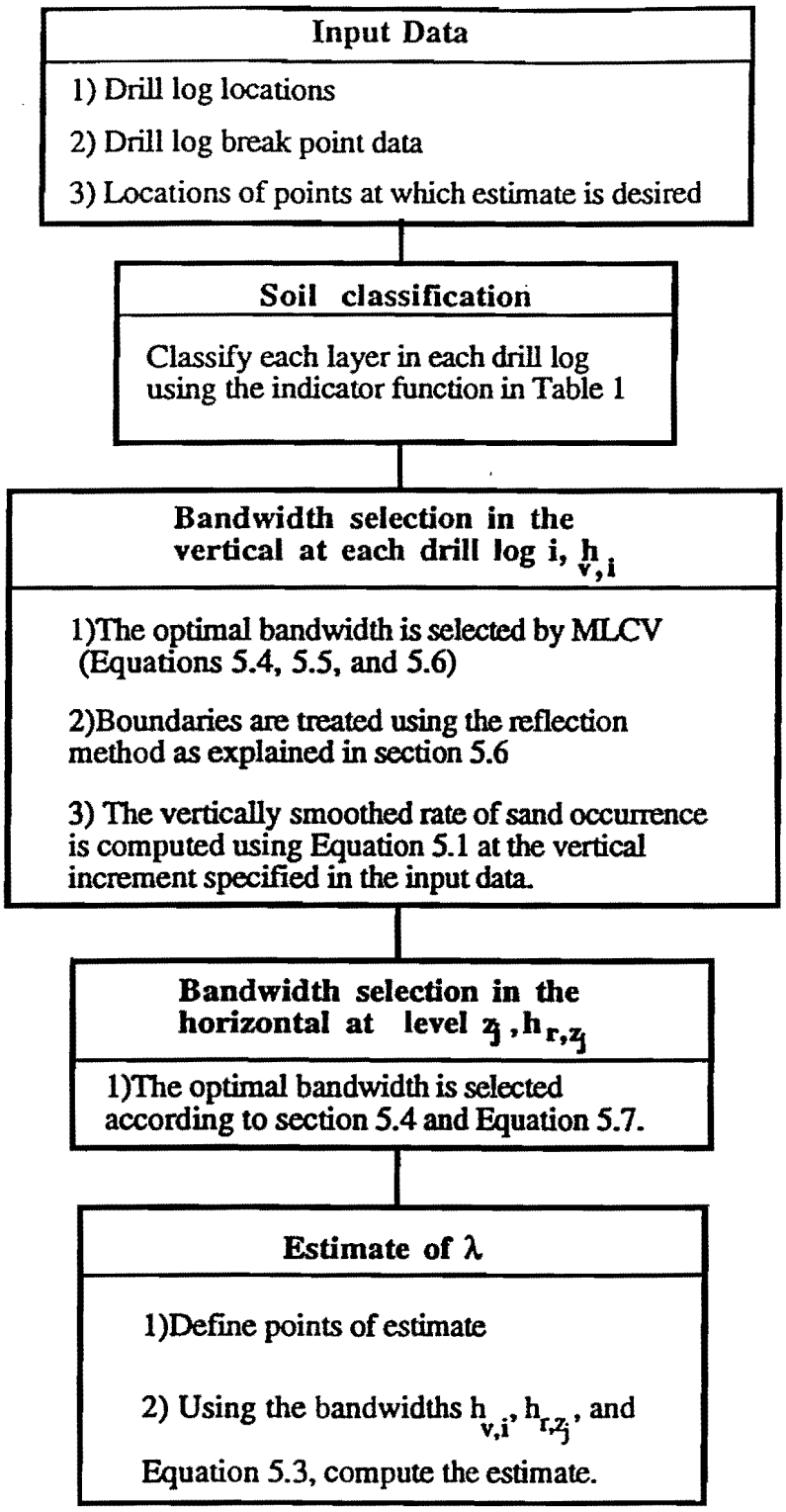



Figure 7. Flowchart that describes the modelling process of the current study.

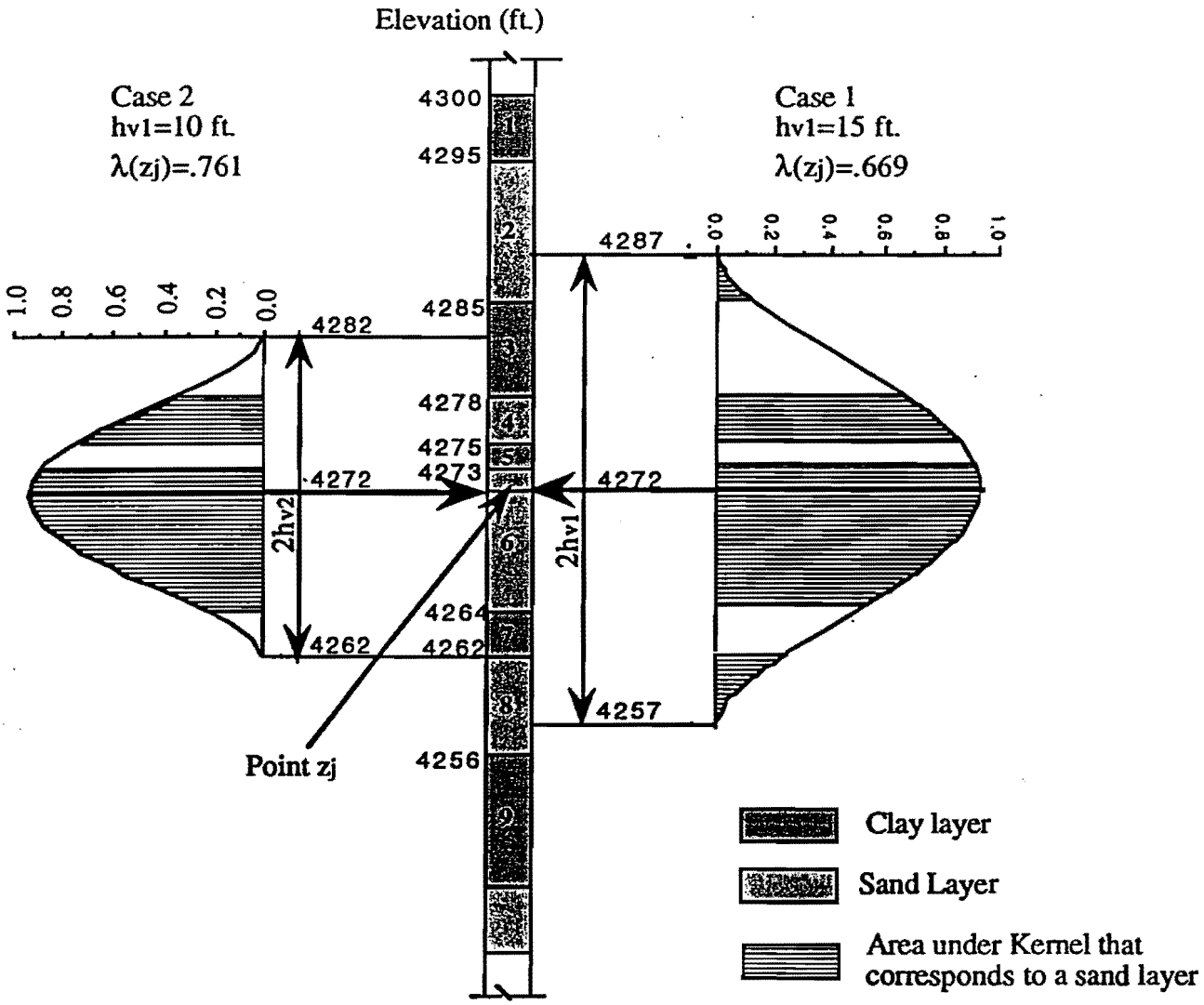
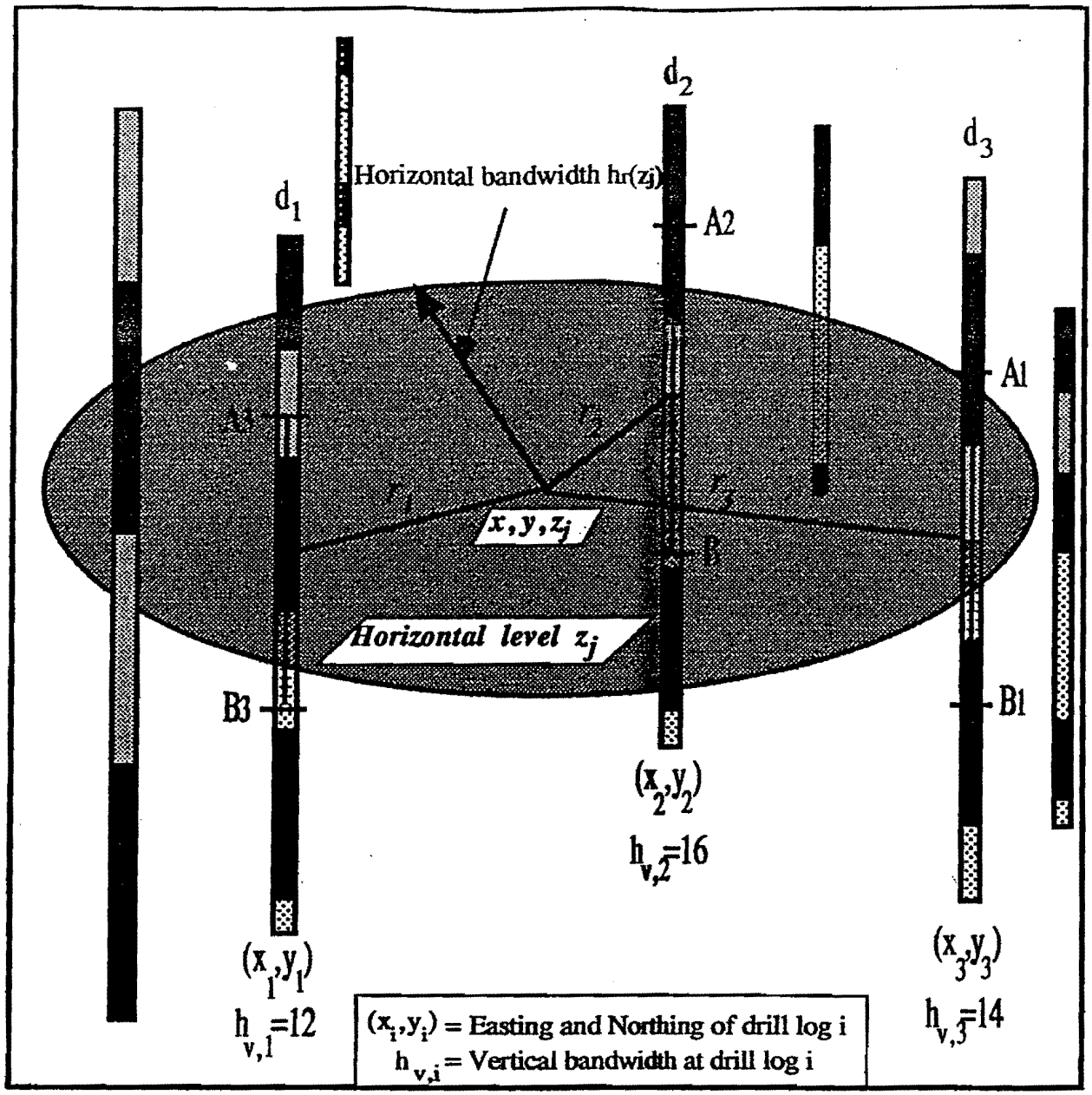



Figure 8a. Estimation of the rate of sand occurrence at point $z_j$ using one-dimensional Kernel, in the vertical at a given drill log $d_i$ using two different bandwidths.




**Note that only the points that lie between [Ai,Bi] at each drill log i that lies within hr from the point of estimate (x,y,zj) contribute to the estimate. In this illustration, 3 drill logs contribute to the estimate.**

Figure 8b. The general scheme for the interpolation of estimated rate between drill logs at a given depth.

iii) Once estimates $\hat{\lambda}_i(z_j)$ are obtained at each drill log i that lies within $h_{r,z_j}$ at elevation $z_j$ (see Figure 8b), a Kernel $K_r(.)$, centered at point $(x_j,y_j,z_j)$ with a radial bandwidth $h_{r,z_j}$, independent of direction, is used to form a moving weighted average in the horizontal {i.e., weights $\omega(r_{ij})$ are assigned to $\hat{\lambda}_i(z)$ to interpolate the estimate at point $(x_j,y_j,z_j)$} The Kernel function $K_r(.)$ in the horizontal is defined as:

$$K_r(u_{i,j}) = \frac{15}{16} * \left(1-u_{i,j}^2\right)^2 \quad -1<u_{i,j}<1 \tag{5.2}$$

where :

$$u_{i,j} = \left(\frac{r_{i,j}}{h_{r,z_j}}\right)$$

$$r_{ij} = \sqrt{(x_j-x_i)^2+(y_j-y_i)^2}$$

= the horizontal distance between point j and drill site i.

$x_j$, $y_j$ = Easting and Northing of point j respectively (point of estimate).

$x_i$, $y_i$ = Easting and Northing of drill site i respectively.

Given that a symmetric Kernel is used, it is expedient to work in radial coordinates with the origin at the point of estimate in order to develop the weight sequence. Note that where there is a high degree of asymmetry in the geometry of the site (in plan view), or of the sampling locations or of the variation in $\lambda(z)$, it may be better to choose different bandwidth $h_x$, and $h_y$. In the interest of parsimony, we shall stay with the single $h_r$ at this stage.

The estimator is defined as:

$$\tilde{\lambda}_j(x,y,z) = \sum_{i=1}^{n_{z_j}} \omega(r_{ij}) * \hat{\lambda}_i(z_j) \tag{5.3}$$

Where: $$\omega_i(r_{ij}) = \frac{K_r(u_{i,j})}{\sum_{i=1}^{n_{z_j}} K_r(u_{i,j})}$$

$n_{z_j}$ = The number of drill logs intersecting the elevation $z_j$.

Table 3 gives a numerical example illustrating the computation process using the 3-d Kernel, and the results for $\tilde{\lambda}$ (x,y,z). From Figure 8b, we observe that three drill logs fall within a radial bandwidth = $h_{r,z_j}$ of the point of estimate.

The previous discussion assumed that the bandwidths $h_{v,i}$ and $h_{r,z}$ are given. Figure 8a and Table 2 show us that changing the bandwidth leads to a change in the estimate. Procedures for selection of the bandwidth in the vertical and horizontal respectively are presented.

**5.3) Vertical bandwidth selection at drill log i**

Referring to Silverman (1986), Kernel probability density and hence intensity estimation is much more sensitive to the value of the bandwidth rather than to the shape of the Kernel. An appropriate choice of the averaging interval or bandwidth depends on the trade-off between variance of estimate, (i.e, the need to get an adequate # of observations to form the MWA) and the bias of the estimate (i.e, the need to restrict the averaging interval to better capture the local variation). Diggle and Marron (1988) show the equivalence between choosing the bandwidth for Kernel intensity and for Kernel probability density estimation. They show that the same bandwidth is optimal for both under a Mean Square Error (MSE) criterion. The bandwidth may also be chosen by maximizing the joint likelihood of the observations with respect to the chosen bandwidth.

Unfortunately, the true function is needed for estimating either the likelihood or the MSE. Therefore, a primary estimate of the true function is needed. Cross Validation may be used to provide the relevant estimates of MSE or likelihood. If we want to get an estimate at an observed point i, we first drop this point and then use the rest of the points to estimate $\lambda$ at this point.

For the selection of the optimal bandwidth ($h_{v,i}$) in the vertical at drill log i, the Maximum Likelihood Cross Validation (MLCV) method was used. In this method, the process rate was estimated, for drill log i, at the center of each sand interval of thickness $\delta$ (e.g., 1 foot) and center elevation z, using all the layers except that interval. Such a rate is called $\hat{\lambda}_{i,-z}(z)$. In other words, if we need to estimate the rate at point z, in a sand layer, we follow the same procedures for

estimating $\hat{\lambda}_i(z)$ except that we drop an interval of thickness $\delta$ centered at point z. The estimate $\hat{\lambda}_{i,-z}(z)$ can be calculated as follows:

$$\hat{\lambda}_{i,-z}(z) = \frac{\hat{\lambda}_i(z) - \Omega_{i,z}}{1 - \Omega_{i,z}} \tag{5.4}$$

where

$\hat{\lambda}_i(z)$ = The rate of sand occurrence computed from Equation 5.1

$$\Omega_{i,z} = \frac{1}{h_{v,i}} \int_{z-\frac{\delta}{2}}^{z+\frac{\delta}{2}} K_v\left(\frac{\zeta - z}{h_{v,i}}\right) d\zeta \tag{5.5}$$

Where: $\Omega_{i,z}$ is the area of interval of thickness $\delta$ centered at point z "the point of estimate".

The log likelihood function ($L_i$) for data from Nonhomogenous Poisson Process at drill log i is the sum of piecewise log integrals of $\hat{\lambda}_{i,-z}(z)$ over the layers of high conductivity at drill log i. This is a measure of the joint probability of occurrence of the observed values under the current choice of the bandwidth. Note that the joint probability or likelihood is given by the product of the $\lambda_i$. Taking the logarithms allows us to write the likelihood function as a sum, which is easier to deal with. This is given as:

$$L_i = \sum_{j=1}^{m} \int_{z_{i,j}}^{z_{i,j+1}} \log\left(\hat{\lambda}_{i,-z}(z)\right) dz \tag{5.6}$$

where m is the number of layers with high conductivity.

The optimal bandwidth is the one that maximizes the value of $L_i$

### 5.4)Horizontal bandwidth selection at depth $z_j$.

As mentioned in the previous section, the optimal bandwidth can be selected by minimizing a MSE Criterion. The Least Square Cross validation method (LSCV) was used to select the optimal bandwidth in the horizontal, since it is more appropriate in the context of interpolation. In this method, the sum of squares of errors was computed at every data point 'i', i.e., the drill log at site i, using all data points except point 'i' and $\lambda_i(z_j)$. The sum of squares of errors function (S) is:

$$S_{z_j} = \sum_{i=1}^{n_{z_j}} \sum_{k=1 \neq i}^{n_{z_j}} \left( \hat{\lambda}_i(z_j) - \tilde{\lambda}_{-i}(z_j) \right)^2_{z_j} \qquad (5.7)$$

where:

$n_{z_j}$ = the number of drill logs observed at level $z_j$.

$\hat{\lambda}_i(z_j)$ = the rate parameter estimated using Equation 5.1 at drill log i and horizontal level $z_j$

$\tilde{\lambda}_{-i}(z_j)$ = the rate parameter estimated using Equation 5.3, in which all data points at the horizontal level $z_j$, except point 'i', were used.

The optimal bandwidth is the one that minimizes the value $S_{z_j}$.

We notice that the selection of the horizontal bandwidth depends on $\hat{\lambda}_i(z_j)$, hence, on the vertical bandwidth.

**5.5)Vertical-Horizontal bandwidths iteration:**

The selection of the vertical bandwidth independently from the horizontal bandwidth may not preserve the variation structure of the domain. Vertical-Horizontal bandwidths iterative procedures may be used in order to correlate the two bandwidths together as follows:

a) Once the horizontal bandwidth is selected according to section 5.5, the vertical bandwidth is reselected according to the procedures in section 5.4 except that the estimate $\hat{\lambda}_i(z)$, computed from Equation 5.1, is replaced by $\tilde{\lambda}$ (x,y,z), computed from Equation 5.3.

b) The new values of the vertical bandwidth are then used to reselect the horizontal bandwidths according to section 5.5.

c)Steps a and b are repeated successively until a convergence with a desired accuracy is reached.

**5.6) Choice of Kernel.**

Scott (1992) shows that the shape of Kernel does not significantly affect the mean square error (bias$^2$+variance) of estimate. Compared to the bandwidth selection, the selection of the Kernel function does not significantly affect the MSE of estimate. However, it is desirable to base the choice of Kernel on other considerations, e.g., the degree of differentiability required or the computational effort involved (Silverman 1986). The Kernel function is usually symmetric to assure a symmetric distribution of the weights on both sides of the point of estimate. It is also positive everywhere. The optimal Kernel, that minimizes the MSE, is the Epanechnikov

Kernel,$\{K(t)=\frac{3}{4}(1-t^2);\ -1<t<1\}$, whose efficiency is said to be 1. However other Kernels have nearly the same efficiency by adjusting their bandwidth such that a similar MSE is obtained. A smoother kernel (differentiable at t=1) is preferred as it gives better differentiability and continuity of the resulting estimate. In this study, the bisquare Kernel $K(t)=\frac{15}{16}*(1-t^2)^2$, whose efficiency is .9939,was used.This function is shown in Figure 5.

**5.7)Boundary effect in the vertical.**

A problem with forming a moving weighted average centered at the point of estimate occurs as one approaches the boundaries of the sample (i.e., drill log). Near the endpoints, the averaging intervals on one side extend across the boundary where there are no observations. Thus, we do not have a symmetrically weighted moving average, leading to the effective center of the average not being at the point of estimate. A variety of solutions to this problem are offered in the literature. We adopted the use of reflection at the boundary. This considers that the rate of change of the target function (i.e. $\lambda(z)$) is small near the endpoints. One generates synthetic observations across the boundary that are a mirror image of the data in the domain (mirror placed at the boundary). These synthetic observations are used only within a bandwidth of the boundary.

**5.8)Boundary effect in the horizontal.**

To solve this problem in the horizontal, we used the Cut and Normalize technique. In this method, we ignore the area of Kernel that extends outside the boundary, where no data points are found. The weights are then normalized to the total sum of the weights of the points that lie under the remaining area of the Kernel.

| Layer # | Case (1), $h_{v1}$=15 ft., $z_j$=4272 ft. | | | | Case (2), $h_{v2}$ =10 ft, $z_j$=4272 ft. | | | |
|---|---|---|---|---|---|---|---|---|
| | Elevation ($z_i$) layer's top | $t=(z_j-z_i)/h_{v1}$ | Area under Kernel (w) | Corresponding Indicator (I) | Elevation ($z_i$) layer's top | $t=(z_j-z_i)/h_{v2}$ | Area under Kernel (w) | Corresponding Indicator (I) |
| 2 | 4287* | -1 | | | | | | |
| | | | .003 | 1 | | | | |
| 3 | 4285 | -.867 | | | 4282** | -.1 | | |
| | | | .16 | 0 | | | .058 | 0 |
| 4 | 4278 | -.4 | | | 4278 | -.6 | | |
| | | | .154 | 1 | | | .177 | 1 |
| 5 | 4275 | -.2 | | | 4275 | -.3 | | |
| | | | .12 | 0 | | | .172 | 0 |
| 6 | 4273 | .067 | | | 4273 | -.1 | | |
| | | | .476 | 1 | | | .584 | 1 |
| 7 | 4264 | .533 | | | 4264 | .8 | | |
| | | | .051 | 0 | | | .009 | 0 |
| 8 | 4262 | .667 | | | 4262** | 1 | | |
| | | | .036 | 1 | | | | |
| 8 | 4257* | 1 | | | | | | |

* Elevation of the higher and lower edges of Kernel in case 1.
** Elevation of the higher and lower edges of Kernel in case 2.

By applying Equation 5.1, the rate in case 1 = .003*1+.16*0+.154*1+.12*0+.476*1+.051*0+.036*1=.669
and the rate in case 2 = .058*0+.177*1+.172*0+.584*1+.009*0=.761

Table 2. Example for computing $\hat{\lambda}_i(z_j)$ at drill log $d_i$ using two different bandwidths (see Figure 8a).

| Drill log # | $\hat{\lambda}_i(z_j)$ * | Distance $r_i$ to point$(x,y,z_j)$ | $u_i(z_j)=\left\langle \frac{r_i}{h_{r,z_j}} \right\rangle$ | $K(u_i(z_j))$ | $\omega_i(x,y)$ ** | $\alpha_i * \hat{\lambda}_i(z_j)$ |
|---|---|---|---|---|---|---|
| 1 | 0.32 | 1000 | 0.667 | 0.289 | 0.308 | 0.099 |
| 2 | 0.85 | 750 | 0.500 | 0.527 | 0.561 | 0.477 |
| 3 | 0.61 | 1200 | 0.800 | 0.122 | 0.13 | 0.079 |

* See Table 3 for the procedures of calculating the vertically smoothed rate $\hat{\lambda}_i(z_j)$.

** $\omega_i(x,y) = \dfrac{K_r(u_i(z_j))}{\sum_{i=1}^{3} K_r(u_i(z_j))}$, where: $K(u_i(z_j)) = (1-u_i(z_j)^2)^2$.

By Applying Equation 5.3, we get : $\tilde{\lambda}(x,y,z_j) = 0.099+.477+0.079 = .655.$

Table 3. Example of estimation of $\tilde{\lambda}(x,y,z_j)$ for a point between drill logs (see Figure 8b).

# 6)USERS MANUAL

## 6.1)Introduction

The procedures described in the methodology section are implemented as a FORTRAN 77 computer code (DLOG3D). Program DLOG3D interprets drill log data in the the three-dimensional space in terms of sand occurrence rate. The computational procedures follow the algorithm shown in Figure 7. These procedures used a nonparametric technique to estimate the rate of sand occurrence at a particular area. Program DLOG3D computes this rate at a fixed point or at nodes of a line, a plane, or a three-dimensional grid. One may use such outputs to map a particular area where the occurrence rate of a high conductivity zone is of his concern. An executable version of this code is available on request. A description of the user interface is provided in the following subsections.

6.2) Input file and data preparation

6.3) Interactive input data

6.4)Warning Messages

6.5) Output files

## 6.2)Input file and data Preparation

Drill log data reports usually contain profile descriptions of the soil from which the drill logs were sampled. At a drill log, the data recorded at each observed layer is its soil type and the Elevation of the top and bottom of this layer. These reports are the source from which the input file is prepared. This input file consists of two parts. The first part is the data needed for the code to create the model. The second part specifies the locations at which estimates are needed. Input instructions for both parts are provided in Table 4.

**Part** I

A drill log report for soil data usually includes top and bottom Elevations for layers as well as the soil type of such layers. The soil type for each layer should be replaced by a binary classification. If a layer has a high permeability soil, soil type should be replaced by 1 and if it has a low permeability soil, soil type should be replaced by 0.

Part I consists of number data sets. This number is the number of drill logs and contained in the line #1. Data set i includes all information about drill log i. In this set, the first line is the number of layers at drill log i. The second line is the Easting and Northing coordinates of drill log

i followed by number of lines equal the number of layers. At each line, the top and bottom Elevation and the binary indicator of the corresponding layer are respectively found. The input instructions of Part I of the input file are found in Table 4.

**Part II**

The data contained in Part II of the same input file is the one necessary to form the grid of estimate. The DLOG3D program can do the required estimate at a point or at a one, two, or three-dimensional grid. However, the computing process is sensitive to the location of such a grid. DLOG3D does not do extrapolation. The more data within the location of the grid of estimate, the more precise and reliable the estimate. Therefore, visualizing the data locations in the vertical and horizontal will be helpful to locate the grid of estimate. The following steps may be followed to get the best possible grid locations:

1) The following plots should be obtained:

a) Easting versus Northing of all drill logs.

b) The projected length of drill logs in the northing direction versus Easting.

c) The projected length of drill logs in the easting direction versus Northing.

2) By examining the three plots, the selection of a grid location can be made by maximizing the number of data points in the vicinities of such locations. The borders of the grid of estimate should be within the borders of data points. The word 'grid' refers to one of the following:

i) A point, in which the maximum and the minimum are the same for all coordinate directions.

ii) A line, in which the maximum and the minimum are the same for two coordinate directions. The remaining direction is the direction of the relevant line.

iii) A plane, in which the maximum and the minimum are the same for one coordinate directions. The remaining two directions are the directions of the length and width of the relevant plan.

iv) A box, in which all the maximum and the minimum are not the same for all coordinate directions.

Part II input instruction is shown in the lower portion of Table 4.

Table 4 shows complete input instructions for program DLOG3D. The table consists of 3

columns. The first column indicates the line number in the input file. The second one shows the variable names, corresponding to this line, as they appear in the source file of program DLOG3D. Column 3 contains a description of the corresponding variables mentioned in column 2. All dimensions are in feet, Unless Other Wise Stated.

**6.3)Interactive input data:**

Program DLOG3D provides interactive inputs for some options. The user will be asked for the following options:

i) If a list of the vertical bandwidths is desired. An output file name will be required if the answer is yes.

ii) If a list of the horizontal bandwidths is desired. An output file name will be required if the answer is Yes.

iii) If the rate of sand occurrence at some drill logs is desired. If the answer is Yes, the answer will be asked to enter the following:

a) The number of drill logs.

b) Specification number for each drill log. This number indicates the rank of such drill log data set in the input file.

c)An output file name.

iv) If a warning message file is desired for the incorrectly located points (section 6.4). An output file name will be required if the answer is Yes. If the answer is No, a flag (-9999) will be given to each of these points instead of a value of estimate.

A sample run for the interaction process.is shown in Figure 9 as it may appear on the computer screen.

**6.4)Warning messages:** DLOG3D is protected against the incorrectly located parts of any given grid (points of estimates). The incorrect locations are those where the distance to the nearest drill log is greater than the optimal bandwidth. If a warning message file is desired (see subsection 6.3.iv), a warning message will appear, written in a file, at the incorrectly located points of estimates. In such a situation, DLOG3D keeps computing but after expanding the optimal bandwidth to the nearest neighbour The warning message gives the amount of oversmoothing in percentage, the percentage of increase of the optimal bandwidth, as well as Easting, Northing and

| Line # | Name | Description |
|---|---|---|
| | | **Part I** |
| 1 | nlogs | Number drill logs |
| 2 | nlayers(1) | Number of layers for drill log #1 |
| 3 | xcord, ycord | Easting and Northing ( in meters) of drill log #1 |
| 4 | nst, nfin, typ | Top and bottom Elevations, and binary indicator of layer #1. |
| 5 | nst, nfin, typ | Top and bottom Elevations, and binary indicator of layer #2 |
| . | .. .. .. | .. .. .. |
| . | .. .. .. | .. .. .. |
| sum(1)* | nst, nfin, typ | Top and bottom Elevations, and binary indicator of last layer |
| sum(1)+1 | nlayers(2) | Number of layers of drill log #2 |
| sum(1)+2 | xcord, ycord | Easting and Northing ( in meters) of drill log #2 |
| sum(1)+3 | nst, nfin, typ | Top and bottom Elevations, and binary indicator of layer #1. |
| sum(1)+4 | nst, nfin, typ | Top and bottom Elevations, and binary indicator of layer #2 |
| . | .. .. .. | .. .. .. |
| . | .. .. .. | .. .. .. |
| sum(2)* | nst, nfin, typ | Top and bottom Elevations, and binary indicator of last layer |
| * | * | * |
| * | * | * |
| sum(nlogs)* | nst, nfin, typ | Top and bottom Elevations, and binary indicator of last layer of last drill log. |
| | | **Part II** |
| sum(nlogs)+1 | incx, incy, incz | Easting, Northing and Elevation increments for estimation. |
| sum(nlogs)+2 | zmax, zmin | Maximum and Minimum possible Elevations for estimate. |
| sum(nlogs)+3 | nplans | Number of grids of estimates. |
| sum(nlogs)+4 | x1,x2,y1,y2,z1,z2 | Coordinates of grid #1 |
| sum(nlogs)+4 | x1,x2,y1,y2,z1,z2 | Coordinates of grid #2 |
| * | | |
| * | | |
| end** | x1,x2,y1,y2,z1,z2 | Coordinates of grid #nplans |

*$\text{sum(k)} = \sum_{i=1}^{k} \text{nlayers(i)}+2*k+1$ **end = sum(nlogs)+nplans+2

Table 4. Input instructions for program DLOG3D.

```
 give the input file name
sample.dat
 Do you want a list of vertical bandwidths ?
 Type 1, if yes. Type 0, if no.
1
 give a file name for the list of vertical bandwidths
vb.out
 Do you want a list of horizontal bandwidths ?
 Type 1, if yes. Type 0, if no.
0
 do you want the intensity to be calculated at some drill logs
 Type 1, if yes. Type 0, if no.
1
 how many drill logs ?
5
 give the drill log numbers
7 15 22 29 32
 do you want to renter the drill log numbers
 Type 1, if yes. Type 0, if no.
0
 give a name for drill logs intensity output
drill.out
 For the mislocated points,
 Do you want warning messages?
 If yes, type 1. A warning message output file will be created for
you
 If no, type 0. A flag (-9999) will be given for each mislocated point
1
```

Figure 9. A computer screen printout showing a sample run for the interaction process.

the Elevation of the point at which such an amount of oversmoothing took place. Should more bandwidth expansion be necessary at a different point at the same Elevation, another warning message will appear showing the new percentage of oversmoothing. Each time a bandwidth is expanded at a given Elevation, the new bandwidth remains for the rest of the points at that Elevation unless another expansion is needed.

### 6.5)Output files

There are two types of the output results of Program DLOG3D:

i)Main output files: They are always produced and contain the values of the rate of sand occurrence at the nodes of the estimating grids. The number of these files equals to the number of the grids of estimate. This number is indicated by "nplans" at line "sum(nlogs)+3" in Table 4.

ii)Optional output files: They are generated optionally based on the interactive directions. (See section 6.3 and Figure 9)the files created due to selecting some options from the interactive inputs.

## 7)APPLICATION

An input file was set up using thirty four drill log data points. Three plots are presented to visualize the data point layout (Figures 10, 11, and 12). DLOG3D was applied to create nine output files for nine different planes (three in each direction). Locations of these planes are shown in Figures 10, 11, and 12. Contour line plots of estimates $\lambda$ for all the nine planes are presented.

### 7.1)Data Preparation

Data used in this example was extracted from USGS report provided by Michelle Lemieux, Division of Water Rights, State of Utah. Figure 10 shows a plot of Northing versus Easting, as well as the locations of the three planes of estimate at fixed Eastings and the three planes at fixed Northing. Figures 11 and 12 show vertical view of drill logs versus Easting and Northing respectively. Figures 11 shows the locations of the three planes of estimate at fixed Eastings and six planes of estimates at fixed Elevations. Figures 12 shows the locations of the three planes of estimate at fixed Northings. The planes of estimates are categorized into three sets as follows:

***a) Three planes in Easting***

They extend between the coordinates 1623000 and 1627000 feet in Northing, and between Elevations 4120 and 4320 in the vertical. The three locations in Easting are 1383000, 1386000, and 1389000 feet respectively.

***b) Three planes in Northing***

They extend between the coordinates 1383000 and 1389000 feet in Easting, and between Elevations 4120 and 4320 in the vertical. The three locations in Northing directions are 1623000, 1625000, 1627000 feet respectively.

***c) Six horizontal planes in the vertical***

They extend between the coordinates 1623000 and 1627000 feet in Northing; and the coordinates 1383000 and 1389000 feet in Easting. The six locations in the vertical directions are the Elevations 4285, 4250, 4225, 4200, 4185, and 4165 feet.

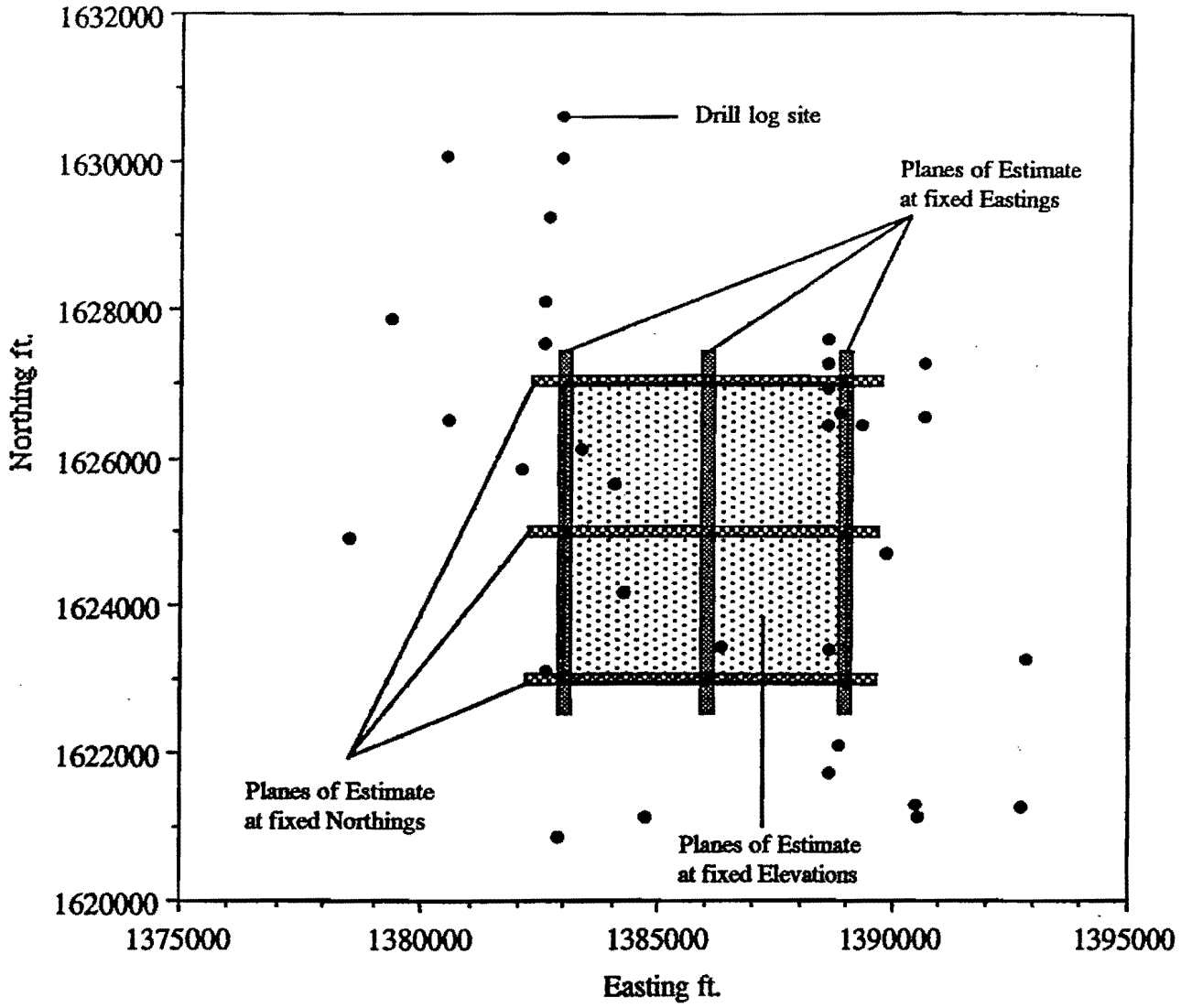



Figure 10. A plan view for the drill log sites and the location of estimation planes.

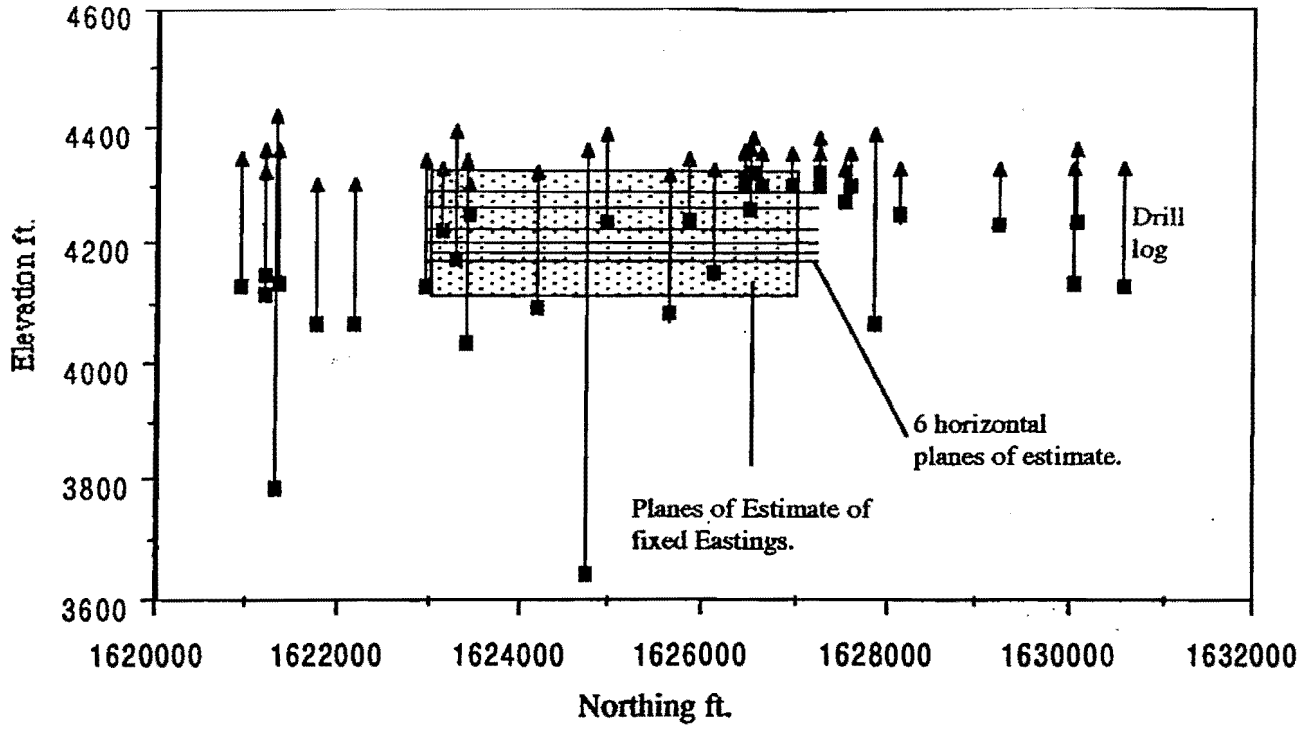



Figure 11 Easting projection for the drill logs and the estimation planes.

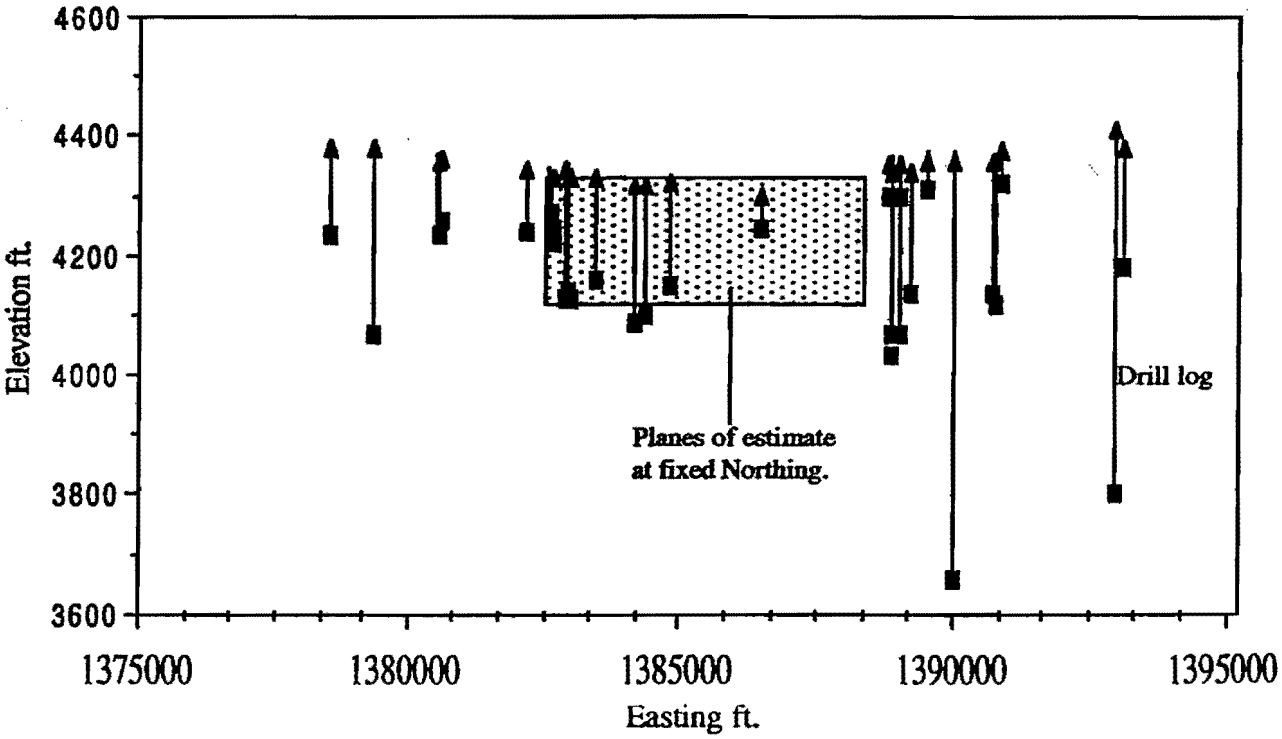



Figure 12 Northing projection for the drill logs and the estimation planes.

## 7.2)Results and Discussions

Nine output files were created by DLOG3D for the nine planes of estimate. Figures 13, 14 show contour plots, indicate the rate of sand occurrence, for sets *a* and *b* respectively. Set *c* is presented through 6 perspective views, Figures 15 to 20, that describe the rate variations of sand occurrence over the 6 horizontal planes mentioned above. All dimensions are in feet.

To interpret the results, we shall look at the rate variation in the vertical through Figures 13 and 14. Perspective views, Figures 15 to 20, are presented to highlight the variations over the horizontal at some elevations that have interesting structures (these elevations are specified above).

From Figures 13 to 20, local variations and global trends of the variations were revealed in the vertical as well as in the horizontal. The general features of the study area are as follows:

1) Highly heterogeneous soil is observed between levels 4270 and 4290. At this level, there is a high probability of observing a sand lens extending in the Northing at Easting = 1386000.

2) There is a high probability, close to 1, of observing sand in the North-East corner above horizontal level =4240.

3) Soil below level = 4225 is less heterogeneous compared to the one above this level.

4) In General, the probability of sand occurrence has an increasing trend towards the Easting direction below level = 4270. The low conductivity soil is likely to be in the west.

The features presented above are revealed from the following examined sections:

a) An extending lens of sand Figures 13 and 14 show an interesting structure of variations between Elevations 4270 and 4300. Figure 15 shows a perspective view for the rate variation over the horizontal level 4285, which lies within this structure. From Figures 13, 14, and 15, we notice the following:

1) Rapid variations in the vertical

2) The rate increases in the North-East direction. It equals 1 in the Northern-Eastern corner

3) This rate increases then decreases in Easting with peak at Easting = 1386000

4) An increase of the rate is observed with Northing. The rate of increase is significantly higher for Easting > 1386000.

5)Low conductivity zone (rate <.5) for Easting < 1385000

The most important feature in this section is the high probability of observing two sand lenses lie between Elevations 4270 and 4290. The first lens extends in the Northing direction and

lies at Easting = 1386500 with width exceeds 1500 feet and thickness of about 20 feet. The second one is a big spot of sand lies in the Eastern-Northern corner (length = 2000, width = 1500 and thickness = 20 ft). We notice from Figure 15 that the two lenses are connected at Easting = 1387000 and Northing = 1626000.

b)Rapid rate increase with Easting The structure revealed between Elevations 4270 and 4240 is significantly different from the one observed in (a) (See Figures 13 and 14.) Figure 16 shows the rate variation over horizontal plane = 4250. The features observed from Figures 13, 14, and 16 are as follows:

1) Rapid increase of the rate with Easting. Mild variation in Northing and in the vertical.

2) For Easting > 1387000, The rate of sand occurrence increases with depth (Compare Figure 15 with Figure 16).

3) Increase of the rate with the East-North direction (Similar to the previous structure).

This section shows an obvious increase of the rate of sand occurrence towards Easting. The persistent increase towards Easting indicates that we are approaching a zone of sand (high conductivity soil).

c) Transition zone in the vertical followed by rate decrease with Easting and invariant rate in Northing Figures 13 and 14 shows a gradual increase, with depth, of the occurrence rate of sand between Elevations 4250 and 4200. Figure 17, plane = 4225, shows a mild variation with Northing. The structure observed in this zone is described as follows:

1) Very mild variation, almost invariant rate, in Northing.

2) Gradual transition (rate increase with depth) in the zone between Elevations 4250 and 4225.

3) Rate decrease with Easting, i.e., increase with Westing, between Elevations 4225 and 4200.

The interesting feature of this section is observing a transition zone in the vertical between Elevations 4250 and 4225. (see Figure 14). This zone leads to a change of the sand zone direction. Figure 17 shows an increasing probability of observing sand with Westing. This is different from the global trend observed along the vertical profile. All other sections exhibit

increasing trend of the rate with Easting.

d)An extending Aquifer The last structure is the one observed below the level 4200. Figures 18, 19, and 20 show three progressive pictures at three different levels. The vertical and the horizontal variations are revealed from these figures as follows:

1) Gradual variation in the vertical.

2) Rapid variation, rate increase, with Easting and a relatively slight variation in Northing.

3) The rate increases with Easting It is greater than .9 for Easting > 1388000.

We can expect a huge zone of high conductivity below Elevation 4200, at Easting > 1387000, that extends from South to North.

**Example I (Contaminant monitoring)**

The situation of interest is to monitor the zones of high potential contaminant transport that may exist in the upper stratum so that the contaminant can be prevented from reaching the deep aquifers. Figure 13 shows a lens of high rate of sand occurrence approximately at Easting = 1387000 and elevation = 4285. Such a rate is increasing with Northing until it reaches 1 at Northing = 1627000. At selected points along the Easting plane of 1387000, several monitoring wells could be installed starting from Northing=1623000 until Northing = 1627000. Also, monitoring wells may be installed in the Eastern-Northern corner. The monitoring wells should reach the flow in the zone between 490 and 4270. The water quality data that can be obtained from those wells will show the progress of the contaminant transport with space and time. The engineer will then decide if extracting ground water in the vicinity of those points should be allowed.

**Example II (Ground water extraction "zones and amount of yield")**

Extracting ground water from deep aquifers increases the vertical down gradient, hence, increases the chance of a contaminant transport. The downward contaminant velocity depends on the hydraulic conductivity and the well discharge. A safe ground water extraction is the one that does not lead to a deep aquifer contamination. The situation of interest is to find the appropriate place and estimate the appropriate well discharge for safe groundwater extraction. For a particular area if the rate of high conductivity zone is given, we can estimate the expected flow paths at this area. Given the expected flow paths, what is the maximum allowable discharge for a well located at this area is the problem of interest. Figure 15, plot of plane = 4285, shows that the area located

within Easting =< 1385000, Northing =< 1625000 (zone where the rate =< .5) may be a good area to install pumping wells in the deep aquifer.

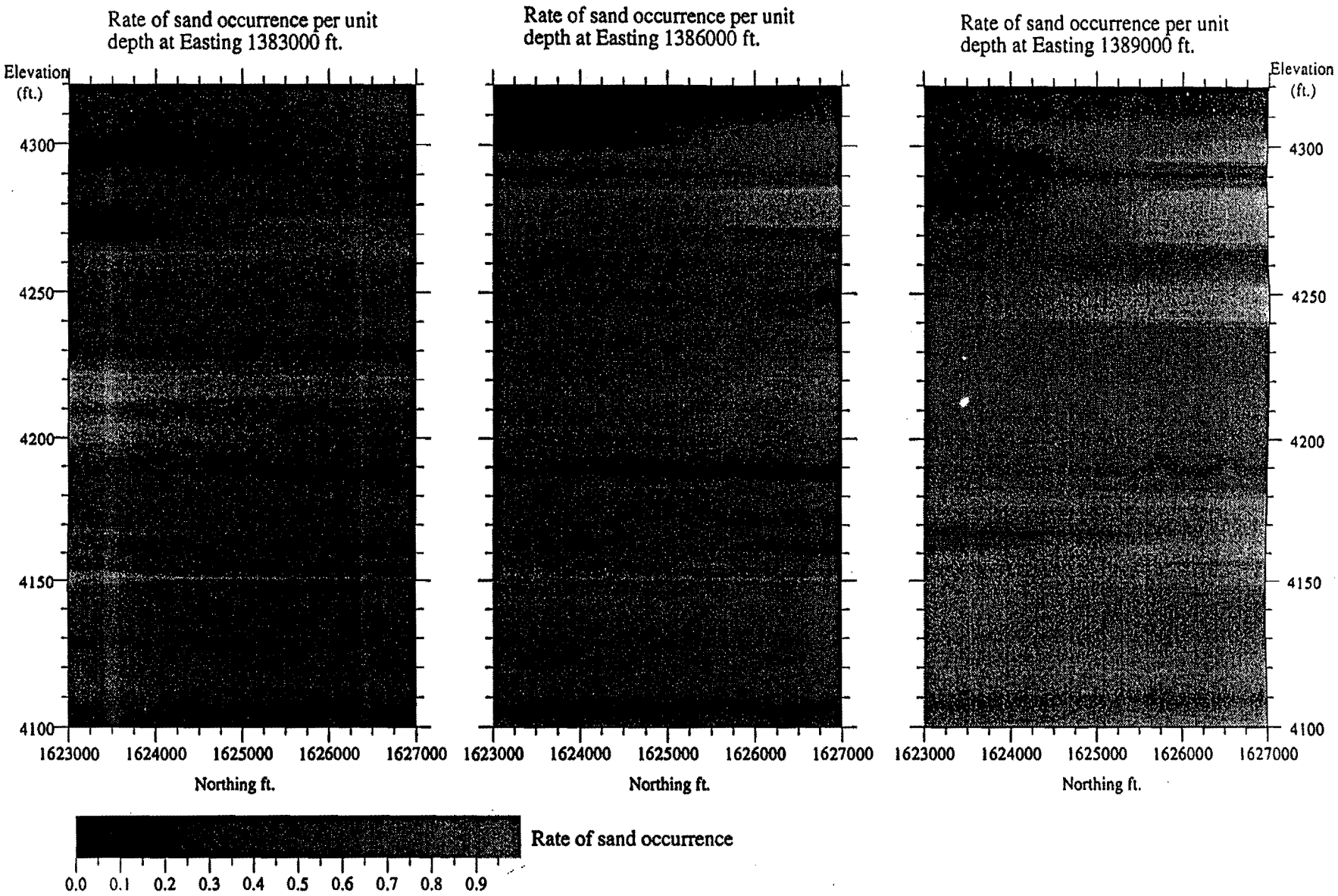



Figure 13. Elevation-Northing contours for the rate of sand occurrence at three different Eastings.

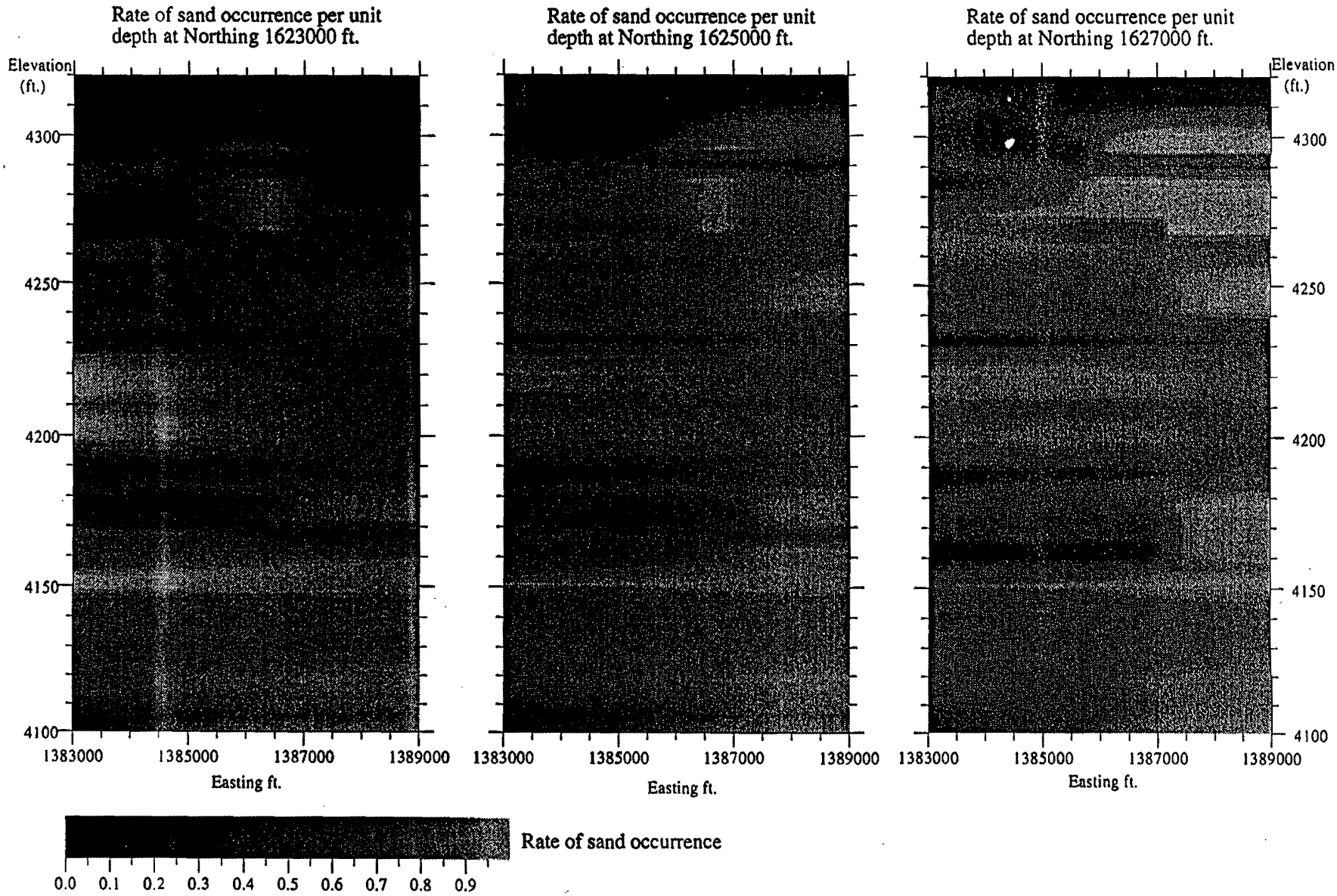



Figure 14. Elevation-Easting contours for the rate of sand occurrence at three different Northings.

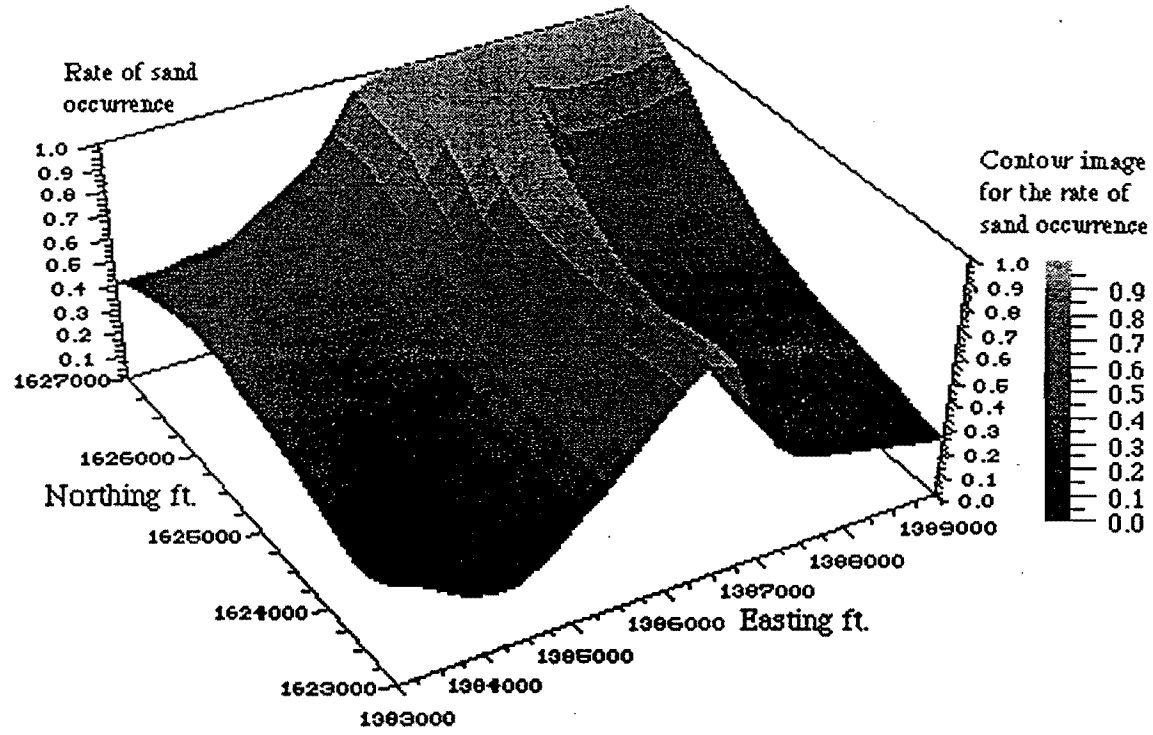



Figure 15. Northing-Easting perspective view for the rate of sand occurrence at Elevation = 4285 ft.

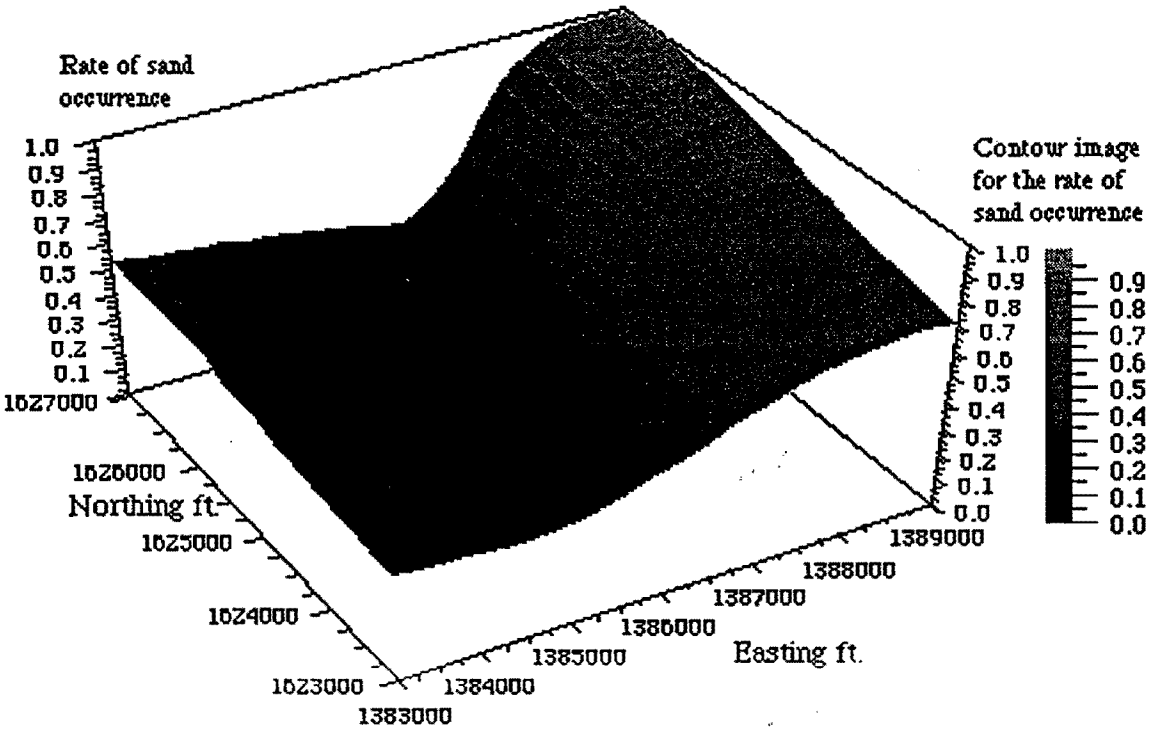



Figure 16. Northing-Easting perspective view for the rate of sand occurrence at Elevation = 4250 ft.

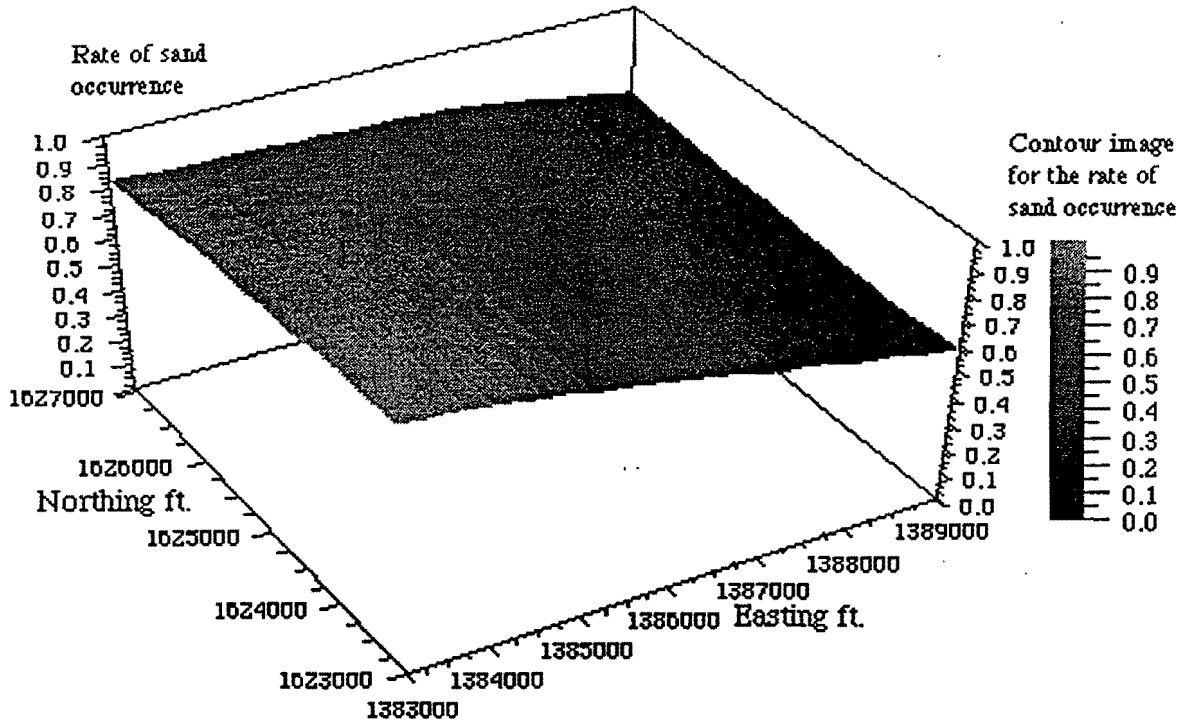



Figure 17. Northing-Easting perspective view for the rate of sand occurrence at Elevation = 4225 ft.

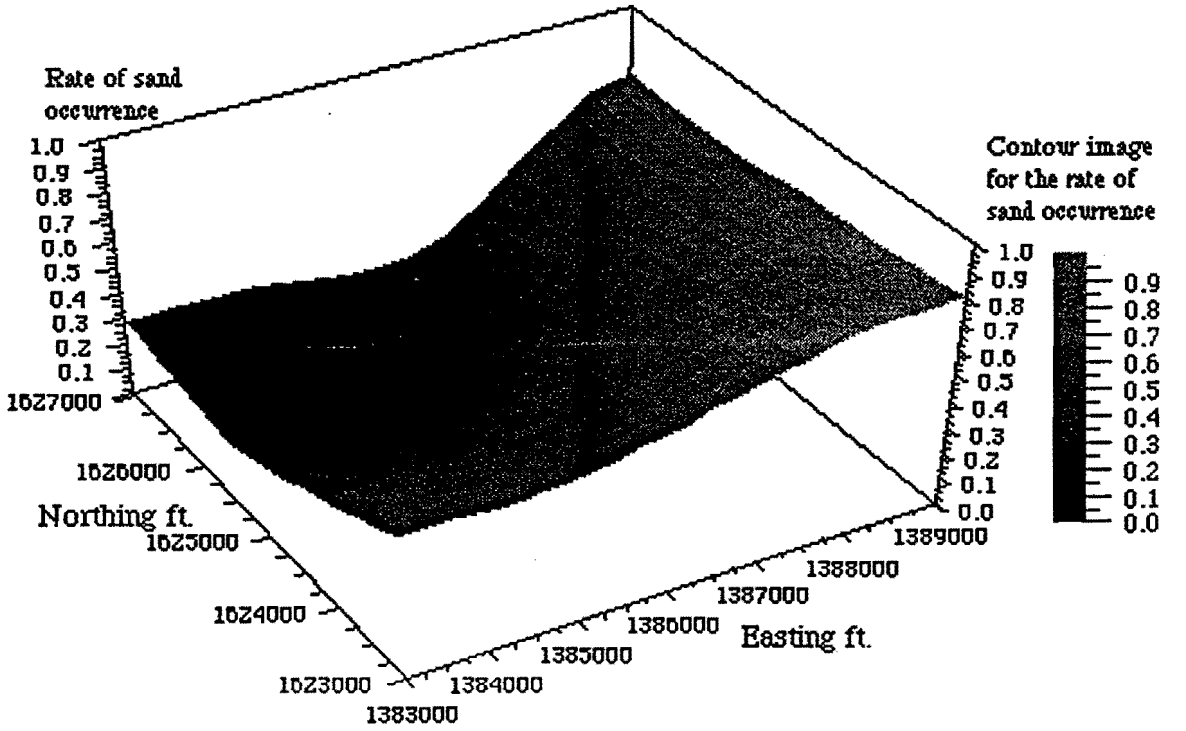



Figure 18. Northing-Easting perspective view for the rate of sand occurrence at Elevation = 4200 ft.

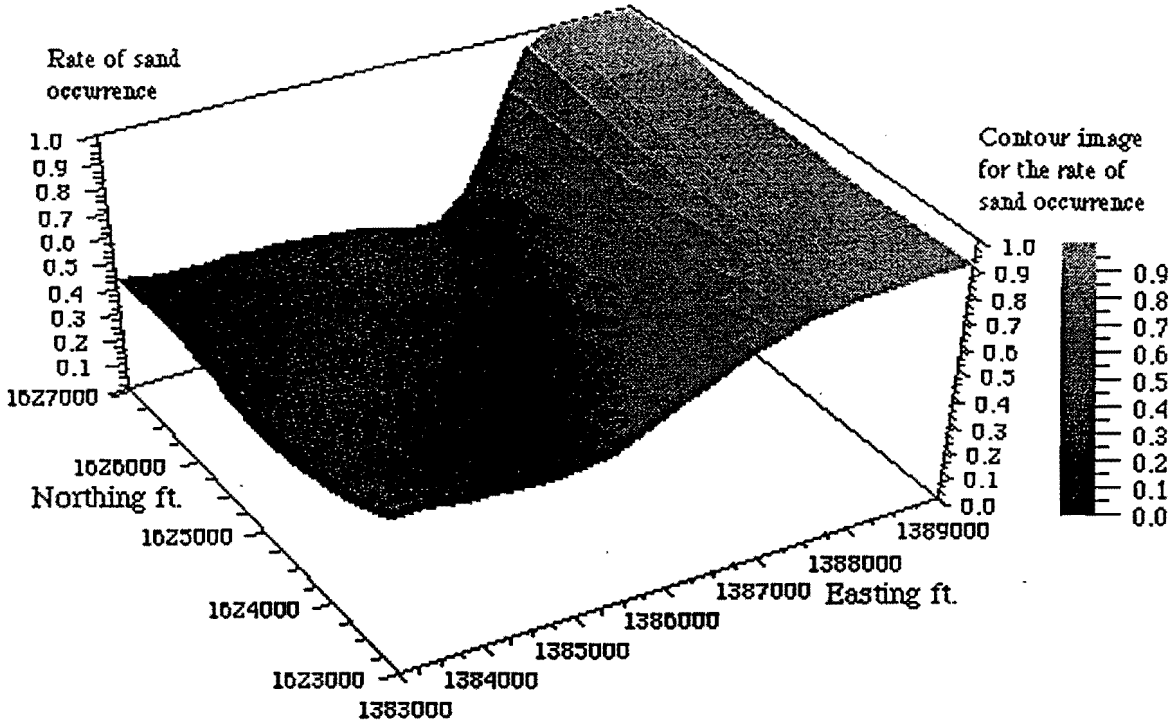



Figure 19. Northing-Easting perspective view for the rate of sand occurrence at Elevation = 4185 ft.

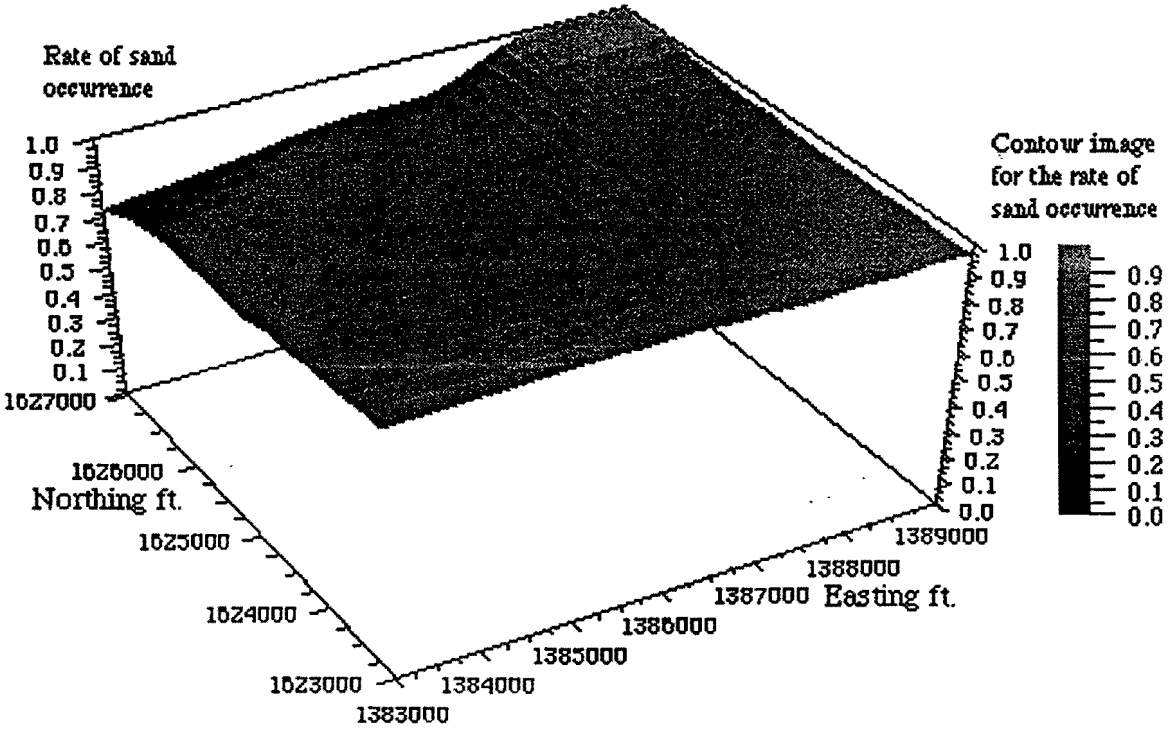



Figure 20. Northing-Easting perspective view for the rate of sand occurrence at Elevation = 4165 ft.

## References


Diggle, P. and Marron, J S. (1988), "Equivalence of smoothing Parameter Selectors in Density and Intensity Estimation", JASA, 83(403), 793-800.

Johnson, N. M. and Dreiss, S.J. (1989), "Hydrostratigraphic Interpretation Using Indicator Geostatistics", Water Resources Research, 25(12), 2501-2510.

Merriam, D. F., (1976), "Random Process in Geology", Springer-Verlag.

Scott, D. W., (1992), "Multivariate Density Estimation", A Wiley-Intersection Publication.

Silverman, B., W., (1986), "Density Estimation for Statistics and Data analysis", Chapman and Hall.

Trivedi, S. K., (1982), "Probability and Statistics with Reliability, Queuing, and Computer Science Applications", Prentice-Hall, INC.