Recognition of units in coarse, unconsolidated braided-stream deposits from geophysical log data with principal components analysis

Warren Barrash Center for Geophysical Investigation of the Shallow Subsurface, Boise State University, Boise, Idaho 83725
Roger H. Morin U.S. Geological Survey, Denver Federal Center, Denver, Colorado 80225

ABSTRACT

Returns from drilling in unconsolidated cobble and sand aquifers commonly do not identify lithologic changes that may be meaningful for hydrogeologic investigations. Vertical resolution of saturated, Quaternary, coarse braided-stream deposits is significantly improved by interpreting natural gamma (G), epithermal neutron (N), and electromagnetically induced resistivity (IR) logs obtained from wells at the Capital Station site in Boise, Idaho. Interpretation of these geophysical logs is simplified because these sediments are derived largely from high-gamma–producing source rocks (granitics of the Boise River drainage), contain few clays, and have undergone little diagenesis. Analysis of G, N, and IR data from these deposits with principal components analysis provides an objective means to determine if units can be recognized within the braided-stream deposits. In particular, performing principal components analysis on G, N, and IR data from eight wells at Capital Station (1) allows the variable system dimensionality to be reduced from three to two by selecting the two eigenvectors with the greatest variance as axes for principal component scatterplots, (2) generates principal components with interpretable physical meanings, (3) distinguishes sand from cobble-dominated units, and (4) provides a means to distinguish between cobble-dominated units.

INTRODUCTION

This paper addresses the general problem of recognizing, in boreholes, variation within coarse, unconsolidated alluvial sediments such as braided-stream deposits. The field characteristics of coarse braided-stream deposits have received considerable attention from examination of modern braided streams (e.g., Williams and Rust, 1969; Boothroyd and Nummedal, 1978), and of recent and ancient deposits in outcrop (e.g., Ore, 1964; Miall, 1984; Jussel et al., 1994). However, in subsurface investigations it is difficult and expensive to recover samples that retain information on variation within unconsolidated cobble and sand deposits. Conventional drilling methods mix and/or grind cobbles with the sand matrix. In a given well, it is difficult to recognize the location or nature of lithologic and associated hydrogeologic or geophysical variations within and between coarse-grained units. This information is valuable for investigations aimed at determining the range and distribution of permeability and for reconstructing sedimentary environments.

We interpret borehole geophysical log data from wells in the unconfined aquifer at the Capital Station site in downtown Boise, Idaho (Fig. 1). The aquifer consists largely of Quaternary alluvial (Qal) deposits above Tertiary Glenns Ferry Formation (Tgf) marginal lacustrine (shoreline–floodplain) sand and silt (Fig. 2). In this paper we focus on finding an objective process for recognizing lithologic variation within the saturated Quaternary alluvial deposits, which are coarse, unconsolidated, braided-stream, glacial outwash from the upper Boise River drainage (Othberg, 1994).

The Quaternary alluvial deposits have favorable characteristics for interpretation based on natural gamma (G), single-detector epithermal neutron (N), and electromagnetically induced resistivity (IR) logs. These deposits are virtually unlithified, unaltered, and uncompacted. Silt or clay lenses have not been observed, although minor amounts of silt and clay occur in some sand and cobble-dominated units (Barrash et al., 1997). The source material for the Quaternary alluvium is almost entirely Cretaceous and Tertiary granitic plutons, and minor Tertiary volcanic material and Paleozoic and Precambrian metasedimentary rocks (e.g., Mitchell and Bennett, 1979). Consequently, factors that can complicate interpretation of G, N, and IR logs in clastic deposits are limited in the Quaternary alluvium from the upper Boise River drainage: Sediment composition is nearly uniform; natural gamma production rates are well above background; cementation or mineralogical changes in place are minimal; and admixtures of clay (i.e., effects of increased presence of radioisotopes due to adsorption and ion exchange, bound water, and electrical conductivity) are minimal or concentrated in few units. As a first approximation, G log data in counts per second (cps) are related to the volume fraction of solids (V_s) which, in the Quaternary alluvium, are high-gamma–producing, clay-poor, uncemented sediment. N cps decrease with increasing porosity ($1 - V_s$) in saturated, clay-poor sediments (or increase with increasing V_s). IR log responses represent some combination of surface area and interconnected porosity of the bulk sediment.

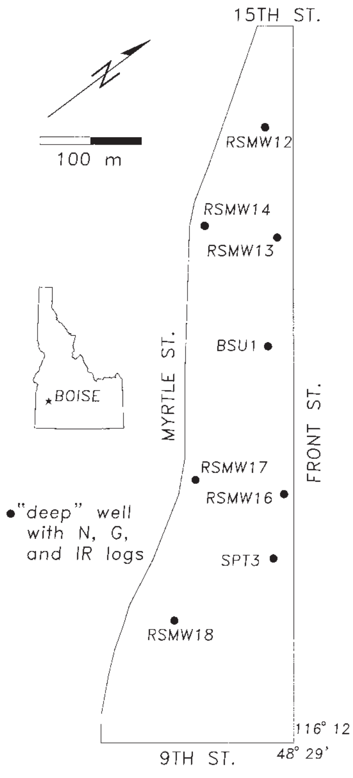

Figure 1. Capital Station site in Boise, Idaho; locations of eight (~20 m deep) monitoring wells investigated with natural gamma (G), epithermal neutron (N), and electromagnetically induced resistivity (IR) logs are shown.

BOREHOLE DATA

Lithologic logs for most monitoring wells at the Capital Station site (United States Pollution Control, Inc., 1992) are highly generalized (Fig. 3). Geophysical logs were obtained in 30 wells, but most of these penetrate only a few me-

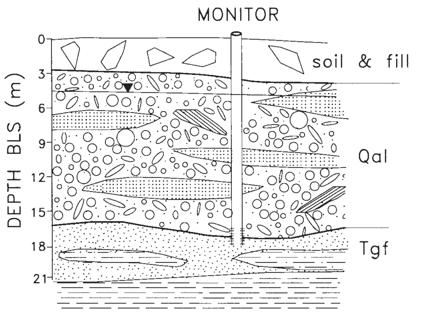

Figure 2. Schematic stratigraphic profile for Capital Station site. Units, from older to younger: Tgf—Tertiary Glenns Ferry Formation (marginal lacustrine) sand and silt; Qal—Quaternary coarse alluvial (braided-stream) deposits; soil & fill—recent soil and industrial fill. Deep monitoring wells generally extend into Tgf. BLS—below land surface.

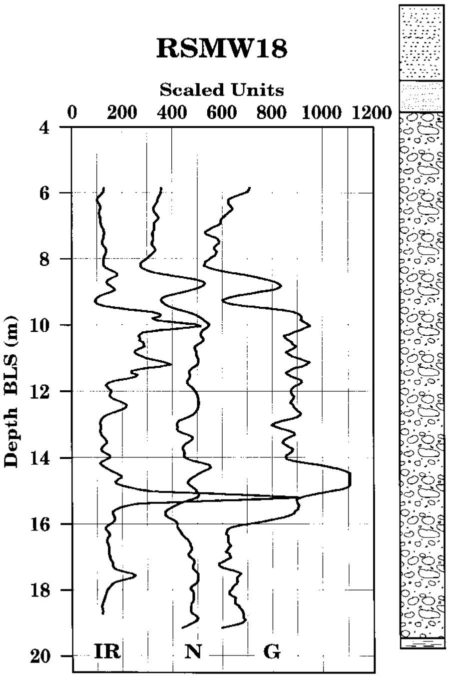

Figure 3. Schematic lithologic log based on drilling returns and geophysical log traces for well RSMW18. Water table was ~5 m below land surface when logs were run. Logs have been smoothed and linearly converted to scaled units for convenience in interpretation and presentation (electromagnetically induced resistivity [IR] values in Ω-m are multiplied by 1.5; epithermal neutron [N] values in cps are at par; natural gamma [G] values in cps are multiplied by 8). Lithologic log descriptions are generalized (United States Pollution Control, Inc., 1992); sand units (i.e., low N and G cps in geophysical logs at <8.4, 8.9–9.6, and >16.2 m) were not recognized during drilling. BLS—below land surface.

A *Variance-Covariance*

	PC1	PC2	PC3
G	1	0.57540	0.39508
N	0.57540	1	0.33169
IR	0.39508	0.33169	1

B *Eigenvalues*

	PC1	PC2	PC3
G	0.41815	0	0
N	0	0.70333	0
IR	0	0	1.87625

C *Eigenvectors*

	PC1	PC2	PC3
G	0.73992	-0.25908	0.62081
N	-0.66082	-0.45261	0.59872
IR	-0.12587	0.85324	0.50610

Figure 4. A: Variance-covariance matrix for original G (gamma log), N (epithermal neutron log), and IR (electromagnetically induced resistivity) data sets after each was transformed to population with zero mean and unit variance. Total system variance is 3.0. B: Eigenvalues from variance-covariance matrix indicate that total variance (still 3.0) is not distributed equally among three combined-variable eigenvectors (principal components, PC). C: Eigenvectors from variance-covariance matrix are defined by loadings of original variables.

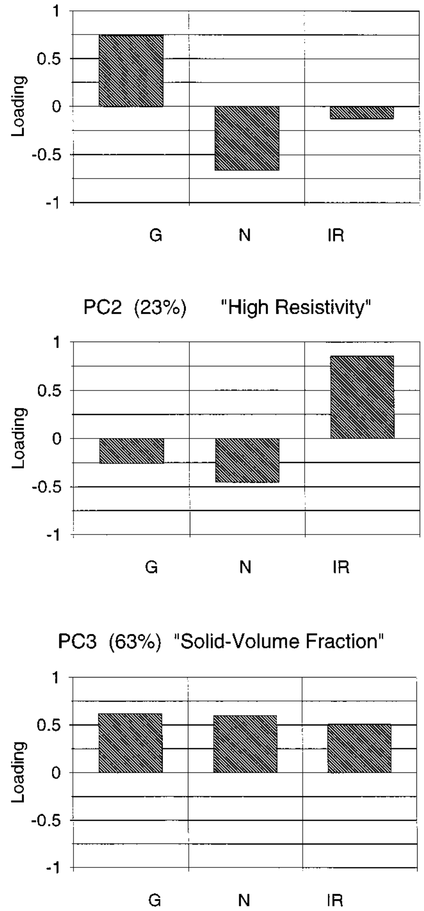

Figure 5. Relative loadings for, and percentages of variance explained by, each principal component (PC). Meanings of principal components, deduced from loadings and percentage of variance associated with each, are discussed in text. G is natural gamma, N is epithermal neutron, and IR is electromagnetically induced resistivity.

tres below the water table. Here we focus on log data obtained in the saturated zone, so the following analysis is limited to eight deeper wells (~17–22 m total depth) that were constructed with 5.1-cm-diameter PVC casing, and that have G, N, and IR logs (Fig. 1).

G and N tools have about the same volume of investigation, a 50-cm-diameter sphere. The IR tool has a toroidal volume of investigation with an ~60 cm central diameter and no response originating immediately at the well (Taylor et al., 1989; Williams et al., 1993). Geophysical log data were digitized every 6.1 cm; individual data points at specific depths include information from above and below due to volume averaging of each tool during data acquisition and smoothing with a moving filter. Typical logs are shown for well RSMW18 in Figure 3 next to the lithologic log for comparison. Sand units not recognized in the lithologic logs may be identified as the low-N and low-G intervals in the geophysical logs, on the basis of log responses checked against cores in recently drilled wells (Barrash et al., 1997). However, as might be expected with braided-stream deposits, the Quaternary alluvium sand or cobble-dominated units commonly cannot be correlated laterally between wells over even 6–10 m at this site.

PRINCIPAL COMPONENTS ANALYSIS

Principal components analysis is a multivariate data analysis technique that combines statistical and data transformation methods (e.g., Davis, 1986), and has been applied successfully in a number of borehole geophysical logging investigations (e.g., Kassenaar, 1991; Moline et al., 1992). Principal components analysis can provide insight into the relationships and variance of three or more variables by recombining the total variance in the system unequally among the principal components, and by reducing the dimensionality of the multivariate space if a subset of the principal components accounts for a large majority of the total variance in the system. In this case, the number of principal components can be reduced from three to two, which permits the examination of responses (that include information from all three original variables) in two-dimensional plots, and thereby simplifies interpretation of the multivariate nature of the material. In addition, the physical meaning of the principal components may be interpreted from the loadings of original variables on each principal component, and from the relative proportions of variance associated with each principal component.

We applied principal components analysis to G, N, and IR data for all Quaternary alluvium units below the water table in eight wells at Capital Station (Fig. 1). Our main objective was to gain resolution within the cobble-dominated intervals of the alluvium. Glenns Ferry Formation sediments were not included because (Barrash et al., 1997): (1) these units already may be distinguished from the alluvium on the basis of petrographic data, stratigraphic position, and separation of the alluvium and Glenns Ferry Formation sand-unit fields in N-G scatterplots, and (2) Glenns Ferry Formation sediments represent a different depositional environment and source area (Snake River drainage), and so would introduce elements of variability in addition to those inherent in the Quaternary alluvium deposits alone.

G, N, and IR responses were recorded independently over similar sample volumes and, although each measurement includes information from above and below, we assume that data values represent the material at the measurement location. This assumption leads to gradational rather than abrupt transitions between lithologic units but, without a priori knowledge of all unit breaks, unit selection would be arbitrary for that sizable fraction of the section where breaks are not obvious. So, the plotted location of each individual data point, or principal components analysis score, is not as important as locations of clusters of scores associated with fine-scale or aggregate units in principal component space.

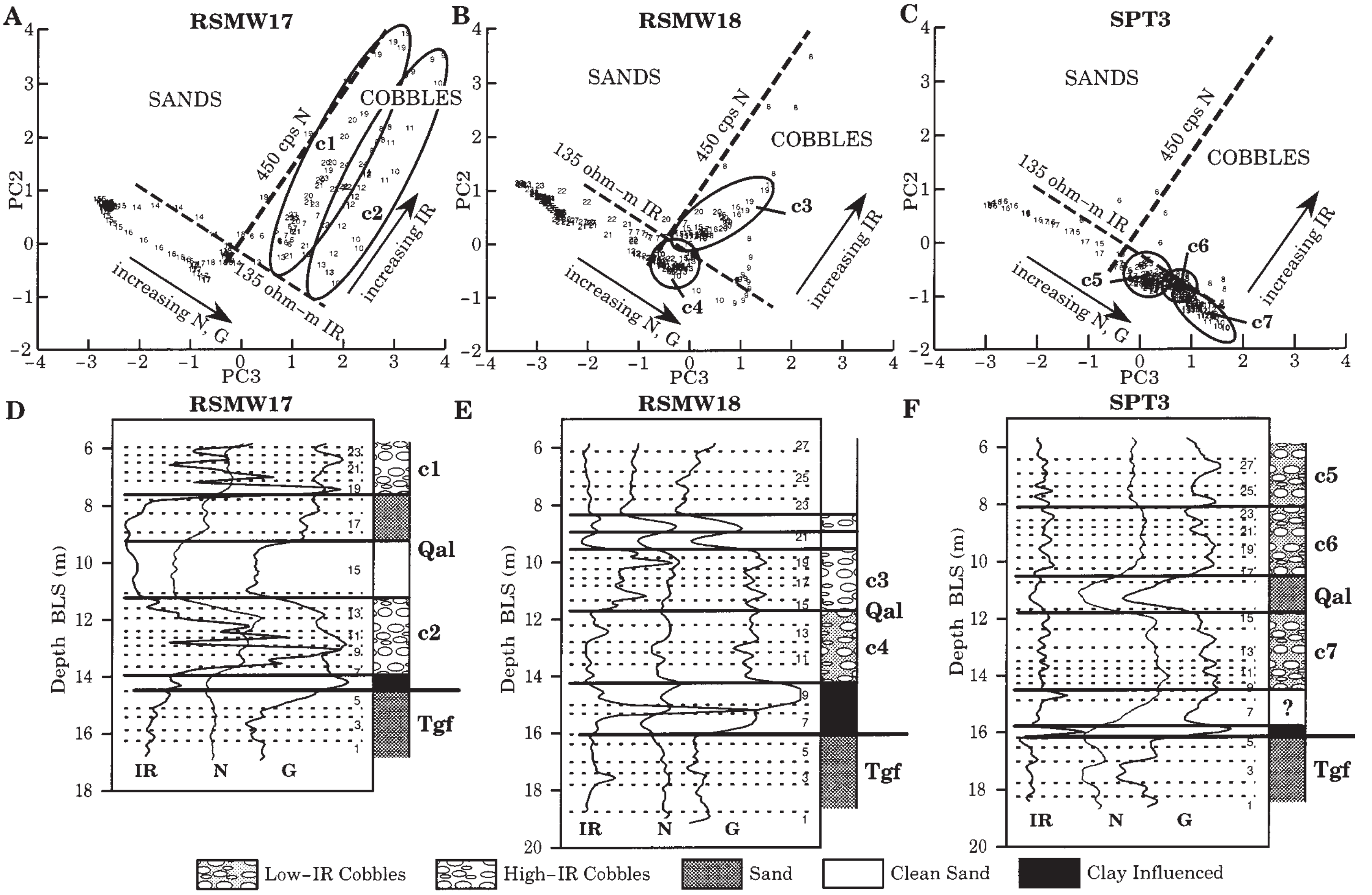

Figure 6. A–C: Principal component (PC3–PC2) scatterplots. D–F: Borehole geophysical log data with lithologic and stratigraphic units for three wells at Capital Station. Reference lines and arrows in scatterplots correspond to log responses in measured units. Data points in A–C are given as numbers of fine-scale units identified in log-lithology diagrams (D–F). Sand and cobble-dominated units plot in fields separated by 450 cps N line. Aggregate cobble units (c1–c7) are identified in scatterplots and log-lithology diagrams, and plot as high- or low-IR (electromagnetically induced resistivity) units relative to the 135 Ω-m IR line. G is gamma, N is epithermal neutron.

Data sets of the measured variables from individual wells were grouped together to form a sample population with 1320 trivariate data points for analysis with principal components analysis. The grouped data set for each variable was transformed to a population with zero mean and unit variance to avoid spurious numerical effects associated with different offsets on a given variable measurement scale. For example, G data are in cps and are offset from zero between 50 and 140 cps, and IR data are in Ω-m and are offset from zero between 30 and 700 Ω-m (Fig. 3).

In applying principal components analysis, first a variance-covariance matrix is generated from the three transformed data sets (Fig. 4A). The eigenvalues (Fig. 4B) and eigenvectors (Fig. 4C) are then determined from the variance-covariance matrix. The eigenvectors are the principal components of the multivariate system under analysis. The percentage of variance explained by each principal component (PC) can be determined by dividing the eigenvalue for that component by the total variance (3.0 in this case). Of the total variance in the system analyzed here, PC1 explains 14%, PC2 explains 23%, and PC3 explains 63%. An important feature of principal components analysis is that the total variance in the system is not changed: the sum of the three original data set variances (trace of matrix in Fig. 4A) equals 3.0, as does the sum of the eigenvalues (trace of matrix in Fig. 4B). The matrix of transformed variable data sets is postmultiplied by the eigenvector matrix to generate scores that may be plotted in principal component space.

Principal Component Axes

The three principal components define three orthogonal axes. The vector elements of each component locate each axis as a transformed-unit vector in the original three-variable space (Davis, 1986). A simple bar graph of the magnitudes of the vector elements, or loadings, for each principal component provides a visual means to evaluate the relative and combined influences of the original variables on each principal component (Fig. 5). The meaning of each principal component is interpreted below from its loadings and relative amount of variance. The order of discussion is from largest to smallest variance: PC3, PC2, then PC1.

Principal Component 3. PC3 encompasses most of the variance in the system (63%) and has strong positive contributions from G, N, and IR; G and N contribute almost equally (Fig. 5C). We interpret PC3 to be the "solid-volume fraction" (V_s) axis because increasing G, N, and IR data values all result from increased V_s of the high-energy, coarse-grained, braided-stream deposits. G and N loadings are nearly identical because both increase with increasing V_s of G-producing feldspathic sediments (i.e., poorer sorting, denser packing) relative to water-filled pore volume. IR typically increases with greater V_s of poorly conductive mineral matter because electrical current is less likely to pass directly through such solids than along mineral-water interfaces or through interconnected water-filled pore spaces (e.g., Kwader, 1985). However, the relationship between IR responses and V_s is less direct than it is for either G or N and V_s, and this difference is reflected in the lower loading for IR compared with G and N in PC3.

Principal Component 2. PC2, the "high resistivity" axis, accounts for 23% of the variance in the system and has a strong positive contribution of IR coupled with lesser negative contributions of N and G (Fig. 5). Observation of raw logs for wells at the Capital Station site indicates that high-IR values occur, with few exceptions, in the cobble-dominated units (Barrash et al., 1997). The meaning of high-IR responses in cobble units is under investigation; preliminary results suggest that large framework clasts (>15 cm largest diameter) are associated with high-IR units.

Principal Component 1. PC1 encompasses the least amount of variance in the system (14%) and includes a strong positive G contribution, a strong negative N contribution, and a minor negative contribution from IR. This type of inverse relationship between G and N is characteristic of units influenced by clay, where increased G is due to gamma-emitting elements taken into clays, and decreased N is due to water in clay-mineral lattices. Low IR is consistent with the electrically conductive nature of clays. The small proportion of variance associated with this clay axis reflects the small amount of clay in Quaternary alluvium (Barrash et al., 1997).

Principal Component Scatterplots

Because the relative percentages of variance explained by the three principal components are unequal such that a large majority of the variance may be explained by two principal components, then the dimensionality of the system may be reduced from three to two for simplification of plotting and interpretation. Together, PC3 (solid-volume fraction) and PC2 (high resistivity) explain 86% of the variance in the Quaternary alluvium deposits. Scores are plotted against these two axes for three wells as examples; the locations of scores in these principal component scatterplots are noted with numbers rather than points (Fig. 6, A–C). The numbers correspond to the fine-scale units associated with particular scores, and thereby facilitate finding their relative positions in, and the original data values from, the original logs (Fig. 6, D–F).

Sand units and cobble-dominated units plot in different fields on the PC3–PC2 scatterplots, as on N-G scatterplots (Barrash et al., 1997). The dividing line between sand and cobble-dominated units occurs at ~450 cps on the N log; this is not evident from the principal components analysis, but can be recognized by referring back to the locations and values of associated fine-scale units (e.g., Fig. 6, D–F) in the original logs. A dividing line between relatively high- and low-IR units occurs at ~135 Ω-m on the IR log. Sand units rarely occur above the 135 Ω-m IR dividing line, but cobble-dominated units commonly occur above this line (Fig. 6, A–C). In addition, relative positions for aggregate or fine-scale cobble units within the high- or low-IR fields can be associated with relative magnitudes of N cps (i.e., porosity) and, to a lesser extent, G cps. For example, the relative locations of three low-IR aggregate cobble units in well SPT3 on both the PC3–PC2 scatterplots (Fig. 6C) and the well logs (Fig. 6F) can be distinguished by relative N responses that are consistent within, but different between, these three aggregate units. Similarly, the relative positions of the two aggregate cobble units in well RSMW17 plot in different regions of the high-IR cobble field in the PC3–PC2 plots (Fig. 6A), more on the basis of relative magnitudes of N than G count rates (Fig. 6D).

SUMMARY AND CONCLUSIONS

1. Saturated, coarse, Quaternary alluvium braided-stream deposits originating in the upper Boise River drainage are particularly amenable to subsurface study with G, N, and IR geophysical logs, despite difficulty in recovering samples from boreholes and limited lateral or vertical continuity of units. The deposits are mostly unaltered, unlithified, and clay poor, and they originate from a largely uniform, high-gamma–producing source. These characteristics combine to simplify interpretation of material responses to the three logs run in wells at the Capital Station site: G and N approximate V_s, and IR responds to some combination of conducting surface area and interconnecting porosity.

2. Principal components analysis provides an objective means for interpreting G, N, and IR log responses to Quaternary alluvium deposits in Boise, Idaho. Interpretations for the principal components are as follows: 63% of total system variance is explained as solid-volume fraction; 23% is explained by a principal component that is sensitive to high resistivity; and 14% is explained by clay-influenced units.

3. Principal components analysis successfully separates (1) sand and cobble-dominated units across a line at ~450 N cps, and (2) cobble-dominated units into high- and low-IR units across a line at ~135 Ω-m. Additional variation between cobble units may be related to differences in porosity.

ACKNOWLEDGMENTS

Supported by U.S. Army Research Office grant DDAH04-94-G-0271. We thank S-Sixteen, Inc. and Anderson Associates Inc. for access to the Capital Station site. This paper was improved by reviews from G. Moline, P. Nelson, and S. Wood. Center for Geophysical Investigation of the Shallow Subsurface Contribution 0061.

REFERENCES CITED

Barrash, W., Morin, R. H., and Gallegos, D. M., 1997, Lithologic, hydrologic and petrophysical characterization of an unconsolidated cobble-and-sand aquifer, Capital Station site, Boise, Idaho: Proceedings of the 32nd Symposium on Engineering Geology and Geotechnical Engineering, March 26–28, 1997, Boise, Idaho, p. 307–323.

Boothroyd, J. C., and Nummedal, D., 1978, Proglacial braided outwash: A model for humid alluvial-fan deposits, *in* Miall, A. D., ed., Fluvial sedimentology: Canadian Society of Petroleum Geologists Memoir 5, p. 641–668.

Davis, J. C., 1986, Statistics and data analysis in geology (second edition): New York, John Wiley & Sons, 646 p.

Jussel, P., Stauffer, F., and Dracos, T., 1994, Transport modeling in heterogeneous aquifers: 1. Statistical description and numerical generation of gravel deposits: Water Resources Research, v. 30, p. 1803–1817.

Kassenaar, J. D. C., 1991, An application of principal components analysis to borehole geophysical data: Proceedings of the Fourth International Symposium on Borehole Geophysics for Minerals, Geotechnical and Groundwater Applications, August 18–22, 1991, Toronto, Ontario, p. 211–218.

Kwader, T., 1985, Estimating aquifer permeability from formation resistivity factors: Ground Water, v. 23, p. 762–766.

Miall, A. D., 1984, Variations in fluvial style in the lower Cenozoic synorogenic sediments of the Canadian Arctic Islands: Sedimentary Geology, v. 38, p. 499–523.

Mitchell, V. E., and Bennett, E. H., 1979, Geologic map of the Boise 1° × 2° quadrangle, Idaho: Idaho Bureau of Mines and Geology, scale 1:250 000.

Moline, G. R., Bahr, J. M., Drzewjecki, P. A., and Shepherd, L. D., 1992, Identification and characterization of pressure seals through the use of wireline logs: A multivariate statistical approach: Log Analyst, v. 34, p. 362–372.

Ore, H. T., 1964, Some criteria for recognition of braided stream deposits: University of Wyoming Contributions to Geology, v. 3, p. 1–14.

Othberg, K. L., 1994, Geology and geomorphology of the Boise Valley and adjoining areas, western Snake River Plain, Idaho: Idaho Geological Survey Bulletin 29, 54 p.

Taylor, K. C., Hess, J. W., and Mazzela, A., 1989, Field evaluation of a slim-hole borehole induction tool: Ground Water Monitoring Review, v. 9, p. 100–104.

United States Pollution Control, Inc., 1992, Interim hydrogeologic site investigation report, Front Street site, Boise, Idaho: Omaha, Nebraska, Report to Union Pacific Railroad, Project Number 96222, 31 p.

Williams, J. H., Lapham, W. W., and Barringer, T. H., 1993, Application of electromagnetic logging to contamination investigations in glacial sand-and-gravel aquifers: Ground Water Monitoring and Remediation Review, v. 13, p. 129–138.

Williams, P. F., and Rust, B. R., 1969, The sedimentology of a braided river: Journal of Sedimentary Petrology, v. 39, p. 649–679.

Manuscript received December 9, 1996
Revised manuscript received May 2, 1997
Manuscript accepted May 7, 1997