Why did Sudicky [1986] find an exponential-like spatial correlation structure for hydraulic conductivity at the Borden research site?

Robert W. Ritzi Jr.¹ and Richelle M. Allen-King²

Received 31 January 2006; revised 30 July 2006; accepted 30 August 2006; published 9 January 2007.

The exponential-like spatial correlation structure for hydraulic conductivity that Sudicky (1986) found at the Borden research site arises from hierarchical sedimentary architecture. The sediments can be described in terms of a hierarchy of stratal unit types wherein larger-scale unit types (hierarchical level II) are composed of smaller-scale unit types (hierarchical level I). There is a hierarchy of permeability populations corresponding to this stratal hierarchy. The shape and effective range of the composite semivariogram is mostly defined by how the proportion of lag transitions crossing these unit types grows as a function of lag distance. Lag transitions across level I units but within the same level II unit type and transitions across level II unit types contribute about equally in defining the composite sample semivariogram in the vertical direction. These two types of cross transitions have shorter and longer length scales, respectively. The composite semivariogram reflects additive contributions from the two length scales. The proportion of cross-transition lags grows with an exponential-like curve because the unit types at either level reoccur with a relatively high variability in length. The shape and range of the semivariogram are modeled from knowing only the univariate statistics for length among the unit types, and the sill is modeled from knowing univariate statistics for hydraulic conductivity populations among the unit types. Understanding the relationship between the semivariogram structure and these quantifiable attributes of the hierarchical stratal architecture helps in reducing the equivocal aspects of choosing a semivariogram model. It also helps in identifying bias in sample semivariograms and in assessing stationarity of second-order spatial bivariate statistics.


1. Introduction

Sudicky [1986] evaluated the Lagrangian-based macrodispersion theory [Dagan, 1986] using the 1982 natural gradient tracer test at the Borden research site. Macrodispersion theory relates the spread of a tracer plume to the spatial variation of log hydraulic conductivity, ln(K), as characterized by a two-point bivariate statistic like the centered covariance or the semivariogram. Sudicky [1986] made 1,279 measurements of ln(K) from sediment cores taken at the Borden site, spaced 1 m laterally and sampled in 5 cm vertical intervals. The spatial bivariate statistics computed from these data (as the centered covariance of Sudicky [1986] and as the semivariogram of Woodbury and Sudicky [1991]) had an asymptotic, exponential-like shape and were fitted with an exponential model. The spreading of the conservative tracer plume observed in the Borden natural-gradient tracer test was explained well by the macrodisperspersion theory using the fitted exponential models.

[1] This article addresses two questions. First, what aspects of the stratal architecture at the Borden site gave rise to the exponential-like spatial correlation in ln(K) which was observed? To address this question, we use information from newer sediment cores taken from the Borden aquifer from which two sets of colocated data were recorded [Divine, 2002; Allen-King et al., 2006]. One data set is a record of sedimentary unit types and the other is a record of ln(K). We analyze these two types of data together with new approaches for geostatistical analyses which link them [Ritzi et al., 2004; Dai et al., 2005].

[2] Sedimentary deposits typically comprise stratal unit types created at different spatial scales, under various processes of deposition and erosion, and organized within a hierarchical framework [Bridge, 2003, Figure 5.42]. Figure 1 depicts part of an exposure at the Borden site delineated with a hierarchy of unit types. Smaller-scale unit types have been defined based on textural attributes [Divine, 2002] and labeled as hierarchical level I unit types. These together form larger-scale unit types defined at hierarchical level II, which together, in turn, form the composite domain. Unit types can be defined based on various criteria (e.g., textural differences, or process of deposition), and are often defined based on combinations of criteria. Thus there typically are alternate hierarchies that can be defined for any specific deposit. The methodology used here does not depend on any specific

¹Department of Earth and Environmental Sciences, Wright State University, Dayton, Ohio, USA.
²Department of Geology, State University of New York at Buffalo, Buffalo, New York, USA.

Copyright 2007 by the American Geophysical Union.
0043-1397/07/2006WR004935$09.00
criteria for the hierarchical classification of unit types and thus can be applied to any of the hierarchies that might be adopted for a given deposit, as shown by Dai et al. [2005].

Ritzi et al. [2004] have shown that the composite \( ln(K) \) semivariogram is related to hierarchical aquifer architecture through the following expression, here written for the example in Figure 1. (We note that it may be written for any number of relevant levels and for any alternate hierarchy, as in the work by Dai et al. [2005].) Consider locations \( x \) and \( y \) which are separated by a lag vector \( h \). Consider that location \( x \) is within level I unit type \( o \) which, in turn, is within level II unit type \( r \) (the location will be referred to as in region type \( ro \) hereafter), and that \( y \) is within level I unit type \( i \) which is within level II unit type \( j \) (region type \( ji \)). Using this hierarchy of subscripts, the composite \( ln(K) \) sample semivariogram can be written with the designation of regions types in which the heads and tails of \( h \) occur:

\[
\hat{\gamma}(h) = \sum_{r} \sum_{o} \sum_{i} \sum_{j} \hat{\gamma}_{ro,oi}(h) \hat{p}_{ro}(h) \hat{t}_{ro,oi}(h)
\]

where the \( \hat{\gamma}_{ro,oi}(h) \) are the autovariogram and cross semivariograms [or = ji and or \( \neq ji \), respectively] defined by level II and level I unit type. The \( t_{ro,ji}(h) \) are the transition probabilities which give the fraction of lags of a particular distance and direction that transition to region type \( ji \) given they start in region type \( ro \). The proportions \( \hat{p}_{ro}(h) \) give the fraction of lags that start in region type \( ro \). We will refer to the four groups of terms on the RHS of equation (1) as follows: the first is the \( \alpha \alpha \) autotransition group of terms, corresponding to lags that are autotransitions at both levels II and I. The second group is the \( \alpha \chi \) cross-transition group (autotransitions at level II, cross transitions at level I). The third group is the \( \chi \alpha \) cross-transition group (cross transitions at level II, autotransitions at level I), and the forth group is the \( \chi \chi \) cross-transition group (cross transitions at both levels II and I).

The contribution of each group in defining the composite sample semivariogram can be studied as by Ritzi et al. [2004] using data representing a fluvial point bar deposit, and by Dai et al. [2005] using data from an alluvial deposit. In both examples, the shape and the range of the composite semivariogram were mostly determined by the cross-transition groups, with negligible contribution from the \( \alpha \alpha \) group. The shape and range of the cross-transition groups were, in turn, mostly determined by the \( \hat{p}_{ro}(h) \hat{t}_{ro,ji}(h) \) terms, which represent the fraction of lags of a specific distance and direction that are of a particular transition type (i.e., the fraction with tail in region type \( ro \) and head in region type \( ji \)). Note that the \( \hat{p}_{ro}(h) \hat{t}_{ro,ji}(h) \) terms are defined entirely by the stratal architecture and are independent of \( ln(K) \). Under these conditions, the \( \hat{\gamma}_{ro,ji}(h) \) were approximated well using the mean \( \hat{m}_{ro} \) and the variance \( \hat{\sigma}_{ro}^2 \) for \( ln(K) \) subpopulations within the following expression for the composite semivariogram:

\[
\hat{\gamma}(h) = \sum_{r} \sum_{o} \sum_{i} \sum_{j} \frac{1}{2} \left[ \hat{\sigma}_{ro}^2 + \hat{\sigma}_{ri}^2 + \left( \hat{m}_{ro} - \hat{m}_{ri} \right)^2 \right] \hat{p}_{ro}(h) \hat{t}_{ro,oi}(h)
\]

\[
+ \sum_{r} \sum_{j} \sum_{i} \sum_{o} \frac{1}{2} \left[ \hat{\sigma}_{ro}^2 + \hat{\sigma}_{jo}^2 + \left( \hat{m}_{ro} - \hat{m}_{jo} \right)^2 \right] \hat{p}_{ro}(h) \hat{t}_{ro,ji}(h)
\]

\[
+ \sum_{r} \sum_{j} \sum_{i} \sum_{o} \frac{1}{2} \left[ \hat{\sigma}_{ro}^2 + \hat{\sigma}_{ji}^2 + \left( \hat{m}_{ro} - \hat{m}_{ji} \right)^2 \right] \hat{p}_{ro}(h) \hat{t}_{ro,ji}(h)
\]

\[
\cdot \hat{p}_{ro}(h) \hat{t}_{ro,ji}(h)
\]

\[\text{(2)}\]

Figure 1. Hierarchy of unit types as exposed in a trench excavated at the Borden site. After field notes by D. Gaylord (personal communication, 2000).
In section 3, we present the analysis of lithologic and ln(K) data collected recently at the Borden site [Divine, 2002; Allen-King et al., 2006] using equations (1) and (2). In doing so, we analyze how the shape and the range of the spatial correlation structure for ln(K) are related to the stratal architecture at the Borden site.

The second question we address is: Does an understanding of the relationship between the ln(K) semivariogram and the stratal architecture provide a rational basis for modeling the semivariogram? If so, it may help us to avoid equivocal choices among competing models (e.g., Gaussian, spherical, or exponential as by Turcke and Kueper [1996]). Without having an underlying process associated with models, the traditional curve fitting exercise just provides a description of the sample semivariogram. Can we bring reason to the structure by identifying an underlying process which provides a physical reason for selecting a model? If the \( p(h) \) and \( t(h) \) mostly govern the shape of the composite semivariogram, as in equation (2), then modeling the semivariogram can be based upon developing a model for the \( p(h) \) and \( t(h) \). Methods for modeling \( t(h) \) from information about stratal architecture have been discussed by Carle [1998], Ritzi [2000], and Dai et al. [2005]. In section 4 we use ideas from these methods and the new lithologic data collected at the Borden site to model \( t(h) \) and, in turn, the \( p(h) \) and the composite semivariogram. In this approach, the \( t(h) \) model is developed from univariate length statistics characterizing the unit types within the stratal hierarchy. The composite semivariogram model is developed independent from the sample semivariogram and without curve fitting.

2. Review of Newer Data From the Borden Site

Lithologic data have been collected in accompany with the collection of permeability data in new studies at the Borden site [Divine, 2002; Allen-King et al., 2006]. The data come from eleven cores collected at the same depth interval (1.5–3.0 m below ground surface) and adjacent to the location of the Stanford-Waterloo (SW) natural-gradient tracer test, as shown in Figure 2. The cores are spaced 1 m apart, as were those taken by Sudicky [1986]. Permeability was measured in 1.5 cm samples taken from these cores (a higher vertical resolution than the sampling by Sudicky). Sudicky’s analyses were presented for ln(K) defined at 22 degrees Celsius. Accordingly, we multiplied permeability by the ratio of specific weight to viscosity for water at 22 degrees Celsius, and analyzed ln(K) for the majority of the analyses below.

The sediments collected in the cores were categorized, based on a set of quantifiable textural attributes, into ten stratal unit types (see Divine [2002, p. 77] for the classification criteria). These are listed as level I unit types in Table 1. As an aside, the Borden aquifer sediment was interpreted as a lakeshore deposit by Bohla [1986], and the planar laminations of the DPL and FPL unit types, which together occur with largest proportion, were related to sediment sorting in beach swash zones. Because reliable interpretations of the depositional history specific to each level I unit type are lacking, depositional categories are not adopted here.

Different level I unit types that have common mean grain size are commonly found adjacent and thus form larger-scale regions that can be delineated based on grain size, as depicted in Figure 1. Accordingly, we further grouped the ten level I unit types into three broader textural groups defined by mean grain size: medium sand, fine sand, and silt. These are listed under level II unit types in Table 1. In dividing the sediments this way, we create a hierarchical framework for subdividing space. The spatial distribution of unit types at each level is shown in Figure 3. The hierarchy of ln(K) populations corresponding to the stratal hierarchy is

---

**Table 1. Hierarchical Classification of Sediments Based on Textural Characteristics Quantified by Divine [2002]**

<table>
<thead>
<tr>
<th>Unit Type</th>
<th>Level II</th>
<th>Level I</th>
<th>Level I Code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unit Name</td>
<td>Unit Code</td>
<td></td>
</tr>
<tr>
<td>Medium sand</td>
<td>high-angle planar cross stratified</td>
<td>HPXS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>massive coarse grained</td>
<td>MCG</td>
<td></td>
</tr>
<tr>
<td>Fine sand</td>
<td>low-angle planar cross stratified</td>
<td>LPXS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>faint planar cross stratified</td>
<td>FPCXS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>complex planar cross stratified</td>
<td>CPXS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>stratified</td>
<td>XGS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>distinct plane laminated</td>
<td>DPL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>faint plane laminated</td>
<td>FPL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>massive fine grained</td>
<td>MFG</td>
<td></td>
</tr>
<tr>
<td>Silt</td>
<td>silt</td>
<td>Z</td>
<td></td>
</tr>
</tbody>
</table>
shown in Figures 4a and 4b. Univariate statistics for these populations are given in Tables 2a and 2b. Divine [2002] showed that mean ln(K) is statistically significantly different among many of the level I unit types. Note that the existence of (statistically) significantly different means among unit types is not assumed in equation (1) and is not strictly required for the general methodology we use here, but makes it more likely that equation (2) will represent the composite semivariogram. A decomposition of the composite variance according to the two-level stratal hierarchy is given in Appendix A.

3. Analysis
3.1. Vertical Direction

[12] The composite vertical semivariogram was computed first [Deutsch and Journel, 1998, equation III.1]. It is plotted in Figure 5a and it has, in general, an exponential-like shape. The ln(K) data were then divided into subpopu-
ations according to the stratal hierarchy and used to compute each $\hat{\gamma}_{\alpha,\beta}(h)$. An integer-indicator data set, colo-
cated with the $ln(K)$ data, was used to compute each $p_{\alpha}(h)$
and $i_{\alpha,\beta}(h)$. The 100 terms of equation 1 were then
computed. The hierarchy adopted for the Borden site gives
rise to 10 terms in the $\alpha\alpha$ group, 42 terms in the $\alpha\chi$ group,
0 terms in the $\alpha\chi$ group, and 48 terms in the $\chi\chi$ group.
Because the hierarchy does not give rise to $\chi\alpha$ terms, that
group will not be referred to hereafter. The sum of all terms,
and the sum of the $\alpha\alpha$, $\alpha\chi$, and $\chi\chi$ groups of terms are also
plotted in Figure 5a. The sum of all terms, by definition,
exactly matches the composite sample semivariogram.

[13] The plot for each group of terms individually in
Figure 5a indicates the group’s contribution to the com-
posite sample semivariogram (i.e., to the total sum in
equation (1)). With the exception of the two shortest lag
distances, the $\alpha\alpha$ autotransition group contributes in a
minor way to the composite semivariogram. The composite
semivariogram has its shape (i.e., the exponential-like structure)
and its correlation range determined to a greater
extent by the $\alpha\chi$ and $\chi\chi$ cross-transition groups. The sum
of the $\alpha\chi$ and $\chi\chi$ groups together (all cross terms) is also
plotted in Figure 5a. There is fairly small and constant
difference between the sum of all cross-transition terms and
the composite sample semivariogram. Thus the shape and
correlation range of the composite sample semivariogram are
essentially determined by cross transitions. The $\alpha\chi$
and $\chi\chi$ cross transitions each contribute fairly equitably.

[14] A better understanding of cross transitions comes
from a study of the sum of $p_{\alpha}(h)$ for $\alpha\chi$ cross
transitions, $\chi\chi$ cross transitions, and all cross transitions,
each plotted in Figure 5b. Each curve defines the fraction of
lag pairs, for a given lag distance, which are of a particular
cross-transition type. Each of the $\alpha\chi$ and $\chi\chi$ curves has an
exponential-like growth with lag. The curve for all cross
transitions is the sum of these two exponential-like curves
with different ranges and is, itself, asymptotic and exponen-
tial-like. Over 80% of lags are cross transitions after
about 0.2 m lag distance.

[15] For any lag distance, the fraction of $\alpha\chi$ lags is
greater than the fraction of $\chi\chi$ lags. Yet, there are greater
$ln(K)$ differences between the tails and heads of $\chi\chi$ lags as
compared to $\alpha\chi$ lags. The net result is that the $\chi\chi$ group in
Figure 5a contributes about equally with the $\alpha\chi$ group in
defining the composite semivariogram.

[16] The curve in Figure 5b for $\chi\chi$ cross transitions has a
greater range than the curve for the $\alpha\chi$ cross transitions.
This reflects the fact that $\chi\chi$ lags span larger-scale level II
unit types. Furthermore, the shape and range of the curves
in Figure 5a for the $\alpha\chi$ and $\chi\chi$ groups closely follow the
shape and range of the respective $\alpha\chi$ and $\chi\chi$ curves in
Figure 5b. The longer correlation range exhibited by the $\chi\chi$
group in Figure 5a corresponds directly to the behavior of
the $\chi\chi$ curve in Figure 5b. The shorter correlation range
exhibited by the $\alpha\chi$ group in Figure 5a corresponds directly
to the behavior of the $\alpha\chi$ curve in Figure 5b. Thus there are
two correlation structures, $\alpha\chi$ and $\chi\chi$, with two different
ranges, lumped together within the composite sample semi-
variogram. Furthermore, these two correlation structures are
strongly related to how the fractions of $\alpha\chi$ and $\chi\chi$ cross
transitions grow with lag distance.

3.2. Semivariogram Structure Controlled by the
Stratal Architecture

[17] The $p_{\alpha}(h)$ for cross transitions and the uni-
variate statistics for $ln(K)$ given in Tables 2a and 2b were
used to compute equation (2). This result is also plotted in
Figure 5a for comparison to the composite sample semi-
variogram. Equation (2) slightly underrepresents the sill
and range of the composite sample semivariogram. The differ-
ences between equation (2) and the composite sample semi-
variogram exist because the small contributions of $\alpha\alpha$
terms are not represented in equation (2), and the
$\hat{\gamma}_{\alpha,\beta}(h)$ are approximated with constants in equation (2).
The small differences become even smaller if we remove a
vertical trend from the $ln(K)$ data.

[18] Allen-King et al. [2006] showed that a statistically
significant vertical trend exists in $ln(K)$ with slope of
$-0.61$ $ln$(cm/s) per meter depth. As shown in Figure 6a,
the result is that mean $ln(K)$ is not stationary with lag among
the tails and heads of lag pairs. When this trend is removed
from the data the mean $ln(K)$ becomes relatively stationary
in the heads and tails of lag pairs, as shown in Figure 6b,
and the unit types still have distinct $ln(K)$ subpopulations
as shown in Figures 4c and 4b. The composite sample
semivariogram computed from the detrended data is shown
in Figure 5c. The basic structure of the composite sample
semivariogram has not changed much; the effect of de-
trending is to flatten the curve beyond 0.3 m lag and to
reduce the sill. Without the trend, the variate statistics in
equation (2) are a better approximation for the $\hat{\gamma}_{\alpha,\beta}(h)$ and,
as a result, equation (2) becomes an even better represen-
tation of the composite semivariogram.

[19] The shape and range of equation (2) are entirely
determined by the $p_{\alpha}(h)$ for $\alpha\chi$ transitions. The results of com-
paring equation (2) to equation (1) in Figures 5a and 5c show
that much of the composite semivariogram structure can be inferred by knowing just the $p_{\alpha}(h)$ for $\alpha\chi$ transitions. We
emphasize that the $p_{\alpha}(h)$ for $\alpha\chi$ transitions are defined by the stratal
architecture, independent of the $ln(K)$ values. The $ln(K)$

<p>| Table 2a. The $ln(K)$ Univariate Statistics for Level II Unit Types |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Level II Unit</th>
<th>Number of Samples</th>
<th>Proportion of Samples</th>
<th>Mean $ln(K)$</th>
<th>Variance $ln(K)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium sand</td>
<td>179</td>
<td>0.21</td>
<td>$-2.90$</td>
<td>0.37</td>
</tr>
<tr>
<td>Fine sand</td>
<td>649</td>
<td>0.77</td>
<td>$-3.47$</td>
<td>0.27</td>
</tr>
<tr>
<td>Silt</td>
<td>15</td>
<td>0.02</td>
<td>$-4.56$</td>
<td>0.38</td>
</tr>
</tbody>
</table>

<p>| Table 2b. The $ln(K)$ Univariate Statistics for Level I Unit Types |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Level I Unit</th>
<th>Number of Samples</th>
<th>Proportion of Samples</th>
<th>Mean $ln(K)$</th>
<th>Variance $ln(K)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium sand</td>
<td>179</td>
<td>0.21</td>
<td>$-2.95$</td>
<td>0.40</td>
</tr>
<tr>
<td>Fine sand</td>
<td>649</td>
<td>0.77</td>
<td>$-3.18$</td>
<td>0.34</td>
</tr>
<tr>
<td>Silt</td>
<td>15</td>
<td>0.02</td>
<td>$-4.56$</td>
<td>0.38</td>
</tr>
</tbody>
</table>

5 of 13
Two additional observations reinforce the notion that the \( \rho(h) \) terms mostly define the shape and range of the composite semivariogram structure, and that \( \ln(K) \) values mostly define the sill. First, Woodbury and Sudicky [1991] observed that the Sudicky [1986] data have some low \( \ln(K) \) values (< 6.5) referred to as outliers, and that these outliers had an influence on the semivariogram. We see from our results that the influence of such outliers should be mostly on the sill, rather than the shape or range of the semivariogram. Removing the outliers does not change the architecture of unit types and should not affect the shape or range of the semivariogram. Removing the outliers should affect the univariate mean and variance of \( \ln(K) \) among unit types and so should affect the sill. Indeed, Figures 3–6 of Woodbury and Sudicky [1991] show that the semivariogram shape and range are essentially the same with and without including the outliers. Removing the outliers decreases the sill.

Second, the permeability (\( k \), without log transformation) and the \( \ln(K) \) semivariograms have almost identical shape and range (compare Figure 5d to Figure 5a), as if they are linearly related. Yet they are not linearly related (the natural log transform is a smoothing filter). A linear-like relationship between the \( k \) semivariogram and the \( \ln(k) \) or \( \ln(K) \) semivariograms (which are equal) would only be expected a priori under a small \( \ln(K) \) variance condition, and the condition is not met with the Borden data. To explain, let \( \alpha = \ln(k) \), with mean \( m_\alpha \), and let \( \alpha' \) be a perturbation term such that \( \alpha = m_\alpha + \alpha' \). A simple derivation shows that the \( k \) semivariogram will equal \( e^{2m_\alpha} \) times the \( \ln(K) \) semivariogram if the variance is small enough so that the series for exponential functions can be truncated as \( e^{\alpha'/2} \). The data give \( e^{2m_\alpha} = 1.13 \times 10^{-13} \), and the ordinate values of the \( k \) semivariogram (Figure 5d) are not \( e^{2m_\alpha} \) times those of the \( \ln(K) \) semivariogram (Figure 5a). Thus the linear-like relationship between semivariograms is not explained by small variance. The ordinate values differ by the ratio variance(\( k \))/variance(\( \ln(K) \)), which has a value twice as large as \( e^{2m_\alpha} \). Furthermore, the linear-like relationship between semivariograms has been observed in an example with even larger variance [Ritzi et al., 2004, p. 9]. The linear-like relationship can be explained, without assuming small variance, through the fact that the shape and range of the semivariograms are largely controlled by the \( \rho(h) \) terms and the architecture of the unit types, as in equation (2), independent of the \( k \) or \( \ln(K) \) values themselves. The \( k \) or \( \ln(K) \) values act mostly to scale the ordinate axes of the semivariograms with an ordinate ratio of...
variance($k$)/variance($\ln(K)$), as is true when computing equation (2) with univariate statistics for $k$ or $\ln(K)$.

[22] The comparison of equation (2) to equation (1) in this section showed that the exponential-like shape and the range of the composite vertical semivariogram are determined mostly by the underlying aquifer architecture, as characterized by the $\hat{p}_n(h)\hat{u}_{\text{ro,ji}}(h)$ terms. Correspondingly, Figures 5a, 5c, and 5d together showed that changing among permeability and $\ln(K)$ data, with or without log transformation or detrending, mostly affects the scaling on the ordinate axis of the composite semivariogram. Changing among them had little effect on the shape or range.

[23] As an aside, we are not suggesting this result is expected at every site. For example, another site could have a much stronger trend, strong enough to give rise to $\hat{u}_{\text{ro,ji}}(h)$ that are so nonstationary that they overpower the influence of the $\hat{p}_n(h)\hat{u}_{\text{ro,ji}}(h)$ terms. The methodology presented here can be used to study these relationships. The results thus far from other examples [Ritzi et al., 2004; Dai et al., 2005] have been consistent with the results presented here for the Borden site in that the $\hat{p}_n(h)\hat{u}_{\text{ro,ji}}(h)$ terms were found to mostly define the shape and range of the hydraulic conductivity correlation structure. However, many well-studied examples are needed in order to understand the spectrum of possibilities in nature.

[24] In summary, we can say that Sudicky [1986] found exponential-like vertical spatial correlation for hydraulic conductivity at the Borden site because of the aquifer architecture. The sedimentary deposit can be delineated into a hierarchy of stratal unit types with a corresponding hierarchy of $\ln(K)$ modes. The exponential-like growth in transitions across the unit types with increasing lag, as represented by $\hat{p}_n(h)\hat{u}_{\text{ro,ji}}(h)$ terms, is the basic structure underlying the exponential-like shape of the $\ln(K)$ semivariogram. These terms are determined entirely by the stratal architecture and are independent of the $\ln(K)$ values. The $\ln(K)$ values in the heads and tails of cross transitions act mostly to scale the ordinate axis and thus define the sill.

4. Semivariogram Modeling Based on Univariate Statistics for Thickness and Permeability

[25] The results suggest that we can model the composite semivariogram through modeling the structure exhibited by $\hat{p}_n(h)\hat{u}_{\text{ro,ji}}(h)$ terms summed according to $\alpha\chi$ and $\chi\chi$ cross transitions. One could simply fit the curves in Figure 5b with standard models from geostatistics, but we would like to move away from curve fitting and adopt methods that are more closely linked to quantifiable attributes of the stratal architecture. Methods for modeling the $\hat{u}_{\text{ro,ji}}(h)$ in these terms, based upon quantifiable attributes of the stratal architecture, have been discussed by Carle [1998], Ritzi [2000], and Dai et al. [2004, 2005]. We combine several ideas from these methods and use stratal thickness data in

![Figure 5](image-url)
order to model the $i_{o,i}(h)$ and, in turn, the $p_h i_{o,i}(h)$ groups and the composite semivariogram.

[26] To develop the approach, we temporarily decouple the hierarchical levels and consider each independently, starting with level I. The $\alpha\chi$ transitions are from unit type $o$ to unit type $i$ (both within the same level II unit type). The shape and range of individual $i_{o,i}(h)$ are assumed to be mostly influenced by the proportions, mean length, variance in length of the level I unit types. Ritzi [2000] showed that sample autotransition probabilities, $i_{o,o}(h)$, can be largely influenced by the variance in length and take on an asymptotic, exponential-like shape if the coefficient of variation for unit length ($cv$, defined as the standard deviation in length normalized by mean length, $l$) among the unit types is close to unity (as is true by definition for an exponentially distributed population of lengths). This was determined in heuristic studies in which the variance in unit lengths was systematically increased among the populations of each unit type. If exponential, transition probabilities can be expressed as:

$$t_{o,i}(h) = p_i + (\delta_{o,i} - p_i) \exp(-3h/a_{o,i})$$  \(3\)

where $\delta_{o,i}$ is the Kronaker delta and $a_{o,i}$ is the effective range. (For semivariogram or transition probability models which reach their sills asymptotically, the effective range is defined as the distance at which the model reaches 0.95 the value of the sill [Deutsch and Journel, 1998]. Only asymptotic models are discussed here and we will simply refer to it as the range hereafter). Furthermore, if unit types occur in equal numbers the range for exponential $t_{o,o}(h)$ was shown to be

$$a_{o,o} = 3\tilde{l}_o(1 - p_o)$$  \(4\)

Probability laws dictate that

$$\sum_i t_{o,i}(h) = 1$$  \(5\)

[Carle and Fogg, 1996]. Following Dai et al. [2005], substituting (3) into (5) gives

$$\sum_i t_{o,i}(h) = 1 + \sum_i p_i \left(\exp(-3h/a_{o,i}) - \exp(-3h/a_{o,i})\right)$$  \(6\)

which shows that for the exponential model system represented in equation (3), probability laws expressed in equation (5) are honored if $a_{o,o} = a_{o,i}$ i.e., the ranges for autotransition and cross-transition probabilities are equal. The $a_{o,o}$ computed with equation (4) are given in Table 3a. They are not strictly equal among level I unit types because the unit types do not occur in equal numbers. Dai et al. [2005] showed that a parsimonious model and good approximation for the system was created by using, in equation (3), the proportion-weighted average of $\tilde{a}_{o,o}$. Here that average is equal to 0.18 m, which, accordingly, is a very good representation for the range of the $\alpha\chi$ curve plotted in Figure 5b.
### Table 3a. Thickness Statistics for Level I Unit Types

<table>
<thead>
<tr>
<th>Level I Unit</th>
<th>Number of Occurrences</th>
<th>Proportion of Occurrences</th>
<th>Mean Thickness, m</th>
<th>(a_{\alpha r})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium sand</td>
<td>HPXS 20</td>
<td>0.10</td>
<td>0.12</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>MCG 13</td>
<td>0.07</td>
<td>0.04</td>
<td>0.12</td>
</tr>
<tr>
<td>Fine sand</td>
<td>LPXS 13</td>
<td>0.05</td>
<td>0.07</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>FPXS 6</td>
<td>0.03</td>
<td>0.08</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>CPXS 27</td>
<td>0.14</td>
<td>0.05</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>XSS 15</td>
<td>0.08</td>
<td>0.04</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>DPL 33</td>
<td>0.17</td>
<td>0.07</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>FPL 48</td>
<td>0.24</td>
<td>0.07</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>MFG 22</td>
<td>0.11</td>
<td>0.07</td>
<td>0.19</td>
</tr>
<tr>
<td>Silt</td>
<td>Z 5</td>
<td>0.03</td>
<td>0.05</td>
<td>0.13</td>
</tr>
</tbody>
</table>

[27] Focusing now on level II, the transition probabilities at level II from unit type \(r\) to unit type \(j\) are related by definition to \(t_{r \rightarrow j}(h)\) through

\[
t_{r \rightarrow j}(h) = \sum_r \sum_j \frac{t_{r \rightarrow j}(h)p_{r \rightarrow j}}{p_r} \tag{7}
\]

[Ritzi et al., 2004]. An exponential system of transition probability models can be written for level II as

\[
t_{r \rightarrow j}(h) = p_j + (\delta_{r,j} - p_j) \exp(-3h/\alpha_r) \tag{8}
\]

Following from the equivalent forms of (5) and (6) written for level II, the range can be written as

\[
\alpha_r = \alpha_{r,j} = 3\alpha_r(1 - p_r) \tag{9}
\]

The \(\alpha_r\) are computed for level II unit types in Table 3b. The proportion-weighted average is 0.31 m, which, accordingly, is a very good representation for the range of the \(\chi\chi\) curve plotted in Figure 5b.

[28] The \(c_r\) for level II and level I unit types are 1.3 and 0.87, respectively. Given that the unit types at each hierarchical level have a \(c_r\), approaching unity, we indeed choose exponential models for the cross-transition probabilities (following Ritzi [2000]) and substitute them into equation (2) giving a model in the form:

\[
\gamma(h) = \sum_r \sum_s \sum_{i,j} \frac{1}{2} \left[ \sigma^2_{r,s} + \sigma^2_{i,j} + (m_{r,s} - m_{i,j})^2 \right] \cdot p_{r \rightarrow s}p_{i \rightarrow j} (1 - \exp(-3h/\alpha_r)) + \sum_r \sum_s \sum_{i,j} \sum_{i',j'} \frac{1}{2} \left[ \sigma^2_{r,s} + \sigma^2_{i',j'} + (m_{r,s} - m_{i',j'})^2 \right] \cdot p_{r \rightarrow s}p_{i \rightarrow j} (1 - \exp(-3h/\alpha_{\chi})) \tag{10}
\]

where \(\alpha_{\alpha r}\) is the range of the \(\alpha\chi\) cross-transition probabilities and \(\alpha_{\chi\chi}\) is the range of the \(\chi\chi\) cross-transition probabilities.

[29] In that \(\alpha\chi\) cross transitions are largely influenced by the mean and variance for thickness among the smaller-scale level I unit types, we use \(\alpha_{\alpha r} = \alpha_{\alpha} = 0.18\) m (as computed above from equation (4)). In that \(\chi\chi\) cross transitions are largely influenced by the mean and variance for thickness among the larger-scale level II unit types, we use \(\alpha_{\chi\chi} = \alpha_{\chi} = 0.31\) m (as computed above from equation (9)). Using these values, the model is plotted in Figure 7 and compared to the composite sample semivariogram and its components. The model semivariogram, developed without fitting, compares quite well to the sample semivariogram. This result demonstrates that the composite semivariogram structure can be modeled through modeling the \(\hat{\rho}_{r \rightarrow j}(h)\) based on length statistics (thus modeling how the proportions of \(\alpha\chi\) and \(\chi\chi\) cross transitions grow with lag distance), and scaling the \(\hat{\rho}_{r \rightarrow j}(h)\) according to univariate statistics for \(ln(K)\) by unit type (representing the variance within and the difference in means across the unit types).

### 5. Discussion

[30] The results above showed how the vertical composite \(ln(K)\) sample semivariogram at the Borden site is related to the hierarchical aquifer architecture. This understanding led to developing a model for the semivariogram based on architectural attributes. This same understanding suggests caution should be used in extrapolating the model beyond the region from which it was developed, and gives grounds for rejecting the horizontal composite sample \(ln(K)\) semivariogram computed from the same data. These issues are discussed in the following two sections.

#### 5.1. Comparison of the Results of Vertical Variography to Other Studies at the Borden Site and Discussion of Stationarity

[31] Sudicky [1986] determined the range of vertical correlation to be 0.36 m. (Note that he expressed it as an integral scale, \(\lambda\), of 0.12 m. What we refer to as range is equal to 3\(\lambda\)). Woodbury and Sudicky [1991] estimated the range of the vertical semivariogram separately along Sudicky’s two different lines of sampling. The estimates were 0.63 m along line A-A’ and 1.0 m along line B-B’; the values are statistically significantly different in these different directions. Given the difference, they point out that second-order stationarity cannot be taken for granted even for aquifers that are thought to be relatively homogeneous. Furthermore, Turcke and Kueper [1996] conducted an analysis of log permeability data from ten cores taken 60 m to the north-northeast of Sudicky’s [1986] site and discussed in the following two sections.

#### 5.2. The Line of Sampling by Allen-King et al. [2006] is Parallel to Line A-A’ of Sudicky [1986] and Tends of Meters South-South-East of it (Figure 2). The Range of the Composite Sample Semivariogram (Figure 2). The Range of the Composite Sample Semivariogram from the Allen-King et al. Data Is Indicated in this Study to Be of the Order of 0.3 m, Which Is at or Just Below the Lower 95% Confidence Interval on Estimates from Fitting an Exponential Model to Sudicky’s Results.
Figure 7. Model for the vertical $ln(K)$ semivariogram compared to the composite sample semivariogram. The model was not fitted but was determined independently based on univariate statistics for unit type thicknesses and for $ln(K)$.

The horizontal composite sample semivariogram is shown in Figure 8a. Figure 8a also shows that the $\alpha\chi$ and $\chi\chi$ terms of equation (1) together (i.e., all cross terms) predominantly define the composite semivariogram, as was true for the vertical direction. Furthermore, Figure 8a shows that equation (2) is again a good approximation, and thus the underlying structure can be inferred from the $\tilde{p}_{ro}(h)\tilde{y}_{ro,ji}(h)$ summations plotted in Figure 8b. The structure of the $\tilde{p}_{ro}(h)\tilde{y}_{ro,ji}(h)$ terms is a quick rise to lag one, and some periodicity beyond that. As per Ritzi [2000], a nugget for these terms is not physically possible. The lines must intercept the ordinate axis at zero.

In Figure 3a, some of the occurrences of medium sand are traceable across the whole 10 meter span of the cores, or across most of the sampled region with only one of the lateral terminations indicated (i.e., terminated on either the right or left but not on both sides). Thus lateral transitions across level II unit types are not represented well over the limited, 10 m extent of the cores, and the sum of $\tilde{p}_{ro}(h)\tilde{y}_{ro,ji}(h)$ for $\chi\chi$ terms in Figure 8b spuriously exhibits a range as small as that for $\alpha\chi$ terms. Turcke and Kueper [1996] also concluded that their 10-m-long line of cores did not extend far enough to allow characterizing the lateral semivariogram. Their lateral sample semivariogram likely had the same underlying problem. Because the sampling does not allow characterizing the mean and variance in lateral length for the populations of level I or level II unit types, the modeling approach in section 4 cannot be used directly. It could be used if the univariate statistics for length were inferred from other geologic information [e.g., Weissmann et al., 1999; Dai et al., 2005].
ours, spanned 19 m laterally. This distance may have been sufficient to reflect the extent and variability in length of the longer occurrences of level II, fine-sand units. This would explain why their sample semivariograms were more exponential-like and indicated a much longer range (8.4 m in the work of Sudicky [1986] and 15.42 m in the work of Woodbury and Sudicky [1991]). However, their 1 m lateral sampling resolution may not have properly characterized mean or variance in length of the medium sand and silt occurrences, nor units defined at level I, in the same way as with the 1 m lateral spaced data presented here in Figure 3. Thus the lateral sample semivariograms they fit may have had some bias, perhaps contributing to why plume spreading is slightly overpredicted in the first year by their model (see their Figure 9).

Although the horizontal sample semivariogram in Figure 8a is probably not representative of the true structure, the new data and its analyses still serve a useful purpose. Figures 8a and 8b clearly show that the growth in proportion of horizontal cross transitions with lag distance, though poorly estimated here, does control what structure is exhibited in the (spurious) sample semivariogram. If we improve the characterization of the way cross transitions grow with lag, we will improve the characterization of the structure in the semivariogram. As shown by Dai et al. [2005], additional lithologic indicator data can be used to better estimate the transition probabilities within equation (2) and, in turn, to better determine the semivariogram structure.

The analysis also indicates the type of sampling required in order to properly characterize how the proportion of cross transitions at both hierarchical levels grows with lag distance. The unit types defined at level I require a shorter, sub-1-m sampling interval in order to better characterize the mean and variance in their length and thus their $i_{ro}(h)$. Both a shorter sampling interval and a substantial increase of the length of the transect are required to fully capture level II dimensions. The occurrences of the medium sand and silt unit types require the shorter sampling interval in order to better characterize the mean and variance in their length populations. The occurrences of fine sand require that the (higher resolution) sampling needs to be conducted over a distance greater than 10 m, far enough in order to better characterize the mean and variance in the longer occurrences of fine sand, and thus to better characterize their $i_{r}(h)$.

6. Conclusions

New data collected at the Borden site show that the sediments have a hierarchical sedimentary architecture. These sediments were divided into stratal unit types, based on texture, within a two-tiered hierarchical framework. A
hierarchy of distinguishable permeability subpopulations exists corresponding to the stratral hierarchy.

The exponential-like vertical spatial correlation for \( \ln(K) \) at the Borden site can be related to the stratral hierarchy. The shape and range of the semivariogram are mostly defined by how the fraction of cross transitions grows with lag distance, as quantified by \( \hat{p}_{ro}(h) \ln \chi_{ro, jo}(h) \). These have an asymptotic, exponential-like growth because the unit types at each hierarchical level occur repeatedly along lines of sampling, with a relatively high variability in thickness.

Transitions across level I units but within the same level II unit \((\alpha \chi)\) and transitions across level II units \((\chi \chi)\) contribute about equally in defining the composite sample vertical semivariogram. These two types of cross transitions have shorter and longer length scales, respectively. In this light, the composite semivariogram reflects additive contributions from two length scales.

Much of the composite semivariogram structure can be understood from understanding the shape of the unit types, as reflected in the cross-transition probabilities, and from knowing the univariate statistics for permeability among the subpopulations corresponding to the unit types. The vertical semivariogram can be modeled from just this information, independent of the sample semivariogram.

Because the shape and range of the semivariogram are determined by the stratral architecture, locally developed semivariogram models cannot be generalized (i.e., assumed to be applicable to other areas) at the Borden site unless the stratral architecture at the local sites can be generalized. As yet, this has not been demonstrated. Locally developed estimates and models for univariate and bivariate statistics developed by Woodbury and Sudicky [1991], Turcke and Kueper [1996], and this study are different, perhaps because the unit types and their proportions, their mean and variance in length, their juxtapositioning relationships, and their permeability subpopulations are not stationary over the hundred-meter-scale distance between sampling sites.

Knowing the relationship between the semivariogram structure and attributes of the stratral architecture helps reduce the equivocal aspects of choosing a semivariogram model. It also shows that field studies must be designed so to the equivocal aspects of choosing a semivariogram can be biased.

Appendix A: Decomposition of the Variance

We are interested in the variance of \( \ln(K) \) and the variance in the thickness of unit types. In either case, the sample variance can be decomposed according to the hierarchical framework according to

\[
\sigma^2 = \sum_r \sum_o \hat{p}_{ro} \sigma^2_{ro} + \frac{1}{2} \sum_r \sum_o \sum_{i \neq o} \hat{p}_{ro} \hat{p}_{jo} (m_{ro} - \hat{m}_o)^2 + \frac{1}{2} \sum_r \sum_{j 
eq r} \sum_o \hat{p}_{ro} \hat{p}_{jo} (m_{ro} - \hat{m}_j)^2 + \frac{1}{2} \sum_r \sum_{j 
eq r} \sum_o \sum_{i \neq o} \hat{p}_{ro} \hat{p}_{jo} (m_{ro} - \hat{m}_o)^2 \]  

(A1)

Table A1. Contributions to the Composite Variance Among the Four Groups of Terms in Equation (A1)

<table>
<thead>
<tr>
<th>Group</th>
<th>( \ln(K) )</th>
<th>Thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>69.8%</td>
<td>88.0%</td>
</tr>
<tr>
<td>Second</td>
<td>11.6%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Fourth</td>
<td>18.6%</td>
<td>8.2%</td>
</tr>
</tbody>
</table>

[Itzi et al., 2004]. There are four groups of terms on the RHS, similar to equation (1). The first group of terms gives the contribution due to variance within the same level I unit types falling within the same level II unit types. The second group gives the contribution arising from differences in the sample mean across level I unit types which always exist within the same level II unit type, the third group gives the contribution arising from differences in the sample mean for the same level I unit type as occurring in different level II unit types (there are no such occurrences in the framework applied to the Borden data), and the fourth group gives the contribution arising from differences in the sample mean across different level I unit types always occurring in different level II unit types.

The breakdown of the \( \ln(K) \) variance is given in Table A1. Most of the variance (69.8%) arises from differences within unit types defined at level I. The difference in mean \( \ln(K) \) across unit types at either level I or II contributes less (11.6% and 18.6% respectively) to the global variance. Even though most of the composite sample variance arises from variance within level I units, we see in the main text that the semivariogram is governed by the \( \alpha \chi \), and \( \chi \chi \) terms in equation 1. In this light, we see that the difference in \( \ln(K) \) among heads and tails of lag pairs spanning across units is mostly governed by the variability in \( \ln(K) \) that exists within each unit rather than from the difference in the mean permeability across the units. As per Table A1, most of the sample variance in the thickness of units (88%) arises from variance in the thickness of units defined at level I. The difference in mean thickness between different level I unit types in the same level II region (e.g., differences between the mean thickness of HPXS and MCG) and differences across level II unit types (e.g., differences between the mean thickness of HPXS and FPX) are small and contribute only 4% and 8% respectively to the composite variance in thickness.

Acknowledgments. This work was supported by the National Science Foundation under grants EAR 96-04890, EAR-00-01125, and EAR-0538037. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect those of the National Science Foundation. The authors acknowledge and thank Zhenxue Dai for sharing computer programs and for provided insightful comments. The detailed work by D. Divine, K. Taylor, J. Moriguea, and K. Baldwin in collecting the data used for these analyses is acknowledged and greatly appreciated. We also acknowledge and thank David Gaylord for contributions to the development of criteria for delineating stratral units within the Borden aquifer.

References


---

R. M. Allen-King, Department of Geology, State University of New York at Buffalo, Buffalo, NY 14260, USA.

R. W. Ritzi Jr., Department of Earth and Environmental Sciences, Wright State University, Dayton, OH 45430, USA. (robert.ritzi@wright.edu)