AGU PUBLICATIONS

Water Resources Research

TECHNICAL NOTE
10.1002/2014WR015566

Key Points:
- The covariance of $Y = \ln K$ is derived as a function of particle size curves
- Workable second-order approximations of such covariance are presented
- A field application is presented

Correspondence to:
M. Riva,
monica.riva@polimi.it

Citation:
Riva, M., X. Sanchez-Vila, and A. Guadagnini (2014), Estimation of spatial covariance of log conductivity from particle size data, *Water Resour. Res.*, *50*, 5298–5308, doi:10.1002/2014WR015566.

Received 10 MAR 2014
Accepted 28 MAY 2014
Accepted article online 30 MAY 2014
Published online 17 JUN 2014

Estimation of spatial covariance of log conductivity from particle size data

Monica Riva[1,2], Xavier Sanchez-Vila[3], and Alberto Guadagnini[1,2]

[1]Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Milano, Italy, [2]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA, [3]GHS, Department of Geotechnical Engineering and Geosciences, UPC-BarcelonaTech, Barcelona, Spain

Abstract We derive analytical relationships between the spatial covariance of the (natural) logarithm of hydraulic conductivity (K) and that of representative soil particle sizes and porosity. The latter quantities can be directly measured during routine sedimentological analyses of soil samples and provide a way of incorporating K estimates into groundwater flow models at a relatively modest experimental cost. Here we rely on widely used empirical formulations requiring measurements of representative particle diameters and, in some cases, of medium porosity. We derive exact formulations relating the spatial covariance of these quantities and K and present workable approximations on the basis of perturbation methods. Our formulations provide a direct link between key geostatistical descriptors of sedimentological and hydraulic parameters of heterogeneous aquifers which can be employed in classical estimation and simulation procedures. The approach and theoretical results are tested on an extensive data set comprising 411 particle size curves collected at 12 boreholes in a small-scale alluvial aquifer.

1. Introduction

The potential of a groundwater flow model to represent reality is highly dependent on our ability to properly map the heterogeneous distribution of hydraulic conductivity, K, at the model grid scale. Mapping local values of K is a complex task since it is routinely based on a limited set of directly measured values, which are often obtained at diverse scales, or on interpreted quantities associated with hydraulic tests that are linked to an often ill-defined measurement scale.

Most of the methods employed to estimate spatial distributions of K are grounded, in one way or another, on geostatistical approaches. For example, these approaches include techniques based on (i) an initial classification of geomaterials deduced from descriptive analysis of sedimentological data collected along a borehole, followed by the application of facies delineation techniques, then complemented by assigning K values to each of the mapped facies [e.g., *Ritzi et al*., 1994; *Guadagnini et al*., 2004; *Wohlberg et al*., 2006; *Tartakovsky et al*., 2007]; (ii) the use of geophysical data which are then employed mostly in coherent soils to reconstruct the spatial field of K through some (mostly empirical) correlations between the primary variable (K) and secondary data types [e.g., *McGrath et al*., 2002; *Mariethoz et al*., 2009; *Niwas and Celik*, 2012]; (iii) the generation of collections of K fields on the basis of structural geology patterns driven by sedimentological processes including uncertainty [e.g., *Koltermann and Gorelick*, 1996; *Michael et al*., 2010; *Maurer et al*., 2013]; (iv) two-point, multiple-point, or nonparametric geostatistics (including indicator or transitional probability simulation techniques) with or without the use of training images [e.g., *Comunian et al*., 2011; *Huysmans and Dassargues*, 2012; *Blouin et al*., 2013].

Particle size data coupled with empirical correlations relating them to local K values are still widely employed to obtain inexpensive Darcy-scale estimates of hydraulic conductivity in diverse field settings. Due to simplicity of implementation, a large body of literature is available on the development and testing of such empirical equations, a comprehensive review being provided by *Vuković and Soro* [1992]. Underlying the approach is the idea that conductivity is related to the size and internal distribution of pore spaces, which is linked to particle size distribution. This concept is also supported by theoretical studies based on, e.g., volume-averaging techniques [*Whitaker*, 1999]. The theoretical elements underpinning such formulations are explored nowadays through detailed flow simulations at the pore scale in explicit pore spaces [e.g., *Hyman et al*., 2013, and references therein].

Pore size distributions are difficult to obtain. They can be inferred from mercury injection curves [e.g., *Paz Ferreiro et al*., 2009, and references therein] or directly measured from detailed X-ray images of millimeter-scale soil and/or rock samples [e.g., *Latief et al*., 2010; *Bijeljic et al*., 2013; *Guadagnini et al*., 2014, and references therein]. On the other hand, particle size curves, PSCs, are routinely measured from soil samples in geotechnical laboratories. The procedure for particle size analysis is simple, inexpensive and typically involves the successive use of a series of sieves of decreasing grid size, which are regulated by appropriate international standards. The PSCs enable one to characterize a number of effective grain diameters, d_e, defined as the representative particle size diameter in terms of percent in mass, corresponding to the e^{th} percentile of a measured PSC. It should be noted that several types of methods are available to extract PSCs from soil samples (e.g., sedigraph; laser diffraction; dry and wet sieving) and they do not always render mutually consistent and compatible results.

A suite of empirical models are available to correlate K values with (one or more) representative grain sizes as summarized by *Vuković and Soro* [1992], the resulting K values being considered as associated with some support volume which is equivalent to that of the soil samples analyzed. *Vienken and Dietrich* [2011] and *Rosas et al*. [2014] presented field evaluations of the most widely used empirical models. Some most recent studies analyze the potential to base estimates of K on the full PSC. In this context, e.g., *Rogiers et al*. [2012] use a site-specific model derived from linear and nonlinear data-driven modeling techniques.

The first empirical models were developed by *Seelheim* [1880] and *Hazen* [1893] who suggested a correlation between K and the representative diameters d_{50} and d_{10}. Since then, diverse empirical formulations have been developed involving different combination of particle diameters and including the dependence of K on additional variables such as porosity or soil type or texture. Amongst these, we focus here on the two widely used formulations of *Carman* [1937] and *Beyer* [1964]. Despite some differences in the values of the constants of proportionality and in the particular relationships adopted, these models can be considered as representative of several empirical formulations. Both models consider d_{10} as representative diameter. While *Carman* [1937] included directly the effect of the total porosity on estimates of K, *Beyer* [1964] enclosed the coefficient of uniformity, defined as $U = d_{60}/d_{10}$, in his model. Alternative formulations can be embedded within the theoretical framework illustrated in section 3.

Our work is motivated by the observation that a combination of direct K measurements and representative grain size data is typically available in field applications. As the number of samples related to direct K observation is usually much lower than that associated with PSC data, the calibration of a variogram model relying on hard K data is particularly problematic. A possible way to cope with this problem would entail the use of empirical or semiempirical formulations to obtain estimates of local K values to complement the available data set. However, this would lead to mixing hard data and empirically based estimates in a unique database the overall quality of which would be questionable, also considering that these two types of information display only a partial degree of correlation [*Barahona-Palomo et al*., 2011; *Vienken and Dietrich*, 2011].

As an alternative, we propose here to base the estimate of low-order statistics, i.e., mean and variogram, of the random spatial field of conductivity on existing PSC data. We do so by deriving analytical relationships between the spatial covariance of (natural) log conductivity and that of representative soil particle sizes and porosity or coefficient of uniformity, as embedded in the Kozeny-Carman [*Carman*, 1937] and Beyer [*Beyer*, 1964] formulations. These geostatistical descriptors can then be employed in classical estimation and simulation procedures which can then be conditioned on available hard data.

The paper is structured as follows. Section 2 summarizes the empirical formulations analyzed; section 3 presents the theoretical derivations conducive to the evaluation of the mean and covariance of log conductivity relying on the Beyer and Kozeny-Carman expressions; section 4 illustrates an application of our theoretical results in a field setting comprising a small-scale alluvial aquifer located in the Neckar river valley, Germany, where an extensive data set comprising 411 PSCs collected at various boreholes and previously employed for the geostatistical characterization of the site is available [*Riva et al*., 2006, 2008].

2. Empirical Formulations Relating Grain-Size Composition to Hydraulic Conductivity

The Kozeny-Carman equation [*Carman*, 1937] and Beyer's model [*Beyer*, 1964], which we consider in this work, can be embedded in the general framework provided by *Vuković and Soro* [1992]. In their

comprehensive work, these authors proposed a class of empirical formulations to obtain estimates of K from a representative particle size diameter, d_e, and total porosity, ϕ. All such formulations can be expressed as

$$K = \frac{g}{\nu}\beta\theta(\phi)d_e^2 \tag{1}$$

where g is the gravitational constant, ν is fluid kinematic viscosity, β is a proportionality constant, and θ is a function of ϕ.

From a physical standpoint all models described by (1) are constrained by the following set of limitations: (i) no electrochemical reactions between the soil particles and the water are considered; this implies that (1) is not appropriate in the presence of a large textural component associated with the fine fraction, as in the case of clayey soils, although it provides acceptable results for nonplastic silts; (ii) as Darcy's law assumes laminar flow and a low pore water velocity, the conditions underlying applicability of Darcy's law embedded in (1) might be violated in the presence of coarse particle diameters, such as in the case of gravels; and (iii) (1) does not explicitly account for anisotropy.

2.1. Beyer's Equation

The formulation proposed by *Beyer* [1964] is a direct extension of the work of *Hazen* [1893] to account for additional information embedded in the particle size distribution. *Hazen* [1893] suggested a direct proportionality between hydraulic conductivity K and d_{10}^2. *Beyer* [1964] postulated that intermediate grain sizes should also provide information regarding K, and incorporated the coefficient of uniformity, $U = d_{60}/d_{10}$ into Hazen's formulation.

Considering (1), Beyer's formulation can be written in consistent units as

$$K = 6\times10^{-4}\frac{g}{\nu}\log\frac{500}{U}d_{10}^2 \tag{2}$$

Note that (2) is recommended only when $1 < U < 20$ and $0.06 < d_{10}[mm] < 0.6$ [e.g., *Beyer*, 1964].

2.2. The Kozeny-Carman Equation

The Kozeny-Carman model (hereafter indicated as K-C) is expressed in the form

$$K = C_{CK}\frac{g}{\nu}\frac{\phi^3}{(1-\phi)^2}d_{10}^2 \tag{3}$$

where a nonlinear dependence of K on porosity is included. The coefficient of proportionality, C_{CK}, depends on several factors such as grains shape and distribution. For flow in capillary tubes or beds of spheres $C_{CK} = 1/180$.

3. Theoretical Developments

3.1. Perturbation Analysis of the Beyer Model

Equation (2) can be rewritten as

$$K = A[B - \ln U]d_{10}^2 \tag{4}$$

where $A = \left[\frac{g}{\nu}\frac{6\times10^{-4}}{2.303}\right]$ and $B = \ln 500$. Taking the natural logarithm of (4) leads to

$$\ln K = \ln A + 2\ln d_{10} + \ln B + \ln\left(1 - \frac{\ln U}{B}\right) \tag{5}$$

The last term in (5) can be further expanded as

$$\ln\left(1-\frac{\ln U}{B}\right)=-\frac{\ln U}{B}-\frac{1}{2}\left(\frac{\ln U}{B}\right)^2+\cdots \tag{6}$$

Considering the limits of applicability of Beyer's formula, the truncation of (6) after two terms results in a relative error in the evaluation of ln K that does not exceed 1%. Substituting (6) into (5) leads to

$$\ln K=\ln A+\ln B+2\ln d_{10}-\frac{1}{B}\ln U-\frac{1}{2B^2}\ln^2 U \tag{7}$$

To simplify the notation, we introduce $Y=\ln K$, $Z=\ln d_{10}$ and

$$\beta=\ln A+\ln B;\quad V=\ln U=\ln d_{60}-\ln d_{10} \tag{8}$$

so that (7) can be rewritten as

$$Y(\mathbf{x})=\beta+2Z(\mathbf{x})-\frac{1}{B}V(\mathbf{x})-\frac{1}{2B^2}V^2(\mathbf{x}) \tag{9}$$

x being the vector of space coordinates of a point in the domain.

The (ensemble) mean value of Y, $\langle Y\rangle$, can be derived by applying the expected value operator (denoted by $\langle\ \rangle$) to (9)

$$\langle Y(\mathbf{x})\rangle=\beta+2\langle Z(\mathbf{x})\rangle-\frac{1}{B}\langle V(\mathbf{x})\rangle-\frac{1}{2B^2}\langle V(\mathbf{x})\rangle^2-\frac{1}{2B^2}\sigma_V^2(\mathbf{x}) \tag{10}$$

where σ_V^2 is the variance of V. In this work, we assume that all random fields involved in the analysis are second-order stationary, so that the expected value is independent of location and the covariance between two points is solely a function of separation distance (or lag). Applying the classical Reynolds' decomposition to (9) and subtracting (10) from the resulting equation leads to the following equation satisfied by the perturbation $Y'=Y-\langle Y\rangle$

$$Y'(\mathbf{x})=2Z'(\mathbf{x})-\frac{1}{B}V'(\mathbf{x})-\frac{1}{2B^2}V^2(\mathbf{x})+\frac{1}{2B^2}\langle V^2(\mathbf{x})\rangle \tag{11}$$

The covariance of Y between points **x** and **y** can then be expressed as

$$\begin{aligned}C_Y(\mathbf{x},\mathbf{y})=\langle Y'(\mathbf{x})Y'(\mathbf{y})\rangle=4C_Z(\mathbf{x},\mathbf{y})+\frac{1}{B^2}C_V(\mathbf{x},\mathbf{y})-\frac{4}{B}C_{ZV}(\mathbf{x},\mathbf{y})-\frac{2}{B^2}\langle Z'(\mathbf{x})V^2(\mathbf{y})\rangle\\+\frac{1}{B^3}\langle V'(\mathbf{x})V^2(\mathbf{y})\rangle+\frac{1}{4B^4}\left(\langle V^2(\mathbf{x})V^2(\mathbf{y})\rangle-\langle V^2\rangle^2\right)\end{aligned} \tag{12}$$

Here $C_Z(\mathbf{x},\mathbf{y})=\langle Z'(\mathbf{x})Z'(\mathbf{y})\rangle$ and $C_V(\mathbf{x},\mathbf{y})=\langle V'(\mathbf{x})V'(\mathbf{y})\rangle$, respectively, are the covariance of Z and V while $C_{ZV}(\mathbf{x},\mathbf{y})=\langle Z'(\mathbf{x})V'(\mathbf{y})\rangle$ is the cross covariance between Z and V. Neglecting terms of order higher than 2 in (12) leads to the following second-order approximation for $C_Y(\mathbf{x},\mathbf{y})$ (see Appendix A for details)

$$C_Y(\mathbf{x},\mathbf{y})=4C_Z(\mathbf{x},\mathbf{y})+\frac{1}{B^2}\left(1+2\frac{\langle V\rangle}{B}+\frac{\langle V\rangle^2}{B^2}\right)C_V(\mathbf{x},\mathbf{y})-\frac{4}{B}\left(1+\frac{\langle V\rangle}{B}\right)C_{ZV}(\mathbf{x},\mathbf{y}) \tag{13}$$

Introducing $D=\ln d_{60}$, one can write $V=\ln U=\ln(d_{60}/d_{10})=D-Z$ and then obtain the following second-order expression for C_Y in terms of the low-order statistics of $\ln d_{10}$ and $\ln d_{60}$

$$C_Y(\mathbf{x},\mathbf{y}) = 4C_Z(\mathbf{x},\mathbf{y}) + \\ + \frac{1}{B^2}\left[C_D(\mathbf{x},\mathbf{y}) + C_Z(\mathbf{x},\mathbf{y}) - 2C_{ZD}(\mathbf{x},\mathbf{y})\right]\left[1 + \frac{\langle D\rangle - \langle Z\rangle}{B}\left(2 + \frac{\langle D\rangle - \langle Z\rangle}{B}\right)\right] \\ - \frac{4}{B}\left[C_{ZD}(\mathbf{x},\mathbf{y}) - C_Z(\mathbf{x},\mathbf{y})\right]\left[1 + \frac{\langle D\rangle - \langle Z\rangle}{B}\right] \tag{14}$$

In some cases, the random fields D and Z can be considered as uncorrelated, as in the field application we present in section 4. Under these conditions, (14) simplifies into

$$C_Y(\mathbf{x},\mathbf{y}) = \frac{C_Z(\mathbf{x},\mathbf{y})}{B^2}\left\{1 + 4B + 4B^2 + \frac{\langle D\rangle - \langle Z\rangle}{B}\left(2 + 4B + \frac{\langle D\rangle - \langle Z\rangle}{B}\right)\right\} \\ + \frac{C_D(\mathbf{x},\mathbf{y})}{B^2}\left[1 + \frac{\langle D\rangle - \langle Z\rangle}{B}\left(2 + \frac{\langle D\rangle - \langle Z\rangle}{B}\right)\right] \tag{15}$$

The corresponding (stationary) variance, σ_Y^2, is evaluated upon setting $\mathbf{x}=\mathbf{y}$ in (15) and considering that $\sigma_Z^2 = C_Z(\mathbf{x},\mathbf{x})$ and $\sigma_D^2 = C_D(\mathbf{x},\mathbf{x})$, respectively, are the variance of Z and D.

In summary, the (second-order approximation of the) covariance function of Y based on Beyer's formula is given by either (14) or (15) depending on whether independence between d_{10} and d_{60} can be invoked or not. If data support independence between these characteristic particle size diameters, then C_Y depends linearly on both C_Z and C_D. One can then interpret (15) as a nested covariance model, and the directional integral scale of Y, defined as $I_Y = \frac{1}{\sigma_Y^2}\int_0^\infty C_Y(h)dh$, with $h = |\mathbf{x}-\mathbf{y}|$, is a linear combination of the integral scales of Z and D, respectively, denoted as I_Z and I_D. In case this independence cannot be invoked, an additional linear dependence appears for C_Y in (14), in the form of the cross-covariance C_{ZD}. Note that any admissible model choice for C_Z, C_D, and C_{ZD} yields an admissible covariance model for Y.

3.2. Perturbation Analysis of the Kozeny-Carman Model

Taking the natural logarithm of (3) leads to

$$\ln K = \ln\left(C_{CK}\frac{g}{\nu}\right) + 2\ln d_{10} + 3\ln\phi - 2\ln(1-\phi) \tag{16}$$

The term $\ln(1-\phi)$ can be expanded as $\ln(1-\phi) = -\phi - \frac{\phi^2}{2} + O(\phi^3)$. When considering a range of porosity values typical of sedimentary aquifers (i.e., $\phi \leq 40\%$) truncation to second order in ϕ yields an approximation of $\ln(1-\phi)$ which is associated with a relative error smaller than 6%. Introducing $\alpha = \ln(C_{CK}g/\nu)$ and $W = \ln\phi$, equation (16) becomes

$$Y(\mathbf{x}) = \alpha + 2Z(\mathbf{x}) + 3W(\mathbf{x}) + 2\phi(\mathbf{x}) + \phi^2(\mathbf{x}) \tag{17}$$

The (ensemble) mean value of Y can be derived by taking expectation of (17)

$$\langle Y(\mathbf{x})\rangle = \alpha + 2\langle Z(\mathbf{x})\rangle + 3\langle W(\mathbf{x})\rangle + 2\langle\phi(\mathbf{x})\rangle + \langle\phi(\mathbf{x})\rangle^2 + \sigma_\phi^2(\mathbf{x}) \tag{18}$$

σ_ϕ^2 being porosity variance. As in section 3.1, we assume that all random fields are second-order stationary. Subtracting (18) from (17) leads to the following expression for Y'

$$Y'(\mathbf{x}) = 2Z'(\mathbf{x}) + 3W'(\mathbf{x}) + 2\phi'(\mathbf{x}) + \phi^2(\mathbf{x}) - \langle\phi^2\rangle \tag{19}$$

The log conductivity covariance can then be derived from (19) as

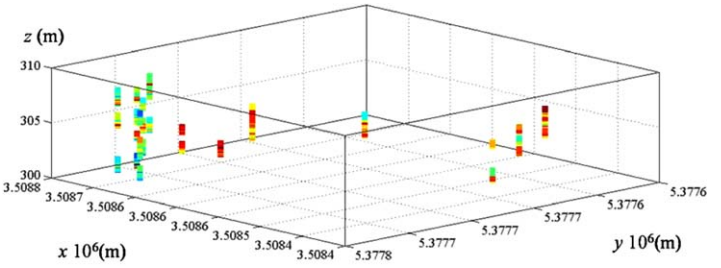

Figure 1. Soil sampling network at the Tuebingen experimental site.

$$C_Y(\mathbf{x},\mathbf{y}) = 4C_Z(\mathbf{x},\mathbf{y}) + 12C_{WZ}(\mathbf{x},\mathbf{y}) + 8C_{\phi Z}(\mathbf{x},\mathbf{y}) + 4\langle\phi^2(\mathbf{y})Z'(\mathbf{x})\rangle + 9C_W(\mathbf{x},\mathbf{y}) + 12C_{\phi W}(\mathbf{x},\mathbf{y}) + 4C_\phi(\mathbf{x},\mathbf{y}) + 6\langle\phi^2(\mathbf{y})W'(\mathbf{x})\rangle + 4\langle\phi^2(\mathbf{y})\phi'(\mathbf{x})\rangle + \langle\phi^2(\mathbf{y})\phi^2(\mathbf{x})\rangle - \langle\phi^2\rangle^2 \quad (20)$$

where C_{WZ}, $C_{\phi Z}$, and $C_{\phi W}$ are cross-covariance terms. Neglecting terms of order higher than 2 leads to (see Appendix B for details)

$$C_Y(\mathbf{x},\mathbf{y}) = 4C_Z(\mathbf{x},\mathbf{y}) + C_W(\mathbf{x},\mathbf{y})\left[25 + 8\phi_G + 4\phi_G^2 + \frac{12}{\phi_G}\right] + 4C_{ZW}(\mathbf{x},\mathbf{y})\left[3 + 2\phi_G + 2\phi_G^2\right] \quad (21)$$

ϕ_G being the geometric mean of ϕ.

If the available data suggest that Z and W are uncorrelated, then (21) is further simplified as

$$C_Y(\mathbf{x},\mathbf{y}) = 4C_Z(\mathbf{x},\mathbf{y}) + C_W(\mathbf{x},\mathbf{y})\left[25 + 8\phi_G + 4\phi_G^2 + \frac{12}{\phi_G}\right] \quad (22)$$

These results show that, similarly to the formulation based on the Beyer's model, C_Y can be expressed (at second order) as a nested model of admissible covariance and cross-covariance functions. Equation (22) relies on the assumption that Z and W be independent. In this context, we note that several empirical formulations relating porosity to particle size diameters are available. In particular, a commonly used equation is [e.g., *Vuković and Soro*, 1992]

$$\phi = 0.255\left[1 + 0.83^U\right] \quad (23)$$

which shows a direct link between ϕ and U, thus suggesting a correlation between these two quantities. However, it is worth noting that the validity of purely empirical equations of the kind (23) is restricted to materials that can be classified as clean sands. Adoption of (23) should then be questioned in the presence of diverse geomaterials.

Table 1. Key Results of the Geostatistical Analyses Performed by *Riva et al.* [2010] on $Z = \ln d_{10}$ and $D = \ln d_{60}$ (d_{iG} is the Geometric Mean d_i, $i = 10, 60$)

	Cluster 1	Cluster 2
Percentage of samples	53%	44%
d_{10G} (mm)	9.63×10^{-1}	3.67×10^{-1}
σ_Z^2	0.53	0.32
Variogram type	Spherical	Spherical
Nugget	0.05	0.05
Sill	0.48	0.27
Horizontal range (m)	28	25
Vertical range (m)	0.70	0.90
d_{60G} (mm)	1.58×10^{1}	1.13×10^{1}
σ_D^2	0.027	0.051
Variogram type	Spherical	Spherical
Nugget	0.005	0.010
Sill	0.0226	0.041
Horizontal range (m)	15	12
Vertical range (m)	0.70	0.70

4. Field-Scale Application

In this section, we apply our theoretical development to the analysis of an extensive data set collected at an experimental site located near the city of Tuebingen, Germany.

The aquifer is composed of alluvial material overlain by stiff silty clay and underlain by hard silty clay. Site characterization is based on detailed information collected at a set of monitoring and pumping wells. The aquifer

Table 2. Results From Our Theoretical Developments and Key Results of the Geostatistical Analyses Performed on $Y = \ln K$ Calculated on the Basis of the Formulation of *Beyer* [1964], as Presented by *Riva et al.* [2006]

	Cluster 1		Cluster 2	
	Riva et al. [2006]	Current Analysis	*Riva et al.* [2006]	Current Analysis
K_G (m/s)	5.92×10^{-3}	6.44×10^{-3} from (10)	0.83×10^{-3}	0.81×10^{-3} from (10)
σ_Y^2	2.41	2.64 from (15)	1.35	1.62 from (15)
Variogram type	Spherical	Equation (15)	Spherical	Equation (15)
Nugget	0.05	0.25	0.05	0.25
Sill	2.36	2.39	1.30	1.37
Horizontal integral scale (m)	3.75	10.50	3.75	9.37
Vertical integral scale (m)	0.33	0.26	0.30	0.34

has a saturated thickness of about 5 m and all boreholes reach the bedrock which forms the impermeable aquifer base. The extensive investigations performed at the site comprise field and laboratory-scale data collection and analysis. Available data include particle size curves, pumping and tracer tests as well as downhole impeller flowmeter measurements. A detailed description of the analyses performed at the site is presented by *Riva et al.* [2006, 2008], to which we refer for additional details. *Neuman et al.* [2007, 2008] applied a stochastic interpretation method to data obtained from cross-hole pumping tests and illustrated the application of a multiscale geostatistical methodology for the characterization of the spatial variability of hydraulic conductivity. *Barahona-Palomo et al.* [2011] analyzed the relationship between hydraulic conductivity estimates obtained through particle size curves and impeller flowmeter measurements at this and other sites and found that these two quantities are partially but not completely correlated.

Here we focus on the available 411 PSCs collected along 12 vertical boreholes at the site which were employed by *Riva et al.* [2006, 2008, 2010] to analyze in a Monte Carlo framework, the data associated with a tracer test and to provide a probabilistic delineation of well-related capture zones. These PSCs were measured on core samples associated with characteristic length ranging from 5 to 26.5 cm. They are reconstructed through grain sieve analysis performed with a set of 12 discrete sieve diameters (i.e., 0.063, 0.125, 0.25, 0.50, 1.0, 2.0, 4.0, 8.0, 16.0, 31.5, 63.0, and 100.0 mm) and suggest the occurrence of heterogeneous and highly conducive alluvial deposits. Figure 1 depicts the three-dimensional structure of the sampling network at the site.

Riva et al. [2006] provide a classification of the spatial distribution of hydrofacies in the system by grouping the available PSCs into three main clusters, i.e., (a) Cluster 1, comprising moderately sorted gravel with about 14% sand and very few fines and representing 53% of the samples; (b) Cluster 2, characterized by poorly sorted gravel with about 24% sand and few fines and representing 44% of the samples; and (c) Cluster 3, characterized by well-sorted sand with very few fines and about 23% gravel and representing 3% of the samples. Characteristic particle diameters estimated from the PSCs have then been employed by *Riva et al.* [2006, 2008, 2010] to estimate hydraulic conductivities for each identified Cluster through Beyer's formula. A detailed variogram analysis of the resulting hydraulic conductivity estimates associated with Cluster 1 and 2 is presented by *Riva et al.* [2006], while the geostatistical analysis of d_{10} and d_{60} is presented by *Riva*

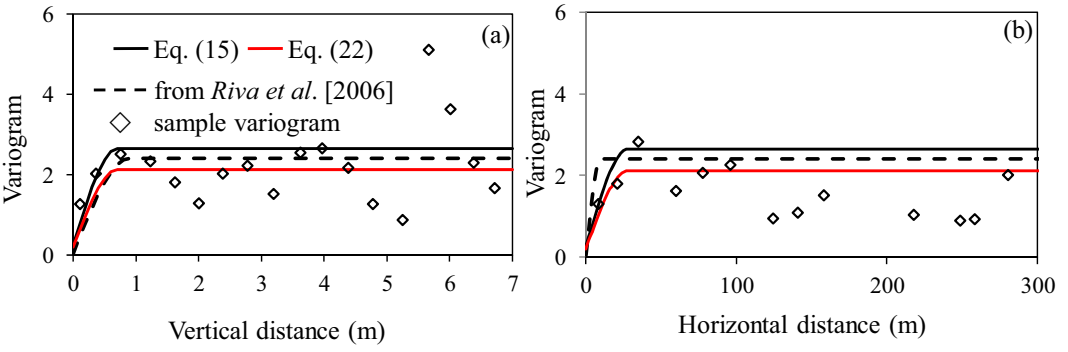

Figure 2. Sample (a) vertical and (b) horizontal variograms of log conductivities associated with Cluster 1 together with the model fitted by *Riva et al.* [2006] and the analytical variogram models derived from (15) and (22).

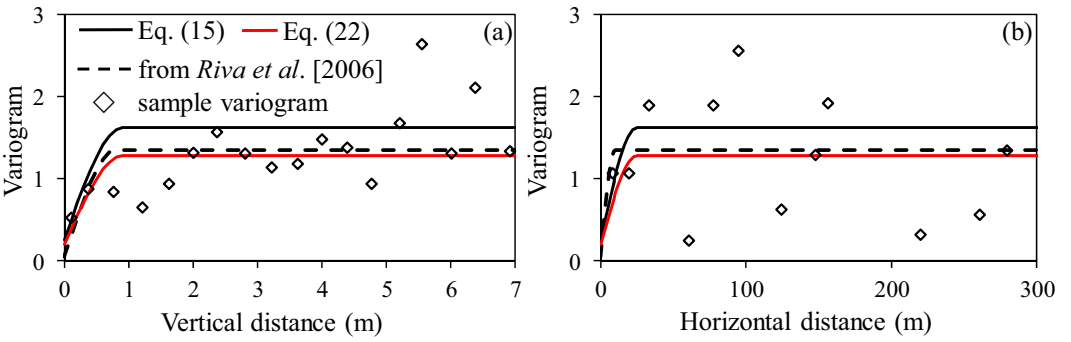

Figure 3. Sample (a) vertical and (b) horizontal variograms of log conductivities associated with Cluster 2 together with the model fitted by *Riva et al*. [2006] and the analytical variogram models derived from (15) and (22).

et al. [2010]. These authors verified that $\ln d_{10}$ and $\ln d_{60}$ associated with each cluster are independent at the site thus allowing us to rely on (15) for the purpose of our analyses. In these works the authors estimated the parameters of the variograms of $\ln d_{10}$, $\ln d_{60}$ and $\ln K$ upon setting the total variogram sill to the sample variance of the data and by relying on the standard least squares method.

Here we rely on these results to apply the theoretical developments presented in section 3.1. Table 1 lists the main results of the variogram analyses performed by *Riva et al*. [2010] on $\ln d_{10}$, $\ln d_{60}$. These are used as input data in (10) and (15) to analytically derive the key statistical parameters of $Y = \ln K$ together with its variogram. These results are listed in Table 2. The latter also lists the corresponding results of *Riva et al*. [2006] which were obtained by estimating K through (2) at measurement locations and then performing a standard varigram analysis as mentioned above.

Figure 2 depicts graphically the sample vertical and horizontal variograms of the log conductivity field associated with Cluster 1 together with the model fitted by *Riva et al*. [2006]. The analytical variogram model derived from (15) is juxtaposed to these results. Figure 3 shows the corresponding depiction for data associated with Cluster 2 at the site. Considering that (a) any empirical formulation yields estimates of K which are associated with an estimation error which is not clearly quantifiable, (b) empirical variograms based on a finite set of data are always prone to uncertainty, and (c) variogram calibration is always associated with parameter estimation errors (with notable uncertainty in the estimation of horizontal ranges), we find the results listed in Table 2 and encapsulated in Figures 2 and 3 to be of acceptable quality for the purpose of our demonstration.

On the basis of the type of information available, which does not include direct porosity measurements, the analytical expression based on the Kozeny-Carman model, i.e., (21) or (22), are not directly applicable to this site. For the sake of comparison and in the absence of ϕ measurements, we assume that porosity is deterministic, i.e., $C_W = 0$ and $Cz_W = 0$. In this case, (21) and (22) reduce to $C_Y(\mathbf{x}, \mathbf{y}) = 4C_Z(\mathbf{x}, \mathbf{y})$ and the variogram of Y is fully defined by the variogram of $\ln d_{10}$. The results obtained with this model are also depicted in Figures 2 and 3, respectively, for Cluster 1 and 2. Obviously, since the variability in porosity is neglected the computed variogram of Y display a smaller sill than the one based on the Beyer model where randomness of ϕ is somehow encapsulated through the spatial variability of U.

5. Conclusions

Our work leads to the following key conclusions.

1. We derive exact formulations relating low-order statistics, i.e., mean and spatial covariance, of (natural) log conductivity and that of representative diameters associated with soil particle size curves (PSCs) and porosity. We base our analytical developments on the widely used empirical formulations of *Beyer* [1964] and Kozeny-Carman [*Carman*, 1937], which, respectively, require the availability of measurements of two representative particle diameters and joint measurements of a representative particle size and system porosity.

2. We present workable approximations of these formulations upon relying on perturbation methods and test our theoretical results on a unique data set comprising 411 particle size curves collected at 12 boreholes in a small-scale experimental alluvial aquifer close to the city of Tuebingen, Germany.

3. Practical implications of our results stem from the ability of our formulation to provide a clear link between the basic geostatistical descriptors of hydraulic conductivity and soil representative particle diameters. This can then be employed in classical geostatistically based estimation and simulation procedures which can then be conditioned on available direct measurements of K.

Appendix A: Derivation of Second-Order Approximation for C_Y Based on the Beyer Model

Here we derive second-order (in σ_Z and σ_V) approximations for C_Y on the basis of the Beyer equation (2). We start by considering the following terms appearing in (12)

$$\langle Z'(\mathbf{x})V^2(\mathbf{y})\rangle = \langle Z'(\mathbf{x})\left[\langle V\rangle + V'(\mathbf{y})\right]^2\rangle = 2\langle V\rangle C_{ZV}(\mathbf{x},\mathbf{y}) + \langle Z'(\mathbf{x})V'^2(\mathbf{y})\rangle \tag{A1}$$

$$\langle V'(\mathbf{x})V^2(\mathbf{y})\rangle = 2\langle V\rangle C_V(\mathbf{x},\mathbf{y}) + \langle V'(\mathbf{x})V'^2(\mathbf{y})\rangle \tag{A2}$$

$$\langle V^2(\mathbf{x})V^2(\mathbf{y})\rangle = \langle V\rangle^4 + 2\langle V\rangle^2\sigma_V^2 + \langle V'^2(\mathbf{x})V'^2(\mathbf{y})\rangle + 4\langle V\rangle\langle V'(\mathbf{x})V'^2(\mathbf{y})\rangle + 4\langle V\rangle^2 C_V(\mathrm{x},\mathbf{y}) \tag{A3}$$

$$\langle V^2\rangle^2 = \langle V\rangle^4 + 2\langle V\rangle^2\sigma_V^2 + \sigma_V^4 \tag{A4}$$

Replacing (A1)–(A4) into (12) yields:

$$\begin{aligned} C_Y(\mathbf{x},\mathbf{y}) = 4C_Z(\mathbf{x},\mathbf{y}) + \left(\frac{1}{B^2} + 2\frac{\langle V\rangle}{B^3} + \frac{\langle V\rangle^2}{B^4}\right)C_V(\mathbf{x},\mathbf{y}) - \left(\frac{4}{B} + \frac{4\langle V\rangle}{B^2}\right)C_{ZV}(\mathbf{x},\mathbf{y}) + \\ - \frac{2}{B^2}\langle Z'(\mathbf{x})V'^2(\mathbf{y})\rangle + \frac{1}{B^3}\left(1 + \frac{\langle V\rangle}{B}\right)\langle V'(\mathbf{x})V'^2(\mathbf{y})\rangle + \\ + \frac{1}{4B^4}\langle V'^2(\mathbf{x})V'^2(\mathbf{y})\rangle - \frac{1}{4B^4}\sigma_V^4 \end{aligned} \tag{A5}$$

Neglecting moments of order larger than two in (A5) leads directly to (13)

Appendix B: Derivation of Second-Order Approximation for C_Y Based on the Kozeny-Carman Model

Here we derive second-order (in σ_Z and σ_W) approximations for C_Y on the basis of the Kozeny-Carman model (3). We start by evaluating terms that include ϕ^2 in (20), i.e.,

$$\begin{aligned} \langle Z'(\mathbf{x})\phi^2(\mathbf{y})\rangle = \left\langle Z'(\mathbf{x})\left[\langle\phi\rangle + \phi'(\mathbf{y})\right]^2\right\rangle = \left\langle Z'(\mathbf{x})\left[\langle\phi\rangle^2 + \phi'^2(\mathbf{y}) + 2\langle\phi\rangle\phi'(\mathbf{y})\right]\right\rangle = \\ = 2\langle\phi\rangle C_{Z\phi}(\mathbf{x},\mathbf{y}) + \langle Z'(\mathbf{x})\phi'^2(\mathbf{y})\rangle \end{aligned} \tag{B1}$$

$$\langle W'(\mathbf{x})\phi^2(\mathbf{y})\rangle = 2\langle\phi\rangle C_{W\phi}(\mathbf{x},\mathbf{y}) + \langle W'(\mathbf{x})\phi'^2(\mathbf{y})\rangle \tag{B2}$$

$$\langle\phi'(\mathbf{x})\phi^2(\mathbf{y})\rangle = 2\langle\phi\rangle C_\phi(\mathbf{x},\mathbf{y}) + \langle\phi'(\mathbf{x})\phi'^2(\mathbf{y})\rangle \tag{B3}$$

$$\begin{aligned} \langle\phi^2(\mathbf{x})\phi^2(\mathbf{y})\rangle = \left\langle\left[\langle\phi\rangle + \phi'(\mathbf{x})\right]^2\left[\langle\phi\rangle + \phi'(\mathbf{y})\right]^2\right\rangle = \\ = \langle\phi\rangle^4 + 2\langle\phi\rangle^2\sigma_\phi^2 + 4\langle\phi\rangle^2 C_\phi(\mathbf{x},\mathbf{y}) + \langle\phi'^2(\mathbf{x})\phi'^2(\mathbf{y})\rangle + 4\langle\phi\rangle\langle\phi'^2(\mathbf{x})\phi'(\mathbf{y})\rangle \end{aligned} \tag{B4}$$

$$\langle\phi^2\rangle^2=\left\langle\left[\langle\phi\rangle+\phi'(\mathbf{x})\right]^2\right\rangle^2=\left[\langle\phi\rangle^2+\langle\phi'^2\rangle\right]^2=\langle\phi\rangle^4+2\langle\phi\rangle^2\sigma_\phi^2+\sigma_\phi^4 \tag{B5}$$

Replacing (B1)–(B5) into (20), after some manipulations leads to

$$\begin{aligned}C_Y(\mathbf{x},\mathbf{y})=&4C_Z(\mathbf{x},\mathbf{y})+9C_W(\mathbf{x},\mathbf{y})+12C_{WZ}(\mathbf{x},\mathbf{y})+8C_{Z\phi}(\mathbf{x},\mathbf{y})[\langle\phi\rangle+1]+\\&+12C_{W\phi}(\mathbf{x},\mathbf{y})[\langle\phi\rangle+1]+4C_\phi(\mathbf{x},\mathbf{y})\left[2\langle\phi\rangle+1+\langle\phi\rangle^2\right]-\sigma_\phi^4+\\&+4\langle Z'(\mathbf{x})\phi'^2(\mathbf{y})\rangle+6\langle W'(\mathbf{x})\phi'^2(\mathbf{y})\rangle+\\&+\langle\phi'^2(\mathbf{x})\phi'^2(\mathbf{y})\rangle+4[\langle\phi\rangle+1]\langle\phi'^2(\mathbf{x})\phi'(\mathbf{y})\rangle\end{aligned} \tag{B6}$$

Recalling that $W=\ln\phi$, allows expanding the cross-covariance $C_{W\phi}$ as

$$\begin{aligned}C_{W\phi}(\mathbf{x},\mathbf{y})=&\langle\phi'(\mathbf{x})W'(\mathbf{y})\rangle=\left\langle\phi'(\mathbf{x})\ln\left(\langle\phi\rangle+\phi'(\mathbf{y})\right)\right\rangle=\\&=\left\langle\phi'(\mathbf{x})\ln\left[\langle\phi\rangle\left(1+\frac{\phi'(\mathbf{y})}{\langle\phi\rangle}\right)\right]\right\rangle=\left\langle\phi'(\mathbf{x})\ln\left[1+\frac{\phi'(\mathbf{y})}{\langle\phi\rangle}\right]\right\rangle=\\&\cong\left\langle\phi'(\mathbf{x})\left[\frac{\phi'(\mathbf{y})}{\langle\phi\rangle}-\frac{1}{2}\frac{\phi'^2(\mathbf{y})}{\langle\phi\rangle^2}+\frac{1}{3}\frac{\phi'^3(\mathbf{y})}{\langle\phi\rangle^3}+.....\right]\right\rangle=\\&\cong\frac{C_\phi(\mathbf{x},\mathbf{y})}{\langle\phi\rangle}-\frac{1}{2}\frac{\langle\phi'^2(\mathbf{y})\phi'(\mathbf{x})\rangle}{\langle\phi\rangle^2}+...\end{aligned} \tag{B7}$$

Replacing (B7) into (B6) and neglecting terms of order higher than two (in σ_Z and σ_W) leads to

$$\begin{aligned}C_Y(\mathbf{x},\mathbf{y})=&4C_Z(\mathbf{x},\mathbf{y})+9C_W(\mathbf{x},\mathbf{y})+12C_{WZ}(\mathbf{x},\mathbf{y})+8C_{Z\phi}(\mathbf{x},\mathbf{y})[\langle\phi\rangle+1]+\\&+4C_\phi(\mathbf{x},\mathbf{y})\left[2\langle\phi\rangle+4+\langle\phi\rangle^2+\frac{3}{\langle\phi\rangle}\right]\end{aligned} \tag{B8}$$

Recalling that $\langle\phi\rangle=\phi_G\langle e^{W'}\rangle$, where $\phi_G=e^{\langle W\rangle}$ is the geometric mean of ϕ, and

$$\phi'(\mathbf{x})=\phi_G\left[e^{W'(\mathbf{x})}-\langle e^{W'(\mathbf{x})}\rangle\right]=\phi_G\left[W'(\mathbf{x})+\frac{W'^2(\mathbf{x})}{2}+...-\frac{\langle W'^2(\mathbf{x})\rangle}{2}+..\right] \tag{B9}$$

at the second order, we obtain

$$C_\phi(\mathbf{x},\mathbf{y})=\langle\phi'(\mathbf{x})\phi'(\mathbf{y})\rangle=\phi_G^2C_W(\mathbf{x},\mathbf{y});\quad C_{Z\phi}(\mathbf{x},\mathbf{y})=\langle Z'(\mathbf{x})\phi'(\mathbf{y})\rangle=\phi_GC_{WZ}(\mathbf{x},\mathbf{y}) \tag{B10}$$

Making use of (B9) and (B10), equation (B8) reduces to (21).

Acknowledgments
The work was partly supported through funding from MIUR (Italian ministry of Education, Universities and Research—PRIN2010-11; project: "Innovative methods for water resources under hydro-climatic uncertainty scenarios"). Partial support was provided by the ICREA Acadèmia Program and the EU (project MARSOL, FP7-ENV-2013, grant 619120).

References

Barahona-Palomo, M., M. Riva, X. Sanchez-Vila, E. Vazquez-Sune, and A. Guadagnini (2011), Quantitative comparison of impeller flowmeter and particle-size distribution techniques for the characterization of hydraulic conductivity variability, *Hydrogeol. J.*, *19*(3), 603–612, doi: 10.1007/s10040-011-0706-5.

Beyer, W. (1964), Zur bestimmung der wasserdurchlässigkeit von kiesen und sanden aus der kornverteilungskurve, *Wasserwirt. Wassertech.*, *14*(6), 165–168.

Bijeljic, B., P. Mostaghimi, and M. J. Blunt (2013), Insights into non-Fickian solute transport in carbonate, *Water Resour. Res.*, *49*, 2714–2728, doi:10.1002/wrcr.20238.

Blouin, M., R. Martel, and E. Gloaguen (2013), Accounting for aquifer heterogeneity from geological data to management tools, *Ground Water*, *51*(3), 421–431.

Carman, P. C. (1937), Fluid flow through granular beds, *Trans. Inst. Chem. Eng.*, *15*, 150–166.

Comunian, A., P. Renard, J. Straubhaar, and P. Bayer (2011), Three-dimensional high resolution fluvio-glacial aquifer analog—Part 2: Geostatistical modeling, *J. Hydrol.*, *405*(1-2), 10–23, doi:10.1016/j.jhydrol.2011.03.037.
Guadagnini, A., M. J. Blunt, M. Riva, and B. Bijeljic (2014), Statistical scaling of geometric characteristics in millimeter scale natural porous media, *Transp. Porous Media*, *101*(3), 465–475, doi:10.1007/s11242-013-0254-7.
Guadagnini, L., A. Guadagnini, and D. M. Tartakovsky (2004), Probabilistic reconstruction of geologic facies, *J. Hydrol.*, *294*, 57–67.
Hazen, A. (1893), Some physical properties of sands and gravels, with special reference to their use in filtration, 24th annual report, pp. 541–556, Doc. 34, Mass., State Board of Health, Boston.
Huysmans, M., and A. Dassargues (2012), Modeling the effect of clay drapes on pumping test response in a cross-bedded aquifer using multiple-point geostatistics, *J. Hydrol.*, *450–451*, 159–167, doi:10.1016/j.jhydrol.2012.05.014.
Hyman, J. D., P. K. Smolarkiewicz, and C. L. Winter (2013), Pedotransfer functions for permeability: A computational study at pore scales, *Water Resour. Res.*, *49*, 2080–2092, doi:10.1002/wrcr.20170.
Koltermann, C. E., and S. M. Gorelick (1996), Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches, *Water Resour. Res.*, *32*(9), 2617–2658.
Latief, F. D. E., B. Biswal, U. Fauzi, and R. Hilfer (2010), Continuum reconstruction of the pore scale microstructure for Fontainebleau sandstone, *Physica A*, *389*, 1607–1618.
Mariethoz, G., P. Renard, and R. Froidevaux (2009), Integrating collocated auxiliary parameters in geostatistical simulations using joint probability distributions and probability aggregation, *Water Resour. Res.*, *45*, W08421, doi:10.1029/2008WR007408.
Maurer, T., A. Schneider, and H. H. Gerke (2013), Scenario-based three-dimensional distributed sediment structures for a constructed hydrological catchment, *Vadose Zone J.*, *12*(4), 1–23, doi:10.2136/vzj2013.02.0047.
McGrath, R. J., P. Styles, E. Thomas, and S. Neale (2002), Integrated high-resolution geophysical investigations as potential tools for water resource investigations in karst terrain, *Environ. Geol.*, *42*(5), 552–557, doi:10.1007/s00254-001-0519-2.
Michael, H. A., H. Li, A. Boucher, T. Sun, J. Caers, and S. M. Gorelick (2010), Combining geologic-process models and geostatistics for conditional simulation of 3-D subsurface heterogeneity, *Water Resour. Res.*, *46*, W05527, doi:10.1029/2009WR008414.
Neuman, S. P., A. Blattstein, M. Riva, D. M. Tartakovsky, A. Guadagnini, and T. Ptak (2007), Type curve interpretation of late-time pumping test data in randomly heterogeneous aquifers, *Water Resour. Res.*, *43*, W10421, doi:10.1029/2007WR005871.
Neuman, S. P., M. Riva, and A. Guadagnini (2008), On the geostatistical characterization of hierarchical media, *Water Resour. Res.*, *44*, W02403, doi:10.1029/2007WR006228.
Niwas, S., and M. Celik (2012), Equation estimation of porosity and hydraulic conductivity of Ruhrtal aquifer in Germany using near surface geophysics, *J. Appl. Geophys.*, *84*, 77–85.
Paz Ferreiro, J., M. Wilson, and E. Vidal Vázquez (2009), Multifractal description of nitrogen adsorption isotherms, *Vadose Zone J.*, *8*, 209–219.
Ritzi, R.W., Jr., D. F. Jayne, D. F., A. J. Zahradnik Jr.,., A. A. Field, and G. E. Fogg (1994), Geostatistical modeling of heterogeneity in glaciofluvial, buried-valley aquifers, *Ground Water*, *32*(4), 666–674.
Riva, M., L. Guadagnini, A. Guadagnini, T. Ptak, and E. Martac (2006), Probabilistic study of well capture zones distribution at the Luswiesen field site, *J. Contam. Hydrol.*, *88*, 92–118.
Riva, M., A. Guadagnini, D. Fernandez-Garcia, X. Sanchez-Vila, and T. Ptak (2008), Relative importance of geostatistical and transport models in describing heavily tailed breakthrough curves at the Lauswiesen site, *J. Contam. Hydrol.*, *101*, 1–13.
Riva, M., L. Guadagnini, and A. Guadagnini (2010), Effects of uncertainty of lithofacies, conductivity and porosity distributions on stochastic interpretations of a field scale tracer test, *Stochastic Environ. Res. Risk Assess.*, *24*, 955–970, doi:10.1007/s00477-010-0399-7.
Rogiers, B., D. Mallants, O. Batelaan, M. Gedeon, M. Huysmans, and A. Dassargues (2012), Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks, *Math. Geosci.*, *44*(6), 739–763.
Rosas, J., O. Lopez, T. M. Missimer, K. M. Coulibaly, A. H. A. Dehwah, K. Sesler, L. R. Lujan, and D. Mantilla (2014), Determination of hydraulic conductivity from grain-size distribution for different depositional environments, *Ground Water*, *52*(3), 399–413.
Seelheim, F. (1880), Methoden zur bestimmung der durschlassigkeit des bodens [in German], *Z. Anal. Chem.*, *19*, 387–418.
Tartakovsky, D. M., B. Wohlberg, and A. Guadagnini (2007), Nearest-neighbor classification for facies delineation, *Water Resour. Res.*, *43*, W07201, doi:10.1029/2007WR005968.
Vienken, T., and P. Dietrich (2011), Field evaluation of methods for determining hydraulic conductivity from grain size data, *J. Hydrol.*, *400*(1-2), 58–71.
Vuković, M., and A. Soro (1992), *Determination of Hydraulic Conductivity of Porous Media From Grain-Size Composition*, Water Resour. Publ., Highlands Ranch, Colo.
Whitaker, S. (1999), *The Method of Volume Averaging*, Kluwer Acad., Dordrecht, Netherlands.
Wohlberg, B., D. M. Tartakovsky, and A. Guadagnini (2006), Subsurface characterization with support vector machines, *IEEE Trans. Geosci. Remote Sens.*, *44*(1), 47–57, doi:10.1109/TGRS.2005.859953.