Title

Probability Distributions in Groundwater Hydrology: Methods and Applications

Journal

Journal of Hydrologic Engineering, 20(5)

ISSN

1084-0699

Author

Loáiciga, Hugo A

Publication Date

2015-05-01

DOI

10.1061/(asce)he.1943-5584.0001061

Peer reviewed

Probability Distributions in Groundwater Hydrology: Methods and Applications

Hugo A. Loáiciga, F.ASCE[1]

Abstract: This paper presents the most frequently used probability-density functions in groundwater hydrology and practical ways to apply them. The paper provides several examples of probability-density functions dealing with (1) their application to various types of groundwater phenomena, (2) the estimation of their parameters by the method of moments, and (3) the implementation of goodness-of-fit tests in probabilistic groundwater hydrology. The versatility of the log-gamma probability-density function to fit highly skewed groundwater data is demonstrated. Important univariate probability-density functions are covered, and the multivariate lognormal probability-density function's applicability to goodness-of-fit testing and synthetic generation of random fields is elucidated. **DOI: 10.1061/(ASCE)HE.1943-5584.0001061.** *© 2014 American Society of Civil Engineers.*

Author keywords: Probability-density function; Groundwater; Aquifer parameters; Goodness-of-fit tests; Random variable.

Introduction

Groundwater hydrology is a discipline of the earth sciences concerned with the quantitative study of water flow, water storage, chemical transport, and related processes in the subsurface. Groundwater hydrologists measure properties of soils and rocks to gain an understanding of subsurface hydrologic processes and to construct predictive models of groundwater phenomena. Those properties include, but are not limited to, porosity, permeability, hydraulic conductivity, specific storage, specific yield, and dispersivity. Because of the complex nature of geologic materials, measurements of these properties exhibit variability even in strata considered to be homogeneous on account of their origin and basic features (such as mineral composition and textural properties). Hydraulic conductivity measurements, for example, made at different locations in an aquifer exhibit substantial variability. This is exemplified in Fig. 1, which plots 201 measurements of hydraulic conductivity made in cohesive sediments of lacustrine origin underlying Mexico City (data from Loáiciga et al. 2006). The measurements of hydraulic conductivity shown in Fig. 1 vary over five orders of magnitude. Those measurements—and those of other aquifer properties—can be analyzed using the laws of probability and statistics to obtain a proper description of the property (or variable) under study that goes beyond the calculation of its average, standard deviation, or other indicators of central tendency, dispersion, and asymmetry. The fitting of an aquifer property with a proper probability-density function (PDF) is a necessary step to arrive at a complete description of its probabilistic characteristics. Analysts can then use the fitted PDF in a variety of analyses and design modes that provide a wider range of options than those available when the property is treated deterministically, i.e., as a nonrandom entity.

It must not be construed from the previous paragraph that all soil and rock properties vary over a wide numerical range. The porosity of soil and rocks, for example, takes values between 0 and 1. Therefore, in the probabilistic analysis of porosity, one must try probability densities functions defined over a finite domain, or use truncated probability functions [see Loáiciga et al. (1992) for an analysis of truncated PDFs in hydrologic applications].

This paper presents (1) several probability-density functions commonly used in groundwater hydrology and (2) examples on how PDFs are used to interpret aquifer properties and groundwater variables in a probabilistic manner. Most of the PDFs used in this work are univariate or bivariate. The coverage of multivariate PDFs (involving more than two random variables) focuses on the lognormal case, elucidating its applicability to goodness-of-fit testing and the synthetic generation of random fields. Several of the examples rely on hydraulic conductivity data because hydraulic conductivity is an aquifer property that controls the movement of groundwater and dissolved chemicals in a fundamental manner. Besides its importance in groundwater hydrology, its variability—as shown in Fig. 1—makes it well-suited for probabilistic analysis. In addition, hydraulic conductivity has been more extensively measured in situ or in the laboratory than any other aquifer property relevant in groundwater hydrology. Thus, data sets that can be analyzed with the methods of this paper are more common for hydraulic conductivity than for any other aquifer property. This makes the hydraulic conductivity an attractive property to work with when describing probabilistic methods amenable for the characterization of aquifer properties. Some of the material presented in this paper has been borrowed from the works of the author and collaborators (Loáiciga 2004; Loáiciga and Leipnik 2005; Loáiciga et al. 2006, 2008a, b, 2010).

The novelty of this paper rests in (1) identifying the PDFs most amenable for the analysis of groundwater data; (2) summarizing the PDFs' properties; (3) describing an efficient method for PDF parameter estimation (the method of moments); (4) providing guidance on selecting PDFs suitable for aquifer properties and conducting goodness-of-fit testing; (5) rendering a consolidated presentation of PDFs and their application in groundwater hydrology; (6) demonstrating the capacity of the three-parameter log-gamma PDF to fit highly skewed groundwater data; and (7) presenting methods to carry out goodness-of-fit testing and synthetic generation of random with the multivariate lognormal PDF.

[1]Professor, Dept. of Geography, Univ. of California, Santa Barbara, CA 93106. E-mail: Hugo.Loaiciga@ucsb.edu

Note. This manuscript was submitted on December 14, 2013; approved on July 15, 2014; published online on August 20, 2014. Discussion period open until January 20, 2015; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699/04014063(13)/$25.00.

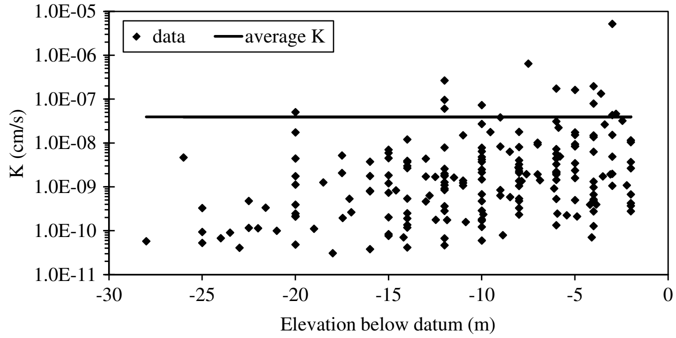

Fig. 1. Measurement of hydraulic conductivity in the lacustrine sediments underlying Mexico City; the horizontal line is the average, 3.94×10^{-8} cm/s

Definitions

The following definitions are building blocks for the material presented in this paper.

Probability-Density Function

A PDF is a mathematical formula that assigns a nonnegative value to any number that is contained in the domain of the PDF. They are functions of the form $f(x)$, in which x denotes any value at which the function f is calculated. The set of x values over which the function f is defined is called the domain of the PDF. The PDF integrated over its entire domain yields a value of 1. When integrated over part of its domain, it produces a probability between 0 and 1. The mathematical formula of a PDF may take many forms. Among the best known and more widely used ones are the uniform, normal (or Gaussian), lognormal, gamma and log-gamma, beta, exponential, Weibull, Gumbel, Student *t*, and chi-square PDFs. There are PDFs in which the x values are strictly integer values. These PDFs are commonly called probability distributions. The binomial, Poisson, and geometric probability distributions are commonly used. Discrete probability distributions and hydrologic applications are exemplified in Loáiciga (2005).

Correlation Coefficient and Covariance

Consider two random variables K_1 and K_2 with expected values (or means) μ_1 and μ_1, and variances σ_1^2 and σ_2^2, respectively, that are correlated with correlation coefficient ρ_{12}. The latter is defined by the following:

$$\rho_{12} = \frac{E[(K_1 - \mu_1)(K_2 - \mu_2)]}{\sigma_1 \sigma_2} \quad (1)$$

where E denotes the expectation operator. The correlation coefficient ρ_{12} is a normalized measure of the degree of statistical association between two random variables. Its magnitude is in the range $[-1, 1]$. A value of -1 indicates perfect negative correlation, a value of $+1$ denotes perfect positive correlation, and a value of zero indicates that the variables K_1 and K_2 are uncorrelated. The covariance σ_{12}^2 between the variables K_1 and K_2 is equal to $\rho_{12}\ \sigma_1\ \sigma_2$.

Spatial Correlation

Spatial correlation is a measure of the degree of statistical association among measurements of an aquifer property made at different locations in an aquifer. Positively correlated measurements occur when the spatial correlation between two measurements of the property K_1 and K_2 made at locations r_1 and r_2, respectively, ranges between 0 and 1. The closer the spatial correlation is to 1, the greater the degree of statistical association between the measurements of K_1 and K_2.

Correlation Scale

Correlation scale is the distance between two locations r_1 and r_2 beyond which the aquifer property K_1 (at r_1) and K_2 (at r_2) cease to be spatially correlated.

Statistical Homogeneity and Independence

Statistical homogeneity and independence of measurements are conditions that must be met when attempting to fit a PDF to a sample of measurements of an aquifer property. Statistical homogeneity implies that the PDF of the property in question is the same everywhere in the aquifer in which measurements are made with a similar device (deployed in the field or applied in the laboratory to core samples). In this case, the measurements exhibit a constant average and a spread of values about the average that are devoid of spatial trends or spatial periodic patterns. Independence of measurements implies that the value of the measured property at any location in an aquifer is not related in a probabilistic sense to any other of its values measured at other locations in the same aquifer. Independent measurements are uncorrelated. Property measurements can be statistically homogeneous and correlated simultaneously. In the latter instance, one must resort to geostatistics, a field of statistics concerned with the study of spatially correlated variables (Journel and Huijbregts 1978; Dagan 1989; Loáiciga et al. 2010). From a physical standpoint, statistical homogeneity is approximated in the field when geological processes produce unconsolidated deposits (clays, silts, sands, gravels, or combinations of these) or consolidated deposits (also called bedrock aquifers) of similar texture, porosity characteristics, and mineral composition. Independence requires physical separations among property measurement locations that ensure the vanishing of any statistical dependence among its values. Measurement locations so chosen produce samples of measurements that are uncorrelated. The minimal spatial separation among measurements must exceed the correlation scale of the saturated hydraulic conductivity. The correlation scale can be estimated by using geostatistical procedures (Loáiciga et al. 2010).

Basic Notation and Key Statistics

A sample of n measurements of an aquifer property K is assumed available for statistical inference. The individual measurements are denoted by $k_1, k_2, \ldots, k_n$, or symbolically by k_j, where $j =$ 1,2, $\ldots, n$. The natural logarithm of K is denoted by $Y = \ln K$. The sample of Y values is denoted by y_j $(= \ln k_j)$, where $j = 1, 2, \ldots, n$. The logarithmic transformation is commonly applied to permeability, hydraulic conductivity, or other aquifer properties that are frequently found to be lognormally distributed. That is, the property is rendered normally distributed upon undergoing the logarithmic transformation. The following subsections introduce several important statistics that describe the central tendency, the degree of spread about a measure of central tendency, and the skewness of data. The statistics are necessary in fitting PDFs to measurements of aquifer properties.

Sample Average and Median

Calculate the sample average of the property K by using the following:

$$\bar{K} = \frac{1}{n}\sum_{j=1}^{n} k_j \tag{2}$$

The sample average $\bar{K}$ is an estimate of the unknown population average of K, μ_K. The sample average is a measure of the central tendency of the data it represents.

The sample average of the log property Y is calculated with the following:

$$\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} y_j \tag{3}$$

The sample average $\bar{Y}$ is an estimate of the unknown population average of Y, μ_Y.

The sample median of a sample of measurements is defined as the value so that half the sample measurements are larger and half the sample measurement values are smaller. It is denoted by $\bar{K}_{0.50}$, where the 0.50 subscript indicates that the median equals the 50% quantile, that is, the probability that the random variable K be equal to or less than $K_{0.50}$ equals 50%.

Geometric Mean

Calculate the sample geometric mean of K with the following:

$$\bar{K}_G = e^{\bar{Y}} \tag{4}$$

The sample geometric mean is an estimate of the unknown population geometric mean, $K_G = \exp(\mu_Y)$. The geometric mean is sometimes used as an effective saturated hydraulic conductivity in groundwater hydrology. The effective saturated hydraulic conductivity is a parameter that relates the average groundwater specific discharge to the average hydraulic gradient.

Standard Deviation and Variance

Calculate the sample standard deviation of the property K as follows:

$$\bar{\sigma}_K = \left[\frac{1}{n-1}\sum_{1}^{n}(k_j - \bar{K})^2\right]^{1/2} \tag{5}$$

The sample's standard deviation $\bar{\sigma}_K$ is an estimate of the unknown population standard deviation of K, σ_K. The sample variance of the property K is equal to $\bar{\sigma}_K^2$. The sample standards deviation measures the spread of the data about its average.

The sample standard deviation of the log property ($\bar{\sigma}_Y$) is calculated as follows:

$$\bar{\sigma}_Y = \left[\frac{1}{n-1}\sum_{1}^{n}(y_j - \bar{Y})^2\right]^{1/2} \tag{6}$$

The sample standard deviation $\bar{\sigma}_Y$ is an estimate of the unknown population standard deviation of Y, σ_Y. The sample variance of log conductivity equals $\bar{\sigma}_Y^2$.

Coefficient of Skew

The sample coefficient of skew measures the degree of asymmetry of a set of measurements of the property K. It may take positive or negative values. The larger the coefficient of skew, the more asymmetric the PDF of the property K. A symmetric PDF, such as the normal PDF, has a coefficient of skew equal to zero. The sample coefficient of skew is calculated by using the following:

$$\bar{C}_{sK} = \frac{n}{(n-1)(n-2)}\sum_{j=1}^{n}\left(\frac{k_j - \bar{K}}{\bar{\sigma}_K}\right)^3 \tag{7}$$

The sample coefficient of skew in Eq. (7) estimates the population coefficient of skew C_{sK}. The sample coefficient of skew of the log property Y is calculated as follows:

$$\bar{C}_{sY} = \frac{n}{(n-1)(n-2)}\sum_{j=1}^{n}\left(\frac{y_j - \bar{Y}}{\bar{\sigma}_Y}\right)^3 \tag{8}$$

The sample coefficient of skew in Eq. (8) estimates the population coefficient of skew C_{sY}. If the log property Y is normally distributed, then its coefficient of skew equals zero. In this instance, the sample coefficient of skew of the log property Y tends to zero. In practice, if $-0.05 \le \bar{C}_{sY} \le 0.05$, then the log property Y can be assumed to be normally distributed, or equivalently, that the property K follows a lognormal PDF. Otherwise, that is, if $|\bar{C}_{sY}| > 0.05$, use a skewed PDF to fit the log property Y.

The average, standard deviation, median, and coefficient of skew can be calculated expeditiously and accurately by using functions available in commercial spreadsheets and numerical software such Microsoft Excel and *MATLAB*.

PDFs Frequently Used in Groundwater Hydrology

This paper presents several PDFs that have been used to model aquifer properties or groundwater processes. Several applications are included in the following sections.

Lognormal PDF

The lognormal PDF has been found to fit well many types of data, including aquifer properties such as permeability and hydraulic conductivity. Freeze (1975) provided early impetus for using the lognormal PDF as a statistical model to fit hydraulic conductivity data. Over time, the lognormal PDF has been accepted as a viable model for describing a variety of aquifer properties [see a discussion of this topic in Loáiciga et al. (2006)]. The following are attractive features of the lognormal PDF in the modeling of some aquifer properties: (1) it can fit positively skewed data; (2) the parameters of a normally distributed log property Y, symbolically $Y \sim N(\mu_Y, \sigma_Y^2)$, are the population mean μ_Y and the population variance σ_Y^2, which are estimable by using the standard sample estimators for the mean and variance introduced previously. Moreover, the quantiles of Y can be obtained straightforwardly from tabulated quantiles of the standard normal PDF $N(0,1)$ or from statistical software. Conversely, the lognormal PDF cannot be used to model either skewed log data or negatively skewed aquifer data. Although the lognormal PDF allows positive lower bounds on aquifer data, it does not allow upper bounds. In contrast, the log-gamma PDF, a generalization of the gamma PDF, can fit skewed data, with upper and lower bounds, or with upper or lower bounds, as shown in the following.

Properties of Lognormal PDF

Let K and θ denote an aquifer property and its lower bound, respectively, and $Y = \ln(K - \theta)$ be the log property. Evidently, $K = \exp(Y) + \theta$. The three-parameter lognormal PDF is given by the following (μ_Y denotes the population mean of the log property Y):

$$f_K(s) = \frac{1}{(s-\theta)\sigma_Y\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left[\frac{\ln(s-\theta)-\mu_Y}{\sigma_Y}\right]^2\right\} \quad s > \theta \tag{9}$$

in which the lower bound θ is, from physical feasibility, nonnegative. The lower bound θ is generally assumed equal to zero in most applications of the lognormal PDF in groundwater hydrology. The lognormal PDF in Eq. (9) implies a number of formulas for the property K, the log property Y, and their parameters. These are presented in the following.

Expected Value of Property K

The expected value of the property K can be shown to be equal to

$$\mu_K = e^{(\mu_Y+\sigma_Y^2)} + \theta \tag{10}$$

The expected value μ_K is estimated by the sample average in Eq. (2).

Median and Geometric Mean of Property K ($K_{0.50}$)

The median is

$$K_{0.50} = e^{\mu_Y} + \theta \tag{11}$$

The sample estimator, $\bar{K}_{0.50}$, of the median $K_{0.50}$ is obtained by replacing μ_Y in Eq. (11) with the sample average $\bar{Y}$ introduced in Eq. (3). The geometric mean of the property K is

$$K_G = \theta + \exp(\mu_Y) \tag{12}$$

usually with $\theta = 0$, in which case the geometric mean and median of lognormally distributed data K are equal to each other. Eq. (11) is convenient to estimate the lower bound θ. To do so, the sample estimator $\bar{K}_{0.50}$ and sample average of log values $\bar{Y}$ are obtained from a sample of measurements, and then $K_{0.50}$ and μ_Y are replaced, respectively, in Eq. (11), which is then solved for an estimate of θ. Alternatively, and preferably, θ could be estimated through maximum likelihood, a parameter estimation method that is not covered in this work. It is commonly assumed in practical applications in groundwater hydrology that $\theta = 0$.

Mode of Property K

The mode (K_M) is the most likely value of K

$$K_M = e^{\mu_Y-\sigma_Y^2} + \theta \tag{13}$$

Eqs. (10), (11), and (13) show that $K_M < K_{0.50} < \mu_K$.

Variance of Property K (σ_K^2)

The following formula provides a relation between the variance of the property K and that of its log property Y:

$$\sigma_K^2 = e^{2\mu_Y+\sigma_Y^2}\cdot(e^{\sigma_Y^2}-1) \tag{14}$$

The variance of K is estimated by the square of the sample standard deviation in Eq. (5).

Coefficient of Variation of K(C_{vK})

$$C_{vK} \equiv \frac{\sigma_K}{\mu_K} = (e^{\sigma_Y^2}-1)^{1/2} \tag{15}$$

The coefficient of variation is a dimensionless ratio that measures the magnitude of the standard deviation of K relative to its mean. The larger the coefficient of variation, the larger the variability of K about its mean. It is calculated from a sample of measurements by the ratio of the sample standard deviation $\bar{\sigma}_K$ over the sample average $\bar{K}$.

Coefficient of Skew of Property K (C_{sK})

$$C_{sK} \equiv \frac{E[K-\mu_K]^3}{\sigma_K^3} = \frac{(e^{3\sigma_Y^2}-3e^{\sigma_Y^2}+2)}{C_{vK}^3} \tag{16}$$

where C_{vK} is given by Eq. (15). The C_{sK} in Eq. (16) is always positive. It is estimated by Eq. (7)

Quantiles of Property K

For $0 < p < 1$, $P(K \leq K_p) = p$ defines the pth quantile (K_p) of the property K. K_p is estimated by the following:

$$\bar{K}_p = \exp(\bar{Y} + z_p\bar{\sigma}_Y) + \bar{\theta} \tag{17}$$

where $\bar{\theta}$ = estimator of the parameter θ; and $z_p = p$th quantile of the standard normal variate with zero mean and unit variance, which is readily obtained with a software such as Excel, in which case by using the function z_p = norm.s.inv(p). The estimate of the quantile K_p can be obtained directly as follows:

$$\bar{K}_p = \exp(\bar{Y}_p) + \bar{\theta} \tag{18}$$

where the pth quantile estimator $\bar{Y}_p$ of the log property Y can be obtained with the norm.inv(p, $\bar{Y}$, $\bar{\sigma}_Y$) function of Excel.

The lower bound $\bar{\theta}$ in Eqs. (17) and (18) is commonly assumed equal to zero in applications.

Gamma PDF and Its Special Case, Exponential PDF

The gamma PDF is a versatile model that is used in many fields of science and engineering, including groundwater hydrology. The gamma PDF offers great flexibility in fitting a wide range of natural phenomena while preserving mathematical simplicity. Loáiciga (2004) proposed the gamma PDF as an alternative to the lognormal PDF in an analysis of stochastic groundwater flow and solute transport. Loáiciga and Leipnik (2005) applied the bivariate correlated gamma PDF to model water-quality variables.

Properties of Gamma PDF

The PDF of a three-parameter gamma-distributed aquifer property K is

$$g_K(s) = \frac{\left(\frac{s-\theta}{\beta}\right)^{\alpha}|s-\theta|^{-1}e^{-[(s-\theta)/\beta]}}{\Gamma(\alpha)}$$
$$s \geq \theta \quad \text{if } \beta > 0, \quad s \leq \theta \quad \text{if } \beta < 0 \tag{19}$$

where α and β = shape and scale parameters, respectively, and $\alpha > 0$; θ = lower bound of the variable K when $\beta > 0$ and upper bound when $\beta < 0$. Most applications in groundwater hydrology assume that $\theta = 0$. Γ denotes the gamma function as follows:

J. Hydrol. Eng., 2015, 20(5): 04014063

$$\Gamma(\alpha) = \int_0^\infty e^{-\nu}\nu^{\alpha-1}d\nu \quad (20)$$

The gamma function is widely tabulated and programmed in commercial software (Excel, *MATLAB*, *Mathematica*). The domain of the gamma PDF is $[-\infty, \theta]$ when $\beta < 0$, which contains negative numbers and thus violates the nonnegativity of positive-valued aquifer properties. When $\theta = 0$, $\alpha = 1$, and $\beta > 0$, the gamma PDF in Eq. (20) becomes the exponential PDF with parameter $\lambda = 1/\beta$. The exponential PDF is

$$h_K(s) = \lambda e^{-\lambda s} \quad s \geq 0; \quad \lambda > 0 \quad (21)$$

The following subsection summarizes the properties of the gamma PDF for positive or negative scale parameter β.

Expected Value (Mean) of Property K

$$\mu_K = \alpha\beta + \theta \quad (22)$$

Median of Property K

$$K_{0.50} = \psi_{0.50}\beta + \theta \quad (23)$$

where $\psi_{0.50}$ must be obtained from the integral equation

$$\frac{1}{\Gamma(\alpha)}\int_0^{\psi_{0.50}} e^{-\nu}\nu^{\alpha-1}d\nu = \frac{1}{2} \quad (24)$$

The integral on the left-hand side of Eq, (24) is called the incomplete gamma function $\gamma(\alpha, \psi_{0.50})$ (e.g., Gradshteyn and Ryzhik 1994), so that Eq. (24) can be shortened to

$$\frac{1}{\Gamma(\alpha)}\gamma(\alpha, \psi_{0.50}) = \frac{1}{2} \quad (25)$$

The left-hand side of Eq. (25) can be evaluated by using the GAMMA.INV(probability, alpha, beta) function in the software Excel, with probability $= 1/2$, alpha $= \alpha$, and beta $= \beta = 1$, which returns the value of $\psi_{0.50}$.

Mode of Property K

When $\alpha > 1$

$$K_M = (\alpha - 1)\cdot\beta + \theta \quad (26)$$

K_M is equal to θ when $0 < \alpha \leq 1$.

Variance of Property K

$$\sigma_K^2 = \alpha\beta^2 \quad (27)$$

Coefficient of Variation of Property K

$$C_{vK} = \frac{|\alpha^{1/2}\beta|}{|\alpha\beta + \theta|} \quad (28)$$

Coefficient of Skew of Property K

$$C_{sK} = \frac{2\alpha\beta^3}{\sigma_K^3} \quad (29)$$

where the sign of the skew is determined by that of the shape parameter β. $C_{sK} > 0$ when $\beta > 0$, in which case the PDF is positively skewed with lower bound θ. $C_{sK} < 0$ when $\beta < 0$, in which case the PDF is negatively skewed with upper bound θ.

Moment Estimators of α, β, and θ Parameters

These are deducible from the various properties of the gamma PDF listed previously. The moment estimators are

$$\bar{\alpha} = \frac{4}{\bar{C}_{sK}^2} \quad (30)$$

$$\bar{\beta} = \frac{\bar{\sigma}_K \bar{C}_{sK}}{2} \quad (31)$$

$$\bar{\theta} = \bar{K} - \frac{2\bar{\sigma}_K}{\bar{C}_{sK}} \quad (32)$$

where $\bar{K}$, $\bar{\sigma}_K$, and $\bar{C}_{sK}$ in Eqs. (30–32) represent the sample estimators of the mean, variance, and coefficient of skewness of the property K, respectively. The estimators in Eqs. (30–32) are estimable also through the maximum likelihood method, which is not covered in this paper.

Quantiles of Property K

For $0 < p < 1$, $P[K \leq K_p] = p$ defines the pth quantile. In particular, $K_{0.50}$ equals the median. In general, K_p is estimated by the following:

$$\bar{K}_p = \bar{K} + \left[\frac{\psi_q \bar{C}_{sK}}{2} - \frac{2}{\bar{C}_{sK}}\right]\bar{\sigma}_K \quad (33)$$

where ψ_q must be obtained from the following integral equation $(0 < p < 1)$:

$$\frac{1}{\Gamma(\bar{\alpha})}\int_0^{\psi_q} e^{-\nu}\nu^{\bar{\alpha}-1}d\nu = p \quad \text{if } \bar{C}_{sK} > 0 \quad (\text{when } \bar{\beta} > 0) \quad (34)$$

with $\bar{\alpha} = 4/\bar{C}_{sK}^2$, or from

$$\frac{1}{\Gamma(\bar{\alpha})}\int_0^{\psi_q} e^{-\nu}\nu^{\bar{\alpha}-1}d\nu = 1 - p \quad \text{if } \bar{C}_{sK} < 0 \quad (\text{when } \bar{\beta} < 0) \quad (35)$$

where $\bar{\alpha} = 4/\bar{C}_{sK}^2$. All the special functions used in the previous equations related to the gamma PDF are available in commercial software, and their calculation is expeditious. In particular, the left-hand side of Eqs. (34) and (35) can be evaluated by using the GAMMA.INV(probability, alpha, beta) function in Excel, with probability $q = p$ (if $\bar{C}_{sK} > 0$) or $1 - p$ (if $\bar{C}_{sK} < 0$), alpha $= \bar{\alpha}$, and beta $= \bar{\beta} = 1$, which returns the value of ψ_q. In the limit $\bar{C}_{sK} \to 0$, the factor within brackets in Eq. (33) tends to the standard normal quantile z_p. Specifically

$$\lim_{\bar{C}_{sK}\to 0}\left[\frac{\psi_q \bar{C}_{sK}}{2} - \frac{2}{\bar{C}_{sK}}\right] \to z_p \quad (36)$$

so that the quantile $\bar{K}_p$ in Eq. (33) becomes $\bar{K}_p = \bar{K} + z_p\bar{\sigma}_K$. In other words, the gamma PDF approaches the normal PDF when the coefficient of skew tends to zero.

Log-Gamm PDF

A variant of the gamma PDF is the log-gamma PDF (also called log-Pearson Type III), which is used by federal agencies in the United States to fit annual streamflow peaks at gauged sites (e.g., USGS 1982). It is particularly well suited for modeling extreme values and highly skewed data. When using the log-gamma PDF, it is assumed that the logarithm of the property K (i.e., $Y = \ln(K)$) follows the gamma PDF in Eq. (19) with the shape and scale parameters replaced by α_Y and β_Y, respectively. In this instance, $\theta = \theta_Y$ in Eq. (19) denotes the lower bound of Y when $\beta_Y > 0$ or its upper bound when $\beta_Y < 0$. The log property Y has the following gamma PDF:

$$g_Y(s) = \frac{\left(\frac{s-\theta_Y}{\beta_Y}\right)^{\alpha_Y} |s-\theta_Y|^{-1} e^{-[(s-\theta_Y)/\beta_Y]}}{\Gamma(\alpha_Y)}$$

$$s \geq \theta_Y \quad \text{if } \beta_Y > 0, \quad s \leq \theta_Y \quad \text{if } \beta_Y < 0 \tag{37}$$

Evidently, $K = \exp(Y) = \exp(Y)$, which is positive with lower or upper bound $\exp(\theta_Y)$ depending on whether $\beta_Y > 0$ or $\beta_Y < 0$, respectively. The PDF of the log-gamma-distributed K is

$$h_K(s) = \frac{\left(\frac{\ln(s)-\theta_Y}{\beta_Y}\right)^{\alpha_Y} |\ln(s)-\theta_Y|^{-1} e^{-[\ln(s)-\theta_Y/\beta_Y]}}{s\Gamma(\alpha_Y)} \tag{38}$$

where $s \geq e^{\theta_Y}$ if $\beta_Y > 0$ or $0 < s \leq e^{\theta_Y}$ if $\beta_Y < 0$. Key properties of the log-gamma-distributed property K are derivable from its PDF in Eq. (38). These are presented next.

Expected Value of Property *K*

$$\mu_K = \frac{e^{\theta_Y}}{(1-\beta_Y)^{\alpha_Y}} \tag{39}$$

Geometric Mean of Property *K*

$$K_G \equiv e^{E(Y)} = e^{\alpha_Y \beta_Y + \theta_Y} \tag{40}$$

Median of Property *K*

$$K_{0.50} = e^{\psi_{0.50}\beta_Y + \theta_Y} \tag{41}$$

where $\psi_{0.50}$ is obtained from the solution of the integral Eq. (25).

Mode of Property *K*

When $\alpha > 1$

$$K_M = e^{[(\alpha_Y - 1)\frac{\beta_Y}{\beta_Y+1} + \theta_Y]} \tag{42}$$

The mode equals e^{θ_Y} if $0 < \alpha_Y \leq 1$.

Variance of Property *K*

$$\sigma_K^2 = \mu_K^2 \cdot \left[\left(\frac{(1-\beta_Y)^2}{(1-2\beta_Y)}\right)^{\alpha_Y} - 1\right] \tag{43}$$

where μ_K (the expected value of K) is given by Eq. (39).

Coefficient of Variation of Property *K*

$$C_{vK} = \left\{\left[\frac{(1-\beta_Y)^2}{(1-2\beta_Y)}\right]^{\alpha_Y} - 1\right\}^{1/2} \tag{44}$$

Coefficient of Skew of Property *K*

$$C_{sK} = \frac{\left[\frac{(1-\beta_Y)^3}{(1-3\beta_Y)}\right]^{\alpha_Y} - 3\left[\frac{(1-\beta_Y)^2}{(1-2\beta_Y)}\right]^{\alpha_Y} + 2}{C_{vK}^3} \tag{45}$$

where C_{vK} is given by Eq. (44).

Moment Estimators of Log Parameters α_Y, β_Y, and θ_Y

Moment estimators are obtained by resorting to the fact that $Y = \ln(K)$ is gamma distributed. Letting $\bar{Y}$, $\bar{\sigma}_Y$, and $\bar{C}_{sY}$ be the estimates of the mean, standard deviation, and coefficient of skew of Y, respectively, one obtains

$$\bar{\alpha}_Y = \frac{4}{\bar{C}_{sY}^2} \tag{46}$$

$$\bar{\beta}_Y = \frac{\bar{\sigma}_Y \bar{C}_{sY}}{2} \tag{47}$$

$$\bar{\theta}_Y = \bar{Y} - \frac{2\bar{\sigma}_Y}{\bar{C}_{sY}} \tag{48}$$

Quantiles of Property *K*

For $0 < p < 1$, $P[K \leq K_p] = p$ defines the pth quantile (K_p) of the property K. In particular, $K_{0.50}$ equals the median. In general, the estimate of K_p is given by the following:

$$\bar{K}_p = \exp\left[\bar{Y} + \left(\frac{\psi_q \bar{C}_{sY}}{2} - \frac{2}{\bar{C}_{sY}}\right)\bar{\sigma}_Y\right] \tag{49}$$

where ψ_q must be obtained from the following integral equations ($0 < p < 1$):

$$\frac{1}{\Gamma(\bar{\alpha}_Y)}\int_0^{\psi_q} e^{-\nu}\nu^{\bar{\alpha}_Y - 1} d\nu = p \quad \text{if } \bar{C}_{sY} > 0 \quad (\text{i.e., } \bar{\beta}_Y > 0) \tag{50}$$

where $\bar{\alpha}_Y = 4/\bar{C}_{sY}^2$, or

$$\frac{1}{\Gamma(\bar{\alpha}_Y)}\int_0^{\psi_q} e^{-\nu}\nu^{\bar{\alpha}_Y - 1} d\nu = 1 - p \qquad \text{if } \bar{C}_{sY} < 0 \quad (\text{i.e., } \bar{\beta}_Y < 0) \tag{51}$$

where $\bar{\alpha}_Y = 4/\bar{C}_{sY}^2$. The left-hand side of Eqs. (50) and (51) can be evaluated by using the GAMMA.INV(probability, alpha, beta) function in the software Excel, with probability $q = p$ (if $\bar{C}_{sY} > 0$) or $1 - p$ (if $\bar{C}_{sY} < 0$), alpha $= \bar{\alpha}_Y$, and beta $= \bar{\beta}_Y = 1$, which returns the value of ψ_q.

In the limit $\bar{C}_{sY} \to 0$, the factor within parentheses on the right-hand side of Eq. (49) tends to the standard normal quantile z_p. Specifically,

$$\lim_{\bar{C}_{sY}\to 0}\left[\frac{\psi_q \bar{C}_{sY}}{2} - \frac{2}{\bar{C}_{sY}}\right] \to z_p \tag{52}$$

so that the estimate of the quantile K_p in Eq. (49) becomes

$$\bar{K}_p = \exp[\bar{Y} + z_p \bar{\sigma}_Y] \tag{53}$$

Therefore, the log-gamma PDF approaches the lognormal PDF when the coefficient of skew tends to zero [compare Eq. (53) with Eq. (17), after setting $\bar{\theta} = 0$ in the latter equation].

Illustrative Examples

The following sections present applications of the lognormal, gamma, log-gamma, and exponential PDFs to various groundwater problems, including fitting aquifer data, groundwater flow, and water quality.

Application of Lognormal PDF to Hydraulic Conductivity Data

The hydraulic conductivity data shown in Fig. 1 have an average of $\bar{K} = 3.94 \times 10^{-8}$ cm/s and a standard deviation of $\bar{\sigma}_K = 3.70 \times 10^{-7}$, which implies an extraordinarily large coefficient of variation of $C_{vK} = 9.4$. The coefficient of skew equals 13.7, a testimony to an acutely right-skewed hydraulic conductivity data. The K data

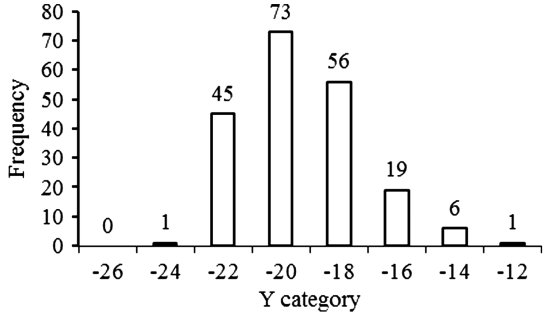

Fig. 2. Log conductivity data Y graphed in Fig. 1

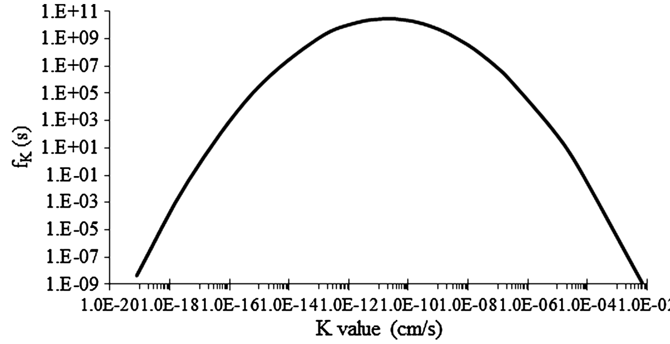

Fig. 3. Normal PDF fitted to the log conductivity data graphed in Fig. 1

were log transformed to produce the log conductivity data Y in an attempt to reduce the asymmetry and facilitate fitting a PDF to the hydraulic conductivity data. The sample average ($\bar{Y}$), standard deviation ($\bar{\sigma}_Y$), and coefficient of skew ($\bar{C}_{sY}$) of the log conductivity data equal −20.30, 2.08, and 0.592, respectively. Although the skew coefficient was reduced by the log transformation, a histogram of the Y data shown in Fig. 2 confirms that it is skewed to the right.

If the log conductivity Y were normally distributed, its PDF would be (setting $\theta = 0$)

$$f_Y(y) = \frac{1}{2.08\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{y-(-20.3)}{2.08}\right]^2\right\} \tag{54}$$

It is customary in applications to assume that the log conductivity Y is sufficiently close to a normal PDF, and then use Eqs. (17) or (18) to estimate quantiles of the hydraulic conductivity K. For example, if the lower quartile ($K_{0.25}$) and upper quartile ($K_{0.75}$) of the hydraulic conductivity were needed in a simulation study of groundwater flow and chemical and heat transport, these two values could be approximated as follows [using Eq. (17) with $\theta = 0$]:

$$\bar{K}_{0.25} = e^{\bar{Y}+z_{0.25}\bar{\sigma}_Y} = e^{-20.3+(-0.6745)2.08} = 3.75 \times 10^{-10}\ \text{cm/s} \tag{55}$$

$$\bar{K}_{0.75} = e^{\bar{Y}+z_{0.75}\bar{\sigma}_Y} = e^{-20.3+0.6745\cdot 2.08} = 6.21 \times 10^{-10}\ \text{cm/s} \tag{56}$$

Under the assumption that hydraulic conductivity K is approximately lognormally distributed, its PDF is

$$f_K(s) = \frac{1}{s(2.08)\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{\ln(s)-(-20.3)}{2.08}\right]^2\right\} \tag{57}$$

Eq. (57) is graphed in Fig. 3. Because of the wide range of the hydraulic conductivity and the complexity of Eq. (57) relative to the normal PDF Eq. (54), the latter formula is easier to work with when making calculations aimed at inferring hydraulic conductivity values.

Application of Log-Gamma PDF to Fit Hydraulic Conductivity Data

The example of the previous section showed how the logarithmic transformation of hydraulic conductivity data can reduce its asymmetry to the point that a normal PDF can be fitted to the log conductivity data reasonably well. One can go one step farther and fit an asymmetric PDF to skewed log conductivity data. Furthermore, one can carry out a formal statistical goodness-of-fit test to ascertain whether the proposed (asymmetric) PDF is an acceptable match to the hydraulic conductivity data. With these two aims, that is, fitting a PDF and testing the fit, the log-gamma PDF in Eq. (38) was fitted to the hydraulic conductivity data graphed in Fig. 1. Recall that the sample average, standard deviation, and coefficient of skew of log conductivity are $\bar{Y} = -20.3$, $\bar{\sigma}_Y = 2.08$, and $\bar{C}_{sY} = 0.592$, respectively. These were used to calculate the log-gamma parameters $\bar{\alpha}_Y = 11.4$, $\bar{\beta}_Y = 0.616$, $\bar{\theta}_Y = -27.3$ by using Eqs. (46–48), respectively. Because $\bar{C}_{sY}$ is positive, the hydraulic conductivity has lower bound equal to $\exp(\bar{\theta}_Y) = 1.39 \times 10^{-12}$. The log-gamma PDF of hydraulic conductivity K is

$$h_K(s) = \frac{\left[\frac{\ln(s)-(-27.3)}{0.616}\right]^{11.4} |\ln(s)-(-27.3)|^{-1} e^{-\left[\frac{\ln(s)-(-27.3)}{0.616}\right]}}{s\Gamma(11.4)} \tag{58}$$

Eq. (58) is graphed in Fig. 4.

Use Eq. (49) to calculate the quantiles of a log-gamma-distributed property. Suppose that the quantiles of log conductivity $K_{0.25}$, $K_{0.50}$, and $K_{0.75}$ are wanted. The values $\Psi_{q=p}$ for $p = 0.25$, 0.50, and 0.75 equal 8.98, 11.1, and 13.5, respectively. Estimate the values of the desired quantiles by using Eq. (49): $\bar{K}_{0.25} = 3.43 \times 10^{-10}$, $\bar{K}_{0.50} = 1.25 \times 10^{-9}$, and $\bar{K}_{0.75} = 5.45 \times 10^{-9}$ cm/s.

Goodness-of-Fit Testing: Chi-Square Test

The chi-square goodness-of-fit test is a formal procedure used to accept or reject a proposed PDF to fit specific data. The procedure can be used for any type of data. Goodness-of-fit tests other than the chi-square test are available. Benjamin and Cornell (1970) and Gilbert (1987) review several goodness-of-fit tests. The following steps must be implemented in applying the chi-square test:

1. Calculate R saturated hydraulic conductivity (or other groundwater variable) quantiles, denoted by $K_{\Delta p} < K_{2\Delta p} < \cdots < K_{R\Delta p}$, using the appropriate equation for quantile calculation. In the notation $K_{r\Delta p}$, the probability corresponding to the quantile is $r \cdot \Delta p$, where $r = 1,2, \ldots, R$, and the probability increment Δp is defined by Eq. (58). A suitable range for R is

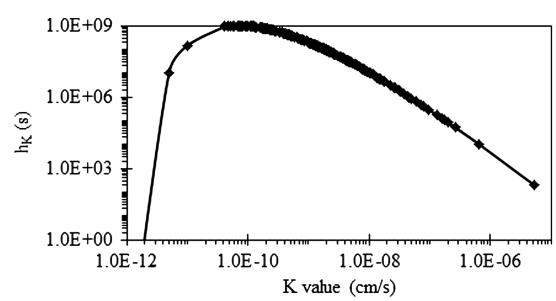

Fig. 4. Log-gamma PDF fitted to the hydraulic conductivity K data shown in Fig. 1

$4 \le R \le 9$. The quantiles $K_{r\Delta p}$ with $r = 1,2, \ldots, R$ are chosen so that they define $R+1$ equal-probability, nonoverlapping intervals of hydraulic conductivity

$$P[K_{r\Delta p} \le K \le K_{(r+1)\Delta p}] = P(K < K_{\Delta p}) = P(K > K_{R\Delta p}) = \Delta p \qquad (59)$$

for $r = 1,2, \ldots, R-1$, where

$$\Delta p = \frac{1}{R+1} \qquad (60)$$

is the probability of each of the $R+1$ intervals of saturated hydraulic conductivity defined by the quantiles $K_{r\Delta p}$, $r = 1,2, \ldots, R$. The quantiles satisfy the probability statement

$$P(K \le K_{r\Delta p}) = r \cdot \Delta p \quad r = 1,2, \ldots, R \qquad (61)$$

2. The expected number of K measurements that fall in any of the $R+1$ (equal-probability) intervals equals $n \cdot \Delta p$, where n is the number of K measurements available. This number compares with the actual number of K measurements observed in the rth interval, n_r, $r = 1,2, \ldots, R+1$. Calculate the test statistic

$$D = \frac{1}{n \cdot \Delta p} \sum_{r=1}^{R+1} (n_r - n \cdot \Delta p)^2 \qquad (62)$$

3. Determine the chi-square critical value associated with a 5% significance level and $R-f$ degrees of freedom, $\chi^2_{0.05,R-f}$. The number of degrees of freedom of the chi-square critical value is customarily R. However, f = two parameters ($\bar{Y}$ and $\bar{\sigma}_Y$) must be estimated from K data for the lognormal PDF (with lower bound equal to zero), and f = three parameters (α_Y, β_Y, θ_Y) must be estimated from data for the log-gamma PDF. Therefore, the number of degrees of freedom of the chi-square critical value becomes $R-f$. The chi-square critical value is tabulated in the technical literature. It can also be obtained by using commercial software. In Excel, the function CHISQ.INV.RTV(0.05, $R-f$) returns the critical value $\chi^2_{0.05,R-f}$. The software *MATLAB* returns the critical value $\chi^2_{0.05,R-f}$ by using the command chi2inv(0.95, $R-f$).
4. If the test statistic D exceeds $\chi^2_{0.05,R-f}$, reject the fitted PDF as a suitable probability model for the K data. Otherwise, accept the fitted PDF.

Calculation Example on How to Fit and Test the Log-Gamma PDF

The goodness-of-fit of the log-gamma PDF to the K data shown in Fig. 1 is assessed. Recall the basic statistics pertaining to the K data are average $\bar{K} = 03.94 \times 10^{-8}$ cm/s, standard deviation $\bar{\sigma}_K = 3.70 \times 10^{-7}$, and coefficient of skew $C_{sK} = 13.7$. Choose $R = 9$. Quantify nine quantiles $K_{\Delta p} < K_{2\Delta p} < \cdots < K_{9\Delta p}$ by using Eq. (49), which define $R+1 = 9+1 = 10$ equal-probability intervals. The probability associated with each interval is $\Delta p = 1/10 = 0.10$ so that the expected number of K measurements in each interval is $n\Delta p = 201 \times 0.10 = 20.1$. The number of measurements observed in each interval is counted from the K sample. The test statistic [D, Eq. (62)] is calculated, and the chi-square critical value determined. Results are summarized in Table 1.

$D = 6.91 < \chi^2(0.05, 9-3=6) = 12.59$; thus, the log-gamma PDF is accepted as a suitable probability model for the K data used in this example. Fig. 5 summarizes the key features of this example.

Table 1. Results of Goodness-of-Fit Test for Data in Fig. 1 and Log-Gamma PDF

Interval number (r)	Upper limit of interval (cm/s)	Expected number ($n \cdot \Delta p$)	Observed number (n_r)	$(n_r - n \cdot \Delta p)^2/(n \cdot \Delta p)$
1	$\bar{K}_{0.10} = 1.248 \times 10^{-10}$	20.10	24	0.757
2	$\bar{K}_{0.20} = 2.559 \times 10^{-10}$	20.10	20	0.000
3	$\bar{K}_{0.30} = 4.494 \times 10^{-10}$	20.10	14	1.851
4	$\bar{K}_{0.40} = 7.505 \times 10^{-10}$	20.10	19	0.060
5	$\bar{K}_{0.50} = 1.245 \times 10^{-9}$	20.10	16	0.836
6	$\bar{K}_{0.60} = 2.119 \times 10^{-9}$	20.10	28	3.105
7	$\bar{K}_{0.70} = 3.859 \times 10^{-9}$	20.10	19	0.060
8	$\bar{K}_{0.80} = 8.097 \times 10^{-9}$	20.10	22	0.180
9	$\bar{K}_{0.90} = 2.433 \times 10^{-8}$	20.10	20	0.000
10	$\propto$	20.10	19	0.060

Note: $D = 6.91$; $\chi^2(0.05, 6) = 12.59$.

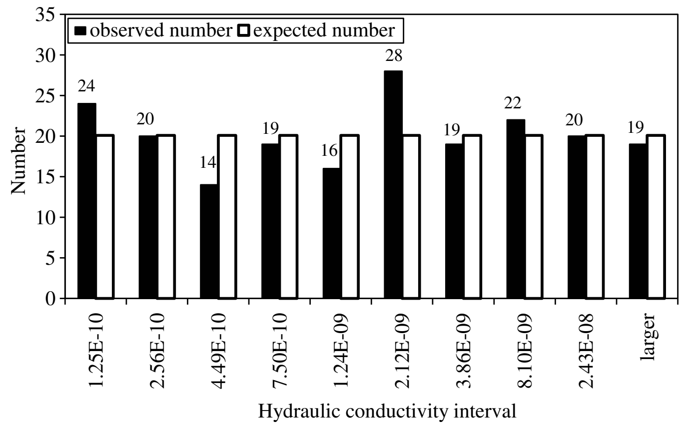

Fig. 5. Observed and expected numbers of K values (cm/s) in 10 equal-probability ($\Delta p = 0.10$) intervals; the expected number of K values equals 20.1 and is represented by the white bars

Application of Exponential Function to Hydraulic Conductivity Data

The exponential PDF has found applications in many fields of inquiry, including groundwater hydrology. Its PDF is

$$h_K(s) = \lambda e^{-\lambda s} \quad s \ge 0, \quad \lambda > 0 \qquad (63)$$

The parameter λ can be estimated from the sample average of the property (e.g., hydraulic conductivity), $\bar{K}$, as follows:

$$\bar{\lambda} = \frac{1}{\bar{K}} \qquad (64)$$

The pth quantile of the exponential PDF is calculated as follows:

$$\bar{K}_p = \frac{1}{\bar{\lambda}} \ln\left(\frac{1}{1-p}\right) \qquad (65)$$

Table 2 lists the hydraulic conductivity values measured with constant-head permeameter in a silty sand derived from weathered sandstone and shale. The exponential PDF was fitted to the K data in Table 2 to yield $h_K(s) = 0.538 \exp(-0.538)$ with sample average $\bar{K} = 1.86$ and parameter $\bar{\lambda} = 0.538$.

Table 3 summarizes the results of the chi-square test implement to assess the goodness-of-fit of the exponential PDF to the

Table 2. Measurements of Hydraulic Conductivity in Silty Sand Obtained with Constant-Head Permeameter (Data from Loáiciga et al. 2008b)

Sample number	Sample identification code	K (m/day)
1	GB1-2	0.14
2	GB1-4	3.52
3	GB1-6	1.12
4	GB1-7	4.58
5	GB2-1.5	2.42
6	GB2-3	0.23
7	GB3-2	5.36
8	GB3-3.5	0.63
9	GB3-5	2.51
10	GB4-3	0.72
11	GB4-4	0.95
12	GB5-4	1.21
13	GB5-6	0.76

Note: Average $\bar{K} = 1.86$; $\bar{\lambda} = 0.538$.

Table 3. Results of Chi-Square Goodness-of-Fit Test Applied to Data in Table 2 and Exponential PDF

Interval number (r)	Interval of K	Δp	$n \cdot \Delta p$	n_r	$(n_r - n \cdot \Delta p)^2/(n \cdot \Delta p)$
1	<0.41	0.2	2.6	2	0.36
2	0.41–0.95	0.2	2.6	4	1.96
3	0.95–1.70	0.2	2.6	2	0.36
4	1.70–2.99	0.2	2.6	2	0.36
5	>2.99	0.2	2.6	3	0.16

Note: $D = 1.23$; $\chi^2(0.05, 3) = 7.81$.

hydraulic conductivity data in Table 2. $D = 1.23 < \chi^2(0.05, 4 - 1 = 3) = 7.81$; the exponential PDF is accepted as a suitable model for the data of Table 2. Fig. 6 graphs the results of the chi-square test.

Application of Gamma PDF to Residence Time and Age of Groundwater

The residence time of a groundwater particle (t_1) extends from the moment when the particle enters a groundwater flow system until it exits it. In contrast, the age of a water particle (t_2) moving in groundwater is the time elapsed since the particle entered the groundwater flow system (e.g., Loáiciga 2005). Let L be the total distance traveled by a groundwater particle in its journey through an aquifer, and let K denote the hydraulic conductivity of the

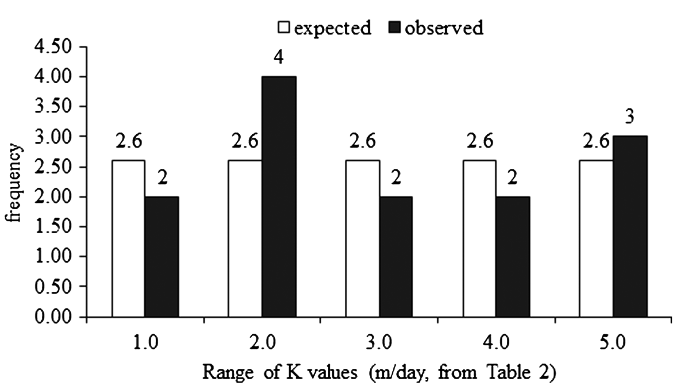

Fig. 6. Example histogram of K data on Table 2, indicative of an exponential PDF

aquifer. It is assumed that groundwater flow takes place under a constant hydraulic gradient g. The aquifer's porosity is n. The residence time of a groundwater particle equals the total travel distance L divided by the average groundwater velocity. The latter is obtained from Darcy's law. The residence time is then given by Eq. (66)

$$t_1 = \frac{n}{g}\frac{L}{K} \quad (66)$$

L and K are independent random variables. A gamma PDF is proposed to characterize the probabilistic characteristics of K. Therefore, the proposed PDF of K is a two-parameter gamma distribution

$$f_K(s) = \frac{s^{a_2-1}e^{-s/b_2}}{\Gamma(a_2)b_2^{a_2}} \quad a_2 > 1; \quad b_2 > 0; \quad s \geq 0 \quad (67)$$

where a_2 and b_2 = shape and scale parameters of the gamma distribution, respectively; and $\Gamma(\cdot)$ = gamma function defined as follows:

$$\Gamma(u) = \int_0^\infty e^{-v}v^{u-1}dv \quad (68)$$

The distribution of the total travel distance L is modeled by a two-parameter gamma distribution

$$f_L(s) = \frac{s^{a_1-1}e^{-s/b_1}}{\Gamma(a_1)b_1^{a_1}} \quad a_1 > 1; \quad b_1 > 0; \quad s \geq 0 \quad (69)$$

The hydraulic conductivity K and total travel distance L have expected values $a_1 b_1$ and $a_2 b_2$, respectively. Eq. (66) implies that the residence time is the scaled ratio of two independent gamma variables, in which the scaling ratio is the constant $n/g \equiv a > 0$. The PDF $f(t_1)$ of the residence time t_1 was derived by Loáiciga (2004)

$$f_1(t) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)}\left(a\frac{b_1}{b_2}\right)^{a_2}\frac{t^{a_1-1}}{(t + a\frac{b_1}{b_2})^{a_1+a_2}} \quad t \geq 0; \quad a = n/g \geq 0 \quad (70)$$

The average residence time is derived from Eq. (70) (letting $\beta = ab_1/b_2$), as follows:

$$T_1 = \int_0^\infty tf_1(t)dt = a\frac{b_1}{b_2}\frac{a_1}{(a_2 - 1)} \quad (71)$$

The average turnover time (T) of groundwater storage is defined as the storage (V) divided by the average rate of aquifer recharge (or discharge, R). Recharge into an aquifer displaces existing groundwater, so that the average time that it takes the recharge to replace the groundwater already in storage equals T. A plausible connection between T and T_1 is intuitive. The key to decipher this connection is the groundwater age. The age of a groundwater particle (t_2) moving through an aquifer is the time elapsed since the particle entered the aquifer. Its PDF is denoted by $f_2(t)$. The latter PDF may not be specified independently of $f_1(t)$, the PDF of the residence time, because t_1 and t_2 are interdependent. Such interdependency is deducible from basic mass-balance considerations, as shown in Loáiciga (2005), who proved that the PDF of the groundwater age, $f_2(t)$ is (with $\beta = ab_1/b_2$)

J. Hydrol. Eng., 2015, 20(5): 04014063

$$f_2(t) = \frac{1}{T}[1 - F_1(t)]$$
$$= \frac{1}{T}\left[1 - \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)}\frac{t^{a_1}}{a_1\beta^{a_1}} \cdot {}_2F_1\left(a_1+a_2, a_1; a_1+1; -\frac{t}{\beta}\right)\right] \quad (72)$$

where $F_1(t)$ = cumulative distribution function of the residence time, which is equal to the integral of $f_1(t)$; ${}_2F_1$ = Gauss' hypergeometric function (Gradshteyn and Ryzhik 1994); T = expected turnover time of groundwater storage introduced previously. Differentiating in Eq. (72) produces the following relationship between $f_1(t)$ and $f_2(t)$:

$$f_1(t) = -T\frac{df_2(t)}{dt} \quad (73)$$

Eq. (73) allows the average residence time to be written as follows:

$$T_1 = \int_0^\infty t f_1(t)dt = -T\int_0^\infty t\frac{df_2(t)}{dt}dt = T \quad (74)$$

Eq. (74) states that the average residence time equals the average turnover time, that is, $T_1 = T$. T_1 was given in Eq. (71). The equality $T_1 = T = V/R$ introduces a constraint involving T_1, V, and R. For example, if the average recharge rate R is known, the storage volume must be $V = T_1 R$. If V and R are known, then, from Eq. (71), $a \cdot b_1 \cdot a_1/[b_2(a_2-1)] = T_1 = V/R$, which imposes a constraint on the gamma parameters a_1, a_2, b_1, and b_2.

Taking into consideration the PDF $f_2(t)$ given in Eq. (72), the expected groundwater age is (with $\beta = ab_1/b_2$, in which $a_2 > 2$ to achieve convergence to a finite T_2)

$$T_2 = \frac{1}{T}\int_0^\infty t\left[1 - \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)}\frac{t^{a_1}}{a_1\beta^{a_1}} \times {}_2F_1\left(a_1+a_2, a_1; a_1+1; -\frac{t}{\beta}\right)\right]dt \quad (75)$$

The average age of groundwater expressed by Eq. (75) must be calculated numerically.

The results presented in this section demonstrate the versatility of the gamma PDF in modeling basic groundwater processes analytically. The next section expands on the power of the gamma PDF to model real-world data.

Application of Gamma PDF to Model Water Quality of Springs: Correlated Gamma Variables

Spring water in Las Palmas Creek, Santa Barbara, California, was tested to study the ratio of fecal coliforms (FC) to fecal streptococcus (FS) in it. Fecal coliforms and fecal streptococcus are enteric bacteria (i.e., they live in the intestinal tract of warm-blooded animals) and are frequently used as indicators of fecal contamination of water bodies (Loáiciga and Leipnik 2005). Loáiciga and Leipnik (2005) fitted FC and FS values with univariate gamma PDFs, allowing for correlation between them. The FC/FS ratio was determined from each pair of FC and FS values obtained from a single water sample. This procedure yielded 38 experimental values of FC/FS. The FC/FS ratio is of interest because, under suitable conditions, it may be used to discern the origin of enteric bacteria. In Las Palmas Creek, a FC/FS ratio in the interval [0, 0.4] was deemed of equine origin, whereas a FC/FS ratio $\geq$ 3.0 was considered to be human in origin. The range 0.4 < FC/FS < 4.0 was associated with mixed origin (i.e., humans, horses, and wildlife, Loáiciga and Leipnik 2005). If the sources of enteric bacteria are correctly identified, management actions are taken to counter the contamination of the spring water. Letting X_1 = FC and X_2 = FS, the correlated (four-parameter) gamma PDFs are

$$f_{X_1}(x_1) = \frac{x_1'^{\gamma\alpha_1-1}e^{-x_1'/b_1}}{\Gamma(\gamma\alpha_1)b_1^{\gamma\alpha_1}} \quad x_1 \geq \xi_1 \quad \text{if } b_1 > 0, \quad x_1 \leq \xi_1 \quad \text{if } b_1 < 0 \quad (76)$$

$$f_{X_2}(x_2) = \frac{x_2'^{\gamma\alpha_2-1}e^{-x_2'/b_2}}{\Gamma(\gamma\alpha_2)b_2^{\gamma\alpha_2}} \quad x_2 \geq \xi_2 \quad \text{if } b_2 > 0, \quad x_2 \leq \xi_2 \quad \text{if } b_2 < 0 \quad (77)$$

where $x_j' = x_j - \xi_j$, $j = 1, 2$; $\gamma\alpha_1$ and $\gamma\alpha_2$ = marginal shapes of the PDFs of X_1 and X_2, respectively; (b_1, b_2) and (ξ_1, ξ_2) = scale and location parameters, respectively; and γ = collective shape parameter of the bivariate distribution of X_1 and X_2. α_1, α_2, and γ are positive. The correlation coefficient ρ is defined in terms of the means (μ_1, μ_2) and variances (σ_1^2,σ_2^2) of X_1 and X_2, respectively, and a parameter β introduced by Loáiciga and Leipnik (2005) to induce statistical dependence between X_1 and X_2

$$\rho \equiv \frac{\mu_{1,1}}{\sigma_1\sigma_2} = \frac{E[(x_1-\mu_1)(x_2-\mu_2)]}{\sigma_1\sigma_2} = \frac{\beta\gamma}{\sigma_1\sigma_2} = \frac{\beta}{b_1 b_2\sqrt{\alpha_1\alpha_2}} \quad (78)$$

The sample estimator of the correlation coefficient is (Priestly 1989)

$$\bar{\rho} = \frac{1}{\bar{\sigma}_1\bar{\sigma}_2} \cdot \frac{1}{n}\sum_{j=1}^{n}(x_{1j} - \bar{X}_1)(x_{2j} - \bar{X}_2) \quad (79)$$

where ($\bar{X}_1$, $\bar{X}_2$) and ($\bar{\sigma}_1$, $\bar{\sigma}_2$) = sample estimators of the means and standard deviations of X_1 = FC and X_2 = FS.

The goal of this application is to present the PDF of the ratio $Z = X_1/X_2$ = FC/FS. This PDF allows a characterization of Z and, thus, of the origin of enteric bacteria in a probabilistic manner. The ratio PDF of the two correlated gamma variables was derived by Loáiciga and Leipnik (2005), as follows (with $\xi_1 = \xi_2 = 0$):

$$g_Z(z) = \sum_{n=0}^{\infty}\sum_{k=0}^{n}\sum_{j=0}^{n}(-1)^{n+k+j}\left(\frac{\beta}{b_1^{\alpha_1}b_2^{\alpha_2}}\right)^n\binom{-\gamma}{n}\binom{n}{k}\binom{n}{j} \times \left[\frac{b_1^{-(\gamma\alpha_1+k)}}{b_2^{\gamma\alpha_2+k}}\right] \cdot \frac{\Gamma(\lambda_{1,2}) \cdot z^{\lambda_1+k-n-1}}{\Gamma(\lambda_1-n+k) \cdot \Gamma(\lambda_2-n+j) \cdot z'^{\lambda_{1,2}}} \quad (80)$$

where $\lambda_j = \alpha_j(n+\gamma)$, $j = 1, 2$; $\lambda_{1,2} = \lambda_1 + \lambda_2 + k + j - 2n$

$$z' = \frac{z}{b_1} + \frac{1}{b_2} \quad (81)$$

and

$$\binom{\lambda}{k} = \frac{\lambda(\lambda-1)\ \cdots\ (\lambda-k+1)}{k!} \quad (82)$$

is the binomial coefficient for any real λ and nonnegative integer k. The estimated parameters in the Las Palmas Creek study were $\hat{b}_1 = \hat{b}_2 = 1.0$; $\hat{\alpha}_1 = 2.471$; $\hat{\alpha}_2 = 8.245$; $\hat{\beta} = 1.417$; $\hat{\gamma} = 0.35$; and $\hat{\rho} = 0.40$.

J. Hydrol. Eng., 2015, 20(5): 04014063

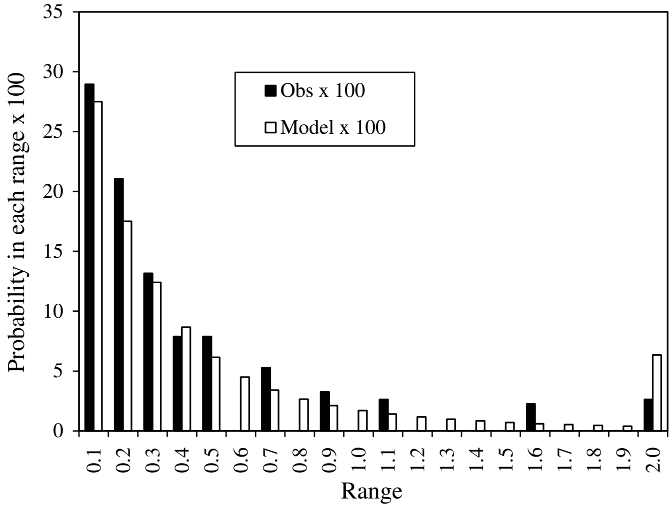

Fig. 7. Empirical (observed × 100) and calculated (model × 100) frequencies of the FC/FS ratio at Las Palmas Creek, Santa Barbara, California (1999–2000)

Fig. 7 shows the empirical (observed × 100) and calculated (model × 100) frequencies of the FC/FS ratio in Las Palmas Creek 2000–1999)). The empirical frequency in each range was calculated by dividing the number of observations within the range by the sample size (= 38), and then scaling it by 100 for ease of interpretation. The model frequency in each range was calculated by integrating Eq. (80) and then scaling it by 100 for ease of interpretation. In Fig. 7, the range labeled 0.1 equals the interval [0.0, 0.1], that labeled 0.2 equals [0.1, 0.2], and so on. The last range is ≥2.0. Fig. 7 shows an overall excellent agreement between the empirical and calculated probabilities. The observed and model-calculated probabilities $P(Z \leq 0.4)$ were 71.1 and 66.1%, respectively, which provides strong evidence of the predominance of equine fecal bacteria in Las Palmas Creek. A chi-square goodness-of-fit test was implement to ascertain the suitability of the PDF $g(z)$ [Eq. (80)] for the ratio $Z = \mathrm{FC/FS}$ to describe the FC/FS data. The chi-square statistic $\chi^2(0.05, 32) = 46.19$ is larger than the test statistic $D = 24.70$. Thus, the null hypothesis of a gamma ratio distribution was not rejected at a 5% significance level. The P value in this case was approximately 0.85, which demonstrates the robustness of the fit of the model probability to the empirical data.

Multivariate Lognormal PDF: Testing Goodness-of-Fit to Data and Synthetic Generation

There are situations in groundwater hydrology when multiple random variables are not independent. In this case, statistical dependence is commonly expressed in terms of spatial correlation. Conforming to physical plausibility, correlation functions are such that the degree of spatial association between random variables decreases with increasing distance separating them. The spatial correlation can be exploited in various ways, a common one being the spatial interpolation using geostatistics (e.g., Journel and Huijbregts 1978; Loáiciga et al. 2010). The application of geostatistics for spatial interpolation does not require the specification of a multivariate PDF for the random variables representing, for example, the value of rock porosity at various locations in a geologic formation. In its simplest form, called ordinary kriging, geostatistical interpolation assumes (1) a constant mean for the regionalized variable (i.e., a random variable defined over geographic space) and (2) a correlation function that expresses the degree of spatial association among the realizations of the random variable at various locations in a geologic formation. With those two assumptions, it is possible to obtain a best linear unbiased estimator (BLUE) of the regionalized variable at locations where measurements are not available using measurements of the regionalized variable at neighboring sites.

Other applications in groundwater hydrology require the specification of a multivariate PDF governing the realization of values of measurable aquifer properties. One example is the goodness-of-fit testing of a data set to a multivariate PDF. Another example is the synthetic generation of random fields (i.e., generation of values in physical space) of hydraulic conductivity and porosity in a geologic formation. The generated random fields may be used as part of Monte Carlo simulations implemented in conjunction with numerical models with the intention of assessing the probabilistic nature of groundwater flow and contaminant transport. The goodness-of-fit testing of the multivariate lognormal PDF and the synthetic generation of variables that have a multivariate lognormal PDF are elaborated upon in the following.

Testing Goodness-of-Fit of Multivariate Lognormal PDF

Assume that the log-transformed variable $Y = \ln(K)$ is normally distributed, where Y could represent an aquifer property, such as log hydraulic conductivity or log porosity. A measurement or realization of K or Y is denoted by k or y, respectively. It has been shown in a previous section that the logarithmic transformation frequently reduces groundwater data to near normally or near symmetrically distributed variables. Assume Y is measured at various locations in a geologic formation, giving rise to a vector of random variables $\underline{\mathbf{Y}} = (Y_1, Y_2, \ldots, Y_n)$ with vector of expected values $\underline{\boldsymbol{\mu}} = (\mu_1, \mu_2, \ldots, \mu_n)$ and covariances σ_{ij}^2 between any two variables Y_i and Y_j, $i, j = 1, 2, \ldots, n$. The elements μ_i of the vector of expected values need not be equal to one another. It is convenient to write the covariances σ_{ij}^2 compactly by defining the covariance matrix$\boldsymbol{\Sigma}$ of the vector of log variables $\underline{\mathbf{Y}}$, as follows:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n}^2 & \sigma_{2n}^2 & \cdots & \sigma_{nn}^2 \end{bmatrix} \tag{83}$$

The multivariate PDF of the log vector $\underline{\mathbf{Y}}$ is as follows (Anderson 1970):

$$f_{\underline{\mathbf{Y}}}(\underline{\mathbf{y}}) = \frac{1}{(2\pi)^{n/2} \cdot |\Sigma|^{1/2}} e^{-1/2 \cdot (\underline{\mathbf{y}} - \underline{\boldsymbol{\mu}})^T \cdot \Sigma^{-1} \cdot (\underline{\mathbf{y}} - \underline{\boldsymbol{\mu}})} \tag{84}$$

where $|\boldsymbol{\Sigma}|$ and $\boldsymbol{\Sigma}^{-1}$ = determinant and inverse of the covariance matrix $\boldsymbol{\Sigma}$, respectively; superscript T = transpose of a vector or matrix; and $\underline{\mathbf{y}}^T = (y_1, y_2, \ldots, y_n)$ = vector of transformed measurements $y_i = \ln(k_i)$. If a multivariate data set $\underline{\mathbf{y}}$ is available and need be tested for multivariate normality, implying the normality of the raw variables $K_i = e^{Y_i}$, one resorts to the fact that there exists a nonsingular matrix $\mathbf{C}$ such that $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T = \mathbf{I}$, where $\mathbf{I}$ denotes the $(n \times n)$ unit or identity matrix (Anderson 1970; Rao 1973; Golub and van Loan 1984). The matrix $\mathbf{C}$ can be obtained by means of the Cholesky decomposition of the inverse matrix $\boldsymbol{\Sigma}^{-1} = \mathbf{C}\mathbf{C}^T$. There are numerically efficient algorithms for calculating the Cholesky matrix $\mathbf{C}$ (e.g., Golub and van Loan 1984). The transformed

J. Hydrol. Eng., 2015, 20(5): 04014063

vector of random variables $\underline{\mathbf{Z}} = \mathbf{C}\underline{\mathbf{Y}}$ is multivariate normal with mean vector $\mathbf{C}\underline{\boldsymbol{\mu}}$ and covariance matrix $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T = \mathbf{I}$; that is, the components Z_i of the transformed random variable $\underline{\mathbf{Z}} = \mathbf{C}\underline{\mathbf{Y}}$ are independently distributed normal variables with unit variances and mean vector $\mathbf{C}\underline{\boldsymbol{\mu}}$.

The test for multivariate lognormality of the variables $K_i = e^{Y_i}$ proceeds as follows: (1) given a data set of measurements $k_1, k_2, \ldots, k_n$ of the raw variable K, log transform them to obtain $\underline{\mathbf{y}}$, and specify the covariance $\boldsymbol{\Sigma}$ of the vector of log variables $\underline{\mathbf{Y}}$; (2) determine the matrix $\mathbf{C}$ such that $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T = \mathbf{I}$, that is, $\boldsymbol{\Sigma}^{-1} = \mathbf{C}\mathbf{C}^T$, and obtain the transformed data set $\underline{\mathbf{z}} = \mathbf{C}\underline{\mathbf{y}}$; (3) subtract the mean vector $\mathbf{C}\underline{\boldsymbol{\mu}}$ from the vector $\underline{\mathbf{z}}$ to obtain a vector $\underline{\mathbf{x}}$ of independent and normally distributed variates with zero mean and unit variance; (4) apply the chi-square test described in a previous section of this paper to the transformed data vector $\underline{\mathbf{x}}^T = (x_1, x_2, \ldots, x_n)$; (4) if the x_is are found to be normally distributed, then the log-transformed variable $\underline{\mathbf{Y}}$ follows the multivariate normal PDF in Eq. (84), and the raw variable $K_1, K_2, \ldots, K_n$ are multivariate lognormally distributed.

Synthetic Generation of Random Fields with Multivariate Lognormal PDF

The synthetic generation of m random variables that are multivariate and lognormally distributed proceeds in a manner that is the inverse analog of that followed in the goodness-of-fit approach of the following subsection. The steps are as follows: (1) use a suitable random number generator algorithm to produce m independent and identically normally distributed variates $\underline{\mathbf{x}}^T = (x_1, x_2, \ldots, x_m)$, with zero means and unit variances; (2) multiply the vector $\underline{\mathbf{x}}^T = (x_1, x_2, \ldots, x_m)$ by the $(m \times m)$ matrix $\mathbf{D}$, where $\mathbf{D}$ satisfies the Cholesky decomposition $\boldsymbol{\Sigma} = \mathbf{D}\mathbf{D}^T$, to produce a zero-mean vector with specified covariance $\boldsymbol{\Sigma}$; (3) add the mean vector $\underline{\boldsymbol{\mu}}$ to the vector $\mathbf{D}\underline{\mathbf{z}}'$ to obtain a vector $\underline{\mathbf{y}}^T = (y_1, y_2, \ldots, y_m)$ of log-transformed variates drawn from a multivariate lognormal PDF with mean $\underline{\boldsymbol{\mu}}$ and covariance matrix $\boldsymbol{\Sigma}$; (4) obtain a random field of raw variates $k_i, i = 1, 2, \ldots, m$, by using the transformation $k_i = e^{y_i}$, which follow a multivariate lognormal PDF as desired. The covariance $\sigma^2_{K,ij}$ between any two variables K_i, K_j is related to the covariance σ^2_{ij} of the transformed variables Y_i, Y_j by the following formula:

$$\sigma^2_{K,ij} = \sigma^2_K \frac{e^{\sigma^2_{ij}} - 1}{e^{\sigma^2_Y} - 1} \tag{85}$$

where σ^2_K and σ^2_Y = variances of the raw variable K and the log-transformed variable Y. Steps 1 through 4 are repeated as many times as necessary following a Monte Carlo design to generate random fields with desired multivariate lognormal PDF. The method for synthetic generation of random fields described in this section is not limited to only one type of random variable, such as hydraulic conductivity. It also includes situations in which the vector of log-transformed random variable includes two or more types of random variable, for example, hydraulic conductivity and porosity. The case of two types of random variable considered jointly gives rise to a vector of log-transformed variables with two component subvectors, $\underline{\mathbf{y}}^T = (\underline{\mathbf{y}}_1^T, \underline{\mathbf{y}}_2^T)$, in which the subvectors $\underline{\mathbf{y}}_1$ and $\underline{\mathbf{y}}_2$ (each with subvectors of mean values $\underline{\boldsymbol{\mu}}_1$ and $\underline{\boldsymbol{\mu}}_2$, respectively) can be defined to represent subsets of hydraulic conductivity and porosity variables, respectively. The procedures described in this section remain unaltered, provided that the covariance matrix $\boldsymbol{\Sigma}$ be redefined to contain the covariances ($\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$) of each subvector and the cross covariances ($\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}$) between the two subvectors.

Current theory for multivariate PDFs other than the lognormal is not well developed and is bedeviled by mathematical complexity, complicating its application to probabilistic, multidimensional problems in groundwater hydrology and other fields of inquiry. This is an active area of research in applied statistics and probability (Genest and Favre 2007).

Conclusion

This paper has demonstrated the power of PDF methodologies to analyze groundwater phenomena. The examples in this paper have shown that probabilistic modeling of asymmetric aquifer properties, such as the hydraulic conductivity, can be accomplished with the log-gamma PDF with succinct flexibility. Other examples in this work illustrate the versatility of PDFs in elucidating the probabilistic nature of aquifer properties. The gamma PDF, for example, was shown to exhibit remarkable versatility to model groundwater processes, such as residence time and age. In the multivariate case, this paper presented algorithms to tackle goodness-of-fit testing and synthetic generation of random fields with desired probabilistic characteristics based on the lognormal PDF.

References

Anderson, T. W. (1970). *An introduction to multivariate statistical analysis*, John Wiley & Sons, New York.

Benjamin, J. R., and Cornell, C. A. (1970). *Probability, statistics, and decisions for civil engineers*, McGraw-Hill, New York.

Dagan, G. (1989). *Flow and transport in porous formations*, Springer, Berlin.

Freeze, R. A. (1975). "A stochastic conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media." *Water Resour. Res.*, 11(5), 725–741.

Genest, C., and Favre, A. C. (2007). "Everything you always wanted to know about copula modeling but were afraid to ask." *J. Hydrol. Eng.*, 10.1061/(ASCE)1084-0699(2007)12:4(347), 347–368.

Gilbert, R. O. (1987). *Statistical methods for environmental pollution monitoring*, Van Nostrand Reinhold, New York.

Golub, G. H., and van Loan, C. F. (1984). *Matrix computations*, Johns Hopkins University Press, Baltimore.

Gradshteyn, I. S., and Ryzhik, I. M. (1994). *Tables of integrals, series, and products*, 5th Ed., Academic, San Diego.

Journel, A., and Huijbregts, C. (1978). *Mining geostatistics*, Academic, New York.

Loáiciga, H. A. (2004). "Residence time, groundwater age, and solute output in steady-state groundwater systems." *Adv. Water Resour.*, 27(7), 681–688.

Loáiciga, H. A. (2005). "On the probability of droughts: The compound renewal model." *Water Resour. Res.*, 41(1), 1029–1037.

Loáiciga, H. A., and Leipnik, R. B. (2005). "Correlated gamma variables in the analysis of microbial densities in water." *Adv. Water Resour.*, 28(4), 329–335.

Loáiciga, H. A., Michaelsen, J., and Hudak, P. F. (1992). "Truncated distribution in hydrologic analysis." *Water Resour. Bull.*, 28(5), 853–863.

Loáiciga, H. A., Yeh, W., and Keyes, C. G. (2008a). "Standard guideline for fitting saturated hydraulic conductivity using probability functions." *ASCE/EWRI Standard 50-2008*, ASCE, Reston, VA.

Loáiciga, H. A., Yeh, W., and Keyes, C. G. (2008b). "Standard guideline for estimating the effective saturated hydraulic conductivity." *ASCE Standard 51-2008*, ASCE, Reston, VA.

Loáiciga, H. A., Yeh, W. G., and Keyes, C. G. (2010). "Saturated hydraulic conductivity." *ASCE/EWRI Standard 54-10*, ASCE, Reston VA.

Loáiciga, H. A., Yeh, W. W.-G., and Ortega-Guerrero, M. A. (2006). "Probability density functions in the analysis of hydraulic conductivity data." *J. Hydr. Eng.*, 10.1061/(ASCE)1084-0699(2006)11:5(442), 442–450.

Mathematica [Computer software]. Wolfram Research, Champaign, IL.

MATLAB [Computer software]. Mathworks, Inc., Natick, MA.

Priestly, M. B. (1989). *Spectral analysis and time series*, Academic, London.

Rao, C. R. (1973). *Linear statistical inference and its applications*, John Wiley & Sons, New York.

USGS. (1982). "Interagency advisory committee on water data." *Bulletin 17B*, Reston, VA.