

# Uncertainty propagation of arbitrary probability density functions applied to upscaling of transmissivities

A. Lourens<sup>1</sup> · F. C. van Geer<sup>1,2</sup>

Published online: 26 May 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** In many fields of study, and certainly in hydrogeology, uncertainty propagation is a recurring subject. Usually, parametrized probability density functions (PDFs) are used to represent data uncertainty, which limits their use to particular distributions. Often, this problem is solved by Monte Carlo simulation, with the disadvantage that one needs a large number of calculations to achieve reliable results. In this paper, a method is proposed based on a piecewise linear approximation of PDFs. The uncertainty propagation with these discretized PDFs is distribution independent. The method is applied to the upscaling of transmissivity data, and carried out in two steps: the vertical upscaling of conductivity values from borehole data to aquifer scale, and the spatial interpolation of the transmissivities. The results of this first step are complete PDFs of the transmissivities at borehole locations reflecting the uncertainties of the conductivities and the layer thicknesses. The second step results in a spatially distributed transmissivity field with a complete PDF at every grid cell. We argue that the proposed method is applicable to a wide range of uncertainty propagation problems.

**Keywords** Error propagation · Probability density function · Piecewise linear · Kriging · Upscaling transmissivity · Monte Carlo simulation

## 1 Introduction

Subsoil parameters are essential data for groundwater flow models. Often, these data originate from borehole descriptions in which thin layers (core scale) are distinguished based on lithological and sedimentological information. The thickness of these layers may vary from a few centimeters up to several meters, depending on the subsoil structure and the drilling method. Typically, the described layers are vertically aggregated to aquifer and aquitard classes at a scale which fits the groundwater model requirements. This scale will be referred to as point scale. The thickness of aquifers typically comes on the order of a few meters to 100 m or up. The core scale layers are normally populated with hydraulic conductivities derived from the literature or estimated in the laboratories. Next, point values of transmissivities and resistances are calculated by vertical integration of the conductivity values. Subsequently, these point values are interpolated to acquire a spatial distributed parameter at model scale. This scale has a lateral block size of about 100–1000 m.

An important issue in the upscaling procedures is the uncertainty of the model parameters. This uncertainty can be divided into two sources. Firstly, the available observations, at core scale, are uncertain, introducing uncertainty in the upscaling to point scale values. In this case, each observation is not treated as one known value but as a random variable (RV). Secondly, there is uncertainty about the spatial distribution of the parameter. At observed locations the point scale parameter values are the upscaled RVs. At unobserved locations, assumptions have to be made about the spatial structure. This spatial structure can be described by regionalized variables (ReV) (Journel and Huijbregts 1978, p. 26).

---

✉ A. Lourens  
a.lourens@uu.nl

<sup>1</sup> Department of Physical Geography, Faculty of Geosciences, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands

<sup>2</sup> TNO Geological Survey of the Netherlands, Princetonlaan 6, 3584 CB Utrecht, The Netherlands

In the Netherlands, a large database (REGIS) exists (Vernes et al. 2005; Vernes and van Doorn 2006), in which all distinguished layers from all boreholes are described at core scale by litho-stratigraphical units. Ranges of possible parameter values for hydraulic conductivity and porosity are assigned to these units. For REGIS, these ranges are obtained from laboratory tests and literature search. When a sufficient amount of data is available for a litho-stratigraphical unit, a probability distribution is derived for the parameter of this unit. In this article, these probability distributions are used as an uncertain value of the hydraulic conductivities at core scale.

As described extensively in the literature, the upscaling of hydraulic parameters is far from trivial and depends highly on: the support scale of the observations, the required model scale, the presence of anisotropy in the hydraulic conductivity, and boundary conditions of the flow problem at hand (Dagan 1986; Bierkens and Weerts 1994; Tran 1996; Fiori et al. 2011). Some clear overviews about these subjects are given by (Cushman et al. 2002; Nøtinger et al. 2005; Sanchez-Vila et al. 2006). Upscaling of hydraulic conductivities needs different approaches in one, two and three dimensions. With an increasing number of dimensions the complexity of the upscaling method increases even more. The upscaled one-dimensional conductivity is calculated by the harmonic mean. In isotropic media with a two-dimensional schematization, the upscaled conductivity can be obtained by the geometric mean (De Wit 1995; Hristopulos 2003). The three-dimensional upscaling is much more complicated and many upscaling methods are proposed in the literature (King 1989; De Wit 1995; Hristopulos and Christakos 1999; Hristopulos 2003; Boschan and Nøtinger 2012). Although in two dimensions the geometric mean yields a usable effective conductivity in isotropic media, in strong heterogeneous media the result may divert too much from realistic values. For the latter case, different solutions are proposed in the literature for strong heterogeneous or binary media (King 1989; Pancaldi et al. 2007; Boschan and Nøtinger 2012). Block kriging on log-conductivity values is equal to geometric upscaling of the two-dimensional situation. If the correlation length is larger than the block size, the within block variability will be low. In this case, the block kriging will yield accurate effective conductivity values. Subsequently, these block average values, the model scale, can be used as a starting point in the above mentioned upscaling methods. In the upscaling literature, this scale is often denoted as the fine scale grid.

In this article, the vertical one-dimensional upscaling is used at point scale, and the lateral two-dimensional upscaling is applied using kriging interpolation. In both cases, the complete parameter distributions of the observation data, as stored in the REGIS database, are used. Herewith,

the probability density functions (PDFs) at each grid cell are calculated. These parameter distributions are assumed to be representative at the model scale.

This article is not meant as a contribution to the problem of scale dependent hydraulic conductivities but as a description of a method to propagate uncertainties. Nevertheless, the proposed method can be used in conjunction with the above mentioned upscaling methods, thus propagating the observation uncertainty, but this is left for future work.

In this article, we will focus on the upscaling of hydraulic conductivities to transmissivities. To be useful to groundwater models, the point scale conductivities, which in fact are RVs, have to be upscaled to spatial distributed transmissivities. Commonly, only one value of this RV (e.g., mean) is used to perform this upscaling. Herewith, only information about the uncertainty of the interpolated mean is obtained, disregarding the uncertainty of the observations. Techniques like Monte Carlo simulation (MC) are often used to obtain results reflecting the data uncertainty. However, a disadvantage of MC is the dependence of the number of calculations, the sampling strategies used (Kyriakidis and Gaganis 2013), and the large number of calculations needed to obtain reasonable results.

The objective of our study is twofold: the derivation of a method to perform calculations with complete PDFs, and the application of this method in the upscaling and spatial interpolation of subsoil parameters. To take full advantage of the prior knowledge of the uncertainty of data, we present a method to propagate this uncertainty throughout all the calculations. Since the RVs are not described by their statistical moments but by numerically discretized PDFs, the proposed method is applicable regardless of the type of distributions used. Although the described technique can be used in conjunction with techniques that account for anisotropy, the proposed methods are applied to homogeneous examples.

The developed method is described in Sect. 2. In Sect. 3 the method is applied to the upscaling of real world borehole data to transmissivities at model scale, using kriging interpolation. The performance of the method is compared with an MC calculation. Section 4 contains the discussion and conclusions.

## 2 Methodology

Parameters obtained from observations are always subject to uncertainty. When this uncertainty contributes significantly to the result of calculations, it should be accounted for. A generally applicable method to propagate the uncertainty of RVs in a wide range of calculations is very attractive. This method should be independent of the

shape of PDFs and supports binary operations (+, −, \*, /) and elementary functions. In this section, we first develop a method to perform calculations with discretized PDFs. Thereafter, this method is implemented in the vertical up-scaling of core scale conductivities. Finally, the method is integrated in the kriging interpolation to obtain the PDF of the spatial distributed transmissivity data reflecting all sources of uncertainty.

### 2.1 Piecewise linear PDFs

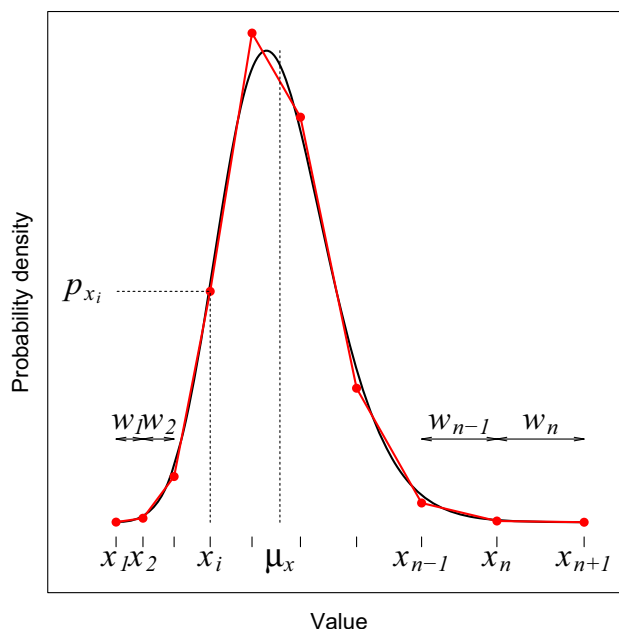
Commonly, parametrized PDFs are used to perform uncertainty calculations analytically. This means that for every possible combination of types of PDFs an analytical solution must be available. When many types of PDFs and operations need to be supported, numerous derivations have to be made. For long chains of calculations, this is highly inefficient. Moreover, the resulting PDFs should be known in closed analytical form, which can not always be achieved (Holmes and Buhr 2007; Silverman et al. 2004).

We aim at a method which is universally applicable and independent of the type of distribution used. To achieve this, a combination of a numerical and an analytical approach is used, that is, the PDFs are described numerically and the arithmetic is performed analytically. A common way to discretize PDFs is to describe them piecewise linear (Kaczynski et al. 2012; Vander Wielen and Vander Wielen in press). Herewith, any probability distribution which can be approximated by a piecewise linear PDF can be used. A drawback of this method is the introduction of inaccuracies by linearization, and the need for truncation of distributions with a one or two sided infinite domain. However, this drawback can largely be overcome by the choice of a sufficient number of discretization points, and discretize large tails when needed. In Fig. 1 an example of a piecewise linear PDF is given. Between two discretization points, the PDF is described by a linear function. This interval is referred to as a bin (Izenman 1991). A calculation method with discretized PDFs is described before in Jaroszewicz and Korzeń (2012) and Korzeń and Jaroszewicz (2014). However, their approach is different from ours which makes both methods applicable in different types of problems. A comparison of both methods is described in Sect. 3.2.

### 2.2 Calculations with PDFs

#### 2.2.1 Binary operations

When the PDF of an RV can be described analytically, the result of a binary operation (+, −, \*, /) can be described analytically as well. Let  $Z$  be the RV formed by the joint distribution of two independent RVs  $X$  and  $Y$ . The general



**Fig. 1** Example of a piecewise linear discretization of a PDF. The discretized PDF (red) is a  $n$  bins discretization of the real PDF (black). At the red points, the cumulative probabilities are equal to those of the real PDF. In this picture is:  $x_i$  the value of the PDF,  $p_{x_i}$  the probability density at value  $x_i$ ,  $w_i$  the width of bin  $i$ , and  $\mu_x$  the average value of the PDF

formulation of the cumulative distribution function (CDF) of  $Z$  can be described as (Papoulis 1991, p. 132ff)

$$F_z(z) = \int \int f_x(x)f_y(y) dx dy, \tag{1}$$

where  $f_x(\cdot)$  and  $f_y(\cdot)$  are the PDFs of  $X$  and  $Y$ , respectively. In this equation, the integration boundaries depend on the value of  $z$  and the binary operation to be calculated. Let  $Z$  be the sum of  $X$  and  $Y$ , then the probability  $\Pr\{Z < z\}$  can be written as

$$F_z(z) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{z-y} f_x(x)f_y(y) dx dy. \tag{2}$$

The integration boundaries for subtraction, multiplication and division are given in Appendix. Unfortunately, for piecewise linear PDFs such analytical formulation can not be solved as one integral. However, the PDF of each bin of the RVs can be described analytically. So for each bin of the marginal distributions, the linear functions  $f_{x,i}(\cdot)$  and  $f_{y,j}(\cdot)$  can be defined as

$$f_{x,i}(x) = p_{x_i} + r_{x_i}(x - x_i) \quad \text{for } x \in \langle x_i, x_{i+1} \rangle \tag{3}$$

$$f_{y,j}(y) = p_{y_j} + r_{y_j}(y - y_j) \quad \text{for } y \in \langle y_j, y_{j+1} \rangle, \tag{4}$$

where  $p_{x_i}$  and  $p_{y_j}$  are the probability densities at the values  $x_i$  and  $y_j$ , respectively. The slopes of these functions are defined as  $r_{x_i} = (p_{x_{i+1}} - p_{x_i}) / (x_{i+1} - x_i)$  and

$r_{y_j} = (p_{y_{j+1}} - p_{y_j}) / (y_{j+1} - y_j)$ . With these functions, we can define the piecewise analytical solution of the CDF of  $Z$  by integration of the probability density of the area inside the joint bin below the line  $z = x + y$ . The integration area is split up into four sub-areas as can be seen in Fig. 2. Because  $X$  and  $Y$  are independent, the probability of the rectangle sub-area  $a$  can be easily defined by the product of its marginal probabilities

$$F_{z,ij,a}(z) = \Pr\{x_i < X \leq x_{i+1}\} \Pr\{y_j < Y \leq y_{j+1}\}. \tag{5}$$

Equivalently, the probabilities of area  $b$  and  $c$  are expressed. The equation of the probability of sub-area  $d$  of joint bin  $(i, j)$  can be written as

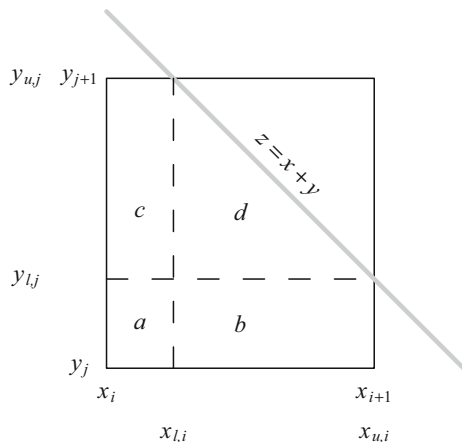
$$F_{z,ij,d}(z) = \int_{y=y_{l,j}}^{y_{u,j}} \int_{x=x_{l,i}}^{z-y} f_{x,i}(x)f_{y,j}(y) \, dx \, dy. \tag{6}$$

The integration boundaries  $y_{l,j}$ ,  $y_{u,j}$ ,  $x_{l,i}$  and  $z - y$  are portrayed in Fig. 2. When  $z > y_{j+1} + x_{i+1}$  or  $z < y_j + x_i$ , the line  $z = x + y$  does not intersect the joint bin  $(i, j)$ . Therefore,  $z_{ij}$  is defined to replace  $z$  in the calculations of joint bin  $(i, j)$ . The value of  $z_{ij}$  is calculated using  $z_{ij} = \min(\max(z, x_i + y_j), x_{i+1} + y_{j+1})$ . Integration of Eq. (6) yields (see Appendix for its derivation)

$$F_{z,ij,d}(z_{ij}) = \frac{1}{2} p_{x_{l,i}} p_{y_{u,j}} (y_{u,j} - y_{l,j})^2 - \frac{1}{3} p_{x_{l,i}} r_{y_j} (y_{u,j} - y_{l,j})^3 + \frac{1}{6} r_{x_i} p_{y_{u,j}} (y_{u,j} - y_{l,j})^3 - \frac{1}{8} r_{x_i} r_{y_j} (y_{u,j} - y_{l,j})^4. \tag{7}$$

To obtain the cumulative probability for a particular value of  $Z$ , a summation of the probabilities of all joint bins is performed

$$F_z(z) = \sum_{j=1}^{n_y} \sum_{i=1}^{n_x} \sum_{A=a,b,c,d} F_{z,ij,A}(z), \tag{8}$$



**Fig. 2** Integration boundaries of the piecewise analytical CDF. Shown is the dependence of the integration boundaries on the position of the line  $z$  in the box of the joint bin  $(i, j)$

where  $n_x$  and  $n_y$  are the numbers of bins of  $X$  and  $Y$ , respectively.

From Eq. (7) the PDF of  $Z$  can be derived by taking the first derivative with respect to  $z$ . The parameters depending on  $z$  have to be rewritten as a function of  $z$  as  $x_{u,i} = z - y_{l,j}$ ,  $y_{u,j} = z - x_{l,i}$  and  $p_{y_{u,j}} = f_{y,j}(z - x_{l,i})$ . Herewith the derivative yields

$$f_{z,ij,d}(z) = p_{x_{l,i}} p_{y_{u,j}} (y_{u,j} - y_{l,j}) - \frac{1}{2} p_{x_{l,i}} r_{y_j} (y_{u,j} - y_{l,j})^2 + \frac{1}{2} r_{x_i} p_{y_{u,j}} (y_{u,j} - y_{l,j})^2 - \frac{1}{3} r_{x_i} r_{y_j} (y_{u,j} - y_{l,j})^3. \tag{9}$$

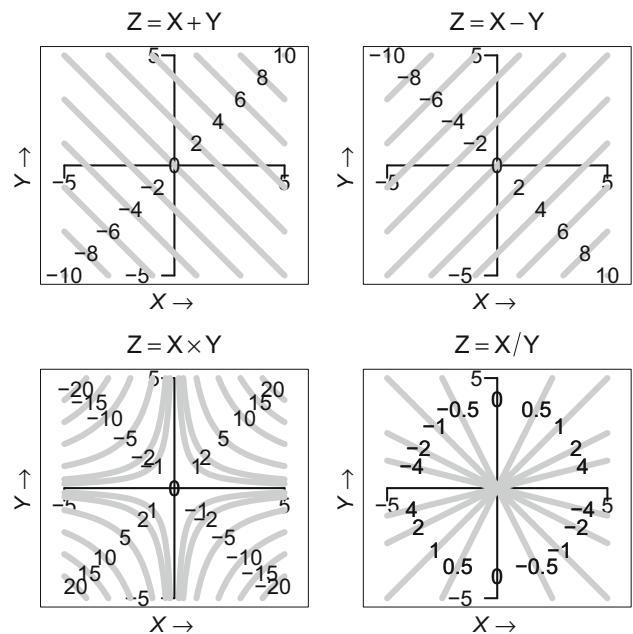
The PDF of all bins writes

$$f_z(z) = \sum_{j=1}^{n_y} \sum_{i=1}^{n_x} f_{z,ij,d}(z). \tag{10}$$

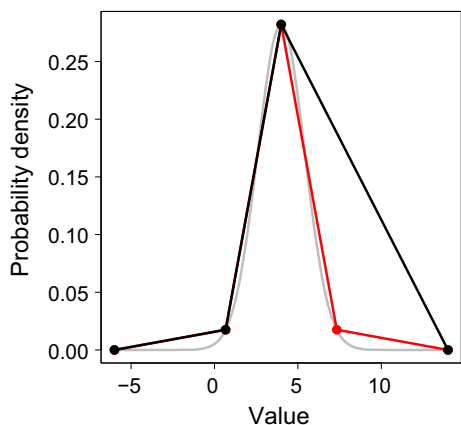
Analogous to the summation, the integration can also be performed for subtraction, multiplication and division. An illustration of the equi  $Z$ -lines of four binary operations is given in Fig. 3. The derivations of the four binary operations can be found in Appendix.

2.2.2 Discretizing unknown variable  $Z$

Performing a binary operation like Eq. (8), raises the need for a proper discretization of the unknown RV  $Z$ . Due to linearization, the integral of this PDF will usually not describe the CDF exactly. This probability error for each bin



**Fig. 3** Example of the graphical representation of CDFs of four binary operations between two independent RVs. The gray lines are the upper boundaries of the integration area of the cumulative probability for a certain value of  $Z$



**Fig. 4** Refining the PDF by adding a Z-value. The gray line is the true solution, the black line shows the 4-point PDF, and the red line shows the effect of adding the 5th defined Z-value

has to be as small as possible without increasing the number of bins too much.

An algorithm is proposed which starts with at least three predefined Z-values (e.g.,  $z_{min}$ ,  $z_{max}$ , and  $z_{mean}$ ). Subsequently, new Z-values are added during calculation. For every Z-value, the cumulative probability (Eq. 8) and the probability density (Eq. 10) are calculated. The probability of each bin can now be calculated in two ways: the difference of the cumulative probability at each edge of the bin, and the integration of the linearized probability density of the bin. Herein, the first probability is the exact solution of the calculations and the second method yields an approximate value. The difference between these probabilities is the error caused by the linearization of the PDF. The bin with the largest absolute probability error will be split up at its center of mass of the probability of the linearized function. This algorithm runs until all probability errors are smaller than a certain threshold, or a predefined maximum number of bins is reached. In Fig. 4, an example of one iteration of the summation of two independent RVs [both  $\mathcal{N}(2, 1)$ ] is illustrated.

### 2.3 Construction of probability fields of transmissivity

This section describes a two step approach of the construction of probability fields of transmissivity. Firstly, the borehole data is upscaled to aquifer scale at point locations. Secondly, these upscaled values are horizontally interpolated using kriging interpolation. Both steps make use of the calculation methods as described in Sect. 2.2.

#### 2.3.1 Vertical upscaling

The transmissivity of a layer at core scale is calculated from borehole data by multiplying the layer thickness by the conductivity

$$T_l = K_l(L_l - L_{l+1}), \tag{11}$$

where index  $l$  denotes the layer number,  $T_l$  is the transmissivity and  $K_l$  the hydraulic conductivity of layer  $l$ , and  $L_l$  the height of the top of layer  $l$ , measured relative to for example Amsterdam Ordnance Datum. The layer numbers increase downwards, so the bottom of layer  $l$  coincides with the top of layer  $l + 1$  (i.e.,  $L_{l+1}$ ). Subsequently, the upscaled aquifer transmissivity at point scale is defined by

$$T = \sum_{l=1}^n T_l, \tag{12}$$

where  $n$  is the number of layers, at core scale, which are combined to one aquifer.

Equation (12) only holds for horizontal flow within an aquifer. As denoted in Sect. 1, we assume the conductivity parameter values appropriate for the scale used after up-scaling. Subjects like anisotropy are beyond the scope of this article.

Both, the layer thickness and the hydraulic conductivity are subject to uncertainty. When transmissivities are up-scaled from consecutive layers, these individual transmissivities are correlated because of the uncertainty of the boundaries between these layers. In order to perform the summation of transmissivities correctly, we need to know the correlation between the layers. The covariance of the transmissivities of two consecutive layers can be calculated as

$$\begin{aligned} \text{cov}(T_l, T_{l+1}) &= \text{cov}(K_l(L_l - L_{l+1}), K_{l+1}(L_{l+1} - L_{l+2})) \\ &= + \text{cov}(K_l L_l, K_{l+1} L_{l+1}) \\ &\quad - \text{cov}(K_l L_l, K_{l+1} L_{l+2}) \\ &\quad - \text{cov}(K_l L_{l+1}, K_{l+1} L_{l+1}) \\ &\quad + \text{cov}(K_l L_{l+1}, K_{l+1} L_{l+2}). \end{aligned} \tag{13}$$

When we assume all variables  $K$  and  $L$  mutually independent, only the third covariance ( $-\text{cov}(K_l L_{l+1}, K_{l+1} L_{l+1})$ ) is not equal to 0.

According to Bohrnstedt and Goldberger (1969) this covariance can be written as

$$\text{cov}(K_l L_{l+1}, K_{l+1} L_{l+1}) = E[K_l]E[K_{l+1}]\text{var}(L_{l+1}). \tag{14}$$

The correlation coefficient can now be written as

$$\rho_{(T_l, T_{l+1})} = - \frac{E[K_l]E[K_{l+1}]\text{var}(L_{l+1})}{\sqrt{\text{var}(T_l)\text{var}(T_{l+1})}}. \tag{15}$$

If the value of  $\rho_{(T_l, T_{l+1})}$  can not be neglected, we have to account for correlations in Eq. (12). When the correlations differ significantly from 0, also in the calculations of Sect. 2.2 the correlations should be taken into account. The correlations as calculated from the observation data are found in Sec. 3.1.



2.3.2 Horizontal upscaling: semivariogram

Sample semivariograms are usually derived from observations which are assumed to be scalar values. Since our point scale observations are RVs, this will cause a different sample semivariogram and the way it is obtained. Our aim is to find a semivariogram based on uncertain observations and to find the PDF of the interpolation. Although the observations are of a different nature than usual (RVs instead of scalars), we assume the intrinsic hypothesis (Journel and Huijbregts 1978, p. 11) still holds.

The definition of the semivariogram is (Goovaerts 1997, p. 96)

$$\gamma(h) = \frac{1}{2} E[(Z(u) - Z(u + h))^2], \tag{16}$$

where  $Z(u)$  is the sample value at location  $u$ , and  $h$  is the spacing between two observation locations.

Equation (16) can be rewritten as

$$\gamma(h) = E \left[ \left( \frac{1}{\sqrt{2}} (Z(u) - Z(u + h)) \right)^2 \right] = E[\Delta_Z(h)^2]. \tag{17}$$

From the intrinsic hypothesis it follows that  $\Delta_Z(h)$  has a symmetrical distribution function with zero mean. So  $\Delta_Z(h)$  is the RV with a probability distribution describing the difference between two observations at lag  $h$ , scaled with factor  $1/\sqrt{2}$ . Equation (17) can now be written as  $\gamma(h) = \text{var}(\Delta_Z(h))$ . The PDF of  $\Delta_Z(h)$  is derived from the observations  $Z(u)$ , which can be either scalar values or RVs. The effect of the observations being RVs, instead of scalars, is shown in Fig. 5. As expected, a nugget effect arises from the use of RVs as observations.

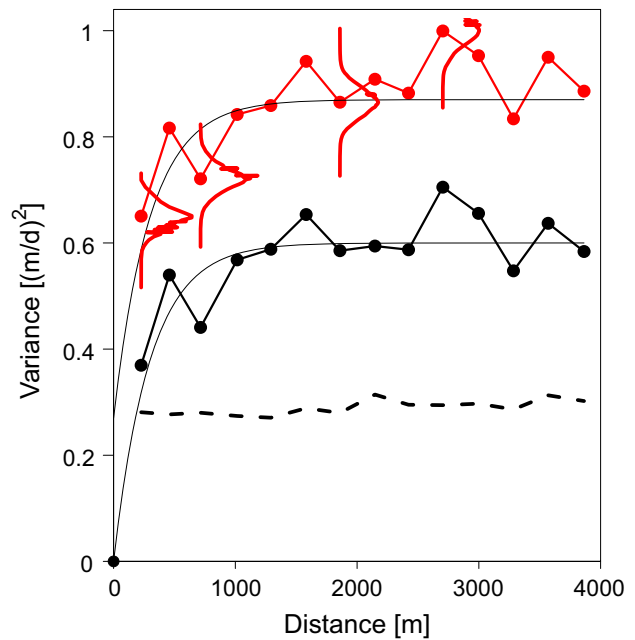
In general,  $\Delta_Z(h)$  is assumed to be normal distributed, which is not always the case (Journel and Huijbregts 1978, p. 50). In the procedure described here, the shape of the distribution is derived from the observations. The assumption we make is that the shape of  $\Delta_Z(h)$  is independent of  $h$ , only the variances differ.

Since we want to use the distribution of  $\Delta_Z(h)$  in the kriging interpolation, we have to relate it to the covariance function. For a stationary random function, the covariance function and the correlogram are directly related to the semivariogram (Journel and Huijbregts 1978, p. 32). The covariance function can be written as

$$C(h) = C(0) - \gamma(h), \tag{18}$$

where  $C(h)$  is the covariance at lag  $h$ , with  $C(0) = \gamma(h \rightarrow \infty) = \text{var}(\Delta_Z(h \rightarrow \infty))$ . For convenience we define  $\Delta_Z = \Delta_Z(h \rightarrow \infty)$ . The correlogram is defined as

$$\rho(h) = \frac{C(h)}{C(0)}, \tag{19}$$



**Fig. 5** Example of a sample semivariogram. The black lines show the result when the observations are treated as scalar values. The red line is the result of observations treated as RVs. The dashed line shows the difference between the red and the black line, which is the expected nugget effect. The smooth black lines are the fitted variogram models. At four points the PDF of  $\Delta_Z(h)$  is drawn from which the variance is derived. The semivariogram is derived from the log-values of the observations

where  $\rho(h)$  is the correlation coefficient at lag  $h$ . From Eq. (19) we can write

$$C(h) = \rho(h)C(0) = \rho(h)\text{var}(\Delta_Z). \tag{20}$$

From this relation we derive that the covariance  $C(h)$  can be calculated as

$$C(h) = \text{var}(\sqrt{\rho(h)}\Delta_Z). \tag{21}$$

The covariance functions must be positive definite (Journel and Huijbregts 1978, p. 34), so  $\rho(h) \geq 0$ .

2.3.3 Horizontal upscaling: interpolation

The vertical upscaled borehole data, as described in Sect. 2.3.1, are used in spatial interpolation. Since these data are subject to uncertainty, an interpolation technique which can handle this kind of data must be chosen. We applied ordinary kriging to perform this interpolation. In this section we describe the way we incorporate the uncertainty of the observations, including the shape of the distributions, in the kriging variance.

Ordinary kriging is based on two equations (Isaaks and Srivastava 1989, p. 280 ff). The interpolation of the observation values is described by

$$\hat{Z}(u_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(u_{\alpha}), \tag{22}$$

where  $\hat{Z}(u_0)$  is the kriging estimate at the unsampled location  $u_0$ ,  $\lambda_{\alpha}$  the weight factor of  $Z(u_{\alpha})$ , and  $n$  the number of sample locations used in the estimate. The variance of  $\hat{Z}(u_0)$  is described by

$$\text{var}(\hat{Z}(u_0)) = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(h_{\alpha\beta}), \tag{23}$$

where  $C(\cdot)$  is the covariance function as discussed in Sect. 2.3.2, and  $h_{\alpha\beta}$  is the distance between location  $u_{\alpha}$  and  $u_{\beta}$ .

In general,  $Z(u_{\alpha})$  represents a scalar value at each location, which yields a scalar value  $\hat{Z}(u_0)$  as well. The variance of  $\hat{Z}(u_0)$  is calculated by Eq. (23), and if probabilities are calculated  $\hat{Z}(u_0)$  is assumed to have a normal distribution. Together, these two results describe the PDF of the interpolation.

Since we have PDFs available at all sample locations we use these PDFs in Eq. (22). This yields an RV for  $\hat{Z}(u_0)$  which honors the uncertainty, including the distribution, of the sample data. Additionally, we want to use the distribution of  $\Delta_Z$  in the uncertainty of the interpolation. In Sect. 2.3.2 we presented a method to obtain the PDF of  $C(\cdot)$ , described in Eq. (21). Inserting Eq. (21) in Eq. (23) yields

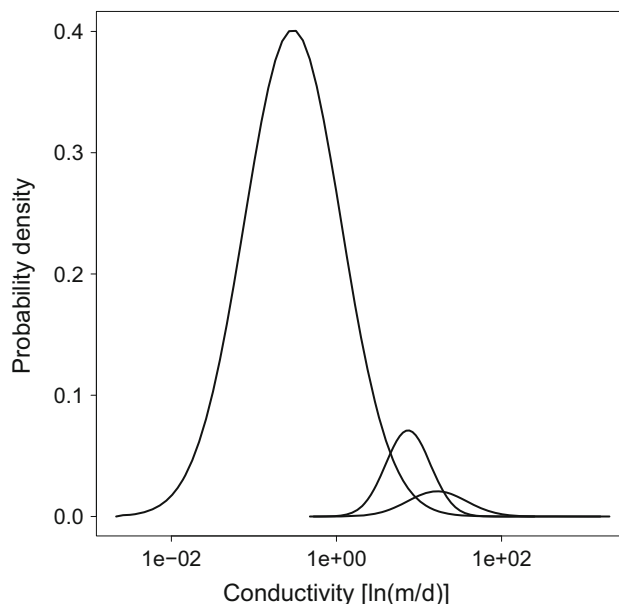
$$\begin{aligned} \text{var}(\hat{Z}(u_0)) &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \text{var}\left(\sqrt{\rho(h_{\alpha\beta})} \Delta_Z\right) \\ &= \text{var}\left(\sum_{\alpha=1}^n \sum_{\beta=1}^n \sqrt{\lambda_{\alpha} \lambda_{\beta} \rho(h_{\alpha\beta})} \Delta_Z\right). \end{aligned} \tag{24}$$

Herein,  $\sum_{\alpha=1}^n \sum_{\beta=1}^n \sqrt{\lambda_{\alpha} \lambda_{\beta} \rho(h_{\alpha\beta})} \Delta_Z$  is the RV describing the uncertainty of the interpolation with a distribution based on  $\Delta_Z$ . When added to  $\hat{Z}(u_0)$ , the resulting RV describes the probability distribution of the interpolation.

### 3 Results

#### 3.1 Application to real world data

This section shows an example of upscaling and interpolation of borehole data, using the proposed methods. From the REGIS database of the Geological Survey of the Netherlands, we used data from the Kiezeloöliet Formation from an area in the south of the Netherlands. The dataset contains about 200 boreholes with data from the second aquifer (Vernes et al. 2005). This aquifer consists mainly of sandy deposits which are divided into three classes with significant different conductivity distributions. Figure 6 shows the PDFs of these distributions.



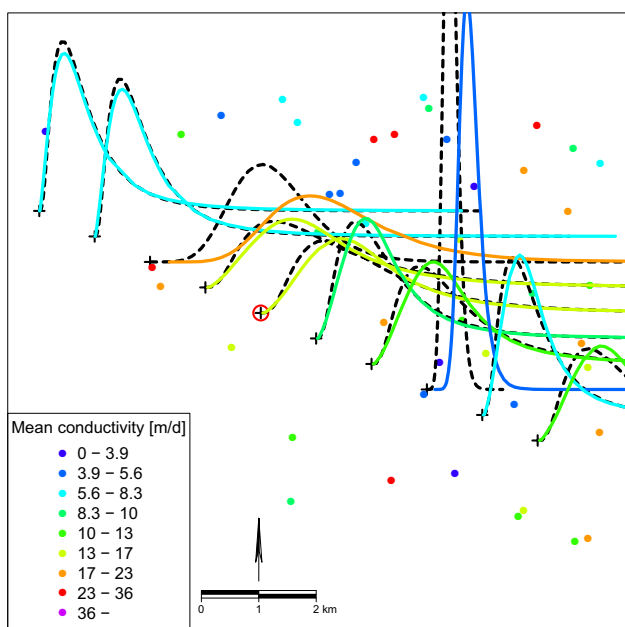
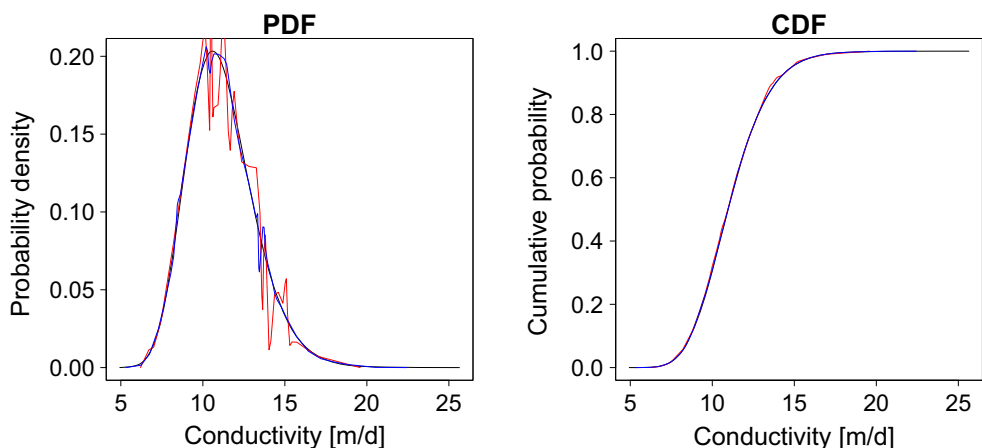
**Fig. 6** PDFs of three classes of sand as used with the upscaling of the borehole data. From left to right: fine sand, medium fine sand, and coarse sand. The horizontal axis is logarithmic which explains the apparent difference in integrated area

The vertical upscaling of the borehole data is performed as described in Sect. 2.3.1. The number of core scale layers at one borehole varied between 1 and 40 layers with an average of about nine layers. During upscaling, we calculated 1645 correlations between consecutive layers using Eq. (15). It appears that almost all (1638) correlations between the transmissivities of consecutive layers have a value between  $-0.05$  and  $0$ , the rest has values between  $-0.085$  and  $-0.05$ . Because of these low correlations, we performed the upscaling without taking the correlations into account.

The variogram model, as shown in Fig. 5, is derived from the upscaled borehole data. The PDFs of the conductivities are log-transformed before kriging (Journel and Huijbregts 1978, p. 570) and the interpolated PDFs are back transformed afterwards. In this example we used an exponential variogram with range 300 m, sill  $0.6 \ln(\text{m/d})^2$ , and nugget  $0.27 \ln(\text{m/d})^2$ .

The performance of the PDF calculation used at interpolation of uncertain data, by using Eq. (22), is compared to a Monte Carlo simulation (MC). For this purpose, we draw a large number of random realizations ( $n_{MC}$ ) of the PDFs of the observations. These random realizations are treated as observations in kriging. Since we assume that the semivariogram does not alter for each realization, the same sets of weight factors,  $\lambda_{\alpha}$ , are used for both, the PDF and the MC calculations. Subsequently, the results of MC are transformed to a CDF and PDF, as displayed in Fig. 7. It can be seen that the CDFs of both MC runs ( $n_{MC} = 1,000$

**Fig. 7** Result of PDF calculations compared to MC. Black PDF calculations, red MC with  $n_{MC} = 1000$ , blue MC with  $n_{MC} = 20,000$ . The black and blue line coincide



**Fig. 8** Map with result of the kriging interpolation of conductivities. The black dashed lines show the result of a standard ordinary kriging, the colored solid lines are the results of the new proposed method. The dots are the observation locations where the color indicates the mean value. The plus signs are the kriging locations

and  $n_{MC} = 20,000$ ) fit quite well with the CDF of the PDF calculations. However, the PDFs of the MC are less smooth than the the PDF of the PDF calculations. The interpolated location in this example is the same location as in Fig. 8 denoted with a red circle.

Some results of the kriging interpolation are shown in Fig. 8. The results in this example are obtained by point kriging. At every kriging location, two PDFs are drawn. The dashed line PDFs are the results of kriging applied on scalar observations, and the solid lines are the kriging results with observations as RVs as described before.

### 3.2 Comparison of calculation methods

In this section, the main differences between the calculation method of Jaroszewicz and Korzeń (2012) and the piecewise linear method as described in this article are discussed.

Both methods divide the PDFs in intervals where the probability densities are approximated by one or more polynomial functions. The piecewise linear method uses only one linear function, where the method of Jaroszewicz and Korzeń uses also higher order polynomials, implemented as Chebyshev polynomials. The latter method has the ability to describe the curve of the PDF much more accurate than the linear functions. Another difference between the two methods is the possibility to describe functions with an infinite domain. The piecewise linear method has to truncate the infinite tails at some finite value, the method of Jaroszewicz and Korzeń is able to support infinite domains by use of exponential tails.

As an example, the summation of ten standard normal distributed RVs is performed. The analytical mean and variance are 0 and 10, respectively. The result of the method of Jaroszewicz and Korzeń is about  $1.2178e-15$  and 10 (with 14 trailing zeros), and the result of the piecewise linear method is  $5.879e-5$  and 10.1049. The piecewise linear PDFs are discretized with 50 bins and truncated at five times the standard deviation.

The higher accuracy is acquired at the cost of calculation time. The calculation of the transmissivity, as described by Eqs. (11) and (12), is used to compare the performance of both methods. In Table 1 the computation time is shown for the addition of one, two and three layers

The calculation time of the method of Jaroszewicz and Korzeń is much higher than the calculation time of the piecewise linear method. Furthermore, the calculation time of the method of Jaroszewicz and Korzeń is not proportional to the number of operations but increases much more. Compared to the vertical upscaling at point scale and



**Table 1** Comparison of the performance of the method of Jaroszewicz and Korzeń to the piecewise linear method

Problem	Jaroszewicz and Korzeń [s]	Piecewise linear [s]
$D_1 * K_1$	1.35	0.00077
$D_1 * K_1 + D_2 * K_2$	24.1	0.0021
$D_1 * K_1 + D_2 * K_2 + D_3 * K_3$	834	0.0033

subsequently the horizontal interpolation in the real world example in this article, this is a very small example.

### 4 Discussion and conclusions

We developed a generic method to propagate the uncertainty of data through calculations and applied it to the upscaling of hydraulic conductivity data. The uncertain data used are represented by piecewise linear PDFs, which can be of any form. A similar calculation method, with a different implementation, has been described before by Jaroszewicz and Korzeń (2012). However, the computation time of their method is so high that it is not easily applicable to the calculations described in this article.

Figure 8 shows that the magnitude of the effect of the proposed method differs between kriging locations. As may be expected, kriging locations close to observations show the largest effects on the interpolated PDFs. The results presented show a good performance of the developed PDF calculations. The implementation in upscaling of borehole data, using kriging interpolation, yields interpolated subsurface parameter data with complete PDFs instead of only the uncertainty of the mean values. Although these PDFs are a common feature of kriging, the propagation of the uncertainty of the basic data in this way throughout the calculations is new. Herewith, any distribution which can be approximated by a piecewise linear PDF can be dealt with. Compared to Monte Carlo simulation (MC), the PDF calculations yield a smoother PDF of the result. The smoothness of the result does not rely on a random number generator or the number of simulations performed.

We performed kriging on the log-values of the PDFs of the observations. This transformation relies on true log-normal distributed values when the RVs are parametrized. When the data is not exactly log-normal distributed, the back transformation may cause a bias in the mean values. Back transformation of the PDFs does not yield a bias in mean value or variance.

Compared to calculations using parametrized PDFs or other analytical solutions, our method takes more computation time. However, we did not perform a benchmark because of the research state of the software. Nevertheless, PDF calculations can be of great value in uncertainty propagation problems where no analytical solutions are applicable. Availability of this method reduces the need for MC.

Compared to the analytical PDFs, the usage of piecewise linear PDFs implies loss of accuracy in the calculated results. So care must be taken when choosing the discretization of a PDF.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### Appendix: PDF arithmetic

#### Probability distributions of binary operations

This appendix describes the derivation of four binary operations (+, −, \*, /) performed on piecewise linear PDFs.

Let  $X$  and  $Y$  be independent RVs and  $Z$  be the result of a binary operation on  $X$  and  $Y$ . The general formulation of the cumulative distribution function (CDF) of  $Z$  can be written as (Papoulis 1991, p. 132 ff)

$$F_z(z) = \int \int f_x(x)f_y(y) dx dy, \tag{25}$$

where  $f_x(\cdot)$  and  $f_y(\cdot)$  are the PDFs of  $X$  and  $Y$ , respectively. These PDFs are linear functions at each bin of the piecewise linear PDFs and are, for bin  $i$  and bin  $j$ , defined as

$$f_{x,i}(x) = p_{x_i} + r_{x_i}(x - x_i) \tag{26}$$

$$f_{y,j}(y) = p_{y_j} + r_{y_j}(y - y_j), \tag{27}$$

where  $p_{x_i}$  and  $p_{y_j}$  are the probability densities at the values  $x_i$  and  $y_j$ , respectively. The slopes of these functions are defined as  $r_{x_i} = (p_{x_{i+1}} - p_{x_i}) / (x_{i+1} - x_i)$  and  $r_{y_j} = (p_{y_{j+1}} - p_{y_j}) / (y_{j+1} - y_j)$ . For convenience, the next variables are defined

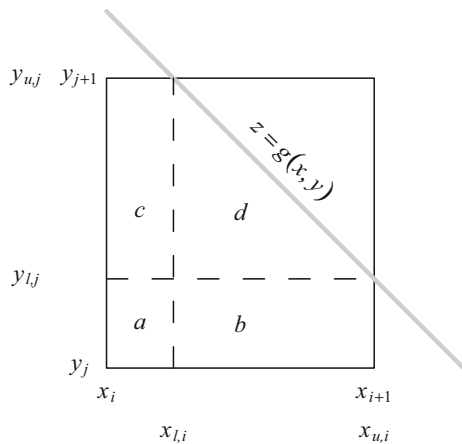
$$p_{0,x_i} = f_{x,i}(0) = p_{x_i} - r_{x_i}x_i \tag{28}$$

$$p_{0,y_j} = f_{y,j}(0) = p_{y_j} - r_{y_j}y_j.$$

Since the functions  $f_{x,i}(\cdot)$  and  $f_{y,j}(\cdot)$  are only continuously within a bin, Eq. (25) has to be defined for each joint bin as

$$F_{z,ij}(z) = \int \int f_{x,i}(x)f_{y,j}(y) dx dy, \tag{29}$$

Furthermore, the integration area of a joint bin is split up into four sub-areas, shown in Fig. 9.



**Fig. 9** Integration boundaries of the piecewise analytical CDF. Shown is the dependence of the integration boundaries on the position of the line  $z$  in the box of the joint bin  $(i, j)$ . The function  $g(\cdot)$  denotes any binary operation

As can be seen, the integration boundaries  $x_{l,i}$ ,  $x_{u,i}$ ,  $y_{l,j}$  and  $y_{u,j}$  depend on the intersection of the line  $z = g(x, y)$  with the lines  $x = x_i$ ,  $x = x_{i+1}$ ,  $y = y_j$  and  $y = y_{j+1}$ . The function  $g(x, y)$  represents a binary operation.

The line  $z = g(x, y)$  for a particular value of  $z$  will not intersect all joint bins. Therefore  $z_{ij}$  is defined as  $z$  but limited to the minimum and maximum value of  $z$  for which  $g(x, y)$  intersects joint bin  $(i, j)$ .

The probabilities of the rectangle sub-areas  $a$ ,  $b$  and  $c$  can be easily defined by the product of their marginal probabilities

$$\begin{aligned} F_{z,ij,a}(z) &= \Pr\{x_i < X \leq x_{l,i}\} \Pr\{y_j < Y \leq y_{l,j}\} \\ F_{z,ij,b}(z) &= \Pr\{x_{l,i} < X \leq x_{u,i}\} \Pr\{y_j < Y \leq y_{l,j}\} \\ F_{z,ij,c}(z) &= \Pr\{x_i < X \leq x_{l,i}\} \Pr\{y_{l,j} < Y \leq y_{u,j}\}. \end{aligned} \tag{30}$$

These three functions hold for the example in Fig. 9, the boundaries may be different for other operations. The function for sub-area  $d$  ( $F_{z,ij,d}(z)$ ) is described by Eq. (29) and is derived for each binary operation separately in the next sections.

The probability of  $Z < z$  for bin  $(i, j)$  for a given value of  $z$  is defined as

$$F_{z,ij}(z) = F_{z,ij,a}(z) + F_{z,ij,b}(z) + F_{z,ij,c}(z) + F_{z,ij,d}(z). \tag{31}$$

To obtain the cumulative probability for a particular value of  $Z$ , a summation of the probabilities of all joint bins has to be performed

$$F_z(z) = \sum_{j=1}^{n_y} \sum_{i=1}^{n_x} F_{z,ij}(z), \tag{32}$$

where  $n_x$  and  $n_y$  are the numbers of bins of  $X$  and  $Y$ , respectively.

Subsequently, the first derivative of  $F_z(z)$  with respect to  $z$  is the corresponding PDF. The PDF is calculated as the derivative of  $F_{z,ij,d}(z)$  only, the probabilities of the areas  $a$ ,  $b$  and  $c$  are constant values in this context.

*Summation*

Let  $Z = X + Y$ . The integration boundaries for joint bin  $(i, j)$  are defined as

$$\begin{aligned} y_{u,j} &= \max(y_j, \min(y_{j+1}, z - x_i)) \\ y_{l,j} &= \max(y_j, \min(y_{j+1}, z - x_{i+1})) \\ x_{u,i} &= \max(x_i, \min(x_{i+1}, z - y_{l,j})) \\ x_{l,i} &= \max(x_i, \min(x_{i+1}, z - y_{u,j})) \\ z_{ij} &= x_{u,i} + y_{l,j} = x_{l,i} + y_{u,j}. \end{aligned} \tag{33}$$

Equation (29) for sub-area  $d$  can be written as

$$\begin{aligned} F_{z,ij,d}(z) &= \int_{y=y_{l,j}}^{y_{u,j}} \int_{x=x_{l,i}}^{z_{ij}-y} f_{x,i}(x)f_{y,j}(y) \, dx \, dy \\ &= \int_{y=y_{l,j}}^{y_{u,j}} \int_{x=x_{l,i}}^{z_{ij}-y} [p_{0,x_i} + r_{x_i}x]f_{y,j}(y) \, dx \, dy. \end{aligned} \tag{34}$$

Integration with respect to  $x$  yields

$$F_{z,ij,d}(z) = \int_{y=y_{l,j}}^{y_{u,j}} \left[ \left( p_{0,x_i}x + \frac{1}{2}r_{x_i}x^2 \right) \right]_{x=x_{l,i}}^{z_{ij}-y} f_{y,j}(y) \, dy. \tag{35}$$

Inserting integration boundaries yields

$$\begin{aligned} F_{z,ij,d}(z) &= \int_{y=y_{l,j}}^{y_{u,j}} [p_{0,x_i}(z_{ij} - y - x_{l,i}) \\ &\quad + \frac{1}{2}r_{x_i}((z_{ij} - y)^2 - x_{l,i}^2)]f_{y,j}(y) \, dy. \end{aligned} \tag{36}$$

Substituting  $((z_{ij} - y)^2 - x_{l,i}^2)$  by  $((z_{ij} - y - x_{l,i})^2 + 2x_{l,i}(z_{ij} - y - x_{l,i}))$ ,  $p_{x_{l,i}} = f_{x,i}(x_{l,i})$  and  $z_{ij} - x_{l,i} = y_{u,j}$  yields

$$\begin{aligned} F_{z,ij,d}(z) &= \int_{y=y_{l,j}}^{y_{u,j}} [p_{x_{l,i}}(y_{u,j} - y) + \frac{1}{2}r_{x_i}(y_{u,j} - y)^2] \\ &\quad \times [p_{0,y_j} + r_{y_j}y] \, dy. \end{aligned} \tag{37}$$

Substituting  $r_{y_j}y = -r_{y_j}(y_{u,j} - y) + r_{y_j}y_{u,j}$ , and  $p_{y_{u,j}} = f_{y,j}(y_{u,j})$  yields

$$\begin{aligned} F_{z,ij,d}(z) &= \int_{y=y_{l,j}}^{y_{u,j}} [p_{x_{l,i}}(y_{u,j} - y) + \frac{1}{2}r_{x_i}(y_{u,j} - y)^2] \\ &\quad \times [p_{y_{u,j}} - r_{y_j}(y_{u,j} - y)] \, dy. \end{aligned} \tag{38}$$

Integration with respect to  $y$  yields

$$\begin{aligned} F_{z,ij,d}(z) &= [p_{x_{l,i}}p_{y_{u,j}} \left( -\frac{1}{2} \right) (y_{u,j} - y)^2 - p_{x_{l,i}}r_{y_j} \left( -\frac{1}{3} \right) (y_{u,j} - y)^3 \\ &\quad + \frac{1}{2}r_{x_i}p_{y_{u,j}} \left( -\frac{1}{3} \right) (y_{u,j} - y)^3 \\ &\quad - \frac{1}{2}r_{x_i}r_{y_j} \left( -\frac{1}{4} \right) (y_{u,j} - y)^4]_{y=y_{l,j}}^{y_{u,j}}. \end{aligned} \tag{39}$$

Inserting integration boundaries yields

$$F_{z,ij,d}(z) = \frac{1}{2} p_{x_{l,i}} p_{y_{u,j}} (y_{u,j} - y_{l,j})^2 - \frac{1}{3} p_{x_{l,i}} r_{y_j} (y_{u,j} - y_{l,j})^3 + \frac{1}{6} r_{x_i} p_{y_{u,j}} (y_{u,j} - y_{l,j})^3 - \frac{1}{8} r_{x_i} r_{y_j} (y_{u,j} - y_{l,j})^4. \tag{40}$$

The first derivative of Eq. (40) with respect to  $z_{ij}$  is its corresponding PDF. The variables dependent on  $z_{ij}$  are  $y_{u,j} = z_{ij} - x_{l,i}$ ,  $x_{u,i} = z_{ij} - y_{l,j}$ , and  $p_{y_{u,j}} = f_{y,j}(y_{u,j}) = p_{0,y_j} + r_{y_j}(z_{ij} - x_{l,i})$ . So the derivative writes

$$f_{z,ij,d}(z) = \frac{1}{2} p_{x_{l,i}} r_{y_j} (z_{ij} - x_{l,i} - y_{l,j})^2 + \frac{1}{2} p_{x_{l,i}} p_{y_{u,j}} 2(z_{ij} - x_{l,i} - y_{l,j}) - \frac{1}{3} p_{x_{l,i}} r_{y_j} 3(z_{ij} - x_{l,i} - y_{l,j})^2 + \frac{1}{6} r_{x_i} r_{y_j} (z_{ij} - x_{l,i} - y_{l,j})^3 + \frac{1}{6} r_{x_i} p_{y_{u,j}} 3(z_{ij} - x_{l,i} - y_{l,j})^2 - \frac{1}{8} r_{x_i} r_{y_j} 4(z_{ij} - x_{l,i} - y_{l,j})^3, \tag{41}$$

and can be rewritten as

$$f_{z,ij,d}(z) = p_{x_{l,i}} p_{y_{u,j}} (y_{u,j} - y_{l,j}) - \frac{1}{2} p_{x_{l,i}} r_{y_j} (y_{u,j} - y_{l,j})^2 + \frac{1}{2} r_{x_i} p_{y_{u,j}} (y_{u,j} - y_{l,j})^2 - \frac{1}{3} r_{x_i} r_{y_j} (y_{u,j} - y_{l,j})^3. \tag{42}$$

**Subtraction**

Let  $Z = X - Y$ . The integration boundaries for joint bin  $(i, j)$  are defined as

$$\begin{aligned} y_{u,j} &= \max(y_j, \min(y_{j+1}, x_{i+1} - z)) \\ y_{l,j} &= \max(y_j, \min(y_{j+1}, x_i - z)) \\ x_{u,i} &= \max(x_i, \min(x_{i+1}, z + y_{u,j})) \\ x_{l,i} &= \max(x_i, \min(x_{i+1}, z + y_{l,j})) \\ z_{ij} &= x_{u,i} - y_{u,j} = x_{l,i} - y_{l,j}. \end{aligned} \tag{43}$$

Equation (29) for sub-area  $d$  can be written as

$$F_{z,ij,d}(z) = \int_{y=y_{l,j}}^{y_{u,j}} \int_{x=x_{l,i}}^{z_{ij}+y} f_{x,i}(x) f_{y,j}(y) dx dy. \tag{44}$$

Integration with respect to  $x$  yields

$$F_{z,ij,d}(z) = \int_{y=y_{l,j}}^{y_{u,j}} \left[ p_{0,x_i}(z_{ij} + y - x_{l,i}) + \frac{1}{2} r_{x_i} \left( (z_{ij} + y)^2 - x_{l,i}^2 \right) \right] f_{y,j}(y) dy. \tag{45}$$

Substituting  $((z_{ij} + y)^2 - x_{l,i}^2)$  by  $((z_{ij} + y - x_{l,i})^2 + 2x_{l,i}(z_{ij} + y - x_{l,i}))$ ,  $p_{x_{l,i}} = f_{x,i}(x_{l,i})$  and  $z_{ij} - x_{l,i} = -y_{l,j}$  yields

$$F_{z,ij,d}(z) = \int_{y=y_{l,j}}^{y_{u,j}} \left[ p_{x_{l,i}} (y - y_{l,j}) + \frac{1}{2} r_{x_i} (y - y_{l,j})^2 \right] \times [(p_{y_j} - r_{y_j} y) + r_{y_j} y] dy. \tag{46}$$

Substituting  $r_{y_j} y = r_{y_j} (y - y_{l,j}) + r_{y_j} y_{l,j}$ , and  $p_{y_{l,j}} = f_{y,j}(y_{l,j})$  yields

$$F_{z,ij,d}(z) = \int_{y=y_{l,j}}^{y_{u,j}} \left[ p_{x_{l,i}} (y - y_{l,j}) + \frac{1}{2} r_{x_i} (y - y_{l,j})^2 \right] \times [p_{y_{l,j}} + r_{y_j} (y - y_{l,j})] dy. \tag{47}$$

Integration with respect to  $y$  yields

$$F_{z,ij,d}(z) = \left[ p_{x_{l,i}} p_{y_{l,j}} \left( \frac{1}{2} \right) (y - y_{l,j})^2 + p_{x_{l,i}} r_{y_j} \left( \frac{1}{3} \right) (y - y_{l,j})^3 + \frac{1}{2} r_{x_i} p_{y_{l,j}} \left( \frac{1}{3} \right) (y - y_{l,j})^3 + \frac{1}{2} r_{x_i} r_{y_j} \left( \frac{1}{4} \right) (y - y_{l,j})^4 \right]_{y=y_{l,j}}^{y_{u,j}}. \tag{48}$$

Inserting integration boundaries yields

$$F_{z,ij,d}(z) = \frac{1}{2} p_{x_{l,i}} p_{y_{l,j}} (y_{u,j} - y_{l,j})^2 + \frac{1}{3} p_{x_{l,i}} r_{y_j} (y_{u,j} - y_{l,j})^3 + \frac{1}{6} r_{x_i} p_{y_{l,j}} (y_{u,j} - y_{l,j})^3 + \frac{1}{8} r_{x_i} r_{y_j} (y_{u,j} - y_{l,j})^4. \tag{49}$$

The first derivative of Eq. (49) with respect to  $z$  is its corresponding PDF. The variables dependent on  $z_{ij}$  are  $x_{u,i} = z_{ij} + y_{u,j}$ ,  $y_{l,j} = x_{l,i} - z_{ij}$  and  $p_{y_{l,j}} = f_{y,j}(y_{l,j}) = p_{0,y_j} + r_{y_j}(x_{l,i} - z_{ij})$ . So the derivative writes

$$f_{z,ij,d}(z) = -\frac{1}{2} p_{x_{l,i}} r_{y_j} (y_{u,j} - y_{l,j})^2 + \frac{2}{2} p_{x_{l,i}} p_{y_{l,j}} (y_{u,j} - y_{l,j}) + \frac{1}{3} p_{x_{l,i}} r_{y_j} (y_{u,j} - y_{l,j})^2 - \frac{1}{6} r_{x_i} r_{y_j} (y_{u,j} - y_{l,j})^3 + \frac{3}{6} r_{x_i} p_{y_{l,j}} (y_{u,j} - y_{l,j})^2 + \frac{4}{8} r_{x_i} r_{y_j} (y_{u,j} - y_{l,j})^3, \tag{50}$$

and can be rewritten as

$$f_{z,ij,d}(z) = p_{x_{l,i}} p_{y_{l,j}} (y_{u,j} - y_{l,j}) + \frac{1}{2} p_{x_{l,i}} r_{y_j} (y_{u,j} - y_{l,j})^2 + \frac{1}{2} r_{x_i} p_{y_{l,j}} (y_{u,j} - y_{l,j})^2 + \frac{1}{3} r_{x_i} r_{y_j} (y_{u,j} - y_{l,j})^3. \tag{51}$$

**Multiplication**

Let  $Z = XY$ . For multiplication integration of probability for joint bins has to be performed separately for each quadrant, as can be seen in Fig. 3. In this section, integration for quadrant 1 ( $z \in \langle 0, \infty \rangle$ ) is derived. The integration boundaries for joint bin  $(i, j)$  are defined as

$$\begin{aligned} x_{l,i} &= \max(x_i, \min(x_{i+1}, z/y_{j+1})) \\ y_{l,j} &= \max(y_j, \min(y_{j+1}, z/x_{i+1})) \\ x_{u,i} &= \max(x_i, \min(x_{i+1}, z/y_{l,j})) \\ y_{u,j} &= \max(y_j, \min(y_{j+1}, z/x_{l,i})) \\ z_{ij} &= x_{u,i} y_{l,j} = x_{l,i} y_{u,j}. \end{aligned} \tag{52}$$

Equation (29) for sub-area  $d$  can be written as

$$F_{z,ij,d}(z) = \int_{y=y_{l,j}}^{y_{u,j}} \int_{x=x_{l,i}}^{z_{ij}/y} f_{x,i}(x)f_{y,j}(y) \, dx \, dy. \tag{53}$$

Integration with respect to  $x$  yields

$$F_{z,ij,d}(z) = \int_{y=y_{l,j}}^{y_{u,j}} [p_{0,x_i}(z_{ij}/y - x_{l,i}) + \frac{1}{2}r_{x_i}((z_{ij}/y)^2 - x_{l,i}^2)]f_{y,j}(y) \, dy. \tag{54}$$

Integration with respect to  $y$  yields

$$F_{z,ij,d}(z) = [p_{0,x_i}p_{0,y_j}(z_{ij} \ln |y| - x_{l,i}y) + p_{0,x_i}r_{y_j}(z_{ij}y - \frac{1}{2}x_{l,i}y^2) + \frac{1}{2}r_{x_i}p_{0,y_j}(-z_{ij}^2/y - x_{l,i}^2y) + \frac{1}{2}r_{x_i}r_{y_j}(z_{ij}^2 \ln |y| - \frac{1}{2}x_{l,i}^2y^2)]_{y=y_{l,j}}^{y_{u,j}}. \tag{55}$$

Inserting integration boundaries yields

$$F_{z,ij,d}(z) = p_{0,x_i}p_{0,y_j}(z_{ij} \ln |y_{u,j}/y_{l,j}| - x_{l,i}(y_{u,j} - y_{l,j})) + p_{0,x_i}r_{y_j}(z_{ij}(y_{u,j} - y_{l,j}) - \frac{1}{2}x_{l,i}(y_{u,j}^2 - y_{l,j}^2)) + \frac{1}{2}r_{x_i}p_{0,y_j}(-z_{ij}^2(y_{u,j}^{-1} - y_{l,j}^{-1}) - x_{l,i}^2(y_{u,j} - y_{l,j})) + \frac{1}{2}r_{x_i}r_{y_j}(z_{ij}^2 \ln |y_{u,j}/y_{l,j}| - \frac{1}{2}x_{l,i}^2(y_{u,j}^2 - y_{l,j}^2)). \tag{56}$$

The first derivative of Eq. (56) with respect to  $z_{ij}$  is its corresponding PDF. The variables dependent on  $z_{ij}$  are  $x_{u,i} = z_{ij}/y_{l,j}$ ,  $y_{u,j} = z_{ij}/x_{l,i}$  and  $\ln |y_{u,j}/y_{l,j}| = \ln |z_{ij}/(x_{l,i}y_{l,j})|$ . So the derivative writes

$$f_{z,ij,d}(z) = p_{0,x_i}p_{0,y_j}(\ln |y_{u,j}/y_{l,j}| + z_{ij}z_{ij}^{-1} - x_{l,i}x_{l,i}^{-1}) + p_{0,x_i}r_{y_j}((y_{u,j} - y_{l,j}) + z_{ij}x_{l,i}^{-1} - \frac{1}{2}x_{l,i}2y_{u,j}x_{l,i}^{-1}) + \frac{1}{2}r_{x_i}p_{0,y_j}(-2z_{ij}(y_{u,j}^{-1} - y_{l,j}^{-1}) + z_{ij}^2x_{l,i}z_{ij}^{-2} - x_{l,i}^2x_{l,i}^{-1}) + \frac{1}{2}r_{x_i}r_{y_j}(2z_{ij} \ln |y_{u,j}/y_{l,j}| + z_{ij}^2z_{ij}^{-1} - \frac{1}{2}x_{l,i}^22y_{u,j}x_{l,i}^{-1}). \tag{57}$$

and can be rewritten as

$$f_{z,ij,d}(z) = p_{0,x_i}p_{0,y_j} \ln |y_{u,j}/y_{l,j}| + p_{0,x_i}r_{y_j}(y_{u,j} - y_{l,j}) - r_{x_i}p_{0,y_j}z_{ij}(y_{u,j}^{-1} - y_{l,j}^{-1}) + r_{x_i}r_{y_j}z_{ij} \ln |y_{u,j}/y_{l,j}|, \tag{58}$$

where  $z_{ij}(y_{u,j}^{-1} - y_{l,j}^{-1})$  can be replaced by  $-(x_{u,i} - x_{l,i})$ .

**Division**

Let  $Z = X/Y$ . For division integration of probability for joint bins has to be performed separately for each quadrant,

as can be seen in Fig. 3. In this section, integration for quadrant 1 ( $z \in \langle 0, \infty \rangle$ ) is derived. The integration boundaries for joint bin  $(i, j)$  are defined as

$$y_{u,j} = \max(y_j, \min(y_{j+1}, x_{i+1}/z)) \\ x_{u,i} = \max(x_i, \min(x_{i+1}, zy_{u,j})) \\ y_{l,j} = \max(y_j, \min(y_{j+1}, x_{l,i}/z)) \\ x_{l,i} = \max(x_i, \min(x_{i+1}, zy_j)) \\ z_{ij} = x_{u,i}/y_{u,j} = x_{l,i}/y_{l,j}. \tag{59}$$

Equation (29) for sub-area  $d$  can be written as

$$F_{z,ij,d}(z) = \int_{y=y_{l,j}}^{y_{u,j}} \int_{x=x_{l,i}}^{z_{ij}y} f_{x,i}(x)f_{y,j}(y) \, dx \, dy. \tag{60}$$

Integration with respect to  $x$  yields

$$F_{z,ij,d}(z) = \int_{y=y_{l,j}}^{y_{u,j}} [p_{0,x_i}(z_{ij}y - x_{l,i}) + \frac{1}{2}r_{x_i}((z_{ij}y)^2 - x_{l,i}^2)]f_{y,j}(y) \, dy. \tag{61}$$

Integration with respect to  $y$  yields

$$F_{z,ij,d}(z) = [p_{0,x_i}p_{0,y_j}(\frac{1}{2}z_{ij}y^2 - x_{l,i}y) + p_{0,x_i}r_{y_j}(\frac{1}{3}z_{ij}y^3 - \frac{1}{2}x_{l,i}y^2) + \frac{1}{2}r_{x_i}p_{0,y_j}(\frac{1}{3}z_{ij}^2y^3 - x_{l,i}^2y) + \frac{1}{2}r_{x_i}r_{y_j}(\frac{1}{4}z_{ij}^2y^4 - \frac{1}{2}x_{l,i}^2y^2)]_{y=y_{l,j}}^{y_{u,j}}. \tag{62}$$

Inserting integration boundaries yields

$$F_{z,ij,d}(z) = p_{0,x_i}p_{0,y_j}(\frac{1}{2}z_{ij}(y_{u,j}^2 - y_{l,j}^2) - x_{l,i}(y_{u,j} - y_{l,j})) + p_{0,x_i}r_{y_j}(\frac{1}{3}z_{ij}(y_{u,j}^3 - y_{l,j}^3) - \frac{1}{2}x_{l,i}(y_{u,j}^2 - y_{l,j}^2)) + \frac{1}{2}r_{x_i}p_{0,y_j}(\frac{1}{3}z_{ij}^2(y_{u,j}^3 - y_{l,j}^3) - x_{l,i}^2(y_{u,j} - y_{l,j})) + \frac{1}{2}r_{x_i}r_{y_j}(\frac{1}{4}z_{ij}^2(y_{u,j}^4 - y_{l,j}^4) - \frac{1}{2}x_{l,i}^2(y_{u,j}^2 - y_{l,j}^2)). \tag{63}$$

The first derivative of Eq. (63) with respect to  $z_{ij}$  is its corresponding PDF. The variables dependent on  $z_{ij}$  are  $x_{u,i} = z_{ij}y_{u,j}$  and  $y_{l,j} = x_{l,i}/z_{ij}$ . So the derivative writes

$$f_{z,ij,d}(z) = p_{0,x_i}p_{0,y_j}(\frac{1}{2}y_{u,j}^2 + \frac{1}{2}x_{l,i}^2z_{ij}^{-2} - x_{l,i}^2z_{ij}^{-2}) + p_{0,x_i}r_{y_j}(\frac{1}{3}y_{u,j}^3 + \frac{1}{3}2x_{l,i}^3z_{ij}^{-3} - \frac{1}{2}2x_{l,i}^3z_{ij}^{-3}) + \frac{1}{2}r_{x_i}p_{0,y_j}(\frac{1}{3}2z_{ij}y_{u,j}^3 + \frac{1}{3}x_{l,i}^3z_{ij}^{-2} - x_{l,i}^3z_{ij}^{-2}) + \frac{1}{2}r_{x_i}r_{y_j}(\frac{1}{4}2z_{ij}y_{u,j}^4 + \frac{1}{4}2x_{l,i}^4z_{ij}^{-3} - \frac{1}{2}2x_{l,i}^4z_{ij}^{-3}). \tag{64}$$

and can be rewritten as

$$f_{z,ij,d}(z) = \frac{1}{2} p_{0,x_i} p_{0,y_j} (y_{u,j}^2 - y_{l,j}^2) + \frac{1}{3} p_{0,x_i} r_{y_j} (y_{u,j}^3 - y_{l,j}^3) + \frac{1}{3} r_{x_i} p_{0,y_j} z_{ij} (y_{u,j}^3 - y_{l,j}^3) + \frac{1}{4} r_{x_i} r_{y_j} z_{ij} (y_{u,j}^4 - y_{l,j}^4). \quad (65)$$

## References

- Bierkens MFP, Weerts HJT (1994) Block hydraulic conductivity of cross-bedded fluvial sediments. *Water Resour Res* 30(10):2665–2678. doi:[10.1029/94WR01049](https://doi.org/10.1029/94WR01049)
- Bohnmstedt GW, Goldberger AS (1969) On the exact covariance of products of random variables. *J Am Stat Assoc* 64(328):1439–1442. doi:[10.2307/2286081](https://doi.org/10.2307/2286081)
- Boschan A, Nøttinger B (2012) Scale dependence of effective hydraulic conductivity distributions in 3d heterogeneous media: a numerical study. *Transp Porous Media* 94(1):101–121. doi:[10.1007/s11242-012-9991-2](https://doi.org/10.1007/s11242-012-9991-2)
- Cushman JH, Bennethum LS, Hu BX (2002) A primer on upscaling tools for porous media. *Adv Water Resour* 25(8–12):1043–1067. doi:[10.1016/S0309-1708\(02\)00047-7](https://doi.org/10.1016/S0309-1708(02)00047-7)
- Dagan G (1986) Statistical theory of groundwater flow and transport: pore to laboratory, laboratory to formation, and formation to regional scale. *Water Resour Res* 22(9 Suppl):120–134. doi:[10.1029/WR022i09Sp0120S](https://doi.org/10.1029/WR022i09Sp0120S)
- De Wit A (1995) Correlation structure dependence of the effective permeability of heterogeneous porous media. *Phys Fluids* 7(11):2553. doi:[10.1063/1.868705](https://doi.org/10.1063/1.868705)
- Fiori A, Dagan G, Jankovic I (2011) Upscaling of steady flow in three-dimensional highly heterogeneous formations. *Multiscale Model Simul* 9(3):1162–1180. doi:[10.1137/110820294](https://doi.org/10.1137/110820294)
- Goovaerts P (1997) *Geostatistics for natural resources evaluation*. Oxford University Press, Oxford
- Holmes DT, Buhr KA (2007) Error propagation in calculated ratios. *Clin Biochem* 40(9–10):728–734. doi:[10.1016/j.clinbiochem.2006.12.014](https://doi.org/10.1016/j.clinbiochem.2006.12.014)
- Hristopulos D, Christakos G (1999) Renormalization group analysis of permeability upscaling. *Stoch Env Res Risk Assess* 13(1–2):131–160. doi:[10.1007/s004770050036](https://doi.org/10.1007/s004770050036)
- Hristopulos DT (2003) Renormalization group methods in subsurface hydrology: overview and applications in hydraulic conductivity upscaling. *Adv Water Resour* 26(12):1279–1308. doi:[10.1016/S0309-1708\(03\)00103-9](https://doi.org/10.1016/S0309-1708(03)00103-9)
- Isaaks EH, Srivastava RM (1989) *An introduction to applied geostatistics*. Oxford University Press, Oxford
- Izenman AJ (1991) Recent developments in nonparametric density estimation. *J Am Stat Assoc* 86(413):205–224. doi:[10.2307/2289732](https://doi.org/10.2307/2289732)
- Jaroszewicz S, Korzeń M (2012) Arithmetic operations on independent random variables: a numerical approach. *SIAM J Sci Comput* 34(3):A1241–A1265. doi:[10.1137/110839680](https://doi.org/10.1137/110839680)
- Journel A, Huijbregts C (1978) *Mining geostatistics*, fifth printing 1991 edn. The Blackburn Press, Caldwell
- Kaczynski W, Leemis L, Loehr N, McQueston J (2012) Nonparametric random variate generation using a piecewise-linear cumulative distribution function. *Commun Stat Simul Comput* 41(4):449–468. doi:[10.1080/03610918.2011.606947](https://doi.org/10.1080/03610918.2011.606947)
- King P (1989) The use of renormalization for calculating effective permeability. *Transp Porous Media* 4(1):37–58. doi:[10.1007/BF00134741](https://doi.org/10.1007/BF00134741)
- Korzeń M, Jaroszewicz S (2014) Pacal: a python package for arithmetic computations with random variables. *J Stat Softw* 57(10):1–34. <http://www.jstatsoft.org/v57/i10>
- Kyriakidis P, Gaganis P (2013) Efficient simulation of (log)normal random fields for hydrogeological applications. *Math Geosci* 45(5):531–556. doi:[10.1007/s11004-013-9470-5](https://doi.org/10.1007/s11004-013-9470-5)
- Nøttinger B, Artus V, Zargar G (2005) The future of stochastic and upscaling methods in hydrogeology. *Hydrogeol J* 13:184–201. doi:[10.1007/s10040-004-0427-0](https://doi.org/10.1007/s10040-004-0427-0)
- Pancaldi V, Christensen K, King P (2007) Permeability up-scaling using haar wavelets. *Transp Porous Media* 67(3):395–412. doi:[10.1007/s11242-006-9032-0](https://doi.org/10.1007/s11242-006-9032-0)
- Papoulis A (1991) *Probability, random variables, and stochastic processes*. McGraw-Hill electrical and electronic engineering series. McGraw-Hill, New York
- Sanchez-Vila X, Guadagnini A, Carrera J (2006) Representative hydraulic conductivities in saturated groundwater flow. *Rev Geophys* 44(3):1–46. doi:[10.1029/2005RG000169](https://doi.org/10.1029/2005RG000169)
- Silverman MP, Strange W, Lipscombe TC (2004) The distribution of composite measurements: how to be certain of the uncertainties in what we measure. *Am J Phys* 72(8):1068–1081. doi:[10.1119/1.1738426](https://doi.org/10.1119/1.1738426)
- Tran T (1996) The ‘missing scale’ and direct simulation of block effective properties. *J Hydrol* 183(1–2):37–56. doi:[10.1016/S0022-1694\(96\)80033-3](https://doi.org/10.1016/S0022-1694(96)80033-3)
- Vander Wielen MJ, Vander Wielen RJ (in press) The general segmented distribution. *Commun Stat Theory Methods*. doi:[10.1080/03610926.2012.758743](https://doi.org/10.1080/03610926.2012.758743)
- Vernes R, van Doorn T, Bierkens M, van Gessel S, de Heer E (2005) Van gidslaag naar hydrogeologisch eenheid, toelichting op de totstandkoming van de dataset REGIS II (in Dutch). Technical report, Nederlands Instituut voor Toegepaste Geowetenschappen TNO—Geological Survey of the Netherlands
- Vernes RW, van Doorn T (2006) REGIS II, a 3D hydrogeological model of The Netherlands. In: *Proceedings of the Philadelphia annual meeting of The Geological Society of America*