Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 281 (2003) 313–324
Journal of Hydrology
www.elsevier.com/locate/jhydrol

Stochastic delineation of capture zones: classical versus Bayesian approach

L. Feyen[a,*], P.J. Ribeiro Jr.[b,c], F. De Smedt[a], P.J. Diggle[c]

[a]*Department of Hydrology and Hydraulic Engineering, Free University Brussels, Pleinlaan 2, Brussels 1050, Belgium*
[b]*Departamento de Estatística, Universidade Federal do Paraná, Curitiba, Brazil*
[c]*Department of Mathematics and Statistics, Lancaster University, Lancaster, UK*

Received 25 July 2001; revised 15 September 2002; accepted 8 May 2003

Abstract

A Bayesian approach to characterize the predictive uncertainty in the delineation of time-related well capture zones in heterogeneous formations is presented and compared with the classical or non-Bayesian approach. The transmissivity field is modelled as a random space function and conditioned on distributed measurements of the transmissivity. In conventional geostatistical methods the mean value of the log transmissivity and the functional form of the covariance and its parameters are estimated from the available measurements, and then entered into the prediction equations as if they are the true values. However, this classical approach accounts only for the uncertainty that stems from the lack of ability to exactly predict the transmissivity at unmeasured locations. In reality, the number of measurements used to infer the statistical properties of the transmissvity field is often limited, which introduces error in the estimation of the structural parameters. The method presented accounts for the uncertainty that originates from the imperfect knowledge of the parameters by treating them as random variables. In particular, we use Bayesian methods of inference so as to make proper allowance for the uncertainty associated with estimating the unknown values of the parameters. The classical and Bayesian approach to stochastic capture zone delineation are detailed and applied to a hypothetical flow field. Two different sampling densities on a regular grid are considered to evaluate the effect of data density in both methods. Results indicate that the predictions of the Bayesian approach are more conservative.

Keywords: Groundwater; Capture zone; Stochastic modelling; Bayesian inference

1. Introduction

A time-related capture zone of an extraction well delineates the area from which water is captured by the well within the specified time interval. Precise delineation of well capture zones is a necessary step in the protection of water supplies from accidental contamination. The size and shape of a capture zone depend on the hydrogeological conditions of the system and the properties of the well. The latter are normally well known but the system geometry, boundary conditions and hydraulic properties of the aquifer are usually difficult to determine and subject to uncertainty.

* Corresponding author. Present address: Department of Hydrogeology and Engineering Geology, Catholic University Leuven, Leuven, Belgium. Tel: + 32-16-326441; Fax: + 32-16-326401. Also at Department of Geological and Environmental Sciences, Stanford University, Stanford University, Stanford, CA 94305-2115, USA.
E-mail address: ifeyen@vub.ac.be (L. Feyen).

0022-1694/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0022-1694(03)00193-8

A number of analytical (e.g. Bear and Jacobs, 1965; Lerner, 1992; Kinzelbach et al., 1992; Jacobson, 2002) and numerical approaches (e.g. Varljen and Shafer, 1991; Cole and Silliman, 1997; Vassolo et al., 1998; Evers and Lerner, 1998; Guadagnini and Franzetti, 1999; van Leeuwen, et al., 2000) have been developed over the years to delineate capture zones in various hydrogeological settings. Although analytical methods provide an exact solution to the mathematical statement of the problem, they are often based on strongly simplifying assumptions, which limit their general use. Numerical approaches provide an approximate solution to the mathematical statement of the problem, but are able to simulate more complex situations. Due to the unabated increase in computer power, the detail and accuracy of these models will further improve. Therefore, numerical simulation is potentially the most accurate method for delineating capture zones. In general, most numerical modelling approaches provide a deterministic best estimate of the time-related capture zone based on a calibrated groundwater model combined with a particle-tracking algorithm. However, it is more realistic to approach the problem from a probabilistic point of view to account for the uncertainty that stems from an imperfect knowledge of the aquifer geometry, boundary conditions and hydrogeological parameters. In this paper we only consider the latter, and more specific we focus on the uncertainty in the transmissivity $T(\mathbf{x})$ $[L^2T^{-1}]$. Of the hydraulic properties relevant to capture zones, transmissivity is considered to be the most important. Its variability in space is considerably higher than that of other properties and it can vary by orders of magnitude over a few meters. In the type of aquifers abstraction wells are usually operative, the porosity is typically more homogeneous and its range of variation is significantly less than that of the transmissivity.

To characterize the spatial variability of the transmissivity the theory of random space functions (RSF), as outlined by Delhomme (1978), is adopted. Varljen and Shafer (1991) were the first to use this technique to delineate capture zones by generating conditional simulations using the Monte Carlo (MC) method. More recently, Franzetti and Guadagnini (1996) and Guadagnini and Franzetti (1999) investigated this approach in greater depth. These authors used the MC approach in conjunction with fast Fourier transform-based spectral methods to generate unconditional simulations considering various degrees of domain heterogeneity. The stability and accuracy of the numerical procedure were examined and an empirical stochastic expression for the location of isochrones was developed. van Leeuwen et al. (1998) investigated the influence of both transmissivity variance and correlation scale in a fully confined and leaky-confined aquifer through statistical evaluation of unconditional MC simulations. van Leeuwen et al. (2000) extended this method to condition on regular grids of transmissivity measurements. Riva et al. (1999) determined time-related capture zones for radial flow in two-dimensional randomly heterogeneous media.

The MC approach to capture zone delineation is based on generating a sample of equiprobable realizations of the hydraulic conductivity or transmissivity field, which are all characterized by the same mean and covariance function. These fields are used as input to solve the groundwater flow equation, resulting in a set of equiprobable head distributions. Subsequently, for each head field the time-related capture zones are calculated using a particle-tracking algorithm. Statistical processing of the sample of equally like capture zones results in a probability distribution of the capture zone. Thus, the resulting capture zone will also be a random space function. However, this predictive capture zone distribution only reflects the uncertainty that stems from the inability to actually predict the transmissivity at unmeasured locations. The method neglects the uncertainty about the parameters used to generate the realizations of the stochastic transmissivity field. In general, these parameters are estimated from a limited number of measurements, and used in the prediction equations without accounting for the error inherent in the estimation process.

We believe that in practice at least two components contribute to the predictive uncertainty: (i) the inherent uncertainty in the true value of the random variable when the stochastic mechanism that generates the data is known; and (ii) the additional uncertainty when the mechanism is unknown. To account for the uncertainty in the parameters of the stochastic model of the log transmissivity, we adopt a Bayesian approach in which the unknown parameters are treated as random variables. We derive a predictive

distribution for the capture zone that accounts for both the natural variability of the transmissivity field and the uncertainty in the parameters of the stochastic model. Kitanidis (1986) was the first to examine in theory the effect of parameter uncertainty on inference of spatial functions in a Bayesian framework. The availability of feasible MC sampling algorithms for a very wide range of statistical models has been an essential step in the development of practical Bayesian inference. Parameter uncertainty was also accounted for by Feyen et al. (2001), when determining stochastic capture zones conditioning on head observations. In this paper we describe the Bayesian method to stochastic capture zone delineation and compare it with the more classical or non-Bayesian approach. We apply both methods to a hypothetical flow field and compare the results.

2. Methodology

2.1. Spatial stochastic approach

When facing the problem of predicting a variable at unmeasured locations, a model must be adopted, either deterministic or stochastic. Most of the parameters and variables dealt with in hydrogeological applications exhibit a large spatial variability and cannot be measured exactly. As a result no deterministic model exists that can accurately explain the spatial variability of hydrogeological variables. The RSF model is a convenient tool to model the spatial variability of such variables. In the spatial stochastic approach, the variable under study is considered as a single realization of an infinite set of possible realizations of a RSF.

Transmissivity values are generally found to be log-normally distributed (Freeze, 1975; Hoeksema and Kitanidis, 1985). Here the log transmissivity $Y(\mathbf{x}) = \log T(\mathbf{x})$, where $\mathbf{x} = (x, z)$ are the two-dimensional spatial coordinates, is modelled as a Gaussian stationary RSF, which implies that it is fully characterized by its first two statistical moments:

$$\langle Y(\mathbf{x})\rangle = \mu \tag{1}$$

$$V(\mathbf{u}) = \langle Y'(\mathbf{x}+\mathbf{u})Y'(\mathbf{x})\rangle = \sigma^2\rho(\mathbf{u}) \tag{2}$$

where $\langle\ \rangle$ is the ensemble averaging operator, μ is the expected value, $Y'(\mathbf{x}) = Y(\mathbf{x}) - \langle Y(\mathbf{x})\rangle$ is the fluctuation, $V(\mathbf{r})$ is the auto-covariance, $\mathbf{u}$ is the lag separation vector, $\sigma^2 = V(\mathbf{u} = 0)$ is the variance, and $\rho(\mathbf{u})$ is the correlation function. In order to apply the assumed model for prediction, the mean and covariance function and its parameters need to be inferred from the measurements. In hydrological applications the number of measurements is often limited. This introduces error in the estimation of the parameters that should be accounted for. The uncertainty associated with the estimation of a sample variogram and the selection of an appropriate model has been addressed by several authors (e.g. Russo and Jury, 1987; Shafer and Varljen, 1990). However, in geostatistical applications to hydrogeology, and more specific in the stochastic delineation of capture zones, this source of uncertainty is still often neglected. In this paper we assume that the mean of the log transmissivity μ is a constant, and that $Y(\mathbf{x})$ is characterized by an isotropic exponential two-point covariance function $V(u) = \sigma^2\exp(-|u|/\varphi)$, with φ the integral scale of the spatial stochastic process and u the Euclidean distance between spatial locations $\mathbf{x}_1$ and $\mathbf{x}_2$. At present we did not account for a nugget effect in the analysis. We thus acknowledge uncertainty in three parameters denoted by $\boldsymbol{\theta} = (\mu, \sigma^2, \varphi)$.

2.2. Approaches to inference

The most fundamental difference between classical and Bayesian inference is that in classical inference the parameter vector $\boldsymbol{\theta}$, whilst not known, is treated as a constant rather than random, as is the case in Bayesian inference. The two approaches to capture zone delineation are described here more in detail. Let $Y(\mathbf{x}) = \log \mathrm{T}(\mathbf{x})$ denote the underlying stochastic transmissivity field and $\mathbf{y} = (y_1, y_2, \ldots, y_{n_y})^{\mathrm{T}}$ the data. In what follows we use the notation $[\cdot]$ for the distribution of the quantity within the square brackets, the notation $p(\cdot)$ for the probability of the quantity within the brackets, and a vertical bar to indicate conditioning.

2.2.1. Non-Bayesian inference

Classical or non-Bayesian inference about the transmissivity is based on the distribution $[Y(\mathbf{x})|\mathbf{y}, \boldsymbol{\theta}]$, i.e. conditioning on the data $\mathbf{y}$ and the parameters $\boldsymbol{\theta}$ of the stochastic model. In practice the parameter vector $\boldsymbol{\theta}$ is unknown and replaced by an estimate $\hat{\boldsymbol{\theta}}$

obtained using a curve-fitting technique based on the variogram or a likelihood based methodology. The curve-fitting methods are based on matching the sample variogram to a theoretical variogram family to find the value of the parameters in $\boldsymbol{\theta}$ that optimise some curve-fitting criterion. This curve fitting can be done 'by eye', or by ordinary (OLS) or weighted (WLS) least squares estimation. Within a Gaussian distributional framework efficient estimation methods based on the likelihood function, e.g. the maximum (ML) or restricted maximum likelihood (RML) estimation method, are available that benefit from the well established and widely applicable optimality properties of likelihood-based methods of parameter estimation.

In classical inference the estimated model parameters are then entered into the prediction equations as if they are the true parameters. In this way, realizations of the transmissivity field $Y(\mathbf{x})$ can be obtained directly from the conditional distribution $[Y(\mathbf{x})|\mathbf{y}, \hat{\boldsymbol{\theta}}]$. For each realization of the transmissivity field the time-related capture zones $\mathrm{CAP}(\mathbf{x}, t)$ are calculated. Statistical processing of the ensemble of capture zones results in the predictive distribution for the capture zones, defined as $[\mathrm{CAP}(\mathbf{x}, t)|\mathbf{y}, \hat{\boldsymbol{\theta}}]$, i.e. conditional on the data and the parameter estimates. It reflects the uncertainty due to the natural variability of the transmissivity field but neglects the parameter uncertainty. As a result, the predictive distribution of the capture zone does not reflect the true predictive uncertainty. In the above-mentioned applications of the MC approach to capture zone delineation (Varljen and Shafer, 1991; Franzetti and Guadagnini, 1996; Guadagnini and Franzetti, 1999; van Leeuwen, et al., 1998, 2000), samples are taken from $[\mathrm{CAP}(\mathbf{x}, t)|\mathbf{y}, \hat{\boldsymbol{\theta}}]$ to assess the uncertainty in the delineation of capture zones, thus ignoring parameter uncertainty. These authors did however emphasis that the estimation of the parameters, and as a consequence also the uncertainty in these estimates, was not subject of their study.

2.2.2. Bayesian inference

In the Bayesian approach, both the transmissivity and the model parameters are considered to be random quantities characterized by a distribution. Bayesian inference starts with formulating a model that is thought to be adequate to describe the situation of interest. As stated above, we define a Gaussian stationary RSF for the underlying log transmissivity field. A prior distribution is specified for the unknown parameters of the stochastic model, $[\boldsymbol{\theta}]$, which expresses the belief about the parameters before the data are introduced. Possible sources of prior information are borehole descriptions, previous studies in similar aquifers, expert judgement, etc. After observing the data $\mathbf{y} = (y_1, y_2, \ldots, y_{n_y})^{\mathrm{T}}$, Bayes' rule (3) is applied to obtain the posterior distribution $[\boldsymbol{\theta}|\mathbf{y}]$ for the unknown parameters by combining the prior information with the information in the data, the latter through the likelihood function $L[\boldsymbol{\theta}|\mathbf{y}] \equiv [\mathbf{y}|\boldsymbol{\theta}]$.

$$[\boldsymbol{\theta}|\mathbf{y}] \propto [\boldsymbol{\theta}][\mathbf{y}|\boldsymbol{\theta}] \tag{3}$$

The posterior distribution of the parameters reflects the uncertainty about the parameters after the data have been observed and generally does not correspond to a standard probability distribution. Therefore, inference by numerical simulation using a MC sampling algorithm is adopted. Hereby, correlation between the unknown parameters is accounted for, as discussed in detail in Feyen et al. (2002)

The basis of Bayesian prediction is the resulting predictive distribution for the log transmissivity, which is given by

$$[Y(\mathbf{x})|\mathbf{y}] = \int [Y(\mathbf{x})|\mathbf{y}, \boldsymbol{\theta}][\boldsymbol{\theta}|\mathbf{y}]\mathrm{d}\boldsymbol{\theta}. \tag{4}$$

Samples from this distribution, i.e. conditional realizations of the log transmissivity field, are obtained through 'conditioning by kriging' using the MC sampled parameter sets from the posterior distribution. From Eq. (4) it can be seen that the Bayesian predictive distribution takes into account the parameter uncertainty by averaging over the parameter space the conditional distribution $[Y(\mathbf{x})|\mathbf{y}, \boldsymbol{\theta}]$, with the weights given by the posterior distribution for the model parameters. Thus, the Bayesian predictive distribution is an average of classical predictive distributions for particular values of $\boldsymbol{\theta}$, weighted according to the posterior distribution $[\boldsymbol{\theta}|\mathbf{y}]$. The effect of the averaging in Eq. (4) is typically that the predictions will be more conservative, in the sense that the variance of $[Y(\mathbf{x})|\mathbf{y}]$ will usually be larger than that of the distribution $[Y(\mathbf{x})|\mathbf{y}, \hat{\boldsymbol{\theta}}]$, obtained by using an estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ in the classical predictive distribution $[Y(\mathbf{x})|\mathbf{y}, \boldsymbol{\theta}]$. In comparison with likelihood-based methods, the Bayesian predictive

distribution takes into account the complete likelihood surface rather than focusing on the maximum likelihood estimates of the parameters.

In the Bayesian approach the predictive capture zone distribution is given by $[\mathrm{CAP}(\mathbf{x},t)|\mathbf{y}]$, i.e. conditional on the data only, and thus the uncertainty about the model parameters is incorporated in the predictions. The Bayesian predictive capture zone distribution can be seen as the transformation of the predictive transmissivity distribution, with the transformation given by the groundwater flow equation and the particle-tracking algorithm.

3. Application

When performing an uncertainty analysis it is desirable to have a reference of which the true conditions are known. Therefore, to represent 'reality', a hypothetical flow field is constructed numerically (Feyen et al., 2001) In the remainder of the paper the capture zones will be presented in dimensionless space (x', z') and time (t') coordinates that result from the following transformations (Bear and Jacobs, 1965):

$$x' = \frac{2\pi q_0}{Q}x,\ z' = \frac{2\pi q_0}{Q}z,\ t' = \frac{2\pi q_0^2}{n_e Q}t, \tag{5}$$

where Q $[\mathrm{L}^2\mathrm{T}^{-1}]$ is the extraction rate per unit thickness of aquifer, $q_0 = -T_\mathrm{G} j_0/D$ $[\mathrm{LT}^{-1}]$ is the Darcy background flow velocity, T_G $[\mathrm{L}^2\mathrm{T}^{-1}]$ is the geometric mean transmissivity, D [L] is the aquifer thickness, j_0 is the background hydraulic gradient, and n_e is the effective porosity. The integral scale of the process φ is also expressed in dimensionless terms and is given by $\varphi' = 2\pi q_0 \varphi/Q$.

The flow domain is presented in Fig. 1. The abstraction well is located at the origin of a Cartesian coordinate system $(x', z') = (0, 0)$. At the north $(z' = 119; x' = -117, 126)$ and south boundaries $(z' = -119; x' = -117, 126)$ Neumann conditions are imposed, the specified flux being zero. Dirichlet conditions are specified at the east $(x' = -117; z' = -119, 119)$ and west $(x' = 126; z' = -119, 119)$ sides of the domain, inducing a mean background gradient j_0. For the central part of the domain $(-6 < x' < 15; -8 < z' < 8)$, consisting of a regular grid with dimensions $\Delta x' = 0.1$, the stochastic transmissivity

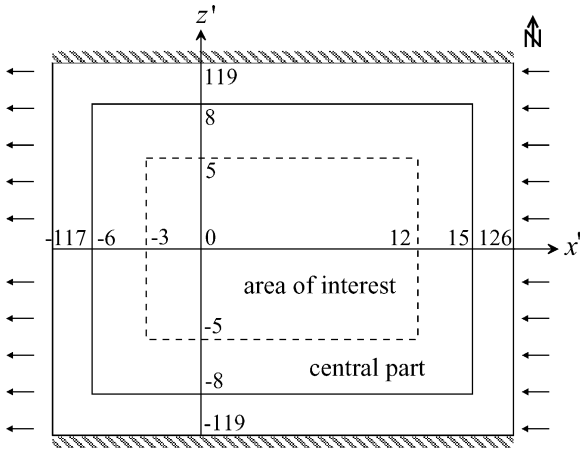

Fig. 1. Plan view of the hypothetical field (not to scale).

fields are generated. In order to minimize interference of the boundaries in the inference of probabilistic capture zones, the surrounding part of the domain consists of 3 rows with increasing dimensions ($\Delta x' = 10^0, 10^1$ and 10^2) and one additional row ($\Delta x' = 0.1$) to impose the boundary conditions. The transmissivity in this part of the flow domain is equal to the geometric mean transmissivity T_G of the central part. A discussion on the effect of grid discretisation and boundary conditions on the convergence of the numerical solution can be found in the work of Franzetti and Guadagnini (1996).

The hypothetical log T field is generated using the sequential Gaussian simulation (sgsim) algorithm of GSLIB (Deutsch and Journel, 1998), and is depicted in Fig. 2. From the area of interest $(-3 < x' < 15; -8 < z' < 8)$ of the hypothetical field, sets of log T measurements are taken positioned in a regular pattern over the area. Two sets of measurements with a different sampling density are taken to serve as measurements in the analysis. The number of measurements in the sets is 13 and 81. The sets are defined such that the smaller set is a subset of the larger one to circumvent the effect of varying measurement locations between the two sampling densities. In a first step, the measurements are used to update the prior distribution for the parameters using Bayes' theorem, yielding the posterior distribution for the parameters. This step is performed using the public domain software geoR (Ribeiro and Diggle, 1999). Next, the measurements are used as conditioning data

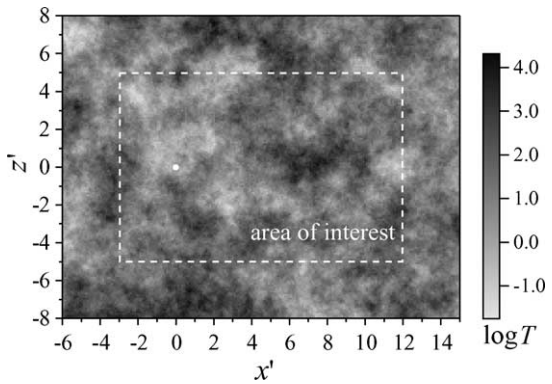

Fig. 2. Hypothetical transmissivity field.

when generating the conditional simulations, i.e. the realizations of the transmissivity fields honour the data at the measurement locations. The method applied to generate the conditional transmissivity fields is 'conditioning by kriging' (Chilès and Delfiner, 1999) using a combination of the kb2d (two-dimensional kriging Program) and sgsim algorithms of GSLIB (Deutsch and Journel, 1998).

Each realization of the transmissivity field is subsequently used to solve the steady-state groundwater flow equation:

$$\frac{\partial}{\partial x}\left[T(x,z)\frac{\partial h}{\partial x}\right]+\frac{\partial}{\partial z}\left[T(x,z)\frac{\partial h}{\partial z}\right]+R=0, \qquad (6)$$

where h is the hydraulic head [L], and R a general sink/source term [L^2T^{-1}], which is limited here to pumping. Eq. (6) is solved with MODFLOW (Mcdonald and Harbaugh, 1989) using a preconditioned conjugate gradient solver. Based on the resulting head field a particle-tracking analysis is performed for the inner area ($-3 < x' < 12; -5 < z' < 5$) of the central part using the semi-analytical particle-tracking algorithm of MODPATH (Pollock, 1989). For any realization the isochrone $\Gamma(x', z', t'_1)$ is defined as the boundary of the capture zone CAP(x', z', t'_1). It envelops all the starting locations (x', z') for which $t' \leq t'_1$, t' being the travel time of a particle towards the well.

Finally, statistical processing of the ensemble of capture zones results in the predictive distribution of the capture zone [CAP($\mathbf{x}', t'|\mathbf{y}$]. The predictive capture zone distribution defines in a point $\mathbf{x}'_1 = (x'_1, z'_1)$ at time step t'_1 the probability $p(\text{CAP}(\mathbf{x}', t')|\mathbf{y})$ that an inert particle released at this point will reach the well within the specified time span. The numerical approximation of this distribution is given by:

$$[\text{CAP}(\mathbf{x}', t')|\mathbf{y}] = \frac{1}{m}\sum_{i=1}^{m}(I(\mathbf{x}', t')|\mathbf{y})_i, \qquad (7)$$

where m is the number of transmissitivity fields generated. The term $(I(\mathbf{x}, t')|\mathbf{y})_i$ inside the summation on the right hand side of Eq. (7) is the probability of intake I by the well and equals one if the particle released in the point $\mathbf{x} = \mathbf{x}_1$ is captured by the well within $t' = t'_1$, and zero otherwise.

4. Results and discussion

4.1. Non-Bayesian inference

Fig. 3 shows the theoretical variogram used to generate the hypothetical field, the sample variogram, and the fitted variograms for the OLS, WLS, ML and RML estimation methods, for $n_y = 13$ (a) and $n_y = 81$ (b). The values of the estimated parameters are given in Table 1. The graphs in Fig. 3 and the values in Table 1 show that, for the field considered, differences exist between the 'true' parameter values of the field and the estimated values, and also between the values of the various estimation methods. This effect is more pronounced when only a few data are available, which is often the case in practice. For the hypothetical field and the selected transmissivity values considered, good estimates for the parameters are obtained for 81 measurements. It must be stated that the mean log transmissivity of the area of interest ($-3 < x' < 12; -5 < z' < 5$) of the hypothetical field is 1.2. This is smaller than the mean of the central part of the domain ($-6 < x' < 15; -8 < z' < 8$), which is equal to 1.48 and close to the mean 1.50 specified when generating the field (see Fig. 2). Since measurements are taken only from the area of interest, this explains the lower estimates of the mean for 81 log T measurements.

In classical inference about capture zones the parameter estimates are used to generate the sample of conditional transmissivity fields and subsequently the capture zones, resulting in the predictive capture zone distribution [CAP($\mathbf{x}, t$)$|\mathbf{y}, \hat{\boldsymbol{\theta}}$], with $\hat{\boldsymbol{\theta}}$ the vector of parameter estimates. Fig. 4 shows contours of [CAP ($\mathbf{x}, t$)$|\mathbf{y}, \hat{\theta}_{\text{avg}}$], with $\hat{\boldsymbol{\theta}}_{\text{avg}}$ the parameter vector with

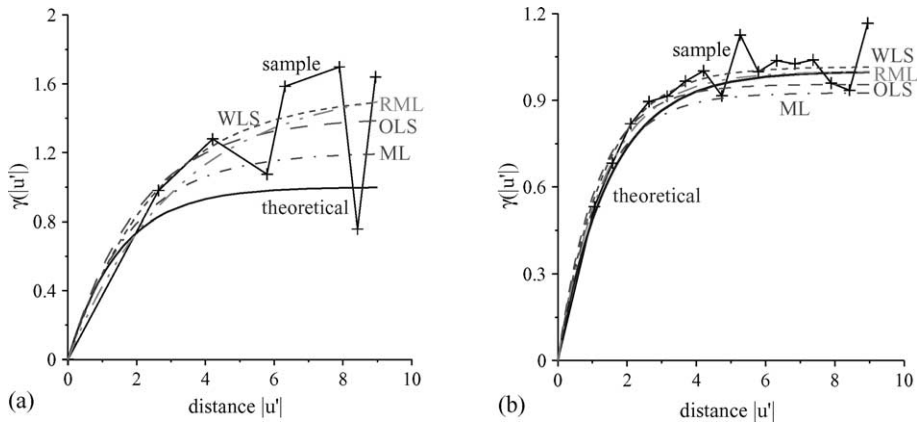

Fig. 3. Theoretical, sample and fitted variograms for the different estimation methods; (a) $n_y = 13$; (b) $n_y = 81$.

the averages of the estimated parameters given in Table 1. Capture zones are determined for 8 time steps, from $t' = 0.5$ to $t' = 4$, with time steps of 0.5. The predictive capture zone distributions for $t' = 1, 2, 3$ and 4 are given in Fig. 4, for 13 (a–d) and 81 (e–h) log T measurements. The zone of uncertainty is the area where $0 < p(\mathrm{CAP}(\mathbf{x}, t)|\mathbf{y}, \hat{\theta}_{\mathrm{avg}}) < 1$, and the 95% uncertainty interval is the area enclosed by the isochrones coinciding with the 2.5 and 97.5 percentiles. The dashed line in the plots of Fig. 4 indicates the location of the reference capture zone for the different time steps.

The contour plots in Fig. 4 show that the zone of uncertainty expands as time increases. Also, the uncertainty is more pronounced in the direction of the regional flow, upstream of the abstraction well. This is a result of the larger travel distances of the particles toward the well in this part of the flow field and as time increases. When more measurements are included in the analysis, the zone of uncertainty decreases and the shape of the distribution approaches the reference capture zone. This is a result of a decrease of the prediction or kriging variance as more conditioning data are available when generating the realizations. The predictive capture zone distribution $[\mathrm{CAP}(\mathbf{x}, t)|\mathbf{y}, \hat{\boldsymbol{\theta}}]$ does not reflect the true predictive uncertainty, as it does only accounts for the uncertainty due to the natural variability of the field, while neglecting the parameter uncertainty. For the case considered, the hypothetical isochrone falls within the uncertainty bounds at all time steps. However, it could happen that for the case where all the parameters are underestimated, i.e. when a smaller mean flow, less variability and little correlation in the field are estimated, this could result in a situation where in distinct areas of the field the reference capture zone is not enveloped by the 95% uncertainty bounds of $[\mathrm{CAP}(\mathbf{x}, t)|\mathbf{y}, \hat{\theta}]$, leading to under-protection of the well.

4.2. Bayesian inference

In the Bayesian approach we must define a prior distribution for the random parameters. For the mean an improper uniform distribution, corresponding to a conjugate Gaussian prior with arbitrarily large variance, is specified. The prior for the variance is also improper and is proportional to $1/\sigma^2$. For the dimensionless correlation parameter a uniform discrete prior between 0 and 10 is specified, with the upper limit defined in order to limit the effects of

Table 1
Parameter values for the hypothetical field and estimated parameters

	μ		σ^2		φ'	
HYPO	1.50		1.00		1.50	
OLS	1.31	1.18	1.41	0.95	2.09	1.11
WLS	1.34	1.17	1.54	1.02	2.56	1.30
ML	1.30	1.18	1.20	0.93	1.86	1.21
RML	1.37	1.17	1.59	0.99	3.23	1.35
	$n_y = 13$	$n_y = 81$	$n_y = 13$	$n_y = 81$	$n_y = 13$	$n_y = 81$

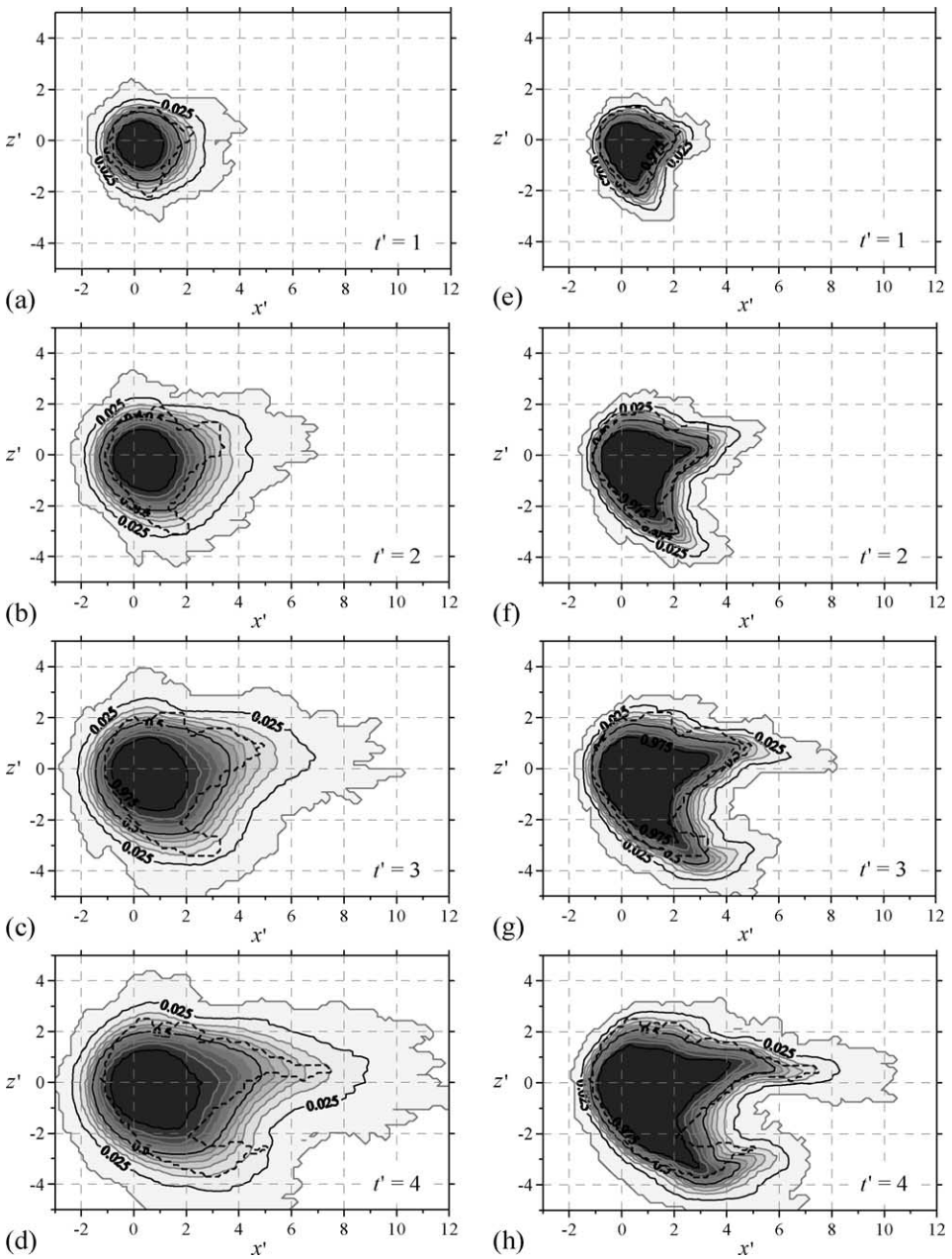

Fig. 4. Stochastic capture zones for the average estimated parameters: (a–d) $n_y = 13$, (e–h) $n_y = 81$. The dashed lines indicate the hypothetical capture zone.

ergodicity. The choice of priors in Bayesian inference is subjective and the main objection to Bayesian inference is that the conclusions will depend on the specific choice of prior. The choices specified here can be interpreted as an expression of prior ignorance, as we assumed that no information about the hydrogeological parameters was present before the data are collected. In real applications, however, prior information about the parameters may be available and more informative priors can be specified.

Using Bayes' theorem, the posterior distribution for the parameters is obtained. This distribution does not resemble a standard probability distribution and MC sampling is used to generate parameter sets from this distribution (Tanner, 1996). Fig. 5 shows samples from the marginal posterior distributions for the mean, sill and integral scale parameter, for $n_y = 13$ (a–c) and $n_y = 81$ (d–f). The histograms show that when more data are incorporated in the conditioning process the variance of the distributions decreases and the central tendency approaches the 'true' values of the model parameters. However, even for relatively large data sets, a considerable degree of uncertainty about the parameters remains, which should be taken into account when delineating capture zones or predicting any other variable of interest.

The sampled parameter sets are used to generate conditional realizations of the transmissivity field for which the capture zones are calculated using the groundwater flow model and particle-tracking algorithm. Statistical analysis of the ensemble of time-related capture zones results in the predictive capture zone distribution $[\mathrm{CAP}(\mathbf{x}, t)|\mathbf{y}]$. The Bayesian predictive distributions for $t' = 1, 2, 3$ and 4 are given in

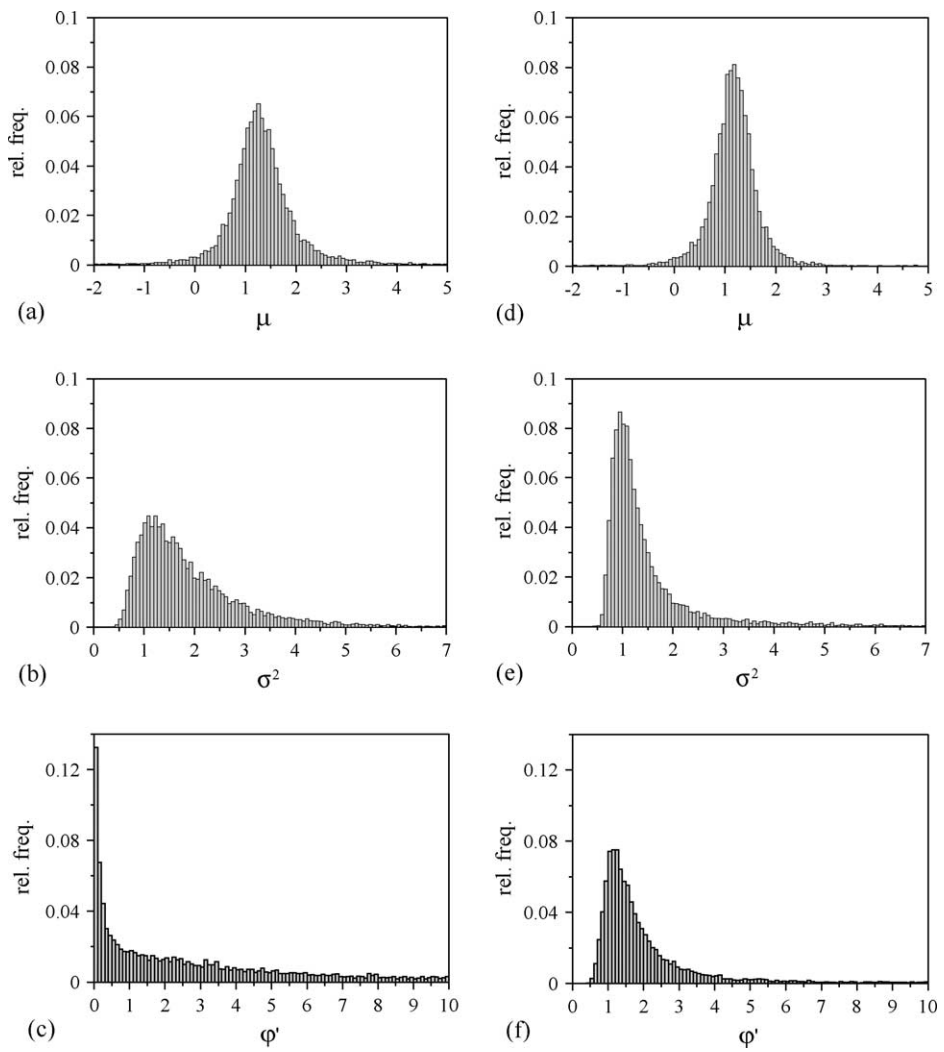

Fig. 5. Samples from the marginal posterior distributions for μ, σ^2 and φ' : (a–c) $n_y = 13$, (d–f) $n_y = 81$.

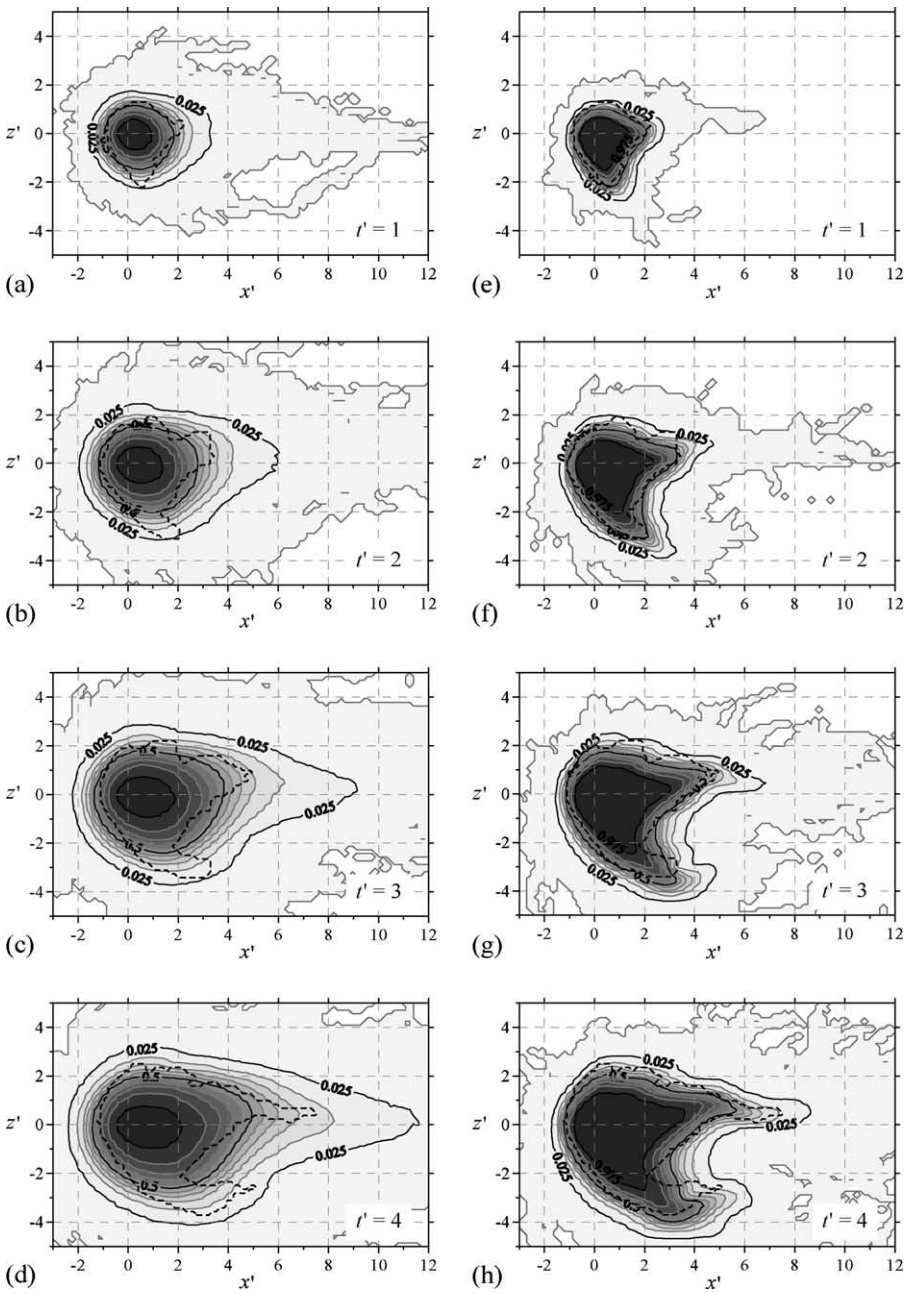

Fig. 6. Stochastic capture zones using Bayesian inference: (a–d) $n_y = 13$, (e–h) $n_y = 81$. The dashed lines indicate the hypothetical capture zone.

Fig. 6, for $n_y = 13$ (a–d) and $n_y = 81$ (e–h). Similar as for the classical predictive distribution, the contour plots show that for an increasing number of conditioning data the zone of uncertainty shrinks and the shape of the distribution approaches the reference capture zone. However, in Bayesian inference this is the result of both the reduction in the variance of the posterior distribution for the model parameters, and the reduction of the kriging variance when more data are used to generate conditional

transmissivity fields. Thus, when more data are incorporated in the analysis, both the effects of parameter uncertainty and natural variability decrease.

Comparing the contour plots in Figs. 4 and 6 shows that, for all time steps and for the two sampling densities considered, there is an increase in the zone of uncertainty and the 95% uncertainty interval when parameter uncertainty is accounted for. However, it is difficult to draw conclusions about the relative contribution of parameter uncertainty to the overall predictive uncertainty, as it depends on the number, location and magnitude of the measurements, and the priors specified. Although a bigger variance of the predictive uncertainty would be expected in the case where the parameters are unknown compared to when the covariance parameters are estimated from the empirical variogram, this is not always true. For a particular data set and priors specified the covariance parameter estimates based on the empirical variogram can be substantially different from the Bayesian estimates, in either direction. This is more likely to occur when the number of measurements is limited. However, the example shows that, even when the variance is overestimated in the classical approach (see Fig. 3a), the Bayesian predictive uncertainty can be larger than the classical predictive uncertainty.

5. Summary and conclusions

In this paper a Bayesian methodology to stochastic capture zone delineation in heterogeneous aquifers was outlined and compared with the classical or non-Bayesian approach. In the latter, the parameters of the stochastic model are usually estimated from a limited number of measurements and then plugged into the prediction equations as if they are the truth. The Bayesian method acknowledges that in practice there are at least two components to predictive uncertainty when using the spatial stochastic approach in prediction: the uncertainty that stems from the inability to exactly predict the transmissivity at unmeasured locations when the stochastic model is known, and the additional uncertainty when the parameters of the stochastic model are unknown. In the Bayesian approach, both the variable of interest and the structural parameters are treated as unknown. Using Bayes' theorem the specified prior distribution for the parameters is updated with the information in the data, yielding the posterior distribution for the parameters. Using a MC sampling strategy parameter sets are obtained from the posterior distribution, which are used to generate conditional simulations of the transmissivity field. Capture zones are determined for each realization by solving the groundwater flow equation and performing a particle-tracking analysis for the flow domain. Statistical analysis of the ensemble of capture zones results in the predictive capture zone distribution, which accounts for both the natural variability and the uncertainty in the parameters of the stochastic model of the hydraulic conductivity. The Bayesian predictive distribution can be seen as an average over the parameter space of the classical predictive distributions, with the weights given by the posterior distribution of the parameters.

Both methodologies have been applied to a hypothetical field. For both methods the predictive uncertainty of the capture zones expands with time. The results also show a reduction of the uncertainty as more measurements are included in the analysis. For the classical approach this is a result of a reduction of the kriging variance when more data are used to generate the conditional transmissivity fields. The reduction in the Bayesian predictive uncertainty results from the smaller variance of the posterior distribution of the model parameters when more data are incorporated, and from the reduction of the kriging variance. Thus, when more data are available in the Bayesian approach, both the effects of parameter uncertainty and natural spatial variability decrease.

Results show that, for all time steps and for the two sampling densities considered, there is an increase in the zone of uncertainty when parameter uncertainty is accounted for. However, it is difficult to draw conclusions about the relative contribution of parameter uncertainty to the overall predictive uncertainty, as it depends on the number, location and magnitude of the measurements, and the priors specified. We argue that predictions based on the fitted parameters do not reflect the true predictive variance. Although we would expect a bigger variance for the Bayesian predictive uncertainty in comparison with the case where the covariance parameters are estimated, this is not always true. For a particular data

set the estimated covariance parameters can be substantially different from the Bayesian estimates, in either direction.

Acknowledgements

The authors wish to thank K.J. Beven and J. Freer for the cooperation, and the access to and aid with the parallel system at the Institute of Environmental Sciences at Lancaster University. The first author wishes to acknowledge the Fund for Scientific Research-Flanders for providing a Research Assistant Scholarship and a travel grant to the first author. The second author wishes to thank CAPES/Brazil, grant BEX 1676/96-2.

References

Bear, J., Jacobs, M., 1965. On the movement of water bodies injected into aquifers. J. Hydrol. 3, 37–57.

Chilès, J.-P., Delfiner, P., 1999. Geostatistics—Modeling Spatial Uncertainty, Wiley, New York, USA.

Cole, B.E., Silliman, S.E., 1997. Capture zones for passive wells in heterogeneous unconfined aquifers. Ground Water 35 (1), 92–99.

Delhomme, J.P., 1978. Kriging in the hydrosciences. Adv. Water Res. 1 (5), 251–266.

Deutsch, C.V., Journel, A.G., 1998. GSLIB, Geostatistical Software Library and User's Guide, Oxford University Press, New York.

Evers, S., Lerner, D.N., 1998. How uncertain is our estimate of a wellhead protection zone? Ground Water 36 (1), 49–57.

Feyen, L., Beven, K.J., De Smedt, F., Freer, J., 2001. Stochastic capture zone delineation within the GLUE-methodology: Conditioning on head observations. Water Resour. Res. 37 (3), 625–638.

Feyen, L., Ribeiro, P.J. Jr., De Smedt, F., Diggle, P.J., 2003. Bayesian methodology to stochastic capture zone determination: Conditioning on transmissivity measurements. Water Resour. Res., 38(9), 10.1029/2001WR000950.

Franzetti, S., Guadagnini, A., 1996. Probabilistic estimation of well catchments in heterogeneous aquifers. J. Hydrol. 174, 149–171.

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

Guadagnini, A., Franzetti, S., 1999. Time-related capture zones for contaminants in randomly heterogeneous formations. Ground Water 37 (2), 253–260.

Hoeksema, R.J., Kitanidis, P.K., 1985. Analysis of the spatial structure of properties of selected aquifers. Water Resour. Res. 21 (4), 563–572.

Jacobson, E., Andricevic, R., Morrice, J., 2002. Ground Water 40 (1), 85–95.

Kinzelbach, W., Marburger, M., Chiang, W.-H., 1992. Determination of groundwater catchment areas in two and three spatial dimensions. J. Hydrol. 134, 221–246.

Kitanidis, P.K., 1986. Parameter uncertainty in estimation of spatial functions: Bayesian analysis. Water Resour. Res. 22 (4), 499–507.

Lerner, D.N., 1992. Well catchments and time-of-travel zones in aquifers with recharge. Water Resour. Res. 28 (10), 2621–2628.

van Leeuwen, M., te Stroet, C.B.M., Butler, A.P., Tompkins, J.A., 1998. Stochastic determination of well capture zones. Water Resour. Res. 34 (9), 2215–2223.

van Leeuwen, M., te Stroet, C.B.M., Butler, A.P., Tompkins, J.A., 2000. Stochastic determination of well capture zones conditioned on regular grids of transmissivity measurements. Water Resour. Res. 36 (4), 949–957.

Mcdonald, M.G., Harbaugh, A.W., 1989. A modular three-dimensional finite-difference ground- water flow model. US Geol. Surv. Open File Rep., 83–875.

Pollock, D.W., 1989. Documentation of computer programs to compute and display pathlines using results from the US geological survey modular three-dimensional finite difference ground-water flow model. US Geol. Surv. Open File Rep., 89–381.

Ribeiro, P.J. Jr., Diggle, P.J., 1999. GeoR/geoS: A geostatistical library for r/s plus. Technical Report ST-99-09. Dept. Maths. and Stats., Lancaster University, Lancaster, UK.

Riva, M., Guadagnini, A., Ballio, F., 1999. Time-related capture zones for radial flow in two dimensional randomly heterogeneous media. Stochastic Environ. Res. Risk. Assess. 13, 217–230.

Russo, D., Jury, W.A., 1987. A theoretical study of the estimation of the correlation scale in spatially variable fields 1, stationary fields. Water Resour. Res. 23 (7), 1257–1268.

Shafer, J.M., Varljen, M.D., 1990. Approximation of confidence limits on sample semivariograms from single realizations of spatially correlated random fields. Water Resour. Res. 26 (8), 1787–1802.

Tanner, M., 1996. Tools for Statistical Inference, Springer, New York.

Varljen, M.D., Shafer, J.M., 1991. Assessment of uncertainty in time-related capture zones using conditional simulation of hydraulic conductivity. Ground Water 29, 737–748.

Vassolo, S., Kinzelbach, W., Schäfer, W., 1998. Determination of a well head protection zone by stochastic inverse modelling. J. Hydrol. 206, 268–280.