

# Application of a Bayesian Approach to Stochastic Delineation of Capture Zones

by L. Feyen<sup>1</sup>, A.M. Dessalegn<sup>2</sup>, F. De Smedt<sup>3</sup>, S. Gebremeskel<sup>4</sup>, and O. Batelaan<sup>5</sup>

---

## Abstract

This paper presents a Bayesian Monte Carlo method for evaluating the uncertainty in the delineation of well capture zones and its application to a wellfield in a heterogeneous, multiaquifer system. In the method presented, Bayes' rule is used to update prior distributions for the unknown parameters of the stochastic model for the hydraulic conductivity, and to calculate probability-based weights for parameter realizations using head residuals. These weights are then assigned to the corresponding capture zones obtained using forward particle tracking. Statistical analysis of the set of weighted protection zones results in a probability distribution for the capture zones. The suitability of the Bayesian stochastic method for a multilayered system is investigated, using the wellfield Het Rot at Nieuwrode, Belgium, located in a three-layered aquifer system, as an example. The hydraulic conductivity of the production aquifer is modeled as a spatially correlated random function with uncertain parameters. The aquitard and overlying unconfined aquifer are assigned random, homogeneous conductivities. The stochastic results are compared with deterministic capture zones obtained with a calibrated model for the area. The predictions of the stochastic approach are more conservative and indicate that parameter uncertainty should be taken into account in the delineation of well capture zones.

---

## Introduction

The awareness of the importance to maintain the quality of ground water in the vicinity of supply wells has resulted in regulations aiming to protect ground water supplies from accidental contamination in both the United States and Europe. The shape and location of the capture

zone for a particular well setting depend on the composite effects of many interacting processes, and on the geometry and hydrogeological properties of the surrounding system. A number of approaches, both analytical and numerical, have been developed to delineate well capture zones for various hydrogeological settings. The application of analytical methods is limited as they are often based on strongly simplifying assumptions. They can be useful for initial estimates of the extent of well catchments or in cases where the assumptions are valid. Numerical approaches, on the other hand, are better suited to simulate more complex systems and, considering the unabated increase in computer power, are potentially the best methods available to delineate the contributing area. In general, most numerical modeling approaches provide a deterministic estimate of the capture zones. However, due to the different sources of error inherent in any modeling exercise, the problem should be approached from a probabilistic point of view. A quantitative measure of the uncertainty associated with the model predictions is needed, which allows the regulatory organizations to implement different degrees of protection for areas with different degrees of uncertainty.

A number of studies have accounted for the uncertainty in the model predictions using different approaches.

---

<sup>1</sup>Formerly with the Dept. of Hydrology and Hydraulic Engineering, Free University Brussels, Pleinlaan 2, 1050 Brussels, Belgium; now with Hydrogeology & Engineering Geology, Catholic University Leuven, Belgium, and the Dept. of Geological and Environmental Sciences, Stanford University, CA; (650) 725-8070; fax (650) 725-0979; lfeyen@stanford.edu

<sup>2</sup>Inter University Program in Water Resources, Free University Brussels-Catholic University Leuven, Belgium; +32-2-6293950; fax +32-2-6293022; amesfind@vub.ac.be

<sup>3</sup>Dept. of Hydrology and Hydraulic Engineering, Free University Brussels, Pleinlaan 2, 1050 Brussels, Belgium; +32-2-6293547; fax +32-2-6293022; fdesmedt@vub.ac.be

<sup>4</sup>Dept. of Hydrology and Hydraulic Engineering, Free University Brussels, Pleinlaan 2, 1050 Brussels, Belgium; +32-2-6293547; fax +32-2-6293022; seifu.gebremeskel@vub.ac.be

<sup>5</sup>Dept. of Hydrology and Hydraulic Engineering, Free University Brussels, Pleinlaan 2, 1050 Brussels, Belgium; +32-2-6293039; fax +32-2-6293022; batelaan@vub.ac.be

Received June 2002, accepted September 2003.

Copyright © 2004 by the National Ground Water Association.

The effect of data variability was investigated by Bhatt (1993), who conducted a parametric analysis to evaluate the individual influence of the main hydrogeological parameters on the delineation process, without quantifying the uncertainty associated with the predicted well catchments for the joint parameters. Bair et al. (1991) used an MC procedure to estimate confidence intervals for isochrones in a homogeneous case. Vassolo et al. (1998) determined capture zone distributions by stochastic inverse modeling using zonation to account for the spatial variability of the hydraulic conductivity and areal recharge. Evers and Lerner (1998) used an MC approach accounting for the uncertainty in five model parameters by sampling parameter values from uniform distributions. They ensured that every model realization produced a realistic simulation by rejecting all simulations that did not meet an arbitrarily chosen head residual criterion. Hunt et al. (2001) presented a similar approach of rejecting unacceptable simulations, combined with a geochemical investigation, to delineate the recharge area of a spring complex in southern Wisconsin. In Levy and Ludy (2000), an approximation of the Gauss-Hermite three-point quadrature approach is presented to more formally quantify the uncertainty associated with the delineation of well catchments.

Another conceptualization, besides the zonation approach, to characterize the spatial variability and uncertainty of ground water flow parameters is based on the theory of random space functions (RSF). Varljen and Shafer (1991) were the first to use this technique to delineate capture zones by generating conditional simulations using an MC approach. Franzetti and Guadagnini (1996) and Guadagnini and Franzetti (1999) used the MC approach in conjunction with fast Fourier transform-based spectral methods to generate unconditional simulations considering various degrees of domain heterogeneity. 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. So far, the only case study that has been reported in which one or more heterogeneous aquifer properties are modeled as a realization of an RSF is the work of van Leeuwen et al. (1999). In that study, capture zones are determined for the Wierden wellfield in the Netherlands, considering uncertainty in both the hydraulic conductivity of the production aquifer and the spatial distribution and thickness of an impeding clay layer situated above the production aquifer. In the work of Frind et al. (2002), the theory of RSF is used to represent the hydraulic conductivity in the four layers of a highly complex glacial moraine system. However, instead of generating multiple realizations, the hydraulic conductivity field is obtained by kriging local  $K$ -values, and then calibrating the model. They use a backward-in-time advective-dispersive transport approach to delineate three-dimensional well capture zones. In this approach, the dispersion term represents heterogeneities not represented explicitly in the kriged hydraulic conductivity field, and results in variations in capture probability due to random uncertainties in the particle travel paths.

Except for the work of Frind et al. (2002), the aforementioned applications of the spatial stochastic approach to capture zone delineation rely on generating a set of equiprobable realizations of the hydraulic conductivity field, for which the capture zone is obtained using a ground water flow model in combination with a particle-tracking algorithm. Statistical analysis of the ensemble of capture zones results in a capture zone probability distribution. However, this approach to capture zone delineation does not take into account the uncertainty in parameter estimates when generating realizations of the stochastic field. Instead, the structural parameters used to generate the simulations are often estimated from a limited number of data and entered into the prediction equations without considering the error inherent in the estimation process. Wingle and Poeter (1993) used a jackknife procedure to estimate the uncertainty of model variogram parameters, and found that realizations of hydraulic conductivity were sensitive to uncertainty in the model variogram. This type of parameter uncertainty was first accounted for in the delineation of capture zones by Feyen et al. (2001) when generating unconditional hydraulic conductivity fields and using head observations within the generalized likelihood uncertainty estimation (GLUE) framework (Beven and Binley 1992) to weight the simulations for prediction. Feyen et al. (2002) present a Bayesian approach to stochastic capture zone delineation, in which the authors condition on regular grids of hydraulic conductivity measurements, hereby accounting for uncertainty in the parameters of the stochastic model for the hydraulic conductivity. In Feyen et al. (2003a, 2003b), this approach is extended to incorporate hydraulic head observations and travel time moments, which are used within a Bayesian framework to assign probability-based weights to realizations of the hydraulic conductivity field.

The novelty of this work is the application of the Bayesian methods presented in Feyen et al. (2002, 2003a) to a wellfield located in a multilayered aquifer system, and the probability weighting of system parameter realizations using head residuals in the three model layers. We account for the spatial variability in the hydraulic conductivity of the production aquifer and for uncertainty in the parameters of the stochastic model used to represent this spatial variability. In addition, we account for the uncertainty in the homogeneous conductivity values of the overlying semi-pervious unit and the top phreatic aquifer of the system. Hydraulic conductivity realizations of the system are weighted based on their ability to reproduce the head observations available in the area. The weights are obtained by applying Bayes' rule and are assigned to the corresponding capture zones for the realization. Ensemble averaging of the set of weighted capture zones results in a probability distribution for the protection zones. The results of the Bayesian approach are compared with the capture zones determined with a deterministic model obtained by a traditional calibration process.

## General Statement of the Problem

The general problem with which we are concerned is to predict the location of the well capture zones from

observed hydrogeological data of the system using a model that simulates the relation between the inputs and outputs, and to quantify the uncertainty of our predictions. Inverse problems arise anywhere data are collected which are related to unknown quantities by a mathematical model. The unknown quantities are the hydrogeological properties of the subsurface, the pumping rates and the boundary conditions and stresses of the system, and the collected data are the conductivity measurements and head observations. The Bayesian approach to the inverse problem, and consequently to capture zone delineation, is the transfer of information from the observations to the unknown quantities to update the prior distributions, yielding posterior distributions for the unknown quantities. Constraining the uncertainty on the ground water model inputs will reduce the uncertainty of the model predictions about the capture zones. In what follows we use the notation  $[.]$  for the distribution of the quantity within the square brackets, the notation  $p(.)$  for the probability of the quantity within the brackets, and a vertical bar to indicate conditioning.

Bayesian inference supposes that the modeler can define a prior distribution for the unknown quantities based on some a priori information. Bayes' rule quantifies how this prior distribution can be modified on the basis of measurements. Consider  $\theta$  as a vector of uncertain parameters and  $\mathbf{d}$  as a vector of measurements, then Bayes' rule states that

$$[\theta|\mathbf{d}] = C_d^{-1}[\theta]L[\theta|\mathbf{d}] \quad (1)$$

where  $C_d = \int [\theta]L[\theta|\mathbf{d}]d\theta$  is a normalizing constant, provided the integral exists. The likelihood function  $L[\theta|\mathbf{d}] \equiv [\mathbf{d}|\theta]$  gives, as a function of the parameter vector  $\theta$ , the probability of observing the data  $\mathbf{d}$  that are actually observed. The term on the left-hand side of Equation 1 is the posterior distribution for the parameters and reflects the conditional probability of each parameter realization to exist after incorporating the observations  $\mathbf{d}$ . Predictions of the capture zones are based on the conditional distribution  $[\theta|\mathbf{d}]$ , resulting in a probability distribution for the capture zones rather than a single deterministic estimate. The capture zone probability distribution reflects the uncertainty in the predictions of the location of the capture zones after incorporating the measurements  $\mathbf{d}$ .

In this work, Bayes' rule is first applied to update the prior distributions of the structural parameters of the production layer conductivity field using the available conductivity measurements, yielding posterior parameter distributions. In a second application of Bayes' theorem, head observations are used to assign probability-based weights to each parameter realization. The latter corresponds to a stochastic field for the production aquifer conductivity field combined with a value for the unknown homogeneous conductivities of the other two model layers. In the section on stochastic inference of well capture zones, we outline how the prior distributions and likelihood functions are defined and show how we can use Bayes' rule to obtain the conditional probabilities associated with the parameter realizations.

## Description of the Study Area: The Wellfield Het Rot

The position of the study area within Belgium is depicted in Figure 1. The wellfield Het Rot is located north-east of Brussels. Land use in the surroundings of the wellfield consists mainly of agriculture, rural residential areas, and nature reserves. The area is hilly with the topography varying between 10 and 80 m above sea level. The wellfield consists of eight extraction wells with an average total discharge of  $\sim 5000 \text{ m}^3/\text{day}$ . The battery of wells is located on a plateau, which forms a topographic water divide. South of the divide, water is drained in the Winge-Molenbeek Valley at  $\sim 2 \text{ km}$  south of the wellfield, whereas water to the north of the divide is drained in the Demer River located  $\sim 3 \text{ km}$  north of the site. The average precipitation in the area is  $780 \text{ mm}/\text{year}$ .

Water is extracted from the Brusseliaan Aquifer, a semiconfined permeable formation consisting of coarse to fine sands. In the lower parts of this aquifer, high permeable zones of very coarse sand occur, which have been deposited in broad northeast-southwest channels several kilometers wide. The wellfield is installed in one such coarse sand channel. The Brusseliaan Aquifer is bounded at the bottom by an impervious clay layer (Kortrijk Formation) and is overlain by a semipervious unit of fine sands with variable clay content (Tongeren Formation) and a phreatic aquifer of coarse to fine glauconitic sands (Diest Formation). A schematic geological profile of the area is shown in Figure 2.

## Deterministic Model for the Study Area

A three-dimensional hydrological model, comprising three layers, is constructed for the area using the U.S. Geological Survey MODFLOW computer code (McDonald and Harbaugh 1988). The conceptual model with indication of the boundaries and the location of the extraction and observation wells is depicted in Figure 3. The horizontal discretization comprises 202 columns and 160 rows, with a constant grid spacing of 50 m, covering an area of  $80.8 \text{ km}^2$  and resulting in a total of 96,960 grid cells in the model. A

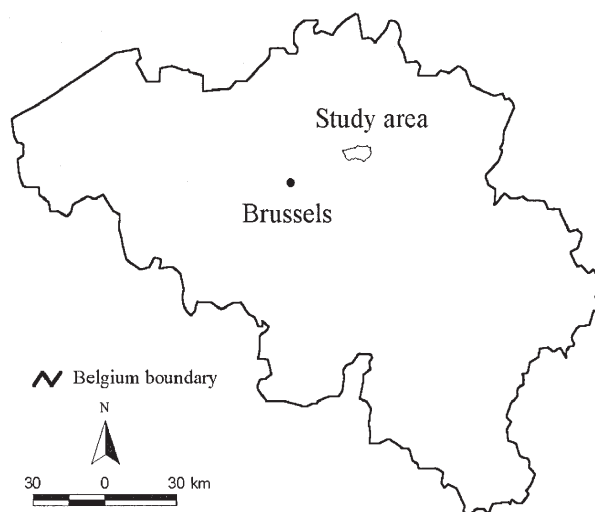


Figure 1. Location of the wellfield Het Rot in Belgium.

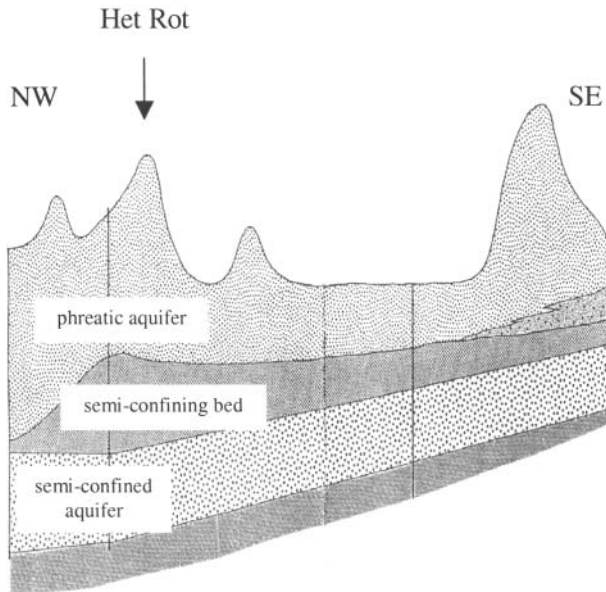


Figure 2. Schematic geological profile of the area (not to scale).

no-flow boundary is defined at the bottom of the model to represent the impervious clay layer. The model is bounded in the north by the Demer River, in the east by the Motte River, and in the south by the Winge-Molenbeek River system. For the west boundary of the model, a constant head boundary is specified based on regional observations of the ground water levels. The river system is modeled in layer 1 using the MODFLOW river package. Other small canals and ditches present in the area are simulated as drains.

Within the area of interest, the production aquifer is characterized by an average thickness of 30 m with an inclination of 0.6% in the north direction. The thickness of the different model layers varies throughout the area and is obtained by interpolating borehole data and geological profiles. The topography of the area is obtained from a digital elevation model of scale 1:10.000. In a site investigation conducted before the actual well implementation, hydraulic conductivity values for the production aquifer have been determined using residual drawdown tests at four locations, indicated in Figure 3 as VB1, VB2, VB3, and VB4. The measured values for the hydraulic conductivity and the aquifer thickness at these locations are given in Table 1.

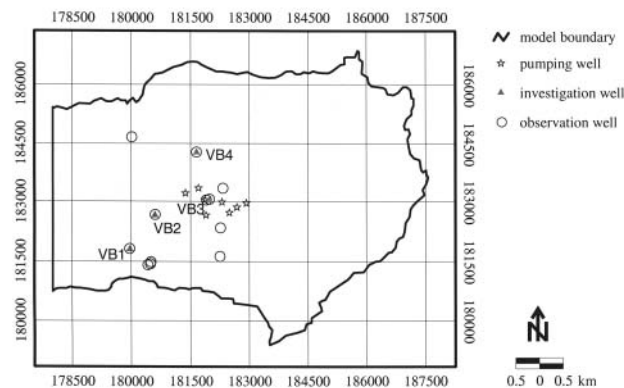


Figure 3. Model boundaries and location of the pumping, investigation, and observation wells.

Well ID	$K$ (m/day)	$D$ (m)
VB1	3	28
VB2	2.5	29
VB3	11.9	27
VB4	6.2	30

The capture zones are determined with the semianalytical particle-tracking code of MODPATH (Pollock 1994). Forward particle tracking is applied to ideal tracer particles placed at the top of the phreatic aquifer and uniformly distributed over the area. Capture zones are not static as their shape and extension depend on time-variable inputs such as recharge and pumping rates. However, generally the temporal variations in these inputs are on a timescale that is considerably smaller than the typical travel time of a particle to the well. Such changes with a higher frequency are averaged out by the inertia of the aquifer (Vassolo et al. 1998). Therefore, the delineation of capture zones is usually based on long-term average conditions using a steady-state model, which is the approach followed in this work. All the time-variable inputs and outputs are averaged over the period 1998 to 2000. The steady-state recharge to the ground water table is obtained with WetSpas (Batelaan and De Smedt 2001), a physically based water balance model for calculating the quasi-steady-state spatially variable evapotranspiration, surface runoff, and ground water recharge.

By manual calibration, the deterministic model is adjusted to closely simulate the averaged observed heads at 12 observation locations, of which some have multiple filters with the screens located in the different model layers, resulting in a total of 17 head calibration data. The parameters subject to calibration are the homogeneous conductivity of the three model layers and the river conductance

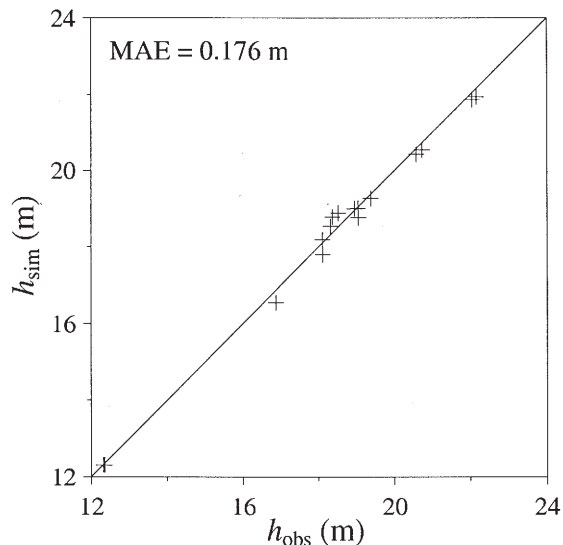


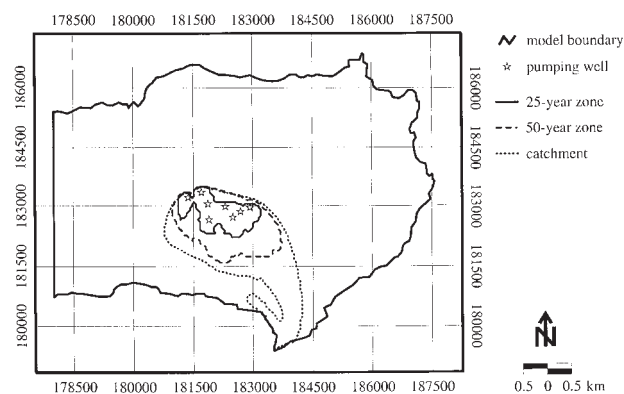
Figure 4. Simulated vs. observed heads for the calibrated deterministic model.

values. However, the influence of the latter on the head distribution in the area of interest is negligible. As no direct measurements are available about the conductivity of the upper two model layers, the ranges evaluated in calibration have been chosen based on values found in the literature (Bronders and De Smedt 1991; Vázquez et al. 2002). The effective porosity values of the model layers are assumed homogeneous and are based on lithology descriptions and values found in the literature.

The results of the calibration are depicted in Figure 4, where a graph of the simulated vs. the observed hydraulic heads is presented. The calibrated parameter values for the conductivities of the different layers are given in Table 2. The mean absolute error of the calibrated model is 0.176 m, and fairly good reproductions of the heads are obtained for all observation data. The model results showed that the flow in the first two layers is mainly vertical. The deterministic 25-year and 50-year capture zones and the well-field catchment obtained with the calibrated model are presented in Figure 5. Flow toward the wellfield is mainly from the southeast direction and the total well catchment reaches the river network in the south of the domain. Due to the nonuniqueness of the inverse problem, multiple parameterizations will be able to closely simulate the hydraulic heads at the observation locations, each yielding a different prediction for the capture zones. Therefore, we believe that the capture zones presented in Figure 5 are a possible representation of the true capture zone, among many other possible representations, and argue that the problem should be approached within a stochastic framework.

## Sources of Uncertainty

A sensitivity analysis is conducted to evaluate the influence of the parameters on the flow model output and the well capture zones. The main sources of uncertainty to delineate the capture zones for the study area are the spatial distribution of the hydraulic conductivity in the Brusselianaan layer and the homogeneous conductivity values of the other model layers. The deposit history of the Brusselianaan Aquifer, resulting in gully structures of high permeable zones (Houthuys 1989) and the variability in the values of the four measurements (Table 1), indicates that a spatially uniform value does not allow for the characterization of the



**Figure 5. Deterministic 25-year and 50-year capture zones and total catchment for the wellfield Het Rot, obtained with the calibrated deterministic model.**

true heterogeneity of the conductivity in the production aquifer. Therefore, the theory of RSF is used to represent the spatial variability of the conductivity in the production aquifer. Notwithstanding the anisotropic nature of the aquifer deposits at the channel scale, no strong anisotropy is observed within the coarse sand channels (Houthuys 1989), and isotropy of the aquifer was assumed at the scale of interest. The conductivity values of the other layers are assumed homogeneous. This implies that uncertainty as a result of the spatial variability of the conductivity in the top two layers is not accounted for in the present study. Van Leeuwen et al. (1999) showed that the variability in the vertical conductance of a confining clay layer strongly affected the shape of the capture zone. In the present study, this variability is partly accounted for by the spatial variation of the thickness of the semiconfining unit.

The rate of natural recharge entering the system is an important factor for the location of the capture zones. However, given the detailed method to determine the average steady-state recharge, at present we do not consider uncertainty in the recharge rate. Pumping rates vary in time and do affect the well capture zones, but the rates are generally well known and, as mentioned earlier, effects of time-varying pumping rates are filtered out on the longer timescale at which the capture zones are determined.

Uncertainty in the effective porosity is not accounted for, as no data are available that allow the effective porosity to be estimated in the inverse analysis. We would like to emphasize that the effective porosity only influences the time-related capture zones. The total well catchment does not depend on the aquifer effective porosity, as it does not influence the head distribution, but only the velocity field, and as a result the timescale of the capture zones.

## Stochastic Inference of Well Capture Zones

### Definition of Prior Distributions for the Unknown Parameters

The stochastic method starts with specifying distributions for the uncertain inputs of the ground water flow model for the area. For the homogeneous conductivity in the upper two layers, lognormal distributions have been specified, with the mean and variance based on borehole descriptions and values found in the literature (Bronders and De Smedt 1991; Vázquez et al. 2002). We denote the unknown parameters of the Diest and Tongeren layers by  $\theta_{D,T} = (K_D, K_T)$ . The prior distributions for the homogeneous parameters are described in Table 2.

The log conductivity  $Y = \log K$  in the Brusselianaan Aquifer is modeled as an isotropic, multi-Gaussian random space function with mean  $\mu$ , variance  $\sigma^2$ , and correlation function  $\rho(u) = \exp[-(u/\varphi)]$ , where  $u$  is the Euclidean distance between two points, and  $\varphi$  is the integral scale. We denote the three structural parameters for the log  $K$  field of the Brusselianaan by  $\theta_B = (\mu, \sigma^2, \varphi)$  and their joint prior distribution by  $[\theta_B]$ . The following functional forms have been used for the priors of the parameters in  $\theta_B$ : (i) for the mean a normal distribution,  $\mu | \sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0)$ , with hyperparameters  $\theta_\mu = (\mu_0, \kappa_0) = (0.778, 1)$ ,  $\kappa_0$  being equivalent to the

<b>Table 2</b> <b>Calibrated Homogeneous Values for the Conductivities of the Layers (Deterministic Model), and Prior Parameter Distributions, with Values for the Mean and Variance, for the Conductivities of the Different Layers (Stochastic Model)</b>		
Geological Unit	Calibrated Values $K$ (m/day)	Prior Distributions (log $K$ )
Diestiaan	7.5	$N(1,0.5)$
Tongeriaan	0.12	$N(-1,0.5)$
Brusseliaan	8	spatial stochastic approach

number of prior measurements; (ii) for the sill, a scaled-inverse  $\chi^2$ -distribution,  $\sigma^2 \sim \text{Sc-Inv-}\chi^2(v_0, \sigma_0^2)$ , with hyperparameters  $\theta_{\sigma^2} = (v_0, \sigma_0^2) = (1, 1)$ ,  $v_0$  being the degrees of freedom and  $\sigma_0^2$  being the scale (Gelman et al. 1995); and (iii) for the integral scale, a locally uniform distribution,  $\varphi \sim U(\varphi_{ll}, \varphi_{ul})$ , with hyperparameters  $\theta_{\varphi} = (\varphi_{ll}, \varphi_{ul}) = (100, 10000)$ ,  $\varphi_{ll}$  and  $\varphi_{ul}$  being the lower and upper limit, respectively. The appearance of  $\sigma^2$  in the prior for the mean parameter means that  $\mu$  and  $\sigma^2$  are dependent in their joint prior density. In this way, prior belief about  $\mu$  is calibrated by the variance  $\sigma^2$  of the measurements and the prior is equivalent to  $\kappa_0$  prior measurements with variance  $\sigma^2$ . The prior distribution for  $\sigma^2$  can be thought of as providing the information equivalent to  $v_0$  observations with average squared deviation  $\sigma_0^2$ . The prior defined for the integral scale can be interpreted as a partial expression of ignorance, as no direct information about the

correlation length was present. The values specified for the hyperparameters are based on values found in the literature about the Brusseliaan Aquifer (Bronders and De Smedt 1991; Vázquez et al. 2002). The choice of priors is inherently subjective and the main objection to Bayesian inference is that the conclusions will depend on the specific choice of a prior distribution. Care should be taken when specifying the prior parameter distributions. If the prior is biased away from the true value, commensurately more data will need to be collected to zero in on the correct parameter values relative to the case of not using prior information.

#### Updating the Prior Distributions with the Conductivity Measurements

The prior distribution  $[\theta_B]$  is updated with the information in the measurements  $\mathbf{y} = (y_1, y_2, y_3, y_4)^T$  using Bayes' rule ( $[\theta_B|\mathbf{y}] \propto [\theta_B]L[\theta_B|\mathbf{y}]$ ), yielding the posterior distribution  $[\theta_B|\mathbf{y}]$  for the parameters of the stochastic model for the log conductivity in the production aquifer. The likelihood  $L[\theta_B|\mathbf{y}] \equiv [\mathbf{y}|\theta_B]$  is a function of the parameter vector  $\theta_B$  and, for the spatial model assumed here for  $Y = \log K$ , has an expression given by the equation of a multivariate normal distribution. The posterior distribution  $[\theta_B|\mathbf{y}]$  reflects the uncertainty about the structural parameters of the conductivity in the Brusseliaan Aquifer after the log  $K$  measurements have been incorporated. This distribution does not correspond to a standard probability distribution and inference by numerical simulation using an MC sampling algorithm is adopted (Diggle et al. 2002; Feyen et al. 2002), taking into account correlation between the parameters. The effect of updating the prior parameter distributions  $[\theta_B]$  with the log  $K$  data is shown in Figure 6, where

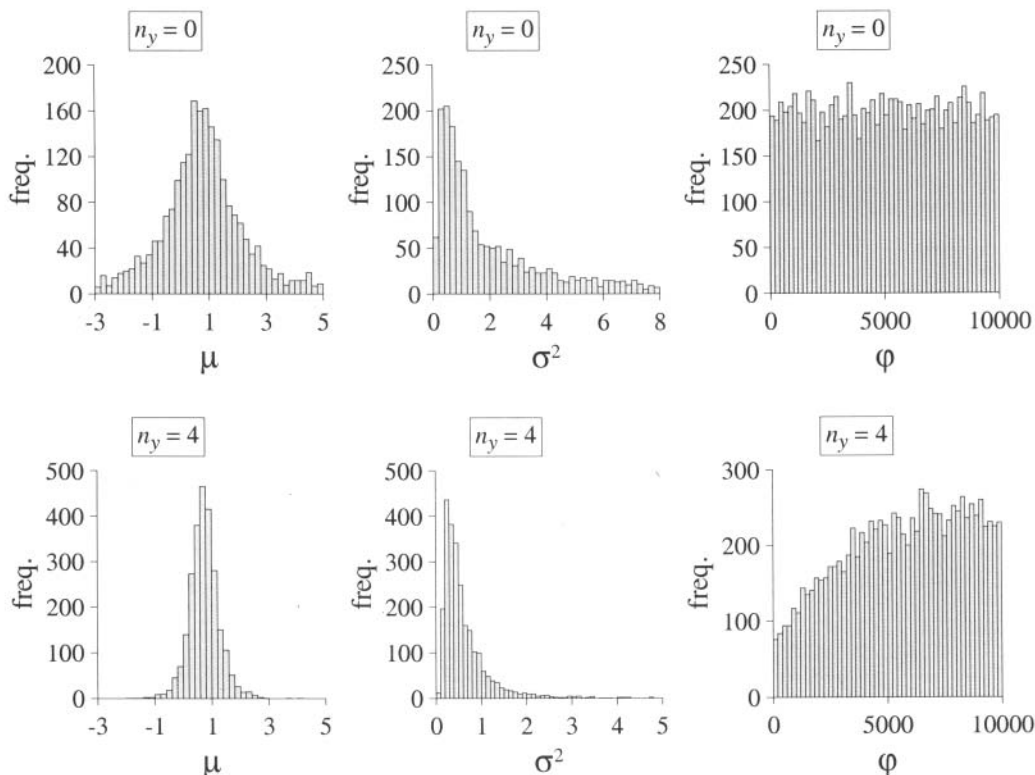


Figure 6. Histograms of sampled parameter sets from parameter distributions  $[\theta|\mathbf{y}]$ , with  $n_y = 0$  and  $n_y = 4$ .

histograms for  $\mu$ ,  $\sigma^2$ , and  $\varphi$ , obtained by sampling from the prior and posterior distributions, are presented. The histograms show that the four measurements considerably reduce the initial uncertainty about the mean and variance of the conductivity field, whereas the uncertainty about the integral scale remains large. Since there are no conductivity measurements for the upper two layers, the prior distributions for the parameters in  $\boldsymbol{\theta}_{D,T}$  cannot be updated with hard data, and parameter values are sampled from the prior distributions  $[\boldsymbol{\theta}_{D,T}]$ .

### Generation of Parameter Realizations

For each parameter set  $\boldsymbol{\theta}_{B,i}$  sampled from the posterior parameter distribution  $[\boldsymbol{\theta}_B|\mathbf{y}]$ , equiprobable realizations of the log  $K$  field honoring the four log  $K$  values at the measurement locations are generated using the sequential Gaussian simulation algorithm of GSLIB (Deutsch and Journel 1998). We use  $Y_{\boldsymbol{\theta}_{B,i},j}$  to denote the  $j$ th realization obtained with the  $i$ th-sampled parameter set  $\boldsymbol{\theta}_{B,i}$ . All log  $K$  realizations generated with  $\boldsymbol{\theta}_{B,i}$  are combined with the  $i$ th-sampled set of values for the parameters in  $\boldsymbol{\theta}_{D,T}$ , resulting in parameter realizations denoted by  $\boldsymbol{\theta}_{ij} = (Y_{\boldsymbol{\theta}_{B,i},j}, \boldsymbol{\theta}_{D,B,i})$ . The parameter realizations are used as input in the deterministic flow model described earlier, resulting in a corresponding head field for each realization  $\boldsymbol{\theta}_{ij}$ .

### Using Head Observations to Assign Probability-Based Weights to Parameter Realizations

Observations of state variables of the system may suggest that some parameter realizations are more likely than others to represent reality. In this study, the observations of state variables of the system are limited to hydraulic head observations, which are used in a Bayesian framework to obtain probability-based weights for each parameter realization. This method is an alternative to the method proposed by Evers and Lerner (1998), but rather than retaining those simulations that yield a fitness value above an arbitrarily chosen threshold, or rather than using the maximum likelihood estimate, we take into account the whole likelihood surface.

The flow models we use are always simplifications of the real processes and the observations of the system are prone to measurement errors. Omitting from the notation the inputs and parameters of the flow model, which are assumed to be known and fixed, e.g., porosity and recharge, the relation between the observed head observations  $\mathbf{h} = (h_1, h_2, \dots, h_{n_h})^T$  and the flow model output  $\mathbf{h}_M = (h_{M,1}, h_{M,2}, \dots, h_{M,n_h})^T$  can be stated as

$$\mathbf{h} = \mathbf{h}_{M(\boldsymbol{\theta})} + \boldsymbol{\epsilon}_h(\boldsymbol{\theta}) \quad (2)$$

where  $\boldsymbol{\epsilon}_h = (\epsilon_{h,1}, \epsilon_{h,2}, \dots, \epsilon_{h,n_h})^T$  is the vector of head residuals, comprising both model and measurement errors, and  $n_h = 17$ . To apply Bayes' theorem, a probability density function with parameters  $\boldsymbol{\psi}$  needs to be specified for the head residuals, which is consistent with the available information about the errors. The joint conditional distribution  $[\boldsymbol{\epsilon}_h|\boldsymbol{\theta}, \boldsymbol{\psi}]$  describes the distribution of the residuals, given  $\boldsymbol{\theta}$  and the parameters  $\boldsymbol{\psi}$  of the assumed error model. This expression, seen as a function of  $\boldsymbol{\theta}$  and  $\boldsymbol{\psi}$ , is called the

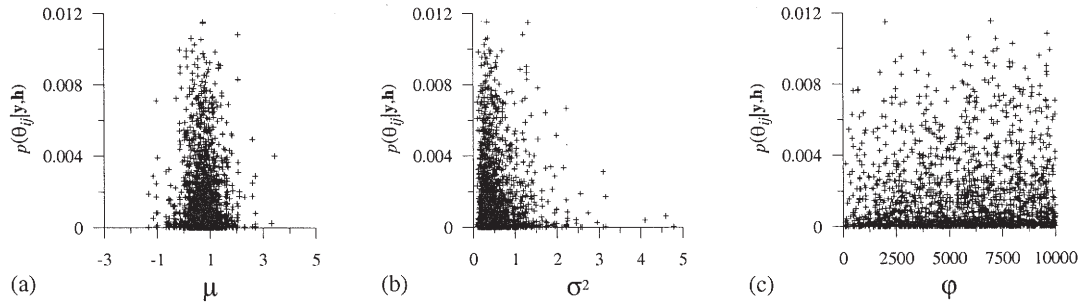
likelihood function of the unknown quantities  $\boldsymbol{\theta}$  and  $\boldsymbol{\psi}$ , and expresses the likelihood of observing the head residuals given the parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\psi}$ . Since the structure of the flow model is known, the likelihood  $L[\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{h}]$  is actually proportional to the probability distribution of the observations  $\mathbf{h}$ .

According to the central limit theorem, in situations where the residuals are composed of linear aggregates of independent component errors of comparable importance, the residuals will be roughly normally distributed, either in the original metric or in some simple transformation (Box and Tiao 1973). Initial model evaluations showed that the residuals are not systematically biased and can be approximated by a normal distribution. Therefore, we assume that the combination of forward modeling and measurement errors is unbiased and Gaussian, in which case the likelihood is given by

$$L[\boldsymbol{\theta}, \boldsymbol{\psi}] = (2\pi)^{-n_h/2} (\det \mathbf{V}_{\boldsymbol{\epsilon}_h})^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{\epsilon}_h(\boldsymbol{\theta}))^T \mathbf{V}_{\boldsymbol{\epsilon}_h}^{-1}(\boldsymbol{\epsilon}_h(\boldsymbol{\theta}))\right) \quad (3)$$

where  $\mathbf{V}_{\boldsymbol{\epsilon}_h}$  is the covariance matrix of dimension  $n_h$  describing forward modeling uncertainties resulting from the measurement and modeling error, and  $\boldsymbol{\psi} = \mathbf{V}_{\boldsymbol{\epsilon}_h}$ . Assuming that the head residuals are uncorrelated between the observation locations, the error covariance matrix  $\mathbf{V}_{\boldsymbol{\epsilon}_h}$  is diagonal, with the terms in the diagonal given by the respective variances  $\sigma_{\boldsymbol{\epsilon}_h,i}^2$  at each location. We treat the error variances  $\sigma_{\boldsymbol{\epsilon}_h,i}^2$  explicitly as unknown and integrate out their effect. The prior distribution for the noise values  $\sigma_{\boldsymbol{\epsilon}_h,i}^2$  is given by a scaled-inverse  $\chi^2$ -distribution,  $\sigma_{\boldsymbol{\epsilon}_h,i}^2 \sim \text{Sc-Inv} - \chi^2(v_{\boldsymbol{\epsilon}_h,0}, \sigma_{\boldsymbol{\epsilon}_h,0}^2)$ , with the degrees of freedom  $v_{\boldsymbol{\epsilon}_h,0}$  equal to one and the scale  $\sigma_{\boldsymbol{\epsilon}_h,0}^2$  of each error component equal to the variance of the time-variable observations at each observation location. Evaluating the likelihood function  $L[\boldsymbol{\theta}, \boldsymbol{\psi}]$  for each parameter realization  $\boldsymbol{\theta}_{ij}$ , integrating out the error variances and normalizing the probabilities, results in the conditional distribution  $[\boldsymbol{\theta}|\mathbf{y}, \mathbf{h}]$ , which reflects the probability  $p(\boldsymbol{\theta}_{ij}|\mathbf{y}, \mathbf{h})$  associated with each parameter realization  $\boldsymbol{\theta}_{ij}$  to represent reality after incorporating the conductivity and head data.

In Figures 7 and 8, scatter plots are presented for all the unknown parameters. Each cross or dot in the scatter plots corresponds to a simulation, where the value on the  $x$ -axis represents the sampled value for the parameter, and the value on the  $y$ -axis is the calculated probability. These marginal scatter plots are obtained by projecting the calculated probabilities in the conditional distribution  $[\boldsymbol{\theta}|\mathbf{y}, \mathbf{h}]$  from the multidimensional parameter space to the parameter axis of interest. For all parameters, a structure is observed in the scatter plots, indicating that the modeling results are sensitive to changes in all the parameters considered in the inverse analysis. The scatter plots also demonstrate the need to approach the problem in a stochastic framework and to consider many possible parameter realizations, which all might be acceptable in simulating the system, given the observations.



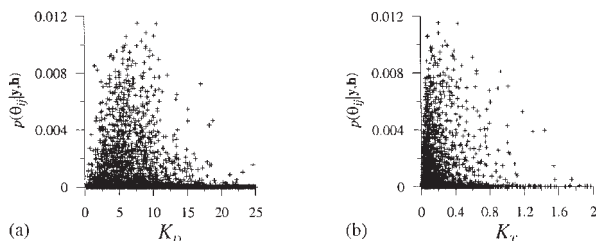
**Figure 7. Marginal scatter plots of calculated conditional probabilities for the structural parameters of the Brusseliana Aquifer.**

### Capture Zone Probability Distributions

The capture zones are determined for each head distribution using forward particle tracking and are weighted by the conditional probability  $p(\theta_{ij}|\mathbf{y},\mathbf{h})$  associated with the corresponding parameter realization  $\theta_{ij}$ . Statistical processing of the set of weighted capture zones results in the numerical approximation of the predictive capture zone distribution, defined by

$$[CAP(\mathbf{x},t)|\mathbf{y},\mathbf{h}] = \frac{\sum_{i=1}^m \sum_{j=1}^n p(\theta_{ij}|\mathbf{y},\mathbf{h}) (I(\mathbf{x},t)|\mathbf{y},\mathbf{h})_{ij}}{\sum_{i=1}^m \sum_{j=1}^n p(\theta_{ij}|\mathbf{y},\mathbf{h})} \quad (4)$$

where  $m$  is the number of parameter sets sampled from  $[\theta_B|\mathbf{y}]$  and  $[\theta_{D,T}]$ ,  $n$  is the number of generated realizations for each sampled parameter set  $\theta_{B,i}$ , and  $p(\theta_{ij}|\mathbf{y},\mathbf{h})$  is the conditional or posterior probability attached to parameter realization  $\theta_{ij}$ . Parameter realizations that better reproduce the head observations will have higher posterior probabilities and therefore have more weight in the predictions. The term  $(I(\mathbf{x},t)|\mathbf{y},\mathbf{h})_{ij}$  inside the summation in the numerator on the right-hand side of Equation 4 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 the specified time interval  $t = t_1$ , and zero otherwise. The predictive capture zone distribution  $[CAP(\mathbf{x},t)|\mathbf{y},\mathbf{h}]$  defines at a point  $\mathbf{x} = \mathbf{x}_1$  and for a time  $t = t_1$  the probability  $p(CAP(\mathbf{x},t)|\mathbf{y},\mathbf{h})$  that an inert particle released at this

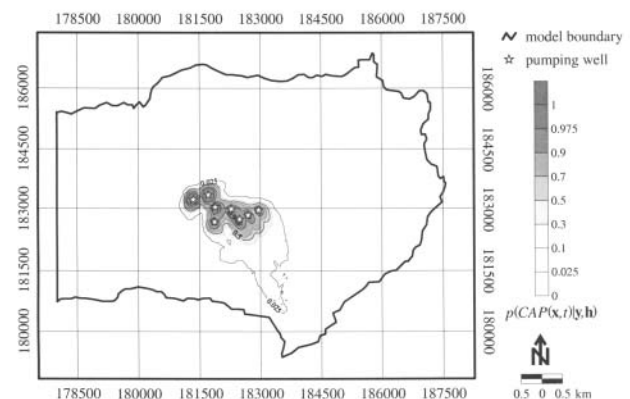


**Figure 8. Marginal scatter plots of calculated conditional probabilities for the homogeneous conductivity of the upper aquifer (Diest Formation) and semipervious unit (Tongeren) (parameters are in m/day).**

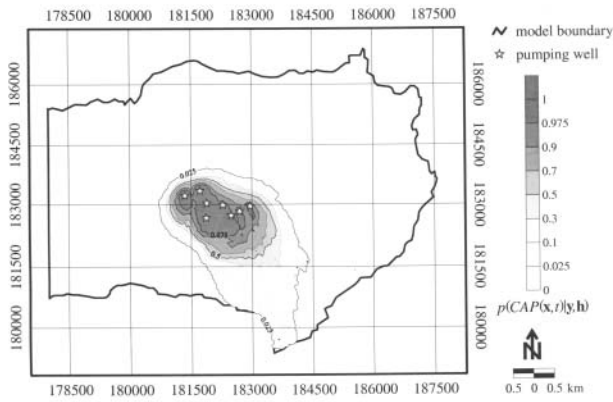
point will reach the well within the specified time span  $t_1$ . The zone of uncertainty is defined as the area where  $0 < p(CAP(\mathbf{x},t)|\mathbf{y},\mathbf{h}) < 1$ , and the 95% uncertainty interval is defined as the area enclosed by the isochrones coinciding with the 2.5 and 97.5 percentiles. The median  $\Gamma_{0.5}(\mathbf{x},t)$  of the predictive distribution is the line for which  $p(CAP(\mathbf{x},t)|\mathbf{y},\mathbf{h}) = 0.5$ .

Figures 9, 10, and 11 show the simulated probability distributions for the 25-year and 50-year capture zones and for the total catchment of the wellfield, respectively. The results presented are based on 10,000 simulations. Darker colors indicate a higher predicted probability to be captured by the wellfield within the specified time span. For a travel time of 25 years, the capture zones of the individual wells can still be distinguished, whereas the capture zones for larger travel times and the total catchment of the individual wells overlap, forming a united capture zone or catchment for the wellfield. The plots for the time-related capture zones and the total catchment exhibit most uncertainty southeast of the wellfield. This is because this is the main contributing area of the wellfield, which results in longer particle pathlengths toward the wells and implies that more heterogeneity is encountered by the particles.

Comparison of the stochastic capture zones in Figures 9, 10, and 11 with the deterministic capture zones in Figure 5 illustrates the advantages of the stochastic approach, as the capture zone probability distribution provides a quantitative measure of the uncertainty associated with the model predictions. Deterministic approaches neglect the uncertainty that is inherent in any modeling practice and



**Figure 9. Stochastic 25-year capture zone for the wellfield Het Rot.**



**Figure 10. Stochastic 50-year capture zones for the wellfield Het Rot.**

underestimate the uncertainty associated with the model predictions. If, for example, the deterministic model parameters result in smaller estimates of the capture zone, e.g., in the case of a smaller estimated mean conductivity, or when neglecting spatial correlation, this could lead to underprotection of the wellfield. Even though the predictions of the stochastic approach are more conservative, they allow the regulatory organizations to implement different degrees of protection for areas with different degrees of uncertainty.

### Summary and Conclusions

In this work, we presented a stochastic approach to delineate the well capture zones for a wellfield located in a multiaquifer system. The approach integrates the theory of RSF, Bayesian inference, MC sampling, inverse modeling, a numerical ground water flow model, and a semianalytical particle-tracking algorithm. We accounted for the spatial variability and uncertainty in the hydraulic conductivity of the three-layered aquifer system. Hydraulic conductivity measurements were used to update prior statistics and to condition conductivity realizations. The hydraulic conductivity realizations of the multiaquifer system, and the corresponding capture zones, were weighted using the head

residuals. Statistical analysis of the ensemble of weighted capture zones resulted in the capture zone probability distribution.

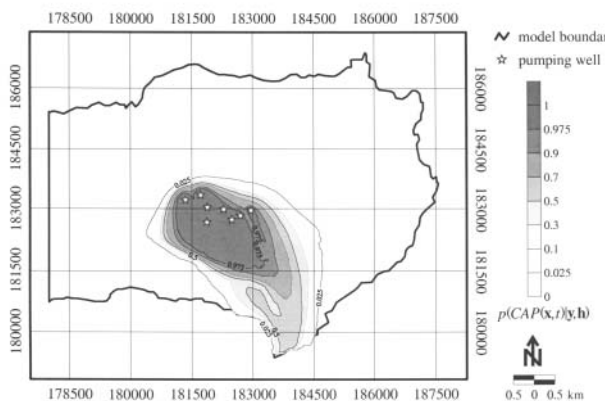
The conductivity in the production aquifer has been modeled as a realization of an RSF. For the conductivity of the semipervious unit overlying the production aquifer and the top phreatic aquifer of the system, homogeneous parameters have been used, which is justified because the flow in the upper two layers is mainly vertical. The statistics of the RSF and the homogeneous parameters have been treated as random variables to account for the fact that they are unknown. Prior distributions have been specified for the unknown parameters. Hard measurements of the conductivity for the production aquifer have been used to update the prior distributions for the structural parameter sets by applying Bayes' theorem. Conditional realizations of the production aquifer conductivity field have been generated using parameter sets obtained by MC sampling from the posterior parameter distributions for the structural parameters. The conditional realizations have been combined with parameter values for the other unknown parameters sampled by MC from the corresponding prior distributions. In a second application of Bayes' theorem, hydraulic head observations have been used to assign probability-based weights to the parameter realizations. For each of the head distributions, the 25-year and 50-year capture zones and the well catchment have been determined using forward particle tracking. Statistical analysis of the ensemble of weighted capture zones resulted in a probability distribution for the capture zones. The stochastic results have been compared with deterministic capture zones obtained with a calibrated model for the area.

Marginal scatter plots of numerically obtained probabilities vs. individual parameter values showed that many different parameter realizations result in acceptable simulations, which should all be taken into account in the prediction of the capture zones. The flow model results showed sensitivity to changes in all the unknown parameters. The capture zone probability distributions showed that the uncertainty in the prediction of the location of the capture zones is most pronounced in the southeast direction, which is the main flow direction toward the wellfield. Comparison of the deterministic isochrones with the capture zone distributions illustrated the advantage of the stochastic approach, as the latter provides a quantitative measure of the uncertainty that is neglected in the deterministic approach.

The stochastic approach presented requires more computational effort than a deterministic approach. However, given the unabated increase in computer power, this should become less problematic in future applications. The predictions are more conservative and indicate that the uncertainty in the predictions should be taken into account in management practices.

### Acknowledgments

The authors wish to thank K.J. Beven and J. Freer for the cooperation, 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



**Figure 11. Stochastic well catchment for the wellfield Het Rot.**

assistant scholarship and a travel grant. The authors wish to thank the Flemish company of water provision (Vlaamse Maatschappij voor Watervoorziening) for providing the data about the wellfield Het Rot. The authors also wish to thank Mary Anderson, Douglas Walker, and an anonymous referee for their valuable comments on the manuscript.

## References

- Bair, E.S., C.M. Safreed, and E.A. Stasny. 1991. A Monte Carlo-based approach for determining travel time-related capture zones of wells using convex hulls as confidence regions. *Ground Water* 29, no. 6: 849–855.
- Batelaan, O., and F. De Smedt. 2001. WetSpa: A flexible, GIS based, distributed recharge methodology for regional groundwater modelling. In *Impact of Human Activity on Groundwater Dynamics*, ed. H. Gehrels, J. Peters, E. Hoehn, K. Jensen, C. Leibundgut, J. Griffioen, B. Webb, and W.-J. Zaadnoordijk, 11–17. Wallingford: IAHS.
- Box, G.E.P., and G.C. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. New York: John Wiley and Sons.
- Beven, K.J., and A.M. Binley. 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes* 6, 279–298.
- Bhatt, K. 1993. Uncertainty in wellhead protection area delineation due to uncertainty in aquifer parameter values. *Journal of Hydrology* 149, 1–8.
- Bronckers, J., and F. De Smedt. 1991. Geostatistical analysis of the hydraulic conductivity of aquifers in the centre of Belgium. *Water* 59, 127–132.
- Deutsch, C.V., and A.G. Journel. 1998. *GSLIB, Geostatistical Software Library and User's Guide*. New York: Oxford University Press.
- Diggle P.J., P.J. Ribeiro Jr., and O.F. Christensen. 2002. An introduction to model based geostatistics. In *Spatial Statistics and Computational Methods*, ed. M.B. Hansen and J. Moller. New York: Springer-Verlag.
- Evers, S., and D.N. Lerner. 1998. How uncertain is our estimate of a wellhead protection zone? *Ground Water* 36, no. 1: 49–57.
- Feyen, L., K.J. Beven, F. De Smedt, and J. Freer. 2001. Stochastic capture zone delineation within the GLUE-methodology: Conditioning on head observations. *Water Resources Research* 37, no. 3: 625–638.
- Feyen, L., P.J. Ribeiro Jr., F. De Smedt, and P.J. Diggle. 2002. Bayesian methodology to stochastic capture zone determination: Conditioning on transmissivity measurements. *Water Resources Research* 38, no. 9: 1164, doi. 10.1029/2001WR000950.
- Feyen, L., P.J. Ribeiro Jr., J.J. Gómez-Hernández, K.J. Beven, and F. De Smedt. 2003a. Bayesian methodology to stochastic capture zone delineation incorporating transmissivity measurements and hydraulic head observations. *Journal of Hydrology* 271, 156–170.
- Feyen, L., J.J. Gómez-Hernández, P.J. Ribeiro Jr., K.J. Beven, and F. De Smedt. 2003b. A Bayesian approach to stochastic capture zone delineation incorporating tracer arrival times, conductivity measurements, and hydraulic head observations. *Water Resources Research* 39, no. 5: 1126, doi. 10.1029/2002WR001544.
- Franzetti, S., and A. Guadagnini. 1996. Probabilistic estimation of well catchments in heterogeneous aquifers. *Journal of Hydrology* 174, 149–171.
- Frind, E.O., D.S. Muhammad, and J.W. Molson. 2002. Delineation of three-dimensional well capture zones for complex multiaquifer systems. *Ground Water* 40, no. 6: 586–598.
- Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 1995. *Bayesian Data Analysis*. London: Chapman and Hall.
- Guadagnini, A., and S. Franzetti. 1999. Time-related capture zones for contaminants in randomly heterogeneous formations. *Ground Water* 37, no. 2: 253–260.
- Houthuys, R. 1989. Vergelijkende studie van de afzettingstructuur van getijdenzanden uit het Eoceen en de huidige Vlaamse banken, Ph.D. dissertation, Department of Geography and Geology, Catholic University, Leuven, Belgium.
- Hunt, R.J., J.J. Steuer, M.T.C. Mansor, and T.D. Bullen. 2001. Delineating a recharge area for a spring using numerical modeling, Monte Carlo techniques, and geochemical investigation. *Ground Water* 39, no. 5: 702–712.
- Levy J., and E.E. Ludy. 2000. Uncertainty quantification for delineation of wellhead protection areas using the Gauss-Hermite quadrature approach. *Ground Water* 38, no. 1: 63–75.
- McDonald, M.G., and A.W. Harbaugh. 1988. A modular three-dimensional finite-difference groundwater flow model. U.S. Geological Survey Techniques of Water-Resources Investigations Book 6, Chapter A1.
- Pollock, D.W. 1994. A particle-tracking post-processing package for MODFLOW, the U.S. Geological Survey finite-difference groundwater flow model. U.S. Geological Survey Open-File Report 88–729.
- van Leeuwen, M., C.B.M. te Stroet, A.P. Butler, and J.A. Tompkins. 1998. Stochastic determination of well capture zones. *Water Resources Research* 34, no. 9: 2215–2223.
- van Leeuwen, M., C.B.M. te Stroet, A.P. Butler, and J.A. Tompkins. 1999. Stochastic determination of the Wierden (Netherlands) capture zones. *Ground Water* 37, no. 1: 8–17.
- van Leeuwen, M., C.B.M. te Stroet, A.P. Butler, and J.A. Tompkins. 2000. Stochastic determination of well capture zones conditioned on regular grids of transmissivity measurements. *Water Resources Research* 36, no. 4: 949–957.
- Varljen, M.D., and J.M. Shafer. 1991. Assessment of uncertainty in time-related capture zones using conditional simulation of hydraulic conductivity. *Ground Water* 29, 737–748.
- Vassolo, S., W. Kinzelbach, and W. Schäfer. 1998. Determination of a well head protection zone by stochastic inverse modelling. *Journal of Hydrology* 206, 268–280.
- Vázquez R.F., L. Feyen, J. Feyen, and J.C. Refsgaard. 2002. Effect of grid size on effective parameters and model performance of the MIKE-SHE code. *Hydrological Processes* 16, no. 2: 355–372.
- Wingle, W.L., and E.P. Poeter. 1993. Uncertainty associated with semivariograms used for site simulation. *Ground Water* 31, no. 5: 725–734.