Characterization of spatial heterogeneity in groundwater applications

PhD Thesis

Department of Geotechnical Engineering and Geo-Sciences (ETCG)

Technical University of Catalonia, UPC

Paolo Trinchero

February 2009

HYDROGEOLOGY GROUP
TECHNICAL UNIVERSITY OF CATALONIA

Chapter 3

A framework for the stochastic Delineation of Connectivity Patterns Honoring Transmissivity, Tracer Data, and Pumping Tests Data*

3.1 Introduction

In the previous chapter we have provided a mathematical framework for the spatial description of transport point-to-point connectivity. In section 2.5 we have presented a potential application of this framework for the delineation of protection areas (i.e. capture zones) around a pumping well used for drinking water production. Nevertheless, in its current formulation the analytical solution is not directly applicable to real applications because it requires a large amount of information that is usually not available and economically unfeasible in standard field campaigns.

* Fernandez-Garcia, D., Trinchero, P., Sanchez-Vila, X., in preparation for submission .

Based on our previous work on point-to-point transport connectivity, in this chapter we present a stochastic framework for the delineation of connectivity patterns using a limited and sparse number of measurements. The lack of complete knowledge of the variables involved in the problem is overcome by treating them as regionalized variables or random functions. The methodology allows to condition the results to three types of data measured over different scales, namely travel times of convergent tracer tests, t_a, estimates of the storage coefficient from pumping tests interpreted using the Cooper-Jacob method, S_{est}, and measurements of transmissivity point values, T.

The ability of the methodology to properly delineate capture zones is assessed through estimations (i.e. ordinary cokriging) and sequential gaussian simulations based on different sets of measurements.

3.2 Problem Description

The objective of this paper is to obtain equally probable realizations of connectivity patterns conditioned to three different types of data, namely:

- Measurements of transmissivity point values, $T(\mathbf{x}_i)$ $(i = 1, ..., N_t)$.

- Measurements of the estimated storage coefficient, $S_{est}(\mathbf{x}_i)$ $(i = 1, ..., N_s)$, obtained from interpreting pumping tests using the Cooper Jacob method (see section 2.2 for a detailed explanation of the method). As observed by *Meier et al.* [62], this method provides estimates of the transmissivity that are somehow representative of the equivalent value of the entire aquifer (i.e., geometric mean of the transmissivity point values). On the contrary, the estimated storage coefficient usually shows a large variability being dependent on the location of the observation point. As explained in detail in section 2.2, this variability is a

direct consequence of the dependence of S_{est} on the particular connecting features existing between the pumping well and the observation location.

- Measurements of mean travel times obtained from convergent-flow tracer tests, $t_a(\mathbf{x}_i)$ ($i = 1, ..., N_a$) estimated as follows

$$t_a(\mathbf{x}) = \int_0^\infty tC(t;\mathbf{x})dt / \int_0^\infty C(t;\mathbf{x})dt \tag{3.1}$$

where $C(t;\mathbf{x})$ is the concentration of a solute injected in $\mathbf{x}$, measured at the well at the time t. In heterogeneous media, t_a is an indicator of the transport point-to-point connectivity between the pumping well and the injection location (see chapter 2 for further details on the physical interpretation of this parameters).

It is worthwile to note that each data type is measured over a different scale and represents crucial field information controlling the shape of the connectivity patterns induced by a pumping well.

3.3 Background

We consider an infinite two-dimensional porous media under quasi-steady-state radial flow conditions induced by a pumping well. Treating transmissivity as a spatially varying property throughout the aquifer while assuming all other flow and transport parameters spatially constant, we have demonstrated (see section 2.3) that

$$t_a(\mathbf{x}_i) = \frac{\pi(r_i^2 - r_w^2)b\phi}{Q_w} \exp\left(\tau_i'\right) \tag{3.2}$$

$$\tau_i' = -\int_0^{r_i} \frac{2r}{r_i^2}\Big(Y_i'(r) - \omega_i'(r) + \omega_i'(r_i)\Big)dr \tag{3.3}$$

$$\omega_i'(r) = -\frac{1}{\pi}\int_{\Re^2} Y'(\mathbf{x})U(\mathbf{x}, \mathbf{x}_i(r))d\mathbf{x} \tag{3.4}$$

where $t_a(\mathbf{x}_i)$ is the mean travel time of a solute injected in $\mathbf{x}_i$, Q_W is the flow rate, r_w is the well radius, Y' is the fluctuation of the natural logarithm of transmissivity around an equivalent value ($Y(\mathbf{x}) = Y_0(\mathbf{x}) + Y'(\mathbf{x})$; $Y_0(\mathbf{x}) = \ln T_0$), ω' is the natural logarithm of the estimated storage coefficient divided by its true value ($\omega' = \ln(S_{est}/S)$), and

$$U(\mathbf{x};\mathbf{y}) = \frac{\|2\mathbf{x} - \mathbf{y} + \mathbf{x}_w\|^2 - \|\mathbf{y} - \mathbf{x}_w\|^2}{4\|\mathbf{x} - \mathbf{x}_w\|^3\|\mathbf{x} - \mathbf{y}\|^2} \tag{3.5}$$

The subscript i denotes the position $\mathbf{x}_i$, and the variable r refers to the distance of a given point from the pumping well centered at the $\mathbf{x}_w$ location. We refer to an intermediate point located on the line between the well and the observation point $\mathbf{x}_i$, using the following notation

$$\mathbf{x}_i(r) = \mathbf{x}_w + \frac{r}{r_i}(\mathbf{x}_i - \mathbf{x}_w) \tag{3.6}$$

Thus, by convention, we have the following equalities,

$$Y_i' = Y'(\mathbf{x}_i) \qquad \omega_i' = \omega'(\mathbf{x}_i) \tag{3.7}$$

$$Y_i'(r) = Y'\left(\mathbf{x}_i(r)\right) \qquad \omega_i'(r) = \omega'\left(\mathbf{x}_i(r)\right) \tag{3.8}$$

3.4 Mathematical Development

Here, we develop a mathematical framework to carry out stochastic conditional estimation/simulation of the natural logarithm of the mean travel time ($\tau = \ln t_a$). This mean travel time is an indicator of transport point-to-point connectivity which can be used afterwards to directly delineate the capture zone of an abstraction well. The method aims at delineating connectivity patterns honoring not only transmissivity data, $T(\mathbf{x}_i)$ $(i = 1, ..., N_t)$, but also tracer data such as $t_a(\mathbf{x}_i)$ $(i = 1, ..., N_a)$ and pumping test data $S_{est}(\mathbf{x}_i)$ $(i = 1, ..., N_s)$. The method is build upon a stochastic model of the spatial dependency of the natural log of transmissivity, $Y(\mathbf{x}) = \ln T(\mathbf{x})$, which is used along with our previous theoretical development on transport connectivity [94]. We consider $Y(\mathbf{x})$ to describe a correlated random function quantified by its expectation, $m_Y = E[Y]$, and the two-point covariance function, $C^{yy}(\mathbf{x}_i, \mathbf{x}_j) = E((Y(\mathbf{x}_i) - m_Y)(Y(\mathbf{x}_j) - m_Y))$, where $E(\cdot)$ denotes the expectation operator.

3.4.1 Conditional Estimation: Kriging

Kriging is a group of geostatistical techniques aimed to get the "best" estimate (and corresponding uncertainty) of an attribute at an unknown location from nearby observations of related attributes. The relative contribution of each observation is based on the the spatial dependency of the attributes. Here, we will consider simple/ordinary linear co-Kriging. Essentially, we want to obtain the best unbias linear estimator of $\tau'_K(x_0)$ at a given location x_0 given the values of Y, τ, and w at some other points $\{Y_1, ..., Y_n; \tau_1, ..., \tau_m; w_1, ..., w_l\}$

$$\tau'_K(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i^Y Y_i + \sum_{i=1}^{m} \lambda_i^\tau \tau_i + \sum_{i=1}^{l} \lambda_i^\omega \omega_i \tag{3.9}$$

where $\tau'_K(\mathbf{x}_0)$ is the Kriging estimator at the $\mathbf{x}_0$ location, $\tau_i = \ln t_a(\mathbf{x}_i)$, and $\omega_i = \ln S_{est}(\mathbf{x}_i)$.

Simple Co-Kriging

In this case, we consider that $Y(\mathbf{x})$ and $\omega(\mathbf{x}) = \ln S_{est}(\mathbf{x})$ are two correlated stationary random functions with known mean, m_Y and m_ω, and covariance function. The natural log of the travel time, $\tau(\mathbf{x}) = \ln t_a(\mathbf{x})$, is a non-stationary random function expressed as the sum of a trend component and a residual, $\tau(\mathbf{x}) = m_\tau(\mathbf{x}) + \tau'(\mathbf{x})$. The residual $\tau'(\mathbf{x})$ is modeled as a stationary random function. Based on this, we can directly obtain an unbias estimator of τ' as

$$\tau'_{SK}(\mathbf{x}_0) = \sum_i \lambda_i^Y (Y_i - m_Y) + \sum_i \lambda_i^\tau (\tau_i - m_\tau(\mathbf{x}_i)) + \sum_i \lambda_i^\omega (\omega_i - m_\omega) \tag{3.10}$$

where $\mathbf{x}_0$ refers to the point of estimation. Defining the random variable

$$Z'_{\alpha i} = \begin{cases} Y'_i = Y_i - m_Y & \text{if } \alpha = Y \\ \tau'_i = \tau_i - m_\tau(\mathbf{x}_i) & \text{if } \alpha = \tau \\ \omega'_i = \omega_i - m_\omega & \text{if } \alpha = \omega \end{cases} \tag{3.11}$$

we write

$$\tau'_{SK}(\mathbf{x}_0) = \sum_\alpha \sum_i \lambda_i^\alpha Z'_{\alpha i} \qquad \text{where } E(Z'_{\alpha i}) = 0 \tag{3.12}$$

Then, the kriging predictor is the one that minimizes the variance of the prediction error written as

$$E\left[(\tau'_{SK} - \tau')^2\right] = \sum_\alpha \sum_\beta \sum_i \sum_j \lambda_i^\alpha \lambda_j^\beta E(Z'_{\alpha i} Z'_{\beta j}) - 2 \sum_\alpha \sum_i \lambda_i^\alpha E(Z'_{\alpha i} \tau'_0) + E(\tau'^2) \tag{3.13}$$

Minimizing the prediction error,

$$\frac{\partial E\left[(\tau'_{SK} - \tau')^2\right]}{\partial \lambda_i^\alpha} = 0 \qquad \forall \alpha = Y, \tau, \omega \quad \forall i = 1, ..., n \tag{3.14}$$

we arrive at the following Kriging linear system of equations

$$\sum_\beta \sum_j \lambda_j^\beta E(Z'_{\alpha i} Z'_{\beta j}) = E(Z'_{\alpha i} \tau'_0) \qquad \forall \alpha = Y, \tau, \omega \quad \forall i = 1, ..., n \tag{3.15}$$

By substituting (3.15) into (3.13) we obtain the variance of the simple Kriging estimate

$$\sigma^2_{SK} = E\left[(\tau'_{SK} - \tau')^2\right] = E(\tau'^2) - \sum_\alpha \sum_i \lambda_i^\alpha E(Z'_{\alpha i} \tau'_0) \tag{3.16}$$

where

$$E(Z'_{\alpha i} Z'_{\beta j}) = \begin{cases} E(Y'_i Y'_j) = C^{yy}(\mathbf{x}_i, \mathbf{x}_j) & \text{if } \alpha = Y, \quad \beta = Y \\ E(Y'_i \tau'_j) = C^{y\tau}(\mathbf{x}_i, \mathbf{x}_j) & \text{if } \alpha = Y, \quad \beta = \tau \\ E(Y'_i \omega'_j) = C^{y\omega}(\mathbf{x}_i, \mathbf{x}_j) & \text{if } \alpha = Y, \quad \beta = \omega \\ E(\tau'_i \tau'_j) = C^{\tau\tau}(\mathbf{x}_i, \mathbf{x}_j) & \text{if } \alpha = \tau, \quad \beta = \tau \\ E(\tau'_i \omega'_j) = C^{\tau\omega}(\mathbf{x}_i, \mathbf{x}_j) & \text{if } \alpha = \tau, \quad \beta = \omega \\ E(\omega'_i \omega'_j) = C^{\omega\omega}(\mathbf{x}_i, \mathbf{x}_j) & \text{if } \alpha = \omega, \quad \beta = \omega \end{cases} \tag{3.17}$$

The number of pumping and tracer test data available is usually limited and largely complicates the inference of the cross-covariance functions between the different random functions (Y', τ', ω'). The approximated analytical solution of S_{est} (Equation 2.3) and t_a (Equation 2.26) allow us to overcome this problem by expressing all the cross-covariance functions in terms of the known

covariance function of the natural log of transmissivity, $C^{YY}(x_i, x_j)$,

$$C^{y\omega}(\mathbf{x}_i, \mathbf{x}_j) = -\frac{1}{\pi}\int_{\Re^2} U(\mathbf{x};\mathbf{x}_j)C^{yy}(\mathbf{x}_i,\mathbf{x})\,d\mathbf{x} \tag{3.18}$$

$$C^{y\tau}(\mathbf{x}_i, \mathbf{x}_j) = -C^{y\omega}(\mathbf{x}_i, \mathbf{x}_j) - \int_0^{r_i}\frac{2r}{r_i^2}\Big(C^{yy}(\mathbf{x}_i, \mathbf{x}_j(r)) - C^{y\omega}(\mathbf{x}_i, \mathbf{x}_j(r))\Big)\,dr \tag{3.19}$$

$$C^{\omega\omega}(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{\pi^2}\int_{\Re^2}\int_{\Re^2} U(\mathbf{x}';\mathbf{x}_i)U(\mathbf{x}'';\mathbf{x}_j)C^{yy}(\mathbf{x}',\mathbf{x}'')\,d\mathbf{x}'\,d\mathbf{x}'' \tag{3.20}$$

$$\begin{aligned} C^{\tau\tau}(\mathbf{x}_i, \mathbf{x}_j) &= C^{\omega\omega}(\mathbf{x}_i, \mathbf{x}_j) + \int_0^{r_i}\frac{2r}{r_i^2}C^{y\omega}(\mathbf{x}_i(r), \mathbf{x}_j)dr - \int_0^{r_i}\frac{2r}{r_i^2}C^{\omega\omega}(\mathbf{x}_i(r), \mathbf{x}_j)\,dr \\ &+ \int_0^{r_j}\frac{2r}{r_j^2}C^{\omega y}(\mathbf{x}_i, \mathbf{x}_j(r))\,dr - \int_0^{r_j}\frac{2r}{r_j^2}C^{\omega\omega}(\mathbf{x}_i, \mathbf{x}_j(r))\,dr + \int_0^{r_i}\int_0^{r_j}\frac{4r'r''}{r_i^2 r_j^2}\Big(C^{yy}(\mathbf{x}_i(r'), \mathbf{x}_j(r'')) \\ &- C^{y\omega}(\mathbf{x}_i(r'), \mathbf{x}_j(r'')) - C^{\omega y}(\mathbf{x}_i(r'), \mathbf{x}_j(r'')) + C^{\omega\omega}(\mathbf{x}_i(r'), \mathbf{x}_j(r''))\Big)\,dr'dr'' \end{aligned} \tag{3.21}$$

$$\begin{aligned} C^{\tau\omega}(\mathbf{x}_i, \mathbf{x}_j) &= \frac{1}{\pi}\int_0^{r_i}\int_{\Re^2}\frac{2r}{r_i^2}U(\mathbf{x};\mathbf{x}_j)\Big(C^{yy}(\mathbf{x}_i(r), \mathbf{x}) - C^{\omega y}(\mathbf{x}_i(r), \mathbf{x})\Big)\,d\mathbf{x}\,dr \\ &+ \frac{1}{\pi}\int_{\Re^2} U(\mathbf{x};\mathbf{x}_j)C^{\omega y}(\mathbf{x}_i, \mathbf{x})\,d\mathbf{x} \end{aligned} \tag{3.22}$$

Ordinary Co-Kriging

Here, the covariance function of $Y(\mathbf{x})$ is known but the mean, m_Y, is uncertain. Let us consider the problem of getting the best unbias linear estimator of the random function $P(\mathbf{x})$ given observations of the random functions $Y(\mathbf{x})$ and $Q(\mathbf{x})$, defined as

$$P(\mathbf{x}) = -\int_0^{r_i} \frac{2r}{r_i^2} \left[Y_i(r) - Q_i(r) + Q_i(r_i) \right] dr \tag{3.23}$$

$$Q_i(r) = -\frac{1}{\pi} \int_{\Re^2} Y(\mathbf{x}) U(\mathbf{x}, \mathbf{x}_i(r)) d\mathbf{x} \tag{3.24}$$

Then, the kriging estimator of $P(\mathbf{x}_0)$, being $\mathbf{x}_0$ an arbitrary location, is expressed as

$$P_{OK}(\mathbf{x}_0) = \sum_i \lambda_i^Y Y_i + \sum_i \lambda_i^P P_i + \sum_i \lambda_i^Q Q_i \tag{3.25}$$

By the definition of $P(\mathbf{x})$ and $Q(\mathbf{x})$ we have

$$E(P) = m_P = -m_Y \qquad E(Q) = m_Q = -2m_Y \tag{3.26}$$

Defining the random variable

$$X_{\alpha i} = \begin{cases} Y_i & \text{if } \alpha = Y \\ P_i & \text{if } \alpha = P \\ Q_i & \text{if } \alpha = Q \end{cases} \tag{3.27}$$

the kriging estimator is written as

$$P_{OK}(\mathbf{x}_0) = \sum_{\alpha} \sum_{i} \lambda_i^{\alpha} X_{\alpha i} \tag{3.28}$$

the co-Kriging system is obtain requiring the unbias condition while minimizing the variance of the estimator error. The unbias constraint gives

$$\sum_{\alpha} m_{\alpha} \sum_{i} \left(\lambda_i^{\alpha}\right) = m_P \tag{3.29}$$

The expectation of the error involved is

$$\sigma_{OK}^2 = E\left[(P_{OK} - P)^2\right] = \sum_{\alpha} \sum_{\beta} \sum_{i} \sum_{j} \lambda_i^{\alpha} \lambda_j^{\beta} E(X_{\alpha i} X_{\beta j}) - 2 \sum_{\alpha} \sum_{i} \lambda_i^{\alpha} E(X_{\alpha i} P) + E(P^2) \tag{3.30}$$

Expressing this in terms of the deviations from the mean, $X'_{\alpha}(\mathbf{x}) = X_{\alpha}(\mathbf{x}) - m_{\alpha}$, and using (3.29), this is reduced to

$$\sigma_{OK}^2 = E\left[(P_{OK} - P)^2\right] = \sum_{\alpha} \sum_{\beta} \sum_{i} \sum_{j} \lambda_i^{\alpha} \lambda_j^{\beta} E(X'_{\alpha i} X'_{\beta j}) - 2 \sum_{\alpha} \sum_{i} \lambda_i^{\alpha} E(X'_{\alpha i} P'_0) + E(P'^2) \tag{3.31}$$

where

$$E(X'_{\alpha i}X'_{\beta j}) = \begin{cases} E(Y'_iY'_j) = C^{yy}(\mathbf{x}_i,\mathbf{x}_j) & \text{if } \alpha = Y, \quad \beta = Y \\ E(Y'_iP'_j) = C^{y\tau}(\mathbf{x}_i,\mathbf{x}_j) & \text{if } \alpha = Y, \quad \beta = P \\ E(Y'_iQ'_j) = C^{y\omega}(\mathbf{x}_i,\mathbf{x}_j) & \text{if } \alpha = Y, \quad \beta = Q \\ E(P'_iP'_j) = C^{\tau\tau}(\mathbf{x}_i,\mathbf{x}_j) & \text{if } \alpha = P, \quad \beta = P \\ E(P'_iQ'_j) = C^{\tau\omega}(\mathbf{x}_i,\mathbf{x}_j) & \text{if } \alpha = P, \quad \beta = Q \\ E(Q'_iQ'_j) = C^{\omega\omega}(\mathbf{x}_i,\mathbf{x}_j) & \text{if } \alpha = Q, \quad \beta = Q \end{cases} \tag{3.32}$$

The minimization of the variance of σ^2_{OK} under the unbias constraint (3.29) results in a linear system of equations involving one "Lagrangian" parameter, μ,

$$\sum_\beta \sum_j \lambda_j^\beta E(X'_{\alpha i}X'_{\beta j}) - \mu\, m_\alpha = E(X'_{\alpha i}\tau'_0) \qquad \forall \alpha = Y,\tau,\omega \quad \forall i = 1,...,n$$

$$\sum_\alpha \sum_i \lambda_i^\alpha m_\alpha = m_P \tag{3.33}$$

Being consistent with the fact that ordinary kriging system is analogous to re-estimate the mean, $\widehat{m}_Y$, within a search neighborhood moving with the estimation location, $\mathbf{x}_0$, as used in simple kriging. We can estimate $\tau'_{OK}(\mathbf{x}_0)$ after solving this system of linear equations simply as

$$\tau'_{OK}(\mathbf{x}_0) = P_{OK} + \widehat{m}_Y = \widehat{m}_Y + \sum_i \lambda_i^Y Y_i + \sum_i \lambda_i^P\,(\tau_i - E(\tau) - \widehat{m}_Y) + \sum_i \lambda_i^\omega\,(\omega_i - E(\omega) - 2\widehat{m}_Y) \tag{3.34}$$

where $\widehat{m}_Y$ is the local mean estimated as

$$\widehat{m}_Y \approx \frac{1}{N}\sum_{i}^{N} Y_i \tag{3.35}$$

N being the number of data within the search neighborhood, and we have used the relationship

$$Q(\mathbf{x}_i) = \omega'(\mathbf{x}_i) + 2m_Y = \omega(\mathbf{x}_i) - E(\omega) - 2m_Y \tag{3.36}$$

$$P(\mathbf{x}_i) = \tau'(\mathbf{x}_i) + m_Y = \tau(\mathbf{x}_i) - E(\tau) - m_Y \tag{3.37}$$

the expected value of τ and ω, can be approximated as

$$E(\tau) \cong \ln \frac{\pi(r_i^2 - r_w^2)b\phi}{Q_w} \tag{3.38}$$

$$E(\omega) \cong \ln S \tag{3.39}$$

The minimized estimation variance of $\tau'_{OK}(\mathbf{x}_0)$ can then be written as

$$\sigma^2_{OK} = E(\tau'^2) - \mu\, m_Y - \sum_{\alpha}\sum_{i} E(\tau'_{\alpha i} P'_0) \tag{3.40}$$

3.4.2 Conditional Stochastic Simulation

Here, we generate alternative, equally probable, realizations of τ' while honoring data values (Y, τ, ω) at different locations. Each realization is denoted with the superscript index m, $\tau^{(m)}$. For simplicity, we will only consider the sequential gaussian simulation approach. This approach

consists in drawing sequentially the value of a variable (τ) from its conditional pdf (probability density function). The attribute τ is consider to be a multiGaussian random variable and thereby its cdf follows a Gaussian distribution. The mean and variance of the distribution is derived from the previous cokriging system with the only consideration that the conditioning needs to include not only all available data values of the attributes (Y, τ, ω) but also the previously simulated τ values.

3.5 Numerical assessment of the method

In the previous chapter we have assessed the robustness of the analytical solution of $\phi_{est}/\phi = \exp(\tau')$ by comparing the analytical results with numerical simulations carried out in a given synthetic aquifer. In particular, we found that the relative contribution of the hydraulic response (S_{est}) to transport connectivity (τ') is minor and if neglected causes only a slight overestimation of low connectivity patterns. This means that the simplified analytical solution (Equation 2.27) gives a conservative estimate of travel times for risk assessment exercises. For this reason and to avoid the burden of a large number of space integrals involved in the Kriging linear system, in this section we have only evaluated the methodology when conditioning only to measurements of point transmissivity and tracer test data (i.e. mean arrival time).

3.5.1 Numerical approach

We assess the performance of our methodology using a synthetic aquifer where only a limited number of measurements of $T(\mathbf{x})$ and $t_a(\mathbf{x})$ is available. The scenarios considered are the following

- Scenario A: 16 equally spaced measurements of local transmissivity ($T(\mathbf{x})$) are known.
- Scenario B: 16 equally spaced measurements of local transmissivity ($T(\mathbf{x})$) and 16 measurements of travel time ($t_a(\mathbf{x})$) are known. Each measurement of travel time refers to a solute injected in one of the sixteen location where the transmissivity has been sampled.

- Scenario C: 100 equally spaced measurements of local transmissivity ($T(\mathbf{x})$).

We consider the same transmissivity field as used by *Sanchez-Vila et al.* [83] and reproduced in Figure 2.5a. The numerical setup of the flow and transport numerical simulations is explained in detail in section 2.4. We consider that groundwater flow is mainly driven by pumping from an abstraction well. Thus, we neglect those flow fluxes induced by a natural gradient. For each scenario we attempt to reproduce the real connectivity patterns using conditional estimation (i.e. Kriging) and carrying out 300 conditional simulations (i.e. sequential Gaussian simulations). Since we assume that the mean of the log-transmissivity field, m_Y, is unknown, in both approaches the estimation/simulations are carried out by solving the ordinary cokriging linear system (Equation 3.34 and 3.40). The covariance function of $Y(\mathbf{x})$ is modelled using an isotropic spherical semivariogram with unit variance and range of 20 m. The results of the conditional estimation/simulations of each scenario are then used to assess the performance of this geostatistical framework for the delineation of capture zones. Following the definition of capture zone provided in section 2.5, a given estimated/simulated point, i, belongs to the protected area if:

$$t_a^*(\mathbf{x}_i) = \frac{\pi r_i^2 b \phi}{Q_\omega} \exp(\tau'^*(\mathbf{x}_i)) \leq 50 day \tag{3.41}$$

where the asterisk refers to the estimated/simulated value of τ'.

3.5.2 Results and discussion

For illustrative purposes, in Figure 3.1 we compare the real connectivity patterns obtained numerically (see section 2.4 for details) with those obtained from an estimation and a single simulation for scenario C. It is worthwhile to note that in this specific scenario both the estimation and the given simulation are able to identify the three zones highly connected. The tendency of the kriging estimator to produce globally smooth surfaces is reflected in the results while the connectivity patterns obtained using sequential simulation are sharper.

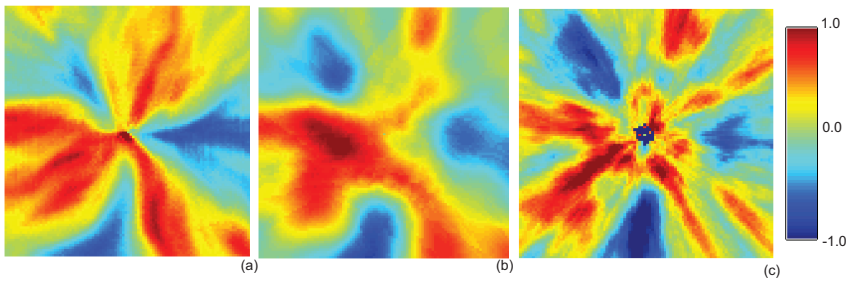


Figure 3.1: (a) Real connectivity patterns (τ') and connectivity patterns estimated/simulated using (b) ordinary cokriging and (c) sequential gaussian simulation (realization 200) for scenario C.

Figure 3.2 shows the isoprobability contour lines of the capture zone for the three scenarios, obtained from the set of 300 sequential gaussian simulations. It is interesting to note that in scenario A, where only a limited number (16) of T measurements is available, the mean shape of the capture zone tends to the circular capture zone given by an equivalent homogeneous T field (i.e. circular capture zone) (Figure 3.2 a). Also, the results show a high variance, with the 0.05 isoline covering a large part of the domain. When a few tracer test data are included (Scenario B) the distance between the isolines diminishes and the mean shape of the simulated capture zone approximates to the real shape (Figure 3.2 b). This is confirmed by Figure 3.3 that shows a comparison of the results for both scenarios. From this figure we can see that the incorporation of the tracer data constrains the simulated t_a (i.e. τ') to values that are closer to the real ones, thus reducing the uncertainty in the delineation of the capture zone. When the number of T measurements available increases by an order of magnitude (Scenario C) the accuracy of the method also increases dramatically (Figure 3.2 c) and a good agreement is observed between the mean shape and the real perimeter of the capture zone. Nevertheless those parts of the protection perimeter that correspond to highly connected zones still show a high uncertainty, probably because of the the lack of tracer test data in Scenario C. As expected, the protection perimeter delineated using the kriging estimation is close to the mean perimeter (isoline 0.5) identified using the 300 conditional simulations.

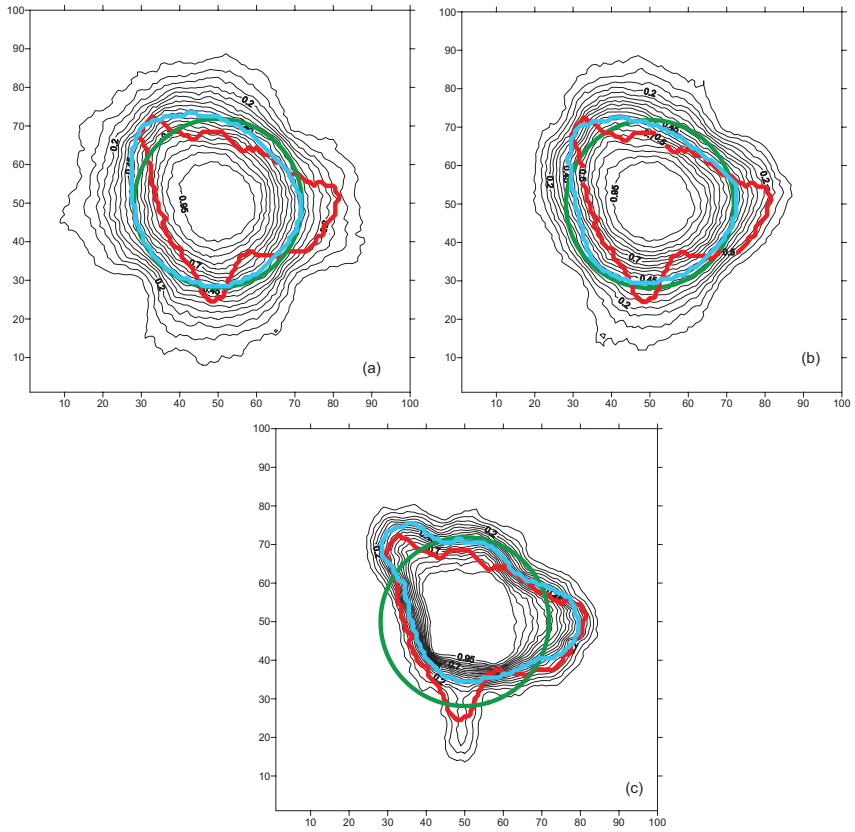


Figure 3.2: Capture zone probability maps. The isoprobability contours are computed using the 300 gaussian sequential simulations for (a) scenario A, (b) scenario B and (c) scenario C. The capture zone estimated using ordinary cokriging (blue line), the real shape of the capture zone (red line) and the capture zone associated to the equivalent homogeneous field ($T = T_G = 1m^2/day$) (green line) are also shown.

To assess the performance of this geostatistical framework for the delineation of capture zones, we define two error norms (i.e. performance indexes) that compare the forecasted protection area, C_{fore}, with the real capture zone, C_{real}. The first index refers to the relative size of the protection area that is not identified by the estimation/simulation, defined as

$$e_{miss} = \frac{A_{miss}}{A_{miss} + A_{fore}} \tag{3.42}$$

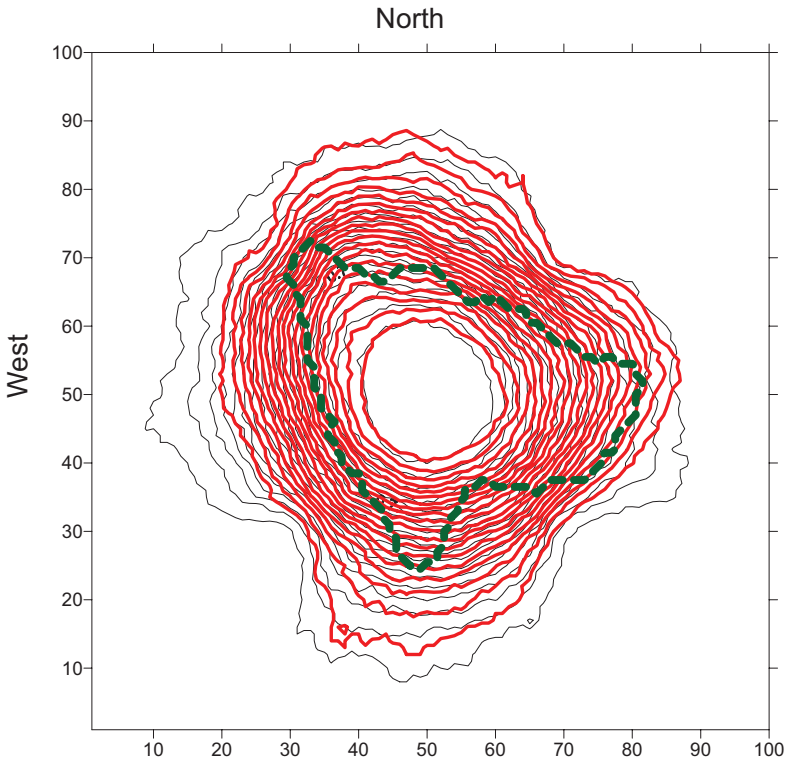


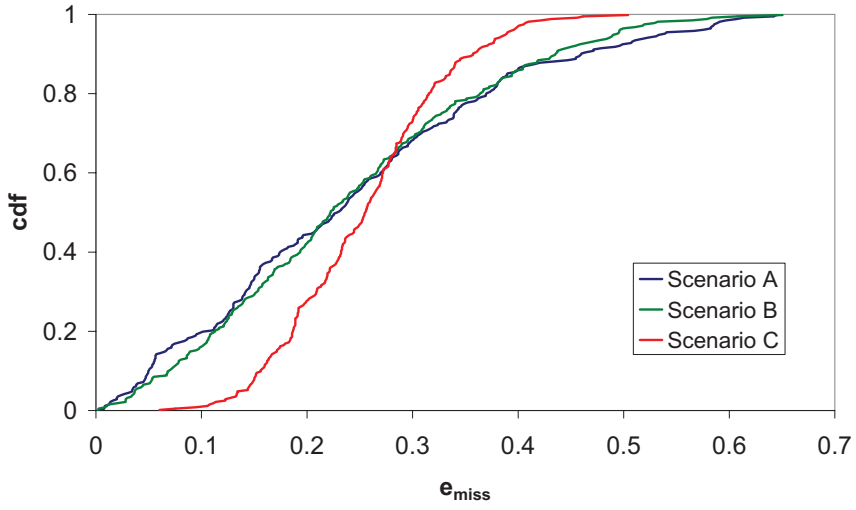
Figure 3.3: Capture zone probability maps for scenario A (gray lines) and scenario B (red lines). The real shape of the capture zone (green line) is also shown.

and the second is related to the relative size of the forecasted area that is unnecessarily protected,

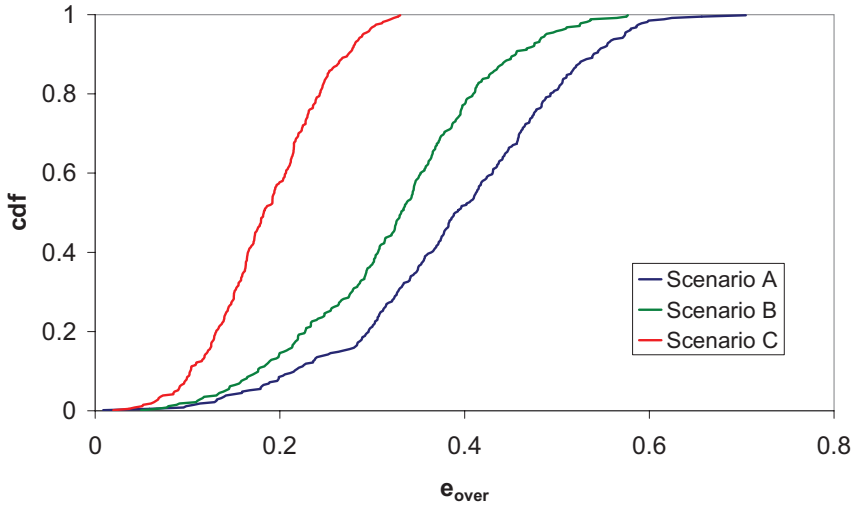
$$e_{over} = \frac{A_{over}}{A_{over} + A_{fore}} \tag{3.43}$$

where A_{miss} $[m^2]$ is the area of the real capture zone that is not identified by the forecast, i.e., the area of $C_{real} \backslash C_{fore}$, A_{fore} $[m^2]$ is the area of the real capture zone that is correctly predicted, i.e., $C_{real} \cap C_{fore}$, and A_{over} $[m^2]$ is the size of the forecasted area that is unnecessarily protected, i.e., $C_{fore} \backslash C_{real}$. In other words, e_{miss} is the percentage of the real capture zone that has not been

identified, whereas e_{over} is the percentage of the forecast that is unnecessarily protected.



(a)



(b)

Figure 3.4: Conditional cumulative density function, ccdf, of (a) e_{miss} (Equation 3.42) and (b) e_{over} (Equation 3.43) for the set of 300 conditional simulations of the three scenarios.

Figure 3.4 (a) and (b) show the conditional cumulative density function, (ccdf, of e_{miss} and e_{over} for the set of 300 conditional simulations of each scenario. From figure 3.4 (a) we see that scenarios A and B display a large number of simulations where the percentage of the real capture zone not identified by the forecast is relatively small. This behavior is visually appreciated in figure 3.4 (b) where we see that both scenarios, in particular scenario A, have a large number of simulations where the forecast largely overestimates the real size of the protection area. In other words, when a limited number of measurements is available, the protected area is highly uncertain and thus many forecasted capture zones largely overestimates the real shape of the protected area. The incorporation of tracer test data improves the forecasts by reducing their uncertainty and thus reducing the percentage of the forecast that is unnecessarily protected. This is even more evident in Scenario C, altough it must be noticed that in this case the improvement of e_{miss} is limited. The explanation lies in the intrinsic bias of the simplified analytical solution (Equation 2.27) already shown in section 2.5.

3.6 Summary and Conclusions

We have developed a geostatistical framework to delineate transport connectivity patterns honoring a limited number of data measured over different scales, namely mean travel times of convergent tracer tests, t_a, estimates of the storage coefficient from pumping tests interpreted using the Cooper-Jacob method, S_{est}, and measurements of transmissivity point values, T. Although the methodology allows to condition the results to the three above-mentioned variables, so far we have only assessed its behavior when conditioning to travel time, t_a, and transmissivity point values, T. We decided to neglect the information provided by S_{est} to avoid the burden of a large number of space integrals involved in the cokriging linear system. This assumption has been extensively assessed in section 2.4 where we showed that the simple analytical solution (Equation 2.27), which does not account for S_{est}, tends to overestimate the low connectivity patterns. In other words, this assumption is conservative from a risk assessment point of view.

Our work leads to the following key conclusions:

1. The limitation of a standard cokriging approach in this kind of application lies in the limited number of measurements usually available. This makes it difficult to infer the covariance function between the variables involved in the problem. Our approach allows to overcome this limitation through the incorporation of the analytical solution of t_a (Equation 2.26) and S_{est} (Equation 2.3).

2. Different scenarios, with different number of T and t_a measurements available, have been assessed. A simple visual inspection of the cokriging estimation and of one conditional simulation for scenario C shows the ability of both methodologies to identify the three zone highly connected. The connectivity patterns delineated using cokriging are smooth due to the well known tendency of this estimator to generate globally smooth surfaces.

3. The ability of the methodology to properly address a standard risk assessment exercise (i.e. delineation of the 50 days travel time capture zone) has been evaluated. When only a limited number of T measurements is available, the isoprobability contour lines of the capture zone, calculated from the 300 conditional simulations, show a high uncertainty with the mean shape of the capture zone that approximates to that of an equivalent homogeneous medium (i.e. circular capture zone). The incorporation of travel time data reduces the distance between the isolines and constrains the simulated results to values of t_a (i.e. τ') closer to the real ones, thus reducing the uncertainty associated to the simulations. When a large number of T measurements is available, the uncertainty in the delineation of the capture zone decreases dramatically and the mean shape of the protection perimeter shows a good agreement with the real one.

4. The performance of each single simulation for each scenario has been evaluated by defining two error norms. The first, e_{miss}, is related to the percentage of the real capture zone that has not been identified and the second, e_{over}, is a function of the percentage of the forecast that

is unnecessarily protected. The conditional cumulative density functions, (ccdf), of e_{miss} and e_{over} show that when a limited number of T measurements is available, the simulated capture zones tend to overestimate the real protection area due to the high uncertainty in the results. The incorporation of tracer test data constrains the simulations to values closer to the real ones, thus reducing the number of simulations where the capture zone is largely overestimated. The reduction of the overestimation is even more evident when a large number of T measurements is available, altough the improvement of e_{miss} is limited due to the intrinsic bias of the simplified analytical solution.