

Simultaneous identification of the pollutant release history and the source location in groundwater by means of a geostatistical approach

Ilaria Butera · Maria Giovanna Tanda ·
Andrea Zanini

Published online: 13 November 2012
© Springer-Verlag Berlin Heidelberg 2012

Abstract This paper describes an innovative procedure that is able to simultaneously identify the release history and the source location of a pollutant injection in a groundwater aquifer (simultaneous release function and source location identification, SRSI). The methodology follows a geostatistical approach: it develops starting from a data set and a reliable numerical flow and transport model of the aquifer. Observations can be concentration data detected at a given time in multiple locations or a time series of concentration measurements collected at multiple locations. The methodology requires a preliminary delineation of a probably source area and results in the identification of both the sub-area where the pollutant injection has most likely originated, and in the contaminant release history. Some weak hypotheses have to be defined about the statistical structure of the unknown release function such as the probability density function and correlation structure. Three case studies are discussed concerning two-dimensional, confined aquifers with strongly non-uniform flow fields. A transfer function approach has been adopted for the numerical definition of the sensitivity matrix and the recent step input function procedure has been successfully applied.

Keywords Geostatistical approach · Transfer function · Source detection · Non uniform flow · Groundwater

I. Butera
DIATI, Politecnico di Torino, Corso Duca degli Abruzzi 24,
10129 Turin, Italy

M. G. Tanda · A. Zanini (✉)
DICATeA, Università degli Studi di Parma, Parco Area delle
Scienze, 181/A, 43124 Parma, Italy
e-mail: andrea.zanini@unipr.it

List of symbols

$C(\mathbf{x}, t)$	Concentration at point \mathbf{x} and time t
t	Time
τ	Time
\mathbf{x}	Position in the domain
\mathbf{x}_0	Source location
\mathbf{u}	Velocity tensor
\mathbf{D}	Dispersion tensor
∇	Nabla operator
$F(t)$	Concentration of the water injected at the source as function of time t
F_0	Constant and known mass rate input function
$f(\mathbf{x}, t)$	Transfer function at position \mathbf{x} and time t
m	Number of observations
n	Number of unknowns
\mathbf{z}	$m \times 1$ Observations
\mathbf{s}	$n \times 1$ Unknowns
$s(t)$	Unknown release function
p	Number of unknown coefficients
$\mathbf{h}(\mathbf{s})$	$m \times 1$ Vector that describes the transport process
\mathbf{v}	$m \times 1$ Measurement errors
\mathbf{R}	$m \times m$ Error covariance matrix
\mathbf{H}	$m \times n$ Sensitivity matrix
T	Sampling time
\mathbf{X}	$n \times p$ Matrix, mean of the unknown process
\mathbf{b}	$p \times 1$ Unknown coefficients
$\mathbf{Q}(\boldsymbol{\theta})$	$n \times n$ Matrix, covariance of the unknown process
$\boldsymbol{\theta}$	Structural parameters of the covariance function
σ_s^2	Variance of the unknown release function $s(t)$
λ_s	Correlation time length of the unknown release function $s(t)$
$\boldsymbol{\Sigma}$	$m \times m$ Dummy matrix
\mathbf{E}	$m \times m$ Dummy matrix

σ_R^2	Variance of the measurement error
$\hat{\mathbf{s}}_{n \times 1}$	Vector of estimated release function
\mathbf{M}	$p \times n$ Multipliers
$\mathbf{\Lambda}$	$n \times m$ Kriging coefficients
\mathbf{V}	$n \times n$ Matrix, covariance of the estimate of the errors
\tilde{s}	Transformed unknown function
α	Positive number
h_D	Head level downstream
h_U	Head level upstream
α_L	Longitudinal dispersivity
α_T	Transversal dispersivity
Q_{in}	Injected flow rate
K	Hydraulic conductivity
Y	Logarithm of the hydraulic conductivity
σ_Y^2	Variance of the log-conductivity field
Z	Normalized log-conductivity
μ_Y	Mean of the log-conductivity field

1 Introduction

General interest in environmental issues has drawn attention to the quality of groundwater resources. Scientific efforts in groundwater flow studies have primarily focused on the flow and transport behavior and on the identification of related parameters. During the final decades of the twentieth century, increasing work has been focused on the problem of source identification and recovering pollutant release histories. The identification of a source location is of paramount importance to identify responsible parties for pollutant releases. The knowledge of the pollutant injection function gives information about the future pollution spread and allows a better planning of remediation actions (Liu and Ball 1999; Snodgrass and Kitanidis 1997; Skaggs and Kabala 1994; Butera and Tanda 2003). Moreover, from a legal and regulatory point of view, it is important to determine the release time, the duration, and the maximum value of the released solute concentration. A reliable release history can also be a useful tool for apportioning remediation costs of a polluted area among the responsible parties.

A geostatistical approach to recovering the release history of a pollutant in groundwater has been applied over the past 30 years. Snodgrass and Kitanidis (1997) proposed this procedure to recover the release history in one-dimensional uniform flow. Several improvements and applications of the geostatistical methodology were proposed (Michalak and Kitanidis 2002, 2003, 2004a, b), also by Butera and Tanda (2003) and Butera et al. (2006).

Due to the linearity of the governing differential equation of the transport problem, Snodgrass and Kitanidis (1997)

developed a geostatistical approach using a transfer function (TF), or Kernel function, (Jury and Roth 1990) that describes the effect in time, at a certain location of the aquifer, of an impulse release of a pollutant at its source. The knowledge of the TF plays an important role in the solution of the direct and inverse transport problems. In the geostatistical procedure, but also using Tikhonov regularization (Skaggs and Kabala 1994), non-regularized nonlinear least squares (Alapati and Kabala 2000) and minimum relative entropy (Woodbury and Ulrych 1996, 1998), TFs are used to concisely describe the response of the aquifer to an impulse excitation. Luo et al. (2006) applied a TF approach to the analysis of a tracer test. The authors showed that the gamma distribution, in their study case, is an efficient parametric model that can take into account a highly non-uniform flow field. This parametric model, interpreted as probability density function (pdf) of the travel times, can be used as prediction of non-linear transport. Fienen et al. (2006) applied a geostatistical TF procedure to identify TFs in the interpretation of a tracer field test.

TFs can be analytically determined if the problem has a simple flow field, but in many practical applications, the characteristics of the groundwater flow field do not allow an analytical formulation of the TFs. This is also the case in non-uniform flows due to complicated boundary conditions, the existence of pumping wells, or high levels of aquifer heterogeneity. To manage these cases through analytical formulas, the technician must resort to an oversimplification of the real problem. As a consequence, a rough approximation in the results can be expected. To overcome this difficulty, a numerical procedure to compute TFs was developed by Butera et al. (2004) and applied to homogeneous and weakly heterogeneous aquifers (Butera et al. 2006).

The previous studies, except for Michalak and Kitanidis (2004b), consider the location of the pollutant source to be known. In the present paper, an innovative procedure is proposed to obtain the simultaneous identification of the release history and the source identification in 2-D non-uniform flow fields. The procedure requires a preliminary delineation of an area where the pollutant source is most likely to be present.

The manuscript is organized in three parts: first the state of the practice is presented then the developments of the geostatistical approach necessary to address our topics are discussed. Finally, three case studies are presented as examples of application of the proposed procedures.

2 State of the practice

In the last 30 years several methods have been developed to identify the release time history of the pollutants and the

location of the contaminant source. A detailed review of these methodologies can be found in the work of Neupauer et al. (2000), Morrison (2000), Atmadja and Bagtzoglou (2001), Michalak and Kitanidis (2004b) and Dokou and Pinder (2009).

The present paper develops a new application of the geostatistical procedure proposed by Snodgrass and Kitanidis (1997) to recover the release history of a pollutant injection; such a procedure presents great capabilities for application to real cases because it is statistically based and useful for a large range of technical topics. Butera and Tanda analyzed the performance of the procedure in 2-D flow field including uncertainty in the observations and in the model parameters (Butera and Tanda 2001, 2003), of areal and multiple sources (Butera and Tanda 2002, 2003) and of heterogeneous but uniform (in the average) flow (Butera and Tanda 2004).

Michalak and Kitanidis (2002, 2004a) applied the geostatistical procedure to real 1-D cases and developed, again in the geostatistical framework, a methodology able to recover the antecedent conditions (position and concentration) of a plume monitored at a certain time in both 2-D homogeneous and heterogeneous case (Michalak and Kitanidis 2004b). The paper shows a method able to also detect the pollutant source location by going backward in time, step-by-step. Michalak and Kitanidis (2004b) used the adjoint state method (Neupauer and Wilson 2001) to obtain the elements of the sensitivity matrix (TF) needed for the inversion process. Milnes and Perrochet (2007), exploiting the properties of the forward and backward location pdf in the linear transport process, proposed a procedure able to recover, backward in time, the plume position and the timing of a pollutant release. The indicated methodology requires the transport simulation on a reversed flow field.

Butera et al. (2004) tackled the topic of numerical identification of the TF in a heterogeneous aquifer; the method derived for 1-D flow fields has been tested in Butera et al. (2006) in homogeneous 2-D fields, comparing the performance of the TFs obtained through the numerical method with that of the TFs obtained analytically. The technique exploits the linearity of the transport process and the associated TF theory (Jury and Roth 1990) and does not require the inversion of the flow field.

It is worthwhile to note that the identification of the TF is useful not only for the application of the geostatistical approach but also for the use of Tikhonov regularization (Skaggs and Kabala 1994), non-regularized nonlinear least squares (Alapati and Kabala 2000) and minimum relative entropy (Woodbury and Urych 1996, 1998). The development of efficient strategies to identify the elements of the sensitivity matrix (or the TF in linear problems) remains a matter of great importance.

3 Mathematical statements

3.1 Groundwater transport

The following Eq. (1) describes the transport process in an aquifer with uniform porosity reacting to the injection of a non-sorbing, non-reactive solute in a point source:

$$\frac{\partial C(\mathbf{x}, t)}{\partial t} = -\nabla[C(\mathbf{x}, t) \mathbf{u}(\mathbf{x}, t)] + \nabla[\mathbf{D}(\mathbf{x}) \cdot \nabla C(\mathbf{x}, t)] + s(t) \delta(\mathbf{x} - \mathbf{x}_0). \tag{1}$$

where \mathbf{x} is a vector describing the point location, $\mathbf{u}(\mathbf{x}, t)$ is the effective velocity, $\mathbf{D}(\mathbf{x})$ the dispersion tensor, $C(\mathbf{x}, t)$ the concentration at location \mathbf{x} and time t , ∇ is the differential operator Nabla and $s(t)$ is the amount of conservative pollutant per unit time injected into the aquifer through the source located at \mathbf{x}_0 .

Equation (1) is a linear differential equation whose solution, when associated with the initial and boundary conditions: $C(\mathbf{x}, 0) = 0$; $C(\infty, t) = 0$, is given by the following integral (Jury and Roth 1990):

$$C(\mathbf{x}, t) = \int_0^t s(\tau) f(\mathbf{x}, t - \tau) d\tau, \tag{2}$$

where $f(\mathbf{x}, t - \tau)$ is the TF, or Kernel function, that describes the effects at \mathbf{x} at time t by an impulse injection occurring at \mathbf{x}_0 at time τ .

3.2 Geostatistical approach

The quasilinear geostatistical methodology used as an inverse procedure is briefly explained in the following. For more details see Kitanidis (1995, 1996) and Snodgrass and Kitanidis (1997).

It is assumed that a set \mathbf{z} of concentration data is available at the monitoring time T , originated from an unknown release process $s(t)$. The observed concentration data can be expressed as function of the release process by the following equation:

$$\mathbf{z} = \mathbf{h}(\mathbf{s}) + \mathbf{v}, \tag{3}$$

where \mathbf{z} is a $m \times 1$ vector of observations, $\mathbf{h}(\mathbf{s})$ is the $n \times 1$ vector containing the time discretization of the unknown release function $s(t)$ and \mathbf{v} is a $m \times 1$ vector of epistemic errors with zero mean and known covariance matrix $\mathbf{R} = \sigma_R^2 \cdot \mathbf{I}$. Even if it can be estimated by the geostatistical procedure (Fienen et al. 2009), in this work it is assigned constant and we will tackle the topic in a future paper. For the case of a conservative solute, the relationship between the observed concentration and the release is linear (see Eqs. (2) and (3)) can be simplified to (Snodgrass and Kitanidis 1997):

$$\mathbf{z} = \mathbf{H} \cdot \mathbf{s} + \mathbf{v}. \quad (4)$$

Equation (4) represents the matrix form of Eq. (2), where the matrix \mathbf{H} contains the values of the TF (f), computed at appropriate times and locations:

$$\mathbf{H} = \Delta t \begin{bmatrix} f(\mathbf{x}_1, T - \Delta t) & \dots & f(\mathbf{x}_1, T - n\Delta t) \\ f(\mathbf{x}_2, T - \Delta t) & \dots & f(\mathbf{x}_2, T - n\Delta t) \\ \dots & \dots & \dots \\ f(\mathbf{x}_m, T - \Delta t) & \dots & f(\mathbf{x}_m, T - n\Delta t) \end{bmatrix}. \quad (5)$$

The transfer matrix \mathbf{H} includes all the characteristics of the flow and transport process and, due to the linearity of the governing differential equation, is also the sensitivity matrix, i.e. the element H_{ij} represents how the observation z_i changes as the release value s_j varies.

Taking into account the unavoidable uncertainties in the processes controlling the pollutant release, \mathbf{s} can be considered random with characteristics of autocorrelation, due to all phenomena (leachate, infiltration, leakage from tanks, etc.) that take place between the pollutant discharge and its arrival into the groundwater system. For this reason, the random vector \mathbf{s} can be characterized by an unknown mean and a covariance function. The mean can be described as $E[\mathbf{s}] = \mathbf{X}\mathbf{b}$ where $E[\cdot]$ denotes the expected value, \mathbf{X} is a $n \times p$ matrix of known functions and \mathbf{b} is a vector of size $p \times 1$ that contains the unknown drift coefficients. In this work, a constant but unknown mean is considered; thus \mathbf{X} is an $n \times 1$ vector filled by 1 and \mathbf{b} is the scalar unknown mean of the function. The covariance among the elements of \mathbf{s} is assumed to be a Gaussian function, represented by the matrix $\mathbf{Q}(\boldsymbol{\theta}) = E[(\mathbf{s} - \mathbf{X}\mathbf{b})(\mathbf{s} - \mathbf{X}\mathbf{b})^T]$, where $\boldsymbol{\theta}$ are the unknown structural parameters, i.e. the variance σ_s^2 and the correlation time length λ_s .

The estimation procedure proposed by Kitanidis (1995) is divided into two parts: first the structural parameters $\boldsymbol{\theta}$ of the selected covariance function are determined, then the unknown release function is estimated by means of a Kriging process.

The identification of the structural parameters follows a restricted maximum likelihood approach. The probability that the random process with parameter $\boldsymbol{\theta}$ reproduces the observation \mathbf{z} can be estimated through the following:

$$p(\mathbf{z}|\boldsymbol{\theta}) \propto |\Sigma|^{-1/2} |\mathbf{X}^T \mathbf{H}^T \Sigma^{-1} \mathbf{H} \mathbf{X}|^{-1/2} \exp \left[-\frac{1}{2} \mathbf{z}^T \Xi^{-1} \mathbf{z} \right], \quad (6)$$

where $\Sigma = \mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R}$ and $\Xi = \Sigma^{-1} - \Sigma^{-1} \mathbf{H} \mathbf{X} (\mathbf{X}^T \mathbf{H}^T \Sigma^{-1} \mathbf{H} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{H}^T \Sigma^{-1}$.

A good estimation of $\boldsymbol{\theta}$ is the one that maximizes the probability $p(\mathbf{z}|\boldsymbol{\theta})$: maximizing Eq. (6) is equivalent to minimizing the negative logarithm of $p(\mathbf{z}|\boldsymbol{\theta})$, resulting in the objective function $L(\boldsymbol{\theta})$:

$$L(\boldsymbol{\theta}) = \frac{1}{2} \ln |\Sigma| + \frac{1}{2} \ln |\mathbf{X}^T \mathbf{H}^T \Sigma^{-1} \mathbf{H} \mathbf{X}| + \frac{1}{2} \mathbf{z}^T \Xi^{-1} \mathbf{z}. \quad (7)$$

The minimization of $L(\boldsymbol{\theta})$ is achieved by setting the derivatives of $L(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ to zero.

Once the structural parameters are computed, the estimation $\hat{\mathbf{s}}$ of the release function $s(t)$ is obtained through Kriging:

$$\hat{\mathbf{s}} = \mathbf{\Lambda} \cdot \mathbf{z}, \quad (8)$$

where the matrix $\mathbf{\Lambda}$ ($n \times m$) of the Kriging weights is calculated by solving the following system obtained from the un-biasedness and minimum variance conditions:

$$\begin{bmatrix} \Sigma & \mathbf{H}\mathbf{X} \\ (\mathbf{H}\mathbf{X})^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^T \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{H}\mathbf{Q} \\ \mathbf{X}^T \end{bmatrix}. \quad (9)$$

In Eq. (9), \mathbf{M} ($p \times m$) is a matrix of Lagrange multipliers. The covariance matrix of the estimation error is:

$$\mathbf{V} = -\mathbf{X}\mathbf{M} + \mathbf{Q} - \mathbf{Q}\mathbf{H}^T \mathbf{\Lambda}^T. \quad (10)$$

This methodology is functional and efficient but does not enforce non-negativity of the estimated concentration (unconstrained case). Box and Cox (1964), with the aim at avoiding this problem, suggest the use of a power transformation of the unknown variable \mathbf{s} . Following Kitanidis and Shen (1996) and Snodgrass and Kitanidis (1997) the new unknown function becomes:

$$\tilde{\mathbf{s}} = \alpha (\mathbf{s}^{1/\alpha} - 1), \quad (11)$$

where α is a positive number and it is chosen as small as possible while ensuring $\tilde{\mathbf{s}} > -\alpha$. For $\tilde{\mathbf{s}} < -\alpha$ imaginary results from Eq. (11) are possible. This transformation is general: it includes the unmodified \mathbf{s} ($\alpha = 1$), the square root of \mathbf{s} ($\alpha = 2$), and in the limit, for large value of α , it reduces to a logarithm transformation. Small values of α cause minor transformations, which helps the method to converge quickly. Alternative methods to enforce the non-negativity of concentration values are presented in Michalak and Kitanidis (2003, 2004c).

When the values of \mathbf{s} are constrained to be positive (constrained case) and they are physically compatible, Eq. (4) becomes:

$$\mathbf{z} = \mathbf{H} \cdot \mathbf{s} + \mathbf{v} = \mathbf{H} \cdot \left(\frac{\tilde{\mathbf{s}} + \alpha}{\alpha} \right) + \mathbf{v} = \mathbf{h}(\tilde{\mathbf{s}}) + \mathbf{v}. \quad (12)$$

In this case, $\mathbf{h}(\tilde{\mathbf{s}})$ is not linear with respect to the new unknown $\tilde{\mathbf{s}}$ and the solution is reached iteratively (for details see Kitanidis 1995, 1996).

The quality of the recovering process depends on the number (n) of the time interval chosen to discretize $s(t)$, as well as on the number, location and sampling times of the m measurements (\mathbf{z}). Without information on the release

history it is impossible to select the best data set of measured concentrations and the optimal discretization of s that make a good compromise between computational effort and the accuracy of the release recovery. If, for example, the s_i data are too close to each other, some rows of the \mathbf{H} matrix are almost identical, \mathbf{H} and the $\mathbf{\Sigma}$ matrix become ill conditioned, and the inversion of $\mathbf{\Sigma}$ could be more difficult. Similarly, if two concentration data z_i are too close (in space or in time), the information they provide is almost the same and the two corresponding rows of the \mathbf{H} matrix become roughly identical.

The quality of the z_i observations influences the estimation error covariance \mathbf{V} , which should be as small as possible for the recovered release history to be considered reliable. Analysis of Eq. (10) shows that \mathbf{V} depends on the covariance of the process $s(t)$ and the relationship between z_i and s_j . An H_{ij} element value of zero can be an important piece of information because it states that there is no link between z_i and s_j . In other words, the release at time t_j does not have effect on z_i . On the other hand, if z_i is zero but H_{ij} is positive, it follows that s_j should be zero.

It is also possible to apply the procedure starting from the data monitored at a few positions but at different times (Butera and Tanda 2004). This is a much more common situation in real cases since periodic monitoring actions at specific locations are usually carried out by public agencies and private companies. It is then possible to also use historical data increasing the observation database and, as a consequence, the reliability of the numerical process.

3.3 Numerical TF computation

In very simple flow conditions TFs can be determined analytically, but in a non-uniform flow field it is necessary to employ numerical strategies.

The stepwise input function (SIF) procedure methodology developed by Butera et al. (2004) is a numerical strategy for TF calculation.

Through a simple variable transformation it is possible to rewrite Eq. (2) as:

$$C(\mathbf{x}, t) = \int_0^t s(t - \tau) \cdot f(\mathbf{x}, t) d\tau. \tag{13}$$

If we assume a stepwise input function $s = F_0 \cdot H(t)$, where $H(t)$ is the Heaviside function, the integral (13) becomes:

$$C(\mathbf{x}, t) = \int_0^t F_0 \cdot H(t - \tau) \cdot f(\mathbf{x}, t) d\tau = F_0 \int_0^t f(\mathbf{x}, t) d\tau. \tag{14}$$

Taking the time derivative of Eq. (14), it results in

$$f(\mathbf{x}, t) = \frac{1}{F_0} \frac{\partial C(\mathbf{x}, t)}{\partial t}. \tag{15}$$

Equation (15) shows that it is possible to compute the TFs at a certain location by processing the concentration history (breakthrough curve) at this location resulting from a stepwise tracer injection.

The application of Eq. (15) in field conditions is rarely possible: in fact, the response of an aquifer due to an infinite step injection at a certain point can develop in a very long time. Nevertheless, a numerical model of flow and transport in the aquifer of interest is often available because it is suitable for the prediction of pollutant fate or the planning of remediation. If a numerical code is available it can be used as a surrogate for a field test. A stepwise injection of solute at known concentration C_0 can be modeled at the source location and at each node of the computation mesh the concentration time series (breakthrough curve) is determined. The normalized time derivative of the breakthrough curve is the TF for that mesh point and the modeled source location.

The choice of using a SIF and then numerically deriving the breakthrough curve rather than directly computing the aquifer response due to an impulse injection, can be justified by the need to reduce numerical errors caused by the discrete time representation of the process. Such errors are unavoidable in numerical computations. In fact, the numerical description of an impulse necessarily has to develop over a small but finite time period and the resulting TFs are incorrect because they are based on a finite injection rather than an impulse. The responses of areas far from the source and less exposed to the contamination can also be underestimated. The study of the transport of a step input of infinite time length is more accurate and not affected by such errors.

At a given measurement point, the breakthrough curve is an output of the numerical model and consists of a series of concentration values recorded in time: the shorter the temporal interval between two computations, the more accurate the computation of the TF by numerical derivative of the breakthrough curve is. By definition, the derivative can be calculated as the rate $\Delta C/\Delta t$, with $\Delta t \rightarrow 0$. Finally, the TF at each observation point is obtained after processing the results obtained through just one run of the forward transport model for each source location.

If there are errors in the processed concentrations (resulting from the numerical model or eventually from the field test), the TF could be poorly determined. Great care should be devoted to numerical model implementation, otherwise the TF can be improved by resorting to parametric functions as Luo et al. (2006) showed.

In the geostatistical approach the TF is needed for the computation of the element of the sensitivity matrix \mathbf{H} . The

elements of \mathbf{H} can be calculated with a perturbation method or using an adjoint state method (Michalak and Kitanidis 2004b). In the perturbation method, it is necessary to run the forward transport model at least as many times as the number of the parameters (elements of the vector \mathbf{s}) plus one ($n + 1$ times). Using the adjoint method, it is necessary to run the transport model on a reversed flow field as many times as there are observations (m times). The SIF procedure presented in Butera et al. (2004) is advantageous, in the case of a source with known location, since it allows the computation of the sensitivity matrix after one run of the forward transport numerical model. The SIF method can be considered as a limiting case of a perturbation approach enabled by the linearity of the transport process: the concentration field due to the Heaviside SIF is a perturbation of the initial zero concentration field.

3.4 Identification of the pollutant source and the release history

Butera and Tanda (2003) showed that the geostatistical approach proposed by Snodgrass and Kitanidis (1997) can be used to identify the true location between two possible source positions. In the present work, we extend this procedure to identify the most probable source location within a suspect area (in short SA). We consider the case that m concentration data in an aquifer are available, but neither the source location nor the release that creates them are known. Hence, the aim of the procedure is twofold: to identify the source location and recover the release history. In the following we will refer to the new technique as the SRSI procedure (SRSI—simultaneous release function and source location identification).

The procedure starts from a hypothesis of the possible source location. An SA for the source location can often be identified from analysis of available information. The SA must be discretized in a finite number (J) of sub-areas, each of which is considered as originating an unknown pollutant release, independent from the others within the SA. The pollutant injection is assumed given by a point source located in the centroid of each sub-area.

Due to the linearity of the advection–dispersion equation, the concentration values in the aquifer can be computed using the superposition method with the following expression:

$$C(\mathbf{x}, t) = \sum_{j=1}^J \int_0^t s_j(\tau) \cdot f_j(\mathbf{x}, t - \tau) d\tau, \quad (16)$$

where j is one of J generic sub-areas inside the SA.

The geostatistical procedure can be now applied since Eq. (4) is still valid. The vector \mathbf{s} of the unknown release

function in (4) is made up by the collection of J sub-vectors s_j , each with dimensions $n_i \times 1$, where n_i is the number of time values used to discretize the release history. The total dimension of \mathbf{s} is: $(n_1 + n_2 + \dots + n_J) \times 1$:

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \dots \\ \mathbf{s}_J \end{bmatrix}. \quad (17)$$

The transfer matrix \mathbf{H} is a block matrix

$$\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2 \ \dots \ \mathbf{H}_J] \quad (18)$$

whose dimensions are $m \times (n_1 + n_2 + \dots + n_J)$. The generic matrix \mathbf{H}_j describes the effects of the pollutant release in the sub-area j on the measured concentration data in the m monitoring point.

The covariance matrix \mathbf{Q} of the \mathbf{s} process, due to the lack of correlation among the release histories, is a block matrix with non-zero elements only in the diagonal blocks:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & 0 & 0 & 0 \\ 0 & \mathbf{Q}_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \mathbf{Q}_J \end{bmatrix}. \quad (19)$$

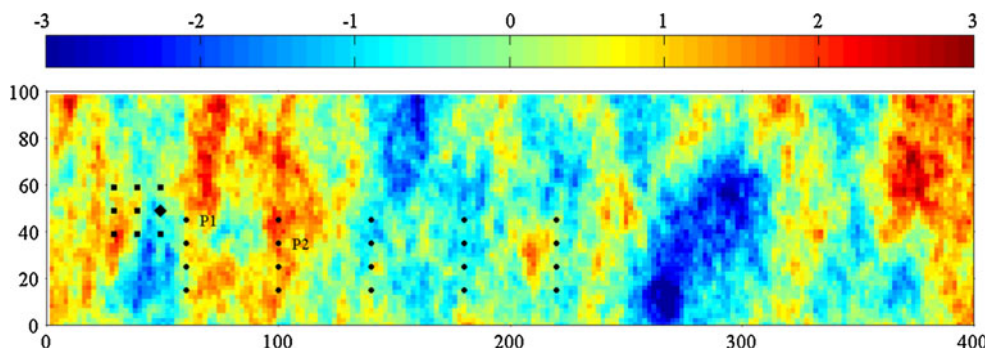
The results of the geostatistical procedure described in Sect. 3 provide the pollutant history in the J hypothetical source locations. The release function in the real source will be substantial, while in the other suspect locations the time histories will be negligible.

Summarizing the above considerations, the source identification procedure, here proposed, can be described with the consecutive steps:

- collect a set of concentration measurements at the same time in multiple locations;
- delineate the SA and discretize it into J sub-areas assuming the origin of the possible sources in the centroid of any sub-area;
- compute (analytically or numerically) the TFs at the monitoring points for each possible source (J runs of the numerical transport model, when the analytical TFs are not available, are needed using the SIF method);
- recover the release histories performing the geostatistical procedure that simultaneously considers all the possible point sources;
- identify the source location as the location from which the highest amount of released pollutant is estimated.

Some heuristic considerations can be used to reduce the number of possible point sources. For instance, it is not advisable to consider, a priori, a source location for which the TFs have very low values: it is unlikely that it is the actual source. From a numerical point of view it can lead to an ill-conditioned \mathbf{H} matrix, resulting in a large number of

Fig. 1 Normalized log-conductivity field ($\sigma_Y^2 = 1.32$). Black dots indicate the measurement points of Cases 1 and 2, P1 and P2 are the measurements points of Case 3 and black squares denote the possible sources in the SA. The black diamond is the actual source location. Distances in meters



iterations and non-convergence problems in the recovery procedure.

4 Applications

4.1 Case 1: Identification of the release history with known source location

Preliminary results of the method presented in Sect. 3 are described in Butera et al. (2006) for uniform or weakly non-uniform flow cases. In this work, the numerical procedure, tested here for two-dimensional strongly non-uniform flow, is presented for a comparison with the results of the Cases 2 and 3.

A numerical model of a 2-D, 1-layer confined aquifer with rectangular shape (400 m long, 100 m wide, 10 m thickness) has been built (Fig. 1) using the MODFLOW code (Harbaugh et al. 2000). The domain is discretized into $2\text{ m} \times 2\text{ m}$ cells. The heterogeneous conductivity field is assumed known and characterized by a lognormal distribution with a mean value equal to 3.2×10^{-4} m/s, a standard deviation of 4.2×10^{-4} m/s and a correlation length equal to 20 m. Figure 1 shows the normalized log conductivity field $Z = (Y - \mu_Y)/\sigma_Y$ where $Y = \log K$ with mean μ_Y and standard deviation σ_Y . The log conductivity field variance is $\sigma_Y^2 = 1.32$. The assumption of a known hydraulic conductivity field is somewhat unrealistic; in fact, in field conditions is not very easy to get detailed information on hydraulic parameters and for this reason there is a huge collection of literature on estimating hydraulic conductivity variability (for example see Fioren et al. 2009 and Zanini and Kitanidis 2009). This paper has the goal of testing the proposed approach, assuming that the hydraulic conductivity field and the transport parameters are known. The covariance of the measurements errors \mathbf{R} is also assumed known with an assigned value of $\sigma_R = 1 \times 10^{-3}$ mg/l.

The boundary conditions are no flow boundaries in the north and south boundaries, constant head in the upstream (west) side $h_U = 24$ m and in the downstream (est) side

$h_D = 20$ m; the resulting flow through the aquifer is about 1.2×10^{-3} m³/s.

As usual, for the simulation of pollution events with uncoupled flow and transport processes, the flow field has been solved and the results have been transferred to the transport package.

On the synthetic field, a transport event has been simulated with the MT3D software (Zheng and Wang 1999). The longitudinal and transversal dispersivity are assumed constant ($\alpha_L = 1$ m and $\alpha_T = 0.1$ m, respectively) and a pollutant source has been located at $x = 49$ m and $y = 49$ m (Fig. 1). The pollutant release is simulated as an injection with a constant water discharge and variable tracer concentration over time. The amount of the conservative pollutant per unit time injected into the aquifer at the source ($s(t)$ of Eq. (2)) is given by:

$$s(t) = Q_{in} \cdot F(t), \tag{20}$$

where Q_{in} is the constant water discharge and $F(t)$ is the concentration history, variable in time, of the injected solution. A unit value was adopted for Q_{in} . Then, following Skaggs and Kabala (1994) and Snodgrass and Kitanidis (1997), we have considered a concentration history with the expression:

$$F(t) = \exp\left(\frac{-\left(\frac{t}{\Delta t} - 130\right)^2}{50}\right) + 0.3 \cdot \exp\left(\frac{-\left(\frac{t}{\Delta t} - 150\right)^2}{200}\right) + 0.5 \cdot \exp\left(\frac{-\left(\frac{t}{\Delta t} - 190\right)^2}{98}\right). \tag{21}$$

Since Q_{in} is of unit value, the identification of the release history $s(t)$ is equivalent to the identification of the concentration history $F(t)$.

The reporting time step Δt is 2 days. The concentration data have been monitored at the time $t = 300\Delta t$ at the 20 nodes of the regular grid shown in Fig. 1.

The SIF method was applied in order to determine the TFs in all the monitoring point locations. Then the geostatistical procedure was applied and the unknown release history $s(t)$ was determined. The results, reported as dimensionless,

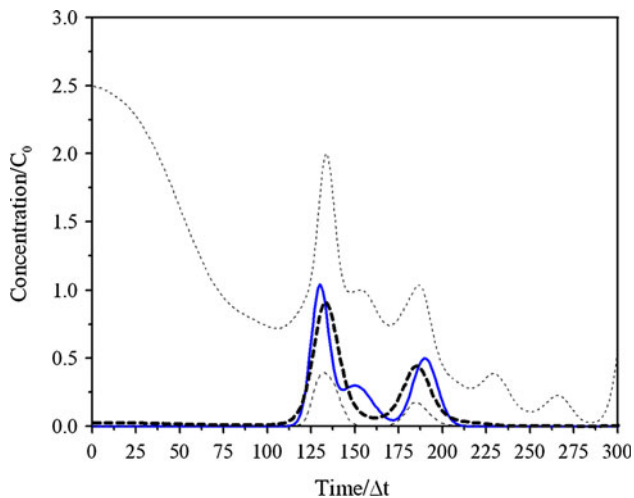


Fig. 2 Case 1: The true solution (solid blue line), best estimate (dashed thick line), and 5–95 % confidence interval (dotted lines). (Color figure online)

are shown in Fig. 2. The times have been divided by the time interval Δt and the concentrations have been divided by the injected concentration C_0 , used to compute numerically the TF. The true solution, the best estimate (given by Eq. 8), and the 5–95 % confidence interval are depicted in Fig. 2. The release is reproduced reasonably and the interquantile range is acceptable. The curves do not match exactly, and the smaller peak has not been reproduced, but the times of the major peaks and the highest concentration have been determined with fair precision.

The results are not as accurate as others presented in Butera et al. (2006) because here, more realistically, the monitoring network is not optimized: only 20 monitoring points have been considered and several of these provide a poor information as one can infer from Fig. 3.

In order to test how closely the recovered release history can reproduce the observation at the monitoring points, the transport model was run with the estimated injection function as the source term. In Fig. 3 the original

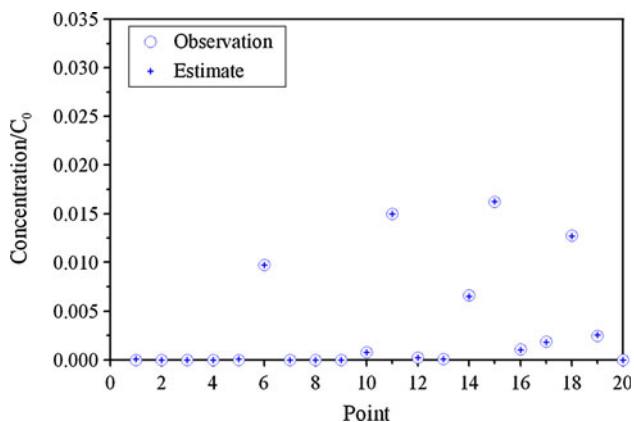


Fig. 3 Case 1: Observed and estimated concentrations

observations and the estimated ones are reported. The values match closely, validating the solution. Solutions of the inverse problem, as the one discussed in this paper, are always affected by non-uniqueness: different release functions, even far from the true one, can satisfactorily fit the observations. In this case (Fig. 3) the good match is a confirmation of the non-uniqueness of the solution.

4.2 Applications of the SRSI procedure

The SRSI methodology for source identification, outlined in Sect. 3, was applied to two illustrative cases simulating a two dimensional aquifer.

The conditions (aquifer characteristics, source position, pollutant release history (20) and (21)) of the numerical experiments are the same as those used for the Case 1, in order to save the observations (concentration values) in the 20 monitoring points.

4.3 Case 2: Analysis with concentration data collected at 20 locations at the same time

The SA has been assumed upgradient from the measurement points in the region $0 \leq x \leq 20$ m and $0 \leq y \leq 20$ m and it is subdivided into nine sub areas as shown in Fig. 1. The centroid of each area represents a possible source location. Then, the TFs relevant to the 20 monitoring points and the nine possible sources have been computed by means of the application of the SIF procedure, requiring nine runs of the forward transport model.

Finally, the SRSI procedure was carried out and the release function for each of the nine possible sources was obtained. The dimensionless results are depicted in Fig. 4.

In all the locations, except in $x = 49$ m $y = 49$ m the release history is null. This result suggests that the source is located in the sub-area with those centroid coordinates (just the ones of the actual source).

A detail of the Fig. 4 is reported (vertical scale exaggerated) in Fig. 5: the recovered release history for the source location $x = 49$ m, $y = 49$ m shows a good reconstruction of the peak times but an underestimation of the peak values.

Comparing Figs. 5 and 2 one can easily see that the recovered release history obtained in the case of a known source position is, as expected, more accurate. Nevertheless, in our opinion, the result of the SRSI is valuable, even if less precise, because in addition to resulting in a reliable release history, it can detect the most probable source location.

A great number of tests, not shown here for brevity, have been carried out to investigate the sensitivity of the SRSI methodology to the discretization degree of the SA.

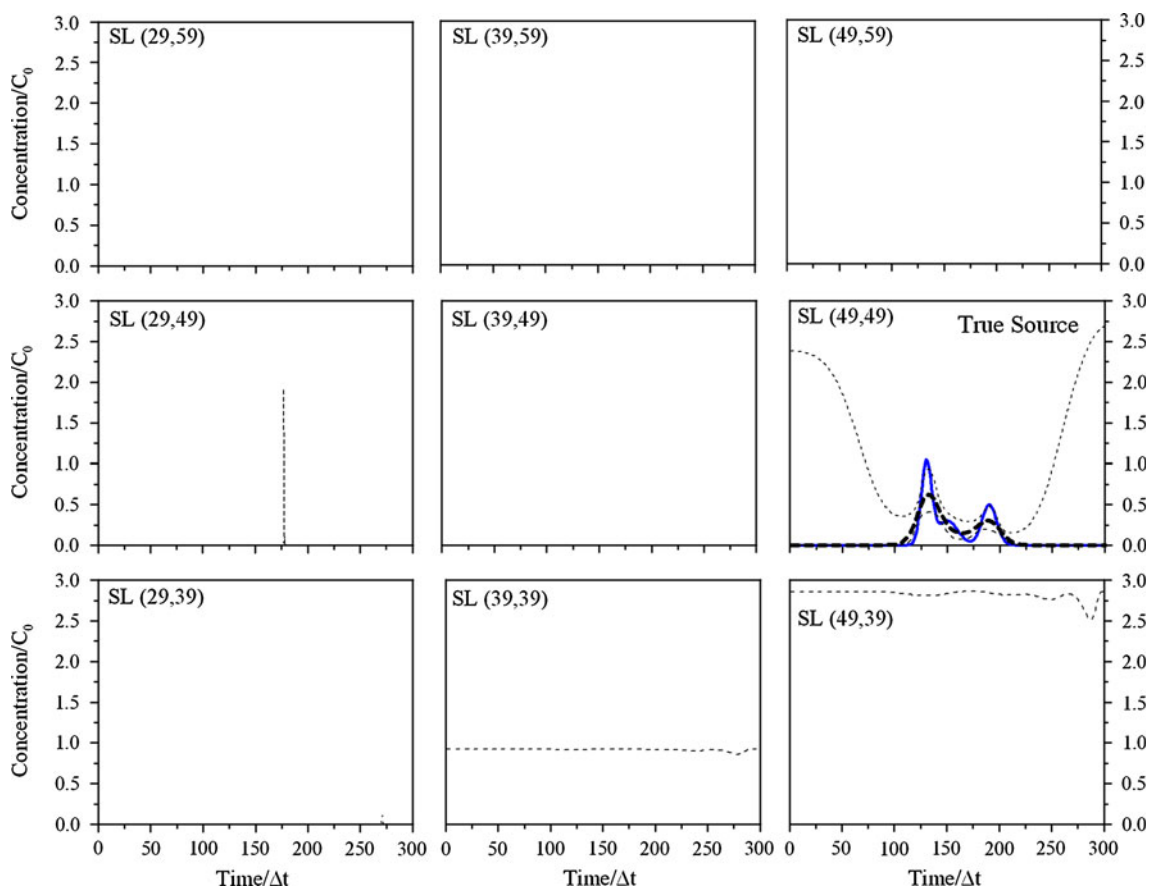


Fig. 4 The recovered release history at the hypothesized source locations of Case 2: the true solution (*solid blue line*), best estimate (*dashed thick line*) and 5–95 % confidence interval (*dotted lines*). (Color figure online)

The conclusions of the tests are that, in case of very dense discretization, the recovered release results can be shared among close subareas, and the source area is identified with fair approximation.

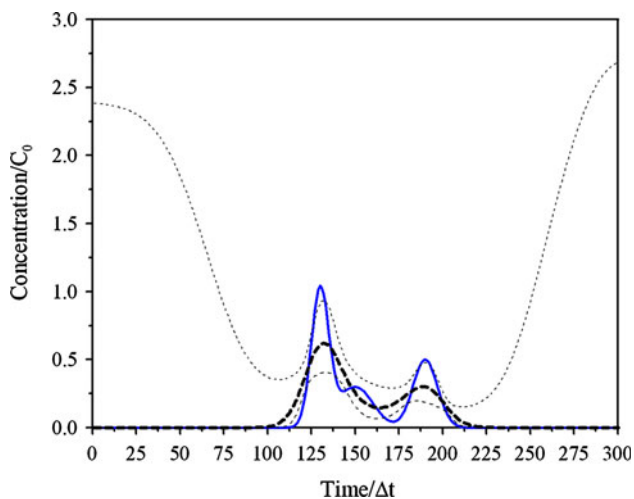


Fig. 5 The recovered release history at the source location of Case 2: the true solution (*solid blue line*), best estimate (*dashed thick line*) and 5–95 % confidence interval (*dotted lines*). (Color figure online)

Another interesting topic, which we will analyze in a future paper, is the case of the true source not overlapping the candidate locations. Moreover, the analysis of multiple sources showed in Butera and Tanda (2003) can be generalized to multiple unknown sources.

4.4 Case 3: Analysis with concentration data collected at two locations at different times

The aim of this application is to investigate whether the processing of the time–concentration history monitored at few spatial locations can give reliable information about the source location. This case is very realistic, as it is common to have only a few monitoring points and several sampled concentration values for each point at different times. This Case 3 was developed on the same aquifer as the previous Cases 1 and 2, the same boundary conditions, release and SA extent. Only two monitoring points (P1 and P2, see Fig. 1) are considered and 25 concentration values are considered available in 600 days, with a time step equal to 12Δt (Fig. 6). The concentrations values monitored at point P1 are lower than those measured at point P2 even though P1 is closer to the source location. This is due to the

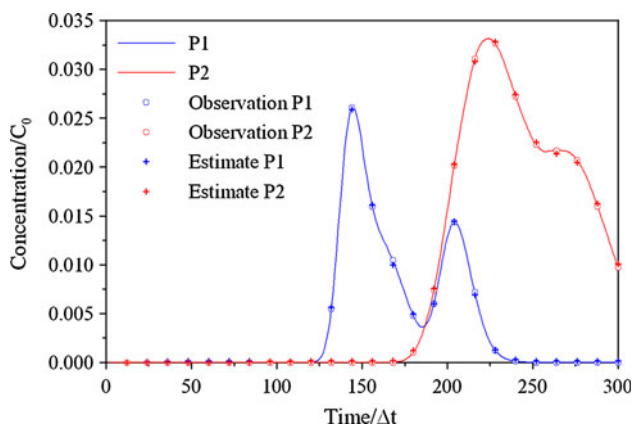


Fig. 6 The sampled time–concentration data for Case 3

heterogeneity of the conductivity field: a zone of low conductivity, where P1 is located, deflects the plume so low concentration values are sampled there.

The TFs for the two observation points have been obtained from the previous computations, after nine runs of the forward transport model. Then the SRSI procedure was applied to recover the release history at the nine possible sources.

In Fig. 7 the recovered releases histories are presented. Again, the actual source location was detected and the recovered release history is very close to the real one. A comparison of the Figs. 5 and 7 again shows that the observations recorded at few monitoring locations (two in Case 3) but at different times can lead to a good reconstruction of the release event. The reliability of the results is comparable with those (Fig. 5) obtained using the more expensive monitoring campaign necessary to obtain the concentration observations at many points (20 points in Case 2).

The accuracy of the Case 3 results (Fig. 7) seems comparable also with the accuracy of the Case 1 results (Fig. 2) obtained with 20 monitoring points and a known source position.

A final test was conducted running the transport model with the source term obtained from the SRSI. The concentration time series was determined in the P1 and P2 monitoring points and in Fig. 6 the values, occurring at the same monitoring times, have been plotted (estimated values) together with the observations. The differences between the actual and the calculated values are very small: again the good match is a confirmation of the non-uniqueness of the solution.

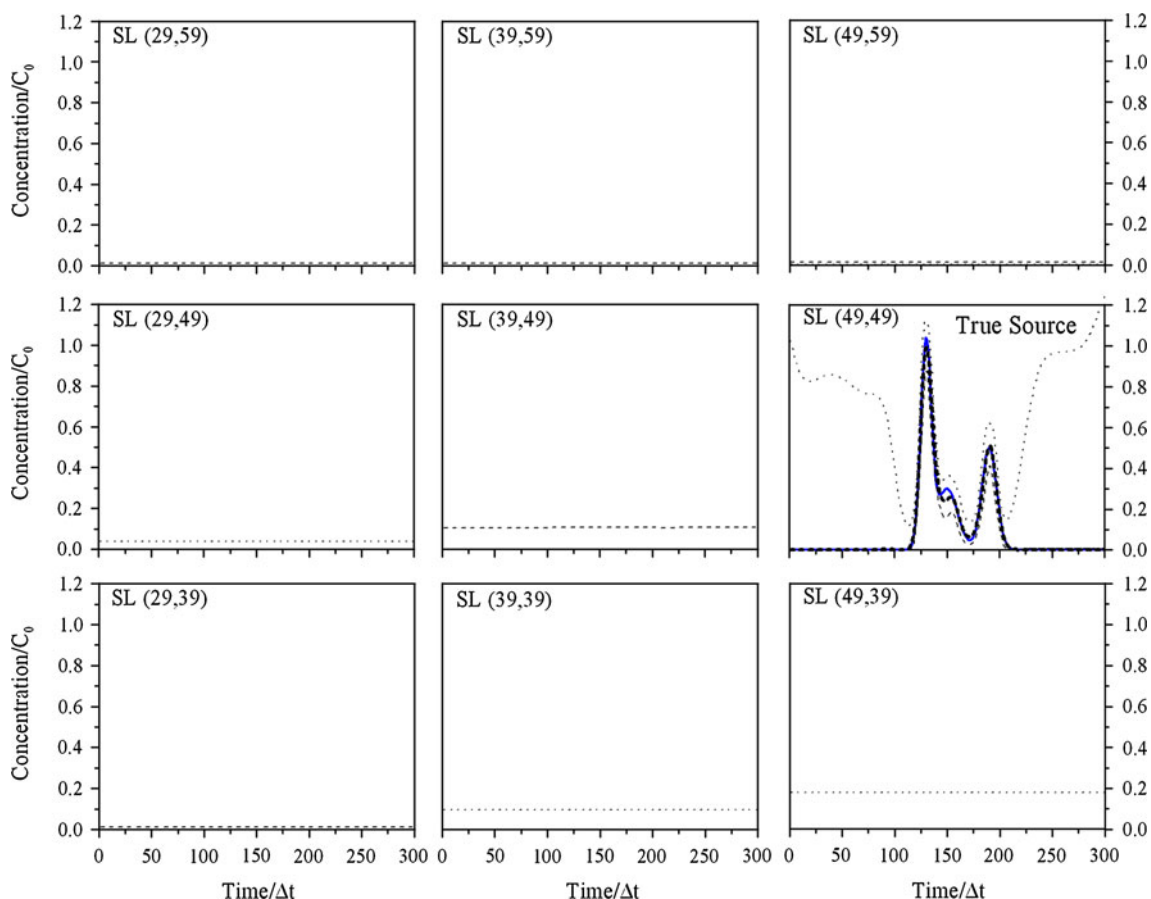


Fig. 7 The recovered release history at the hypothesized source locations of Case 3: the true solution (*solid blue line*), best estimate (*dashed thick line*) and 5–95 % confidence interval (*dotted lines*). (Color figure online)

5 Discussion and conclusions

A new procedure (SRSI) useful for the simultaneous identification of the release function and the source location has been presented in this manuscript. It is based on the geostatistical approach introduced by Snodgrass and Kitanidis (1997) and can be applied in groundwater aquifers with non-uniform flow field (due to heterogeneous media, presence of wells, natural boundaries) after the identification of the sensitivity matrix required by the numerical procedure. In the present paper the sensitivity has been defined using a TF, i.e. the function that describes the effect in time, at a certain location in the aquifer, of an impulse release of pollutant at the source. The TFs were computed with the innovative SIF procedure (Butera et al. 2004, 2006) with good results. The SIF procedure requires only one run of the transport model of the actual flow field per investigated source location: it presents efficient performance in comparison with other numerical methods as the perturbation (finite difference) or adjoint state. In fact, given m observations and n unknowns, the perturbation method needs $(n + 1)$ runs of the transport model and the adjoint state approach requires m runs on an inverted flow field with modified boundary and source conditions. Then, for the problem of recovering the release history, with known source location, the SIF procedure seems the most advantageous in terms of computation time. Moreover, the SIF procedure is simple to apply and the results of the tests here reported are satisfactory and encourage the application to real cases.

The SRSI procedure here proposed is able to simultaneously recover the release function and identify the source location. The procedure follows these steps: (i) delineate the suspect area, SA, and discretize it in sub-areas assuming the origin of the possible sources in the centroid of the sub-areas; (ii) compute the TFs (SIF procedure) at the monitoring points for each possible source; (iii) recover the release histories performing the geostatistical procedure that considers simultaneously all the possible point sources; (iv) identify the source location as the location where the highest amount of released pollutant is estimated.

The numerical tests here reported considered concentration data distributed in space (Case 2) and in time (Case 3). The results show that the SRSI procedure is able to identify the source location and to recover the release in all the considered cases and the accuracy of the results is comparable to those obtained with known source location (Case 1).

The number of preliminary runs of the forward transport model necessary to obtain the numerical TFs is equal to the number of the suspect source locations, resulting in a low computation burden.

We thought this paper as the first with the explanation of the methodology and the applications to simple study cases. In our future research we will analyze the performance of SRSI procedure considering more severe conditions such as the presence of measurements errors of unknown variance, the case of multiple sources and situations in which none of the candidate locations overlaps with the true source.

A crucial response about the reliability of the recovery process outlined in this manuscript can be obtained from an experimental activity to be carried out in controlled field tests or in medium scale laboratory tests. In these conditions it can be possible to recognize the role of the different factors involved in the inversion procedure and the research can be addressed to improve the most meaningful steps.

Acknowledgments We warmly thank Michael Cardiff and Michael Fienen for their valuable review that greatly improved the manuscript.

References

- Alapati S, Kabala ZJ (2000) Recovering the release history of a groundwater contaminant via the non linear least-squares estimation. *Hydrol Process* 14(6):1003–1016
- Atmadja J, Bagtzoglou AC (2001) State of the art report on mathematical methods for groundwater pollution source identification. *Environ Forensics* 2(3):205–214
- Box GEP, Cox DR (1964) An analysis of transformations. *J R Stat Soc B* 26:353–360
- Butera I, Tanda MG (2001) L'approccio geostatistico per la ricostruzione della storia di rilascio di inquinanti in falda. *L'Acqua* 4:27–36 (in Italian)
- Butera I, Tanda MG (2002), Ricostruzione della storia temporale del rilascio di un inquinante in falda mediante l'approccio geostatistico. In: Proceedings of "XXVIII Convegno di Idraulica e Costruzioni Idrauliche", Potenza (I), 16–19 September 2002 (in Italian)
- Butera I, Tanda MG (2003) A geostatistical approach to recover the release history of groundwater pollutants. *Water Resour Res* 39(12):1372. doi:10.1029/2003WR002314
- Butera I, Tanda MG (2004) La ricostruzione della storia del rilascio di inquinanti negli acquiferi. Impiego di rilevazioni a tempi diversi e impatto dell'eterogeneità dell'acquifero. *L'Acqua* 3:27–36 (in Italian)
- Butera I, Tanda MG, Zanini A (2004) La ricostruzione della storia del rilascio di inquinanti in acquiferi sede di moto non uniforme mediante approccio geostatistico. In: Proceedings of "XXIX Convegno di Idraulica e Costruzioni Idrauliche", Trento (I), 7–10 September 2004 (in Italian)
- Butera I, Tanda MG, Zanini A (2006) Use of numerical modelling to identify the transfer function and application to the geostatistical procedure in the solution of inverse problems in groundwater. *J Inv Ill-Posed Problems* 14(6):547–572. doi:10.1163/156939406778474532
- Dokou Z, Pinder GF (2009) Optimal search strategy for the definition of a DNAPL source. *J Hydrol* 376:542–556
- Fienen MN, Luo J, Kitanidis PK (2006) A Bayesian geostatistical transfer function approach to tracer test analysis. *Water Resour Res* 42:W07426. doi:10.1029/2005WR004576

- Fiene MN, Hunt R, Krabbenhoft D, Clemo T (2009) Obtaining parsimonious hydraulic conductivity fields using head and transport observations: a Bayesian geostatistical parameter estimation approach. *Water Resour Res* 45:W08405. doi:10.1029/2008WR007431
- Harbaugh AW, Banta ER, Hill MC, McDonald MG (2000) MODFLOW-2000, the U.S. Geological Survey modular ground-water model—user guide to modularization concepts and the ground-water flow process. U.S. Geological Survey Open-File Report 00-92, p 121
- Jury WA, Roth K (1990) Transfer functions and solute movement through soil: theory and applications. Birkhauser, Boston
- Kitanidis PK (1995) Quasi-linear geostatistical theory for inversing. *Water Resour Res* 31(10):2411–2419
- Kitanidis PK (1996) On the geostatistical approach to the inverse problem. *Adv Water Resour* 19(6):333–342
- Kitanidis PK, Shen KF (1996) Geostatistical estimation of chemical concentration. *Adv Water Resour* 19(6):369–378
- Liu C, Ball WP (1999) Application of inverse methods to contaminant source identification from aquitard diffusion profiles at Dover AFB, Delaware. *Water Resour Res* 35(7):1975–1985
- Luo J, Cirpka OA, Fiene MN, Wu W, Mehlhorn TL, Carley J, Jardine PM, Criddle CS, Kitanidis PK (2006) A parametric transfer function concept for analyzing reactive transport in nonuniform flow. *J Contam Hydrol* 83(1–2):27–41
- Michalak AM, Kitanidis PK (2002) Application of Bayesian inference methods to inverse modeling for contaminant source identification at Gloucester Landfill, Canada. In: Hassanizadeh SM, Schotting RJ, Gray WG, Pinder GF (eds) *Computational methods in water resources XIV*, vol 2. Elsevier, Amsterdam, pp 1259–1266
- Michalak AM, Kitanidis PK (2003) A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification. *Water Resour Res* 39(2):SBH 7-1–SBH 7-14
- Michalak AM, Kitanidis PK (2004a) Application of geostatistical inverse modelling to contaminant source identification at Dover AFB, Delaware. *J Hydraul Res* 42(extra issue):9–18
- Michalak AM, Kitanidis PK (2004b) Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling. *Water Resour Res* 40:W08302
- Michalak AM, Kitanidis PK (2004c) A method for the interpolation of nonnegative functions with an application to contaminant load estimation. *Stoch Environ Res* 18:1–16. doi:10.1007/s00477-004-0189-1
- Milnes E, Perrochet P (2007) Simultaneous identification of a single pollution point-source location and contamination time under known flow field conditions. *Adv Water Resour* 30(12):2439–2446
- Morrison RD (2000) Application of forensic techniques for age dating and source identification in environmental litigation. *Environ Forensics* 1(3):131–153
- Neupauer RM, Wilson J (2001) Adjoint-derived location and travel time probabilities for a multi-dimensional groundwater system. *Water Resour Res* 36(6):1657–1668
- Neupauer RM, Borchers B, Wilson J (2000) Comparison of inverse methods for reconstructing the release history of a groundwater contamination source. *Water Resour Res* 36(9):2469–2475
- Skaggs TH, Kabala ZJ (1994) Recovering the release history of a groundwater contaminant. *Water Resour Res* 30:71–79
- Snodgrass MF, Kitanidis PK (1997) A geostatistical approach to contaminant source identification. *Water Resour Res* 33:537–546
- Woodbury AD, Ulrych TJ (1996) Minimum relative entropy inversion: theory and application to recovering the release history of a groundwater contaminant. *Water Resour Res* 32:2671–2681
- Woodbury AD, Ulrych TJ (1998) Reply to Comment on: minimum relative entropy inversion: theory and application to recovering the release history of a groundwater contaminant. *Water Resour Res* 34(8):2081–2084
- Zanini A, Kitanidis PK (2009) Geostatistical inversing for large-contrast transmissivity fields. *Stoch Env Res Risk Assess* 23:565–577. doi:10.1007/s00477-008-0241-7
- Zheng C, Wang PP (1999) MT3DMS: a modular three-dimensional multispecies transport models; documentation and user's guide, contract report SERDP-99-1. U.S. Army Engineer Research and Development Center, Vicksburg