University of Parma Research Repository

Laboratory sandbox validation of pollutant source location methods

note finali coverpage

(Article begins on next page)

13 October 2024

1   Cupola Fausto, Tanda Maria Giovanna, Zanini Andrea

2   # Laboratory sandbox validation of pollutant source

3   # location methods

4   *DICATeA, Università degli Studi di Parma, Parco Area delle Scienze, 181/A, 43124 Parma*

5   e-mail: andrea.zanini@unipr.it

6   Phone: +39 0521 905931

7   Fax: +39 0521 905924.

8
9   **Abstract**

10   Inverse methods can be used to recover the pollutant source location from concentration data. In this paper, the relative

11   effectiveness of two proposed methods, simultaneous release function and source location identification (SRSI) and

12   backward probability model based on adjoint state method (BPM-ASM) are evaluated using real data collected by using

13   experimental equipment. The device is a sandbox that reproduces an unconfined aquifer in which all the variables are

14   controlled. A numerical model was calibrated using experimental observations. The SRSI is a stochastic procedure

15   which finds the source location and the release history by means of a Bayesian geostatistical approach. The BPM-ASM

16   provides the backward probability location of the pollutant detected at a monitoring point by means of a reverse

17   transport simulation. The results show that both methods perform well. While the simultaneous release function and

18   source location identification method requires a preliminary delineation of a probable source area and some weak

19   hypotheses about the statistical structure of the unknown release function, the backward probability model requires

20   some hypothesis about the contaminant release time. A case study was performed using two observation points only,

21   and despite the scarcity of data, both methodologies were able to accurately reconstruct the true source location. The

22   geostatistical approach has the advantage to recover the release history function too, whilst the backward probability

23   model works well with fewer data. If there are many observations, both methodologies may be computationally heavy.

24   A transfer function approach has been adopted for the numerical definition of the sensitivity matrix in the SRSI method.

25   The reliability of the experimental equipment was tested in previous laboratory works, conducted under several

26   different conditions.

27   *Keywords: Geostatistical approach; Transfer function; Source detection; Backward Location PDF; Sandbox*

28   **List of symbols**

29   **b** $p \times 1$ unknown coefficients

30    $C(\mathbf{x}, t)$ concentration at point **x** and time t

31   C_0 solution concentration

32    $\hat{C}_i$ observed concentration

33   C_w initial condition in backward model

34   **D** dispersion tensor

35    $F(t)$ concentration of the water injected at the source as function of time t

36   F_0 constant and known mass rate input function

76 **X** $n{\times}p$ matrix, mean of the unknown process

77 $\mathbf{x}_w$ observation location

78 **z** $m{\times}1$ observations

79 z_{inj} vertical coordinate of injector

80 α positive number

81 α_L longitudinal dipersivity

82 α_T transversal dispersivity

83 β_x normalization factor

84 Γ_i boundaries

85 δ Dirac delta function

86 Δt numerical model time step

87 Δx longitudinal size of numerical cell grid

88 Δy transversal size of numerical cell grid

89 Δz vertical size of numerical cell grid

90 ε_i ith measurement error

91 η time

92 $\boldsymbol{\theta}$ structural parameters of the covariance function

93 λ_s correlation time length of the unknown release function $s(t)$

94 $\boldsymbol{\Lambda}$ $n{\times}m$ Kriging coefficients

95 $\boldsymbol{\Xi}$ $m{\times}m$ dummy matrix

96 σ^2_R variance of the measurement error

97 σ^2_S variance of the unknown release function $s(t)$

98 $\boldsymbol{\Sigma}$ $m{\times}m$ dummy matrix

99 τ backward time

100 τ_w backward sampling time

101 φ porosity

102 ψ^* adjoint state

103 ∇ Nabla operator

1 Introduction

Source identification and recovery of the pollutant release history in groundwater have received much attention in recent years. The identification of a source location could allow to identify the true cause of the contamination and to foresee the future pollution spread, while the release time, the duration, and the maximum value of the released solute concentration could allow to apportion remediation costs among the responsible parties.

The problems in recovering the release history and/or the source location were studied extensively in the past. In this paper two different approaches are compared with respect to the same test case: simultaneous release function and source location (in short SRSI) (Butera et al. 2013), and backward location model based on adjoint state method (briefly BPM-ASM) (Neupauer and Wilson 1999). Both methodologies were tested on experimental data collected in a laboratory sandbox that reproduces an unconfined aquifer (for details see Citarella et al. 2010). This allows to validate

114   the methodologies in a real test case in which all variables are measured and controlled. This is the first time that a well-

115   known experimental dataset has been used to test and validate these procedures.

116   The first method, developed by Butera et al. (2013), is based on the geostatistical approach proposed by Snodgrass and

117   Kitanidis (1997) for the one-dimensional uniform flow. Several improvements and applications of the geostatistical

118   methodology were proposed by Michalak and Kitanidis 2002, 2003, 2004a, 2004b, and by Butera and Tanda (2003) and

119   Butera et al. (2006, 2013). Given the linearity of the governing differential equation, the approach uses the transfer

120   functions (TFs) to describe the effect in time, at a certain location of the aquifer, of an impulse release of a pollutant at a

121   known source. Although TFs can be analytically determined if the problem has a simple flow field, the characteristics

122   of the groundwater flow field do not allow this solution in many practical applications. To overcome this difficulty, a

123   numerical procedure to compute TFs was developed by Butera et al. (2004) and applied to homogeneous and weakly

124   heterogeneous aquifers (Butera at al. 2006). The source identification procedure requires a preliminary delineation of an

125   area where the pollutant source is most likely to be present, but it allows to obtain the simultaneous identification of the

126   release history and the source location. An application of this methodology, in a complex real case study, was

127   implemented by Gzyl at al. (2014).

128   The second methodology was developed by Neupauer and Wilson (1999, 2001). In their works, these authors showed

129   that the backward location and travel time probability density functions (PDF) are related to adjoint states of

130   concentration, and they developed a technique for obtaining the governing equation of the backward model using the

131   adjoint theory. The backward location PDF describes the possible former positions of the observed contamination at a

132   certain time before the detection, while the backward travel time PDF describes the possible travel time of the

133   contaminant from a selected upgradient position to the observation location. By using an adjoint model, an

134   instantaneous point source of an adjoint state (related to the PDFs) is released at the observation location at an observed

135   time. The adjoint state is thus transported upgradient and backward in time, following the same processes that occur in

136   forward contaminant transport modeling. The resulting spatial distribution of the concentration in the domain is related

137   to the backward location PDF. The work by Neupauer and Wilson (1999, 2001) was improved by considering non-

138   uniform flow field (Neupauer and Wilson 2002), sorbing solutes (Neupauer and Wilson 2004a, 2004b, 2009), multiple

139   observations (Neupauer and Wilson 2005), measured concentration (Neupauer and Lin 2006), and sorption and decay

140   phenomena (Neupauer et al. 2007). By using a previously calibrated numerical model, the single-observation and

141   multiple-observation backward location PDF can be computed. This PDF was then conditioned on the concentration

142   measurement, by reducing the variance of the backward location PDF and improving the results. However, this

143   methodology was developed to manage plumes originated from a single instantaneous point source; nevertheless, it

144   seems an interesting approach and in this work its performance has been tested on a continuous release source.

145   The manuscript is organized in three parts: first, the mathematical statements concerning the two approaches are

146   presented, then a brief description of the experimental equipment is reported; finally, the results of the application of the

147   two methodologies are presented and discussed.

148   **2 Mathematical Statements**

149   **2.1 Groundwater Transport**

150   Equation (1) describes the transport process in an aquifer corresponding to the injection of a non-sorbing, non-reactive

151   solute in a point source (Bear and Verruijt 1987):

152    $$\varphi \frac{\partial (C(\mathbf{x},t))}{\partial t} = \nabla [\varphi \cdot \mathbf{D}(\mathbf{x}) \cdot \nabla C(\mathbf{x},t)] - \nabla [\varphi \cdot \mathbf{u}(\mathbf{x},t) \cdot C(\mathbf{x},t)] + s(\mathbf{x_0},t) \cdot \delta (\mathbf{x} - \mathbf{x_0}) \qquad (1)$$

153   where: φ [-] is the effective porosity (taken as spatially variable, but constant in time), $\mathbf{u}(\mathbf{x},t)$ [L·T^{-1}] is the effective

154   velocity at location $\mathbf{x}$ and time t, $\mathbf{D}(\mathbf{x})$ [L^2 T^{-1}] the dispersion tensor, $C(\mathbf{x},t)$ [M·L^{-3}] the concentration at location $\mathbf{x}$ and

155   time t, $s(\mathbf{x_0},t)=C_0(\mathbf{x_0},t)\cdot q_0(\mathbf{x_0},t)$ [M·T^{-1}] is the amount of pollutant per time unit injected into the aquifer through the

156   source located at $\mathbf{x_0}$, $C_0(\mathbf{x_0},t)$ is the concentration injected and $q_0(\mathbf{x_0},t)$ [L^3·T^{-1}] is the injection flow rate at $\mathbf{x_0}$ at time t

157   [T].

158   The solution of Equation (1), by considering uniform porosity, when associated with the initial and boundary

159   conditions: $C(\mathbf{x},0) = 0$; $C(\infty,t) = 0$, is given by the following integral (Jury and Roth 1990):

160
$$C(\mathbf{x},t) = \int_0^t s(\mathbf{x_0},\eta) f(\mathbf{x},t-\eta)d\eta \qquad (2)$$

161   where $f(\mathbf{x},t-\eta)$ [L^{-3}] is the transfer function (TF) that describes the effects at $\mathbf{x}$ at time t by an impulse injection occurring

162   at $\mathbf{x_0}$ at time η [T].

163   **2.2 Geostatistical approach**

164   The source position was estimated through the procedure developed by Butera et al. (2013). The method, called

165   simultaneous release function and source location identification (SRSI), allows to identify not only the source position

166   but also its release history in time.

167   The SRSI procedure can be summarized with the following steps:

168   • collect a set of concentration measurements in space and/or time;

169   • delineate the suspect area (SA) for the source location and discretize it into J sub-areas assuming the origin of

170   the possible sources in the centroid of any sub-area;

171   • compute the transfer functions at the monitoring points for each possible source (J runs of the numerical

172   transport model);

173   • recover the release histories performing the geostatistical procedure which simultaneously considers all the

174   possible point sources (superposition effect) by means of $C(\mathbf{x},t) = \sum_{j=1}^{J} \int_0^t s_j(\eta)\cdot f_j(\mathbf{x},t-\eta)d\eta$ where j is one of

175   J generic sub-areas within the SA;

176   • identify the source location as the location from which the highest amount of released pollutant is estimated.

177   TFs can be determined easily in simple flow conditions (such as homogeneous and isotropic porous media in absence of

178   interferences) by means of analytical formulations (Bear and Verruijt 1987). When considering a non-uniform flow

179   field (for instance, heterogeneous porous media, or presence of a withdrawal) it is necessary to develop numerical

180   strategies. Butera et al. (2006, 2013) developed the Stepwise Input Function procedure (SIF) for TF calculation. The

181   procedure basically consists in making the time derivative of equation (2) by considering a constant and known input

182   function $s(\mathbf{x_0},t)=F_0 \cdot H_{SF}(t)$, where $H_{SF}(t)$ is the Heaviside step function and $F_0=C_0\cdot q_0$ is the amount of pollutant

183   injected into the aquifer with constant and known concentration C_0 at flow rate q_0.

184   The TF results in:

185
$$f(\mathbf{x},t) = \frac{1}{F_0}\frac{\partial C(\mathbf{x},t)}{\partial t} \qquad (3)$$

5

186   Equation (3), coupled with a numerical flow and transport model known from calibration or expert elicitation, allows to

187   easily determine the TFs at a generic point **x** by processing the concentration history (breakthrough curve) at the same

188   location due to a stepwise tracer injection at $\mathbf{x_0}$.

189   The core of the method is the quasilinear geostatistical approach developed by Kitanidis (1995, 1996) and Snodgrass

190   and Kitanidis (1997), which is briefly summarized in the following.

191   The observed concentration data at a known time $\bar{t}$ can be expressed as a function of the release process by the

192   following equation:

193   $\mathbf{z} = \mathbf{h}(\mathbf{s}) + \mathbf{v}$ (4)

194   where **z** is a $m \times 1$ vector of observations, **h(s)** is the $n \times 1$ vector containing the time discretization of the unknown release

195   function $s(t)$ and **v** is a $m \times 1$ vector of epistemic errors with zero mean and known covariance matrix $\mathbf{R} = \sigma_R^2 \cdot \mathbf{I}$.

196   In the case of a conservative solute, the relationship between the observed concentration and the release is linear, and

197   equation (4) can be simplified to (Snodgrass and Kitanidis, 1997):

198   $\mathbf{z} = \mathbf{H} \cdot \mathbf{s} + \mathbf{v}$ (5)

199   Equation (5) represents the matrix form of equation (2), where matrix **H** contains the values of the TFs (f), computed at

200   appropriate times and locations:

201
$$\mathbf{H} = \Delta t \begin{bmatrix} f(\mathbf{x}_1, \bar{t} - \Delta t) & ... & f(\mathbf{x}_1, \bar{t} - n\Delta t) \\ f(\mathbf{x}_2, \bar{t} - \Delta t) & ... & f(\mathbf{x}_2, \bar{t} - n\Delta t) \\ ... & ... & ... \\ f(\mathbf{x}_m, \bar{t} - \Delta t) & ... & f(\mathbf{x}_m, \bar{t} - n\Delta t) \end{bmatrix} \quad (6)$$

202   The transfer matrix **H** includes all the characteristics of the flow and transport processes. Vector **s** can be considered

203   random and characterized by an unknown mean $E[\mathbf{s}] = \mathbf{Xb}$ and a covariance function $\mathbf{Q}(\boldsymbol{\theta}) = E[(\mathbf{s}-\mathbf{Xb})(\mathbf{s}-\mathbf{Xb})^T]$, where

204   $E[]$ denotes the expected value, **X** is a $n \times p$ matrix of known functions, **b** is a vector of size $p \times 1$ that contains the

205   unknown drift coefficients and $\boldsymbol{\theta}$ are the unknown structural parameters. In this work, a constant but unknown mean is

206   considered; thus **X** is an $n \times 1$ vector filled by 1 and **b** is the scalar unknown mean of the function; moreover a Gaussian

207   covariance function has been considered, so $\boldsymbol{\theta}$ are the variance σ_s^2 and the correlation time length λ_s.

208   The estimation procedure proposed by Kitanidis (1995) is divided into two parts: first the structural parameters $\boldsymbol{\theta}$ of the

209   selected covariance function are determined, then the unknown release function is estimated by means of a Kriging

210   process. The identification of the structural parameters follows a Restricted Maximum Likelihood approach. The

211   probability that the random process with parameter $\boldsymbol{\theta}$ reproduces observation **z** can be estimated through the following:

212   $p(\mathbf{z} | \boldsymbol{\theta}) \propto |\Sigma|^{-1/2} |\mathbf{X}^T \mathbf{H}^T \Sigma^{-1} \mathbf{H}\mathbf{X}|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{z}^T \Xi^{-1} \mathbf{z} \right]$ (7)

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

214   Once the structural parameters are computed, by maximizing the probability (Eq. (7)), estimation $\hat{\mathbf{s}}$ of the release

215   function $s(t)$ is obtained through Kriging:

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

6

217 where matrix $\boldsymbol{\Lambda}$ ($n\times m$) of the Kriging weights is calculated by solving the following system obtained from the un-

218 biasedness and minimum variance conditions:

219 $$\begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{HX} \\ (\mathbf{HX})^T & \mathbf{0} \end{bmatrix}\begin{bmatrix} \boldsymbol{\Lambda}^T \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{HQ} \\ \mathbf{X}^T \end{bmatrix} \qquad (9)$$

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

221 $$\mathbf{V} = -\mathbf{XM} + \mathbf{Q} - \mathbf{QH}^T\boldsymbol{\Lambda}^T \qquad (10)$$

222 A transformation (Box and Cox 1964) of variable $\mathbf{s}$ was considered in order to enforce non-negativity of the estimated

223 concentration; the new unknown becomes:

224 $$\tilde{\mathbf{s}} = \alpha\left(\mathbf{s}^{1/\alpha} - 1\right) \qquad (11)$$

225 where α is a positive number and is chosen as small as possible while ensuring $\tilde{\mathbf{s}} > -\alpha$.

226 When the values of $\mathbf{s}$ are constrained to be positive and are physically compatible, equation (4) becomes:

227 $$\mathbf{z} = \mathbf{H}\cdot\mathbf{s} + \mathbf{v} = \mathbf{H}\cdot\left(\frac{\tilde{\mathbf{s}} + \alpha}{\alpha}\right)^{\alpha} + \mathbf{v} = \mathbf{h}(\tilde{\mathbf{s}}) + \mathbf{v} \quad (12)$$

228 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

229 Kitanidis 1995, 1996).

230 For the SRSI procedure, vector $\mathbf{s}$ of the unknown release function in (4) is made up by the collection of J sub-vectors $\mathbf{s}_j$,

231 each with dimensions $n_i\times 1$, where n_i is the number of time values used to discretize the release history. The total

232 dimension of $\mathbf{s}$ is: $(n_1 + n_2 + \ldots + n_J)\times 1$:

233 $$\mathbf{s} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \ldots \\ \mathbf{s}_J \end{bmatrix} \qquad (13)$$

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

235 $$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \ldots\ldots & \mathbf{H}_J \end{bmatrix} \qquad (14)$$

236 whose dimensions are $m\times(n_1 + n_2 + \ldots + n_J)$. The generic matrix $\mathbf{H}_j$ describes the effects of the pollutant release in the

237 sub-area j on the concentration data measured at the m monitoring points. The release history of each source location is

238 assumed to be independent of the others. For this reason covariance matrix $\mathbf{Q}$ of the $\mathbf{s}$ process is a block matrix with

239 non-zero elements in diagonal blocks only.

240 $$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & 0 & 0 & 0 \\ 0 & \mathbf{Q}_2 & 0 & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \mathbf{Q}_J \end{bmatrix} \quad (15)$$

241 The results of the geostatistical procedure described in this section provide the pollutant history in the J hypothetical

242 source locations. The release function in the real source will be substantial, while in the other suspect locations the time

243 histories will be negligible.

244 **2.3 Backward probability model based on adjoint state method (BPM-ASM)**

245 The use of backward location probability density function to identify the source location is briefly explained in the

246 following. This approach was developed (Neupauer and Wilson 1999) only for an instantaneous point source and it is

247 assumed that release time τ is known. To calculate backward location PDF a numerical model calibrated on the

248 experimental data is often necessary. So, every observation point needs to be considered as an instantaneous source in

249 the adjoint equation. The adjoint equations can be solved with the same flow and transport software used in forward

250 simulation, on a domain with modified boundary conditions that produce reverse flow direction. In the test case

251 explained in the following, one independent backward simulation (flow and transport) for each observation point was

252 performed. The initial condition was

253
$$C_w(\mathbf{x}_w,0) = \frac{1}{Q_w \cdot \Delta t} \qquad (16)$$

254 where $\mathbf{x}_w$ is the position of the observation point, considered now as source, Q_w [L^3T^{-1}] is the instantaneous discharge

255 of tracer solution injected, negligible compared to the background flow, and Δt [T] is the time step length. So the mass

256 injected in the single time step, is

257
$$M_{injected} = C_w \cdot Q_w \cdot \Delta t = 1 \quad (17)$$

258 The evolution of the plume in the backward model provides the backward location PDF and represents the probability

259 that the contaminant exists at that location and given backward time.

2.3.1 Single-Observation Backward Location PDF

261 For more details see Neupauer and Wilson (1999, 2002, 2003, 2004b). Let us consider the advection dispersion

262 equation (Eq. 1) with the following boundary and initial conditions:

263 $\quad C(\mathbf{x},t_0) = \dfrac{m_0}{\varphi}\,\delta_0(\mathbf{x}-\mathbf{x}_0)$

264 $\quad C(\mathbf{x},t) = g_1(t)$ on Γ_1

265 $\quad \left[\mathbf{D}\dfrac{\partial C(\mathbf{x},t)}{\partial x_j}\right]\cdot\mathbf{n} = g_2(t)$ on Γ_2 $\qquad (18)$

266 $\quad \left[\mathbf{u}C(\mathbf{x},t) - \mathbf{D}\nabla C(\mathbf{x},t)\right]\cdot\mathbf{n} = g_3(t)$ on Γ_3

267 where $C(\mathbf{x},t)$ is the concentration, $\mathbf{D}$ is the dispersion tensor, $\mathbf{u}$ is the effective velocity, t_0 is the source release time, $\mathbf{x}_0$ is

268 the source location, m_0 [M] is the source mass, $\delta(\cdot)$ is the Dirac delta function, $g_i(t)$ are known boundary functions, and

269 Γ_i are the boundaries.

270 The adjoint of the equation (1) is given by (Neupauer and Wilson 2002)

271 $\quad \varphi\dfrac{\partial\psi*}{\partial\tau} = \nabla(\varphi\mathbf{D}\nabla\psi*) + \nabla(\varphi\mathbf{u}\psi*) + q_I\psi*\cdot\delta(\mathbf{x}-\mathbf{x}_w)\cdot\delta(\tau-\tau_w) + H_L(\mathbf{x},\tau)$ $\qquad (19)$

272 $\quad \psi*(\mathbf{x},0) = 0$

273 $\quad \psi*(\mathbf{x},\tau) = 0$ on Γ_1

274 $\quad \left[\mathbf{D}\dfrac{\partial\psi*}{\partial x_j} + \mathbf{u}\psi*\right]\cdot\mathbf{n} = 0$ on Γ_2

275 $\quad \left[\mathbf{D}\dfrac{\partial\psi*}{\partial x_j}\right]\cdot\mathbf{n} = 0$ on Γ_3

276   where q_I [T^{-1}] is the source inflow rate per unit volume and $\psi*$ [L^{-3}] is an adjoint state that is also (Neupauer and

277   Wilson 1999) the marginal sensitivity of the concentration to source mass. τ is the backward time, while $\mathbf{x}_w$ and τ_w are

278   the observation location and backward sampling time, respectively. By using samples taken at monitoring wells

279    $$H_L(\mathbf{x},\tau) = \left[\frac{u_1 \alpha_L}{|\mathbf{u}(\mathbf{x})|} \delta'_{x_1}\left(x_1 - x_{1_w}\right)\delta\left(x_2 - x_{2_w}\right) + \frac{u_2 \alpha_L}{|\mathbf{u}(\mathbf{x})|}\delta'_{x_2}\left(x_1 - x_{1_w}\right)\delta\left(x_2 - x_{2_w}\right)\right]\delta(\tau - \tau_w) \quad (20)$$

280   where the sample location is $\mathbf{x}_w = \left(x_{1_w}, x_{2_w}\right)$, $\delta'_{x_i} = \left(x_i - x_{i_w}\right)$ is the derivative of the Dirac delta function with respect to

281   x_i, vertical bars denote magnitude, and α_L [L] is the longitudinal dispersivity.

282   The relationship between the adjoint state from (19) and the backward location probability density function is given by

283   (Neupauer and Wilson 2002)

284    $$f_{\mathbf{x}}(\mathbf{x};\tau,\mathbf{x}_w,\tau_w) = \varphi \cdot \psi*(\mathbf{x};\tau,\mathbf{x}_w,\tau_w) \quad (21)$$

285   where $f_{\mathbf{x}}(\mathbf{x};\tau,\mathbf{x}_w,\tau_w)$ is the backward location PDF concerning position $\mathbf{x}$ at backward time τ of a contaminant particle

286   that was observed at $\mathbf{x}_w$ at backward time τ_w, and $\psi*$ is the adjoint state obtained from (19).

2.3.2 Multiple Observations Backward Location PDF

288   If several observations are available, multiple-observation backward location density function (Neupauer and Wilson

289   2005) can be obtained by first calculating the single-observation PDF for each observation, then combining them. Each

290   observation gives additional information to characterize the former position of contamination, thereby reducing the

291   uncertainty or variance of the backward location PDF. In this step only the presence of the pollutant at monitoring

292   points is considered, and not the concentration values. Let N be the number of observations, and let $\{\mathbf{x}_w\}$ and $\{\tau_w\}$ be N-

293   length vectors of sampling locations and backward sampling times, respectively. The multiple-observation backward

294   location PDF, $f_{\mathbf{X}}(\mathbf{x};\tau,\{\mathbf{x}_w\},\{\tau_w\})$, describes the possible former positions of all contaminant particles observed, given

295   that at a previous time τ, they were at the same location, i.e. source location. It is calculated as

296    $$f_{\mathbf{X}}(\mathbf{x};\tau,\{\mathbf{x}_w\},\{\tau_w\}) = \frac{\prod_{i=1}^{N} f_{\mathbf{X}}(\mathbf{x};\tau,\mathbf{x}_{wi},\tau_{wi})}{\int \prod_{i=1}^{N} f_{\mathbf{X}}(\mathbf{x};\tau,\mathbf{x}_{wi},\tau_{wi})d\mathbf{x}} \quad (22)$$

297   where $f_{\mathbf{X}}(\mathbf{x};\tau,\mathbf{x}_{wi},\tau_{wi})$ is the backward location PDF for the ith sample, obtained from (21). The sum of multiple-

298   observation backward location PDF, calculated on the whole domain, is equal to 1. An important aspect of this method

299   is that only the presence or the absence of the pollutant is considered and the concentration values are not taken into

300   account.

2.3.3 Conditioned Backward Location PDF

302   Neupauer and Lin (2006) improved the backward location probability density function conditioning on measured

303   concentrations collected after an instantaneous release originated from a point source located in $\mathbf{x_0}$.

304   In this case the source mass and the source location are unknown but the measured concentrations are known. Neupauer

305   and Lin (2006) suggested using Bayes' theorem with the aim of constraining source mass M and random source

306   location $\mathbf{x}_0$ on the specific concentrations measured.

307   Let $\hat{C}_i = \hat{C}(\mathbf{x}_{wi},\tau_{wi})$, $i = 1, 2, ..., N$, the observed concentrations, where $\mathbf{x}_{wi}$ is the location and τ_{wi} is the backward time

308   at which sample i was taken. $\hat{C}_i$ is considered as a random variable with true value $C(x_{wi},\tau_{wi}|m_0,\mathbf{x}_0,\tau)$. So, ε_i is the

309   error of the ith measurement and it is normally-distributed with zero mean and variance σ_i^2. Therefore measured

310 concentrations, $\hat{C}_i$ are normally distributed with a mean equal to the true concentration and variance σ_R^2, and its PDF is

311 given by

312
$$f_{\hat{C}_i}\left(\hat{C}_i\middle|m_0,x_0,\tau\right)=\frac{1}{\sqrt{2\pi\sigma_R^2}}\exp\left\{-\frac{\left[\hat{C}_i-C\left(x_{wi},\tau_{wi}\middle|m_0,x_0,\tau\right)\right]^2}{2\sigma_R^2}\right\} \qquad (23)$$

313 If ε_i is considered independent from any other, for a known source mass m_0 and a source location $\mathbf{x}_0$, the joint PDF on

314 all N measured concentrations is simply the product of the PDFs for the individual observations given by

315
$$f_{\hat{\mathbf{C}}}\left(\hat{\mathbf{C}}\middle|m_0,\mathbf{x}_0,\tau\right)=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma_R^2}}\exp\left\{-\frac{\left[\hat{C}_i-C\left(x_{wi},\tau_{wi}\middle|m_0,x_0,\tau\right)\right]^2}{2\sigma_R^2}\right\} \qquad (24)$$

316 Assuming that the source mass is independent of source location in the absence of any concentration information, the

317 final result is (for more details see Neupauer and Lin 2006)

318
$$f_{\mathbf{X}_0|\hat{\mathbf{C}}}\left(\mathbf{x}_0\middle|\hat{\mathbf{C}};\tau\right)=\beta_x\int\prod_{i=1}^{N}\left[f_{\hat{C}_i|M,X_0}\left(\hat{C}_i\middle|m_0,\mathbf{x}_0;\tau\right)\cdot f_x\left(\mathbf{x};\tau,\mathbf{x}_{wi},\tau_{wi}\right)\right]\cdot dm_0 \quad (25)$$

319 where $f_{\hat{C}_i|M,X_0}\left(\hat{C}_i\middle|m_0,\mathbf{x}_0;\tau\right)$ is the distribution of $\hat{C}_i$, calculated for each observation for a range of possible source

320 masses and a range of possible source locations that includes the entire spatial domain of the model, $f_{\mathbf{X}}\left(\mathbf{x};\tau,\mathbf{x}_{wi},\tau_{wi}\right)$ is

321 the unconditioned backward location probability density function and β_x is a normalization factor.

322 **3 Test case**

323 **3.1 Description of the experimental equipment**

324 The transport experiments were performed in a laboratory device (sandbox) built with polymethyl methacrylate

325 (PMMA) plates. The sandbox reproduces an unconfined aquifer governed by two levels (upstream and downstream).

326 The external dimensions of the sandbox are 1.20m × 0.14m × 0.73m. Along the longest axis x, the sandbox is made of

327 three parts (Fig. 1): two tanks (upstream and downstream), which allow the regulation of the water level and, as a

328 consequence, of the flux, and a central chamber (0.95 m × 0.10m × 0.70m) which contains the porous medium. The

329 water discharge is monitored with a flow meter. The porous medium consists of glass beads with diameter in the range

330 between 0.75 and 1 mm; the hydraulic parameters are summarized in Table 1. An injector was positioned in the

331 upstream part of the sandbox (see Fig. 1), and fluorescein sodium salt was chosen as tracer because, when mixed with

332 water and excited with blue light (λ = 490 nm), it irradiates in longer wavelength (green light, λ = 520 nm). The

333 experimental device was placed in a darkroom to avoid all external light contamination and lightened by 8

334 monochromatic blue LEDs. All variables, such as upstream and downstream level, injected discharge, temperature,

335 background discharge, start and ending of injection, were acquired by means of a data acquisition system. The

336 luminosity at each point of the sandbox was recorded by a digital camera and then converted in concentration through

337 an imaging technique. The mass released by the injector and the one estimated through the image processing were

338 compared to evaluate the reliability of data collected. Another confirmation of the validity of the data used was obtained

339 by comparing the mass rate that flows through the sandbox with the known injected one. During the device calibration,

340 the maximum measurement error of concentration σ_R was estimated as less than 3 mg/l. A detailed description of the

341 sandbox can be found in (Citarella et al. 2010).

342 **3.2 Description of the test**

343   A tracer solution with variable mass rate was injected by using the experimental device described in the previous

344   section. The average background flow rate was measured as 25 ml/s, obtained with upstream and downstream constant

345   heads of 59.9 cm and 53.6 cm, respectively. The injector was located at the coordinates x_{inj}= 14.25 cm, z_{inj} = 32.75 cm,

346   and it was as wide as the central chamber.

347   The test had a time length of 2,200 s. The injection started at time t_{start} = 310 s and finished at t_{end} = 1,800 s, so it had a

348   duration of 1,490 s. The mass rate varied in a range between 0 and approximately 60 μg/s. Through a photographic

349   survey, the tracer concentration was estimated at every point of the domain. Two observation points only were

350   considered for the application of the two methodologies discussed in this paper (Figs. 3 and 4). At such points the

351   concentration was estimated every 5 s for the whole duration of the test (Fig. 4). The contaminant release history, in

352   terms of mass rate, presented three peaks of different magnitude (Fig. 5). In Figs. 4 and 5 the time scale is referred at

353   t_{start}. Two monitoring points only were considered so as to simulate a real case, in which there are generally few points

354   and more observations in time, only. Indeed, the realization of a large number of monitoring wells can be very

355   expensive; moreover, a case with a small number of observation points is more severe.

356   **3.3 Description of the numerical model**

357   The sandbox was represented by a numerical flow and transport model in order to provide an essential tool for the two

358   methodologies.

359   The main assumption is that the flow and transport phenomena are uncoupled and the flow has mainly two-dimensional

360   components in the vertical plane; the porous medium was considered homogeneous and isotropic. The groundwater

361   flow was reproduced with MODFLOW 2000 (Harbaugh et al. 2000) and the transport process using MT3DMS with

362   TVD as advection solver package (Zheng and Wang 1999).

363   The sandbox was described by using a finite difference grid of 192 columns (representing the longitudinal size), 140

364   layers (describing the sandbox height) and one row only (the depth size), resulting in 26,880 computation nodes. The

365   size of each numerical cell was $(\Delta x, \Delta y, \Delta z) = (0.5, 10.0, 0.5)$ cm. The hydraulic parameters (see Table 1) of the

366   numerical flow model were set up by comparing the measured and computed flows in steady state condition (without

367   any injection), assuming that the porous medium is homogeneous and isotropic. The transport parameters (see Table 1)

368   were calibrated by fitting the experimental and numerical breakthrough curves of different experiments (with various

369   injection rates and concentration values) at several monitoring points. As an example, Figure 6 shows the good

370   agreement between the numerical and the observed breakthrough curves at three monitoring points located at different

371   distances (14.75, 33.75, 57.75 cm) downstream the injection point of the present test case.

372   Once the numerical model was validated, it was used for both inverse methodologies: to compute the TFs numerically

373   (for more details see Butera et al 2013), and to build the backward probability model (Neupauer and Wilson 1999).

374   **3.4 Results of the geostatistical approach (GA)**

375   The SA was assumed upgradient from the measurement points in the region $11.25 \leq x \leq 17.25$ cm and

376    $29.75 \leq z \leq 35.75$ cm, and it was subdivided into 9 sub areas as shown in Figure 3. The centroid of each area represents

377   a possible source location. The suspected area is a 6x6 cm square, which is quite large compared to the pore dimension

378   (1 mm) and the dispersivity values (1.6 mm). In fact it is about 60 times the grain dimension and it is comparable with

379   the plume's transverse dimension. At first, the TFs relevant to the two monitoring points and the nine possible sources

380   were computed by applying the SIF procedure, requiring nine runs of the forward flow and transport models. In the test

381   case considered the mass rate released in time and source position are the unknowns. For each monitoring point, 32

382   concentration values were considered available throughout the duration of the injection with a time step equal to 70 s

383 (Fig. 4). Uncertainty associated with each concentration value was considered constant as described in section 3.1. A

384 Gaussian covariance function was chosen $Q(t_i, t_j | \boldsymbol{\theta}) = \sigma_s^2 \cdot \exp(-(t_i - t_j)/\lambda_s^2)$, so variance σ_s^2 and correlation length

385 λ_s^2 were estimated for each of the nine suspected point source. Finally, the SRSI procedure was carried out and the

386 release function for each of the nine possible sources was obtained (Fig. 5).

387 The release history assumes negligible values in all the locations, except in x = 14.25 cm, z = 32.75 cm. This result

388 indicates that the source is located in the sub-area with those centroid coordinates (just the ones of the actual source).

389 The release history recovered for the source location x = 14.25 cm, z = 32.75 cm shows a good agreement between the

390 observed and the peak times recovered; moreover the whole release history is included in the 5-95% confidence

391 interval.

392 A drawback of methodology is that if the number of SA is too high, the methodology could be computationally

393 ineffective, because it requires one run of the numerical model for each suspected source.

3.5 Results of backward model based on adjoint state method (BPM -ASM)

3.5.1 Unconditioned single-observation backward location PDF

396 At first, the single unconditioned backward location probability density functions were calculated at the single

397 observation points P1 and P2 (see Fig. 3). At backward time $\tau = 0$, a tracer solution discharge equal to Q_w and a

398 concentration given by (16), was injected for one time step only.

399 The travel times between the true source and observation points P1 and P2 are respectively $\Delta \tau_{P1} = 340\,s$ and

400 $\Delta \tau_{P2} = 400\,s$. So, the evolution of the backward location PDF, at backward time $\tau = 400\,s$ (Figs. 7a and 7b) was

401 computed. Whereas the porous medium is homogeneous and isotropic, the shapes of the contours are symmetric as

402 shown in Figs. 7a and 7b. Moreover, Fig. 7a shows that for a closer observation point, the backward location PDF is

403 higher, as expected. This spreading out of probability density function is caused by dispersion, similar to the processes

404 that occur in a forward model.

3.5.2 Multiple-observation backward location PDF

406 In this case, two observations of contamination collected at the two observation points are used to calculate the

407 multiple-observation location PDF, given by (22). It was assumed that the particles observed in P1 and P2 were

408 originated by a unique release at the same point source location at the same backward time. As shown in Figure 7c, the

409 dispersion of the backward location is smaller than the two single-observation location PDF (Figs. 7a and 7b). This

410 approach does not allow to consider the concentration measurement, but only the presence or the absence of the

411 contaminant. So, one point detected in a marginal position of the plume could compromise the result because the

412 observations cannot identify the plume shape. In fact, in the backward model, the backward location PDF simply

413 follows the same processes that occur in the forward model, and the resulting multiple-observation backward location

414 PDF is not able to give a weight to observations with a higher concentration value.

3.5.3 Backward location PDF conditioned on measured concentrations

416 In the previous sections the concentration measured at the two observation points was not considered. A lot of

417 information about the concentration in time at both points P1 and P2 was available through the images collected during

418 the laboratory test. However the method requires only the concentration values at the monitoring points collected at a

419 time $\bar{t}$. Several sample times were considered during preliminary studies and all yielded very similar results. In this

420 case, the results due to the concentration sampled at $\bar{t}$ = 590 s after t_{start} are shown. Since the release mass is considered

12

421   unknown, as the source location, $f_{\hat{C}_i|M,X_0}\left(\hat{C}_i|m_0,\mathbf{x}_0;\tau\right)$ was calculated for each observation for a range of possible

422   source masses (50 mg < m_0 < 70 mg) and a range of possible source locations that include the entire spatial domain. An

423   iterative procedure was developed with Matlab (Mathworks 2010) in order to change the source position in every cell,

424   and run the flow and transport models. This approach could be computationally challenging if the number of cells is too

425   high, due either to an excessively extensive investigated area or a high spatial resolution. This could be improved by

426   analyzing the delineated suspected area only, as in the SRSI method. The concentrations observed at the monitoring

427   points used to condition the backward location PDF were 16.97 mg/l for P1 and 3.88 mg/l for P2 (Fig. 4). As described

428   in section 3.1, the measurement error is taken as being constant and equal to 3 mg/l, although this value represents the

429   maximum error, and for small concentrations it is smaller than the one considered. The source location, based on the

430   backward location probability density function, conditioned on concentrations measured, is very well identified. The

431   conditioned backward probability PDF plume has a long and narrow shape, as shown in Figure 7d. This improvement

432   reduces the backward location variance and it allows to increase information on the source.

3.5.4 Sensitivity of BPM-ASM on the assumed release time

434   An assumption on the release time is required when the backward probability model is used, so a study about the effect

435   of this parameter on the results was considered. In the previous sections, the backward release time was fixed at τ =

436   400 s. An error of 10% and 20% of $\Delta\tau_{P2}$ was considered with the aim of studying the results with a wrong time

437   release. Thus, backward release times $\tau_{-10\%}$ = 360 s, and $\tau_{-20\%}$ = 320 s were hypothesized, and the results are shown in

438   Figures 8 and 9. While the results obtained with a 10% error remain acceptable, a 20% error does not allow to identify

439   the true source location. Moreover, by considering an error of +10% ($\tau_{+10\%}$ = 440 s) or +20% ($\tau_{+20\%}$ = 480 s), the source

440   location will be estimated upstream the true one and it will present a larger backward location PDF plume than the one

441   estimated with -10% or -20%.

4. Discussion and Conclusions

443   A comparison between two methods, simultaneous release function and source location identification (SRSI) and

444   backward probability model based on adjoint state method (BPM-ASM), was carried out. For the application of these

445   methodologies, experimental data collected through laboratory equipment under controlled conditions were used. This

446   allowed to perform both methodologies in a real test case in which the true solution was known a priori. A numerical

447   model was calibrated on experimental data with the aim of computing the transfer functions (Butera et al. 2004, 2006)

448   and building the backward model for the BPM-ASM (Neupauer and Wilson 1999, 2000). The SRSI procedure is able to

449   simultaneously recover the release function and identify the source location, but it requires a preliminary delineation of

450   the suspected areas and some weak hypothesis about the statistical structure of the unknown release function. The

451   number of preliminary runs of the forward transport model needed to obtain the numerical TFs is equal to the number of

452   the suspect source locations.

453   The backward location PDF describes the possible former positions of the observed contamination at a fixed time in the

454   past. By using an adjoint model, an instantaneous point source of an adjoint state (related to the backward location

455   PDFs) is released at the observation location. Given the simplicity of the flow field, the exchange of the boundary

456   conditions was sufficient to obtain the backward model. At first, the unconditioned backward-location PDFs were

457   calculated for each monitoring point, by considering the known time release. The multiple-observation PDF was then

458   calculated and finally it was conditioned on the experimental concentration measurement. This method identifies the

459   true source very well. As with the SRSI method, two observation points were used, but while in the SRSI multiple times

were considered (32 sampling times for each monitoring point in this test case), in the BPM-ASM only two observations at a specific time, were required.

In conclusion, both methods allow to identify the true source; the SRSI also allows to recover the release history. While the SRSI method requires the definition of a suspected area, the BPM-ASM requires a known release time: both hypotheses could be strong in certain cases. Finally both methodologies could be computationally inefficient, and if the SRSI requires a forward run for each suspected source, the conditioning on the measured concentrations of multiple-observation PDF requires a forward run for each grid node of the numerical model. The number of the forward runs in the BPM-ASM could be reduced by analyzing the suspected area only. Another important difference between the two methods is that the SRSI works with multiple sources too, while the BPM-ASM is able to recognize one point source only. In fact the multiple-observation PDF is based on the hypothesis that all particles observed at a specific time were originated at the same location from an instantaneous release. It is important to notice that the BPM-ASM performed well in the test case considered too, in which the observations were originated from a complex release history.

In order to overcome the limits of the two methods, a new procedure which includes the best performance of both approaches could be developed: for instance the BPM-ASM could be preliminarily used to identify the suspected areas, by considering several backward times, and then the SRSI can be applied to estimate the true source location and its release function.

Acknowledgements

References

Bear J, Verruijt A (1987) Modeling Groundwater Flow and Pollution. Springer

Box GEP, Cox DR (1964) An Analysis of Transformations. Journal of the Royal Statistical Society Series B (Methodological) 26 (2):211-252

Butera I, Tanda MG (2003) A geostatistical approach to recover the release history of groundwater pollutants. Water Resour Research 39 (12):1372. doi:10.1029/2003WR002314

Butera I, Tanda MG, Zanini A La ricostruzione della storia del rilascio di inquinanti in acquiferi sede di moto non uniforme mediante approccio geostatistico. In: XXIX Convegno di Idraulica e Costruzioni Idrauliche, Trento, 7-10 Settembre 2004 2004. Bios, Cosenza

Butera I, Tanda MG, Zanini A (2006) Use of numerical modeling to identify the transfer function and application to the geostatistical procedure in the solution of inverse problems in groundwater. Journal of Inverse and Ill-Posed Problems 14 (6):547-572. doi:10.1163/156939406778474532

Butera I, Tanda MG, Zanini A (2013) Simultaneous identification of the pollutant release history and the source location in groundwater by means of a geostatistical approach. Stochastic Environmental Research and Risk Assessment 27 (5):1269-1280. doi:10.1007/s00477-012-0662-1

Citarella D, Tanda MG, Zanini A Setup and calibration of an experimental device aimed at the validation of geostatistical procedures. In: IAHR International Groundwater Symposium, Valencia (E), 22-24 September 2010 2010.

Gzyl G, Zanini A, Frączekc R, Kura K (2014) Contaminant source and release history identification in groundwater: a multi step approach. Journal of Contaminant Hydrology, Vol 157: 59–72. ISSN 0169-7722. doi: 10.1016/j.jconhyd.2013.11.006

498   Harbaugh AW, Banta EW, Hill MC, McDonald MG (2000) MODFLOW-2000, the U.S. Geological Survey Modular

499   Ground-Water Model--User Guide to Modularization Concepts and the Ground-Water Flow Process. United States

500   Geological Survey: Open File Report 00-92

501   Jury WA, Roth K (1990) Transfer Functions and Solute Movement through Soil: Theory and Applications. Birkhäuser

502   Verlag, Basel; Boston

503   Kitanidis PK (1995) Quasi-linear geostatistical theory for inversing. Water Resources Research 31 (10):2411-2419.

504   doi:10.1029/95WR01945

505   Kitanidis PK (1996) On the geostatistical approach to the inverse problem. Advances in Water Resources 19 (6):333-

506   342. doi:10.1016/0309-1708(96)00005-X

507   Mathworks (2010) MATLAB 7.10.0. 7.10.0 edn. The MathWorks Inc., Natick, MA

508   Michalak AM, Kitanidis PK (2002) Application of Bayesian inference methods to inverse modeling for contaminant

509   source identification at Gloucester Landfill, Canada. In: Hassanizadeh SM, Schotting RJ, Gray WG, Pinder GF (eds)

510   Computational Methods in Water Resources XIV, vol 2. Elsevier, Amsterdam, pp 1259-1266

511   Michalak AM, Kitanidis PK (2003) A method for enforcing parameter nonnegativity in Bayesian inverse problems with

512   an application to contaminant source identification. Water Resources Research 39 (2):1033.

513   doi:10.1029/2002WR001480

514   Michalak AM, Kitanidis PK (2004a) Estimation of historical groundwater contaminant distribution using the adjoint

515   state method applied to geostatistical inverse modeling. Water Resources Research 40 (8):W08302.

516   doi:10.1029/2004wr003214

517   Michalak AM, Kitanidis PK (2004b) Application of geostatistical inverse modeling to contaminant source identification

518   at Dover AFB, Delaware. Journal of Hydraulic Research 42 EXTRA ISSUE:9-18

519   Neupauer RM (2002) Probabilistic Identification of Groundwater Contamination Sources.

520   Neupauer RM, Lin R (2006) Identifying sources of a conservative groundwater contaminant using backward

521   probabilities conditioned on measured concentrations. Water Resour Research 42:W03424.

522   doi:10.1029/2005WR004115

523   Neupauer RM, Lin R, O'Shea H (2007) Conditioned Backward Probability Modeling to Identify Sources of Reactive

524   Groundwater Contaminants. Water Resour Research 43:W11403. doi:10.1029/2006WR005580

525   Neupauer RM, Wilson JL (1999) Adjoint method for obtaining backward-in-time location and travel time probabilities

526   of a conservative groundwater contaminant. Water Resources Research 35 (11):3389-3398

527   Neupauer RM, Wilson JL (2001) Adjoint-derived location and travel time probabilities for a multidimensional

528   groundwater system. Water Resources Research 37 (6):1657-1668. doi:10.1029/2000wr900388

529   Neupauer RM, Wilson JL (2002) Backward probabilistic model of groundwater contamination in non-uniform and

530   transient flow. Advances in Water Resources 25 (7):733-746. doi:http://dx.doi.org/10.1016/S0309-1708(02)00073-8

531   Neupauer RM, Wilson JL (2003) Backward location and travel time probabilities for a decaying contaminant in an

532   aquifer. Journal of Contaminant Hydrology 66 (1):39-58. doi:http://dx.doi.org/10.1016/S0169-7722(03)00024-X

533   Neupauer RM, Wilson JL (2004) Forward and backward location probabilities for sorbing solutes in groundwater.

534   Advances in Water Resources 27 (7):689-705. doi:http://dx.doi.org/10.1016/j.advwatres.2004.05.003

535   Neupauer RM, Wilson JL (2004) Numerical Implementation of a Backward Probabilistic Model of Ground Water

536   Contamination. Ground Water 42 (2):175-189. doi:10.1111/j.1745-6584.2004.tb02666.x

537   Neupauer RM, Wilson JL (2005) Backward probability model using multiple observations of contamination to identify

538   groundwater contamination sources at the Massachusetts Military Reservation. Water Resour Research 41:W02015.

539   doi:10.1029/2003WR002974

540   Neupauer RM, Wilson JL, Bhaskar A (2009) Forward and backward temporal probability distributions of sorbing

541   solutes in groundwater. Water Resources Research 45 (1):W01420. doi:10.1029/2008wr007058

542   Snodgrass MF, Kitanidis PK (1997) A geostatistical approach to contaminant source identification. Water Resources

543   Research 33 (4):537-546

544   Zheng C, Wang PP (1999) MT3DMS: A modular three-dimensional multispecies transport model for simulation of

545   advection, dispersion, and chemical reactions of contaminants in groundwater systems; Documentation and user's

546   guide. U.S. Army Engineer Research and Development Center No. SERDP-99-1, Vicksburg, MS

547

Table 1 Transport and hydraulic parameters of the numerical model.

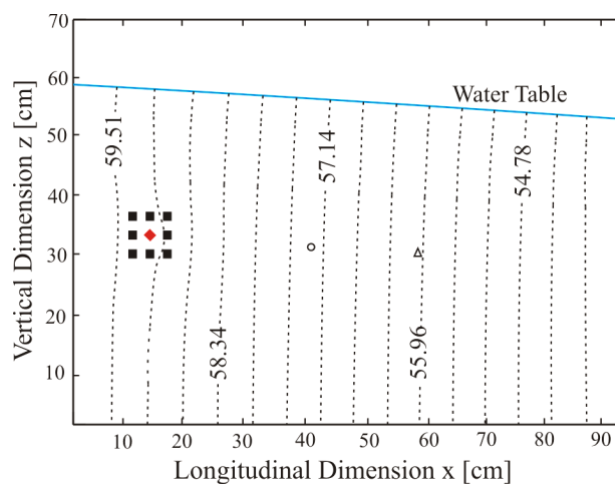|	
Hydraulic Conductivity [cm/s]	0.652
Porosity	0.37
Specific Storage Coefficient [cm^{-1}]	10^{-4}
Longitudinal Dispersivity [cm]	0.16
Transverse Dispersivity [cm]	0.05
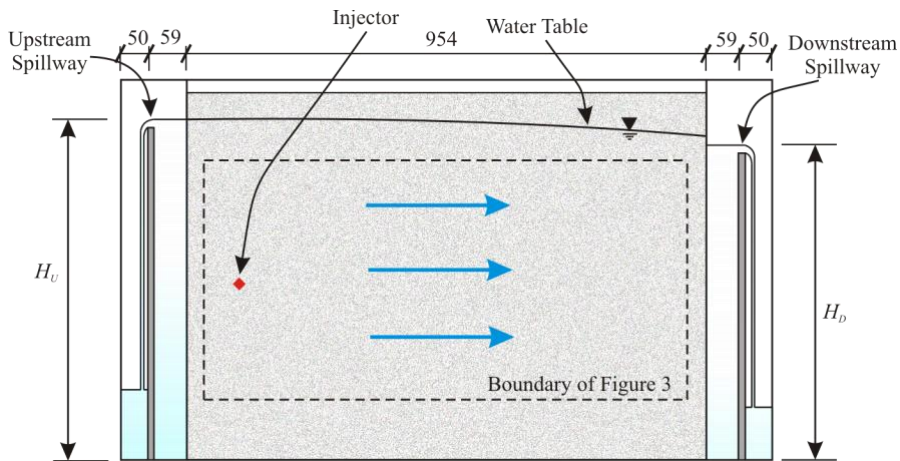
Figures

Figure 1 Sketch of the experimental device (lateral view): constant head boundaries upstream (H_U) and downstream (H_D); the red diamond is the source location. Dimensions are in mm.

Figure 2 Hydraulic head distribution (in cm) from the numerical model. The red diamond is the source location. The filled squares denote the possible sources in the SA. The open circle and the triangle represent observation points P1 and P2, respectively. The experimental equipment reproduced an unconfined aquifer.

17
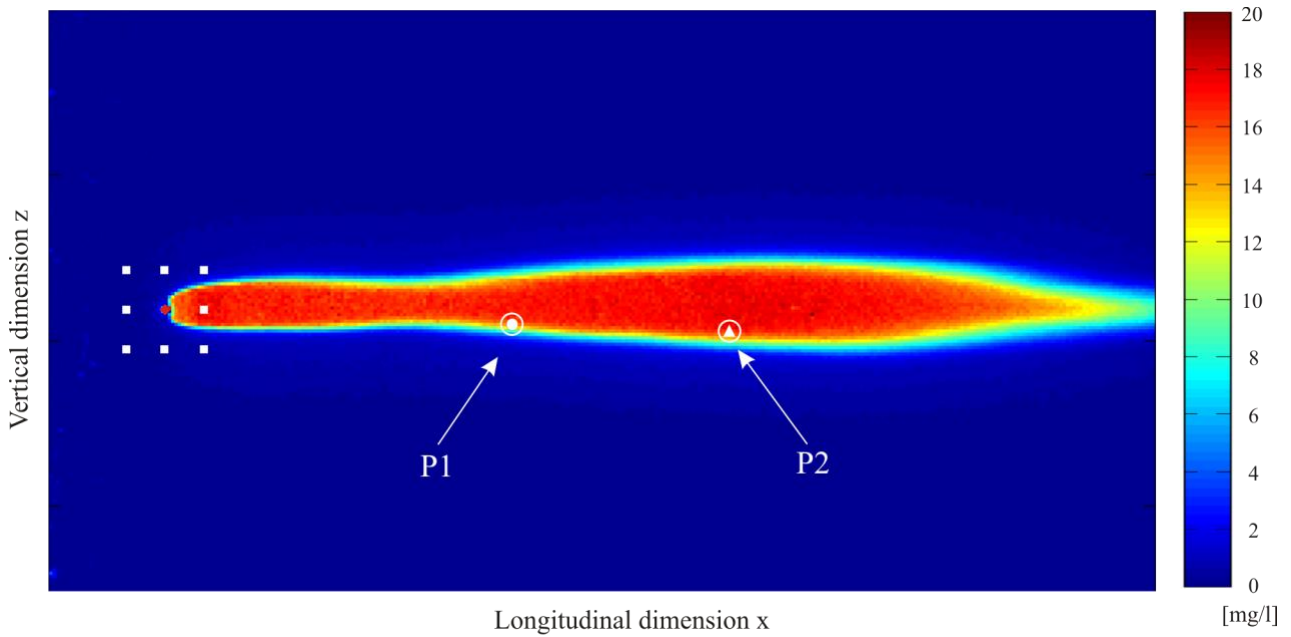
562

563 **Figure 3** Concentration field estimated through the analysis of the image collected 690 s after the start of the injection.

564 The concentration field depicted (85.28 x 44.15 cm) corresponds to the rectangle indicated in Figure 1. The red

565 diamond is the source location. The white filled squares denote the possible sources in the SA.
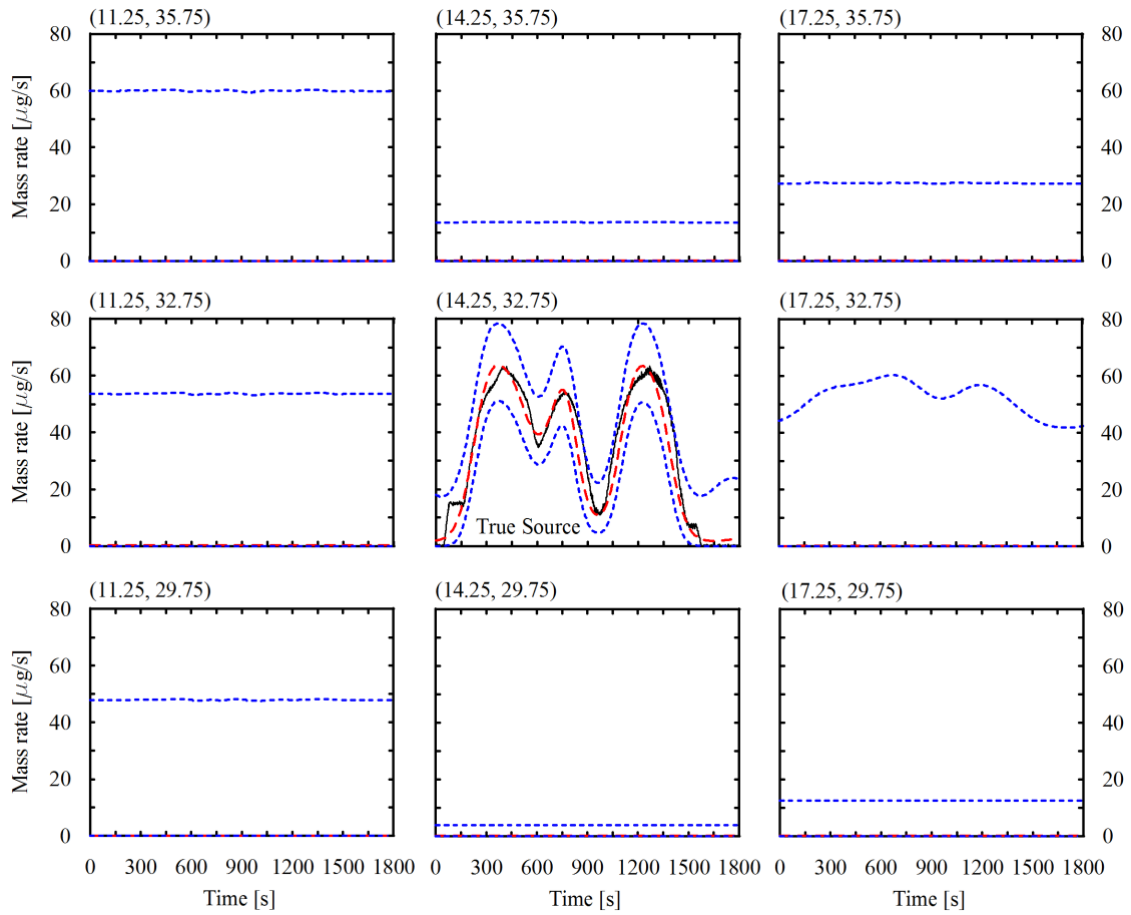
566
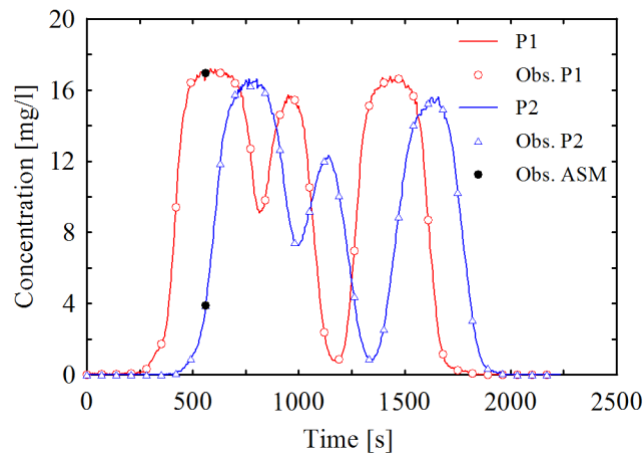
567

568 **Figure 4** Concentration observed at the two monitoring points (solid line). Black dots are the observations used to

569 condition the multiple-observation PDF in BPM-ASM method. Time 0 s represents the time at which injection starts

570 (t_{start}).

Figure 5 The release history recovered at the hypothesized source locations: the true solution (solid black line), best estimate (red dashed line) and 5-95 % confidence interval (blue dotted line). Coordinates of the sources in cm. Time 0 s represent the time at which injection starts (t_{start}).
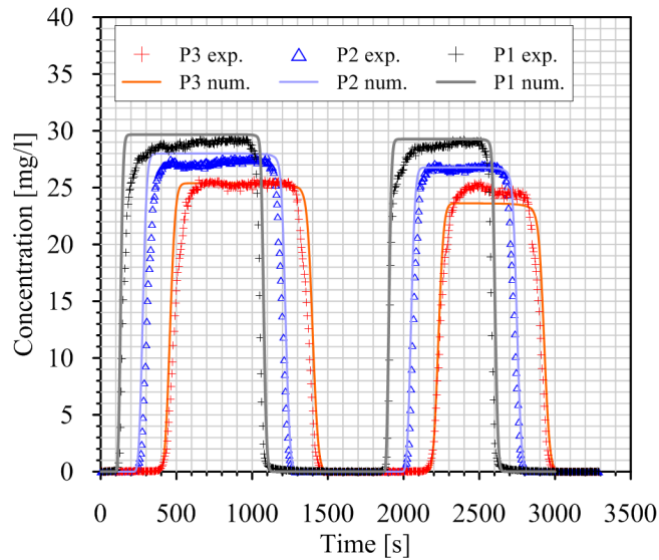
**Figure 6**. Comparison between numerical and experimental breakthrough curves.
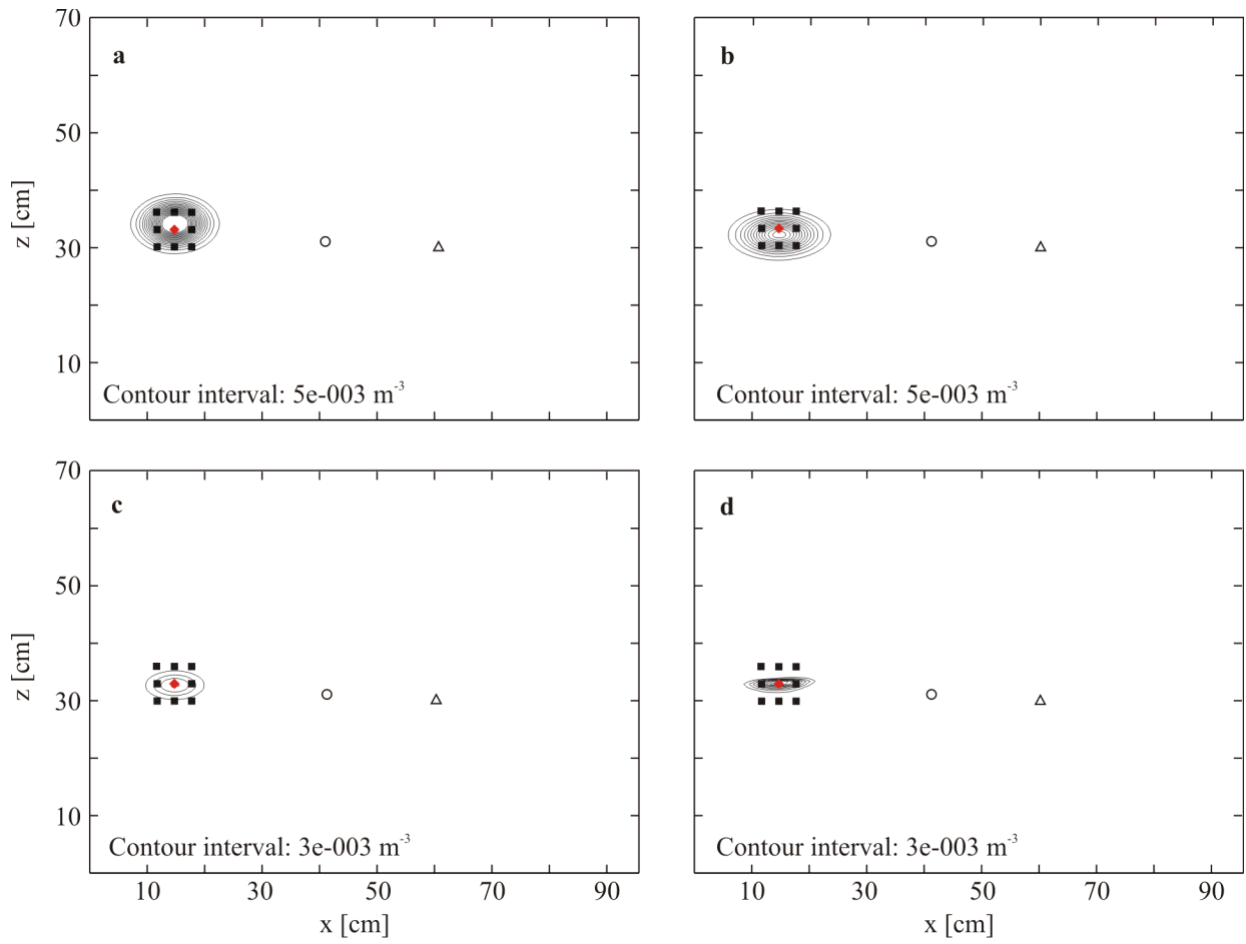
Figure 7 Backward location probability density function at τ = 400 s. (a) Unconditioned backward location PDF: observation point P1. (b) Unconditioned backward location PDF: observation point P2. (c) Multiple-observation (P1 and P2) backward location PDF (d) Conditioned Backward Location PDF on concentrations.
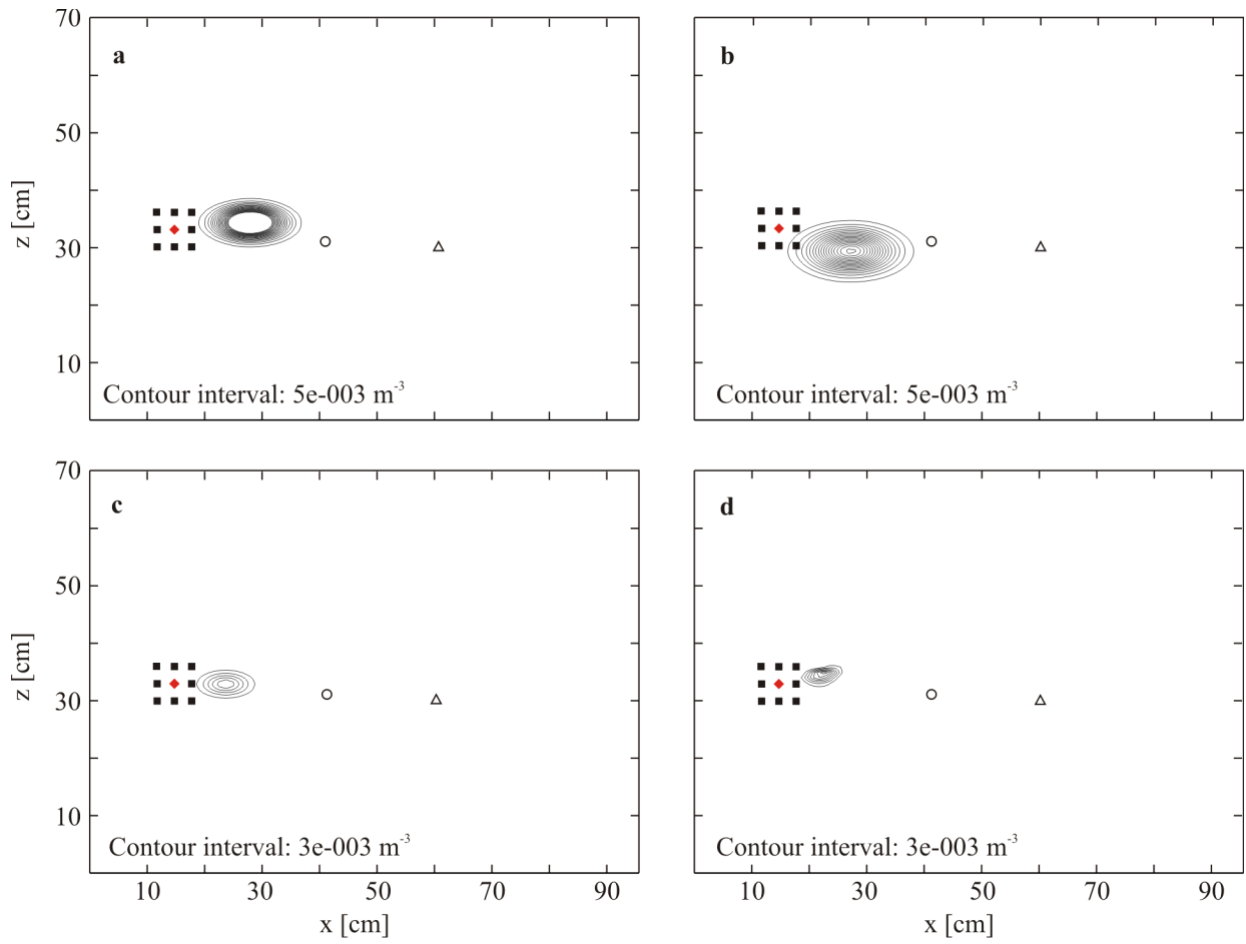
582
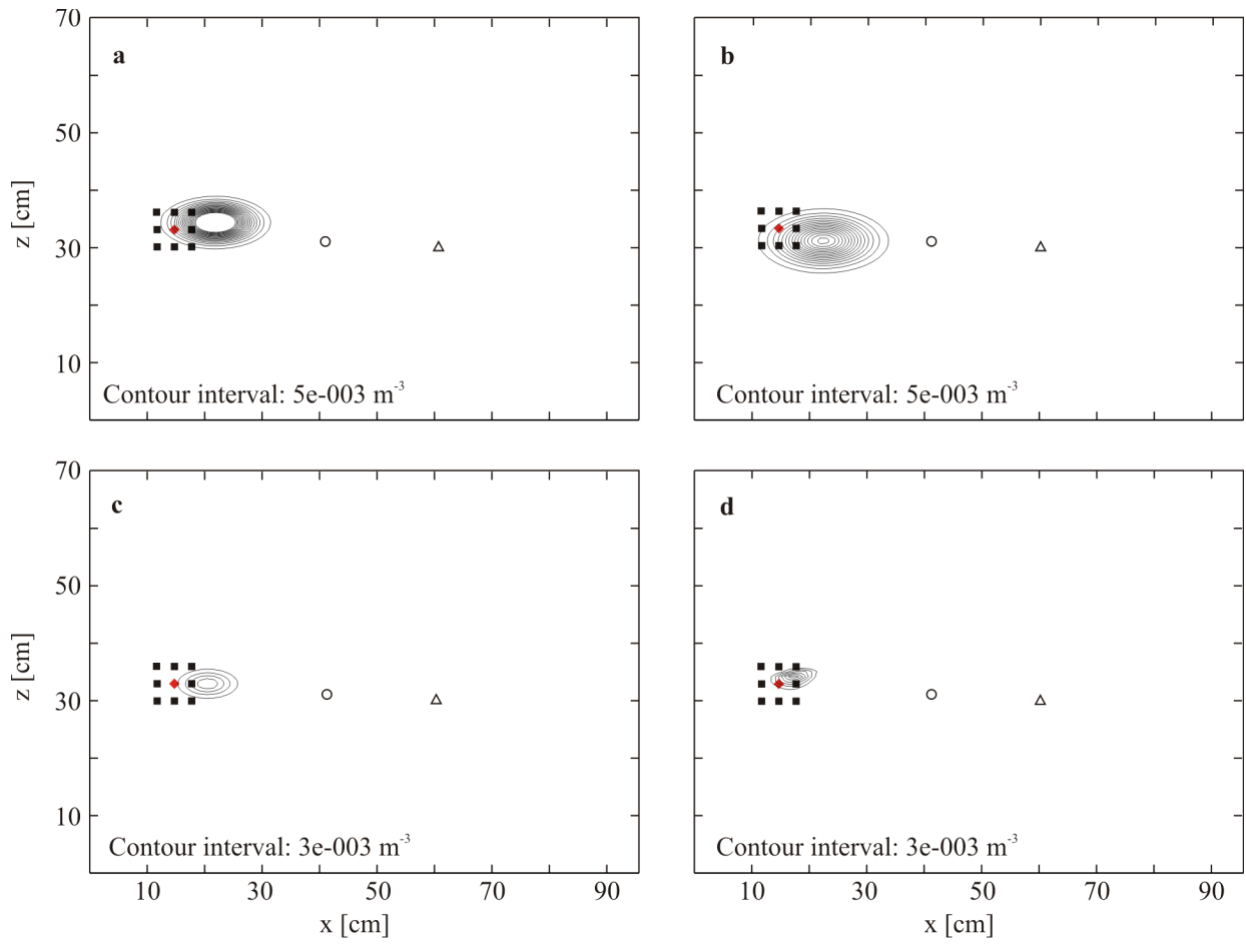
Figure 8 Backward location probability density function at $\tau_{-10\%} = 360$ s. (a) Unconditioned backward location

PDF: observation point P1. (b) Unconditioned backward location PDF: observation point P2. (c) Multiple-observation

(P1 and P2) backward location PDF (d) Conditioned Backward Location PDF on concentrations.

586

Figure 9 Backward location probability density function at $\tau_{-20\%} = 320$ s. (a) Unconditioned backward location
PDF: observation point P1. (b) Unconditioned backward location PDF: observation point P2. (c) Multiple-observation
(P1 and P2) backward location PDF (d) Conditioned Backward Location PDF on concentrations.