A data completion method for identifying pollution intrusion in aquifers

Lamia Guellouz1,2 and Faten Khayat2

The data completion method combined with the modeling of water flow and solute transport in saturated porous media is used twice to identify the extent of polluted water intrusion from the downstream boundary of an aquifer. The aim of this work is to solve the problem even though data are missing on some of the boundaries, by exploiting the fact that there are over specified data on the other boundaries. Indeed, it is assumed that the flow and the water head are known on the upstream boundary while the pollutant concentrations are known on the downstream boundary. The method developed in this work allows to determine the flows, water heads and pollutant concentrations in the whole domain. The method is applied in different aquifer configurations with pumping wells, the results are satisfactory. The model developed has shown its effectiveness in detecting the intrusion of polluted water and can be used in real cases.

Groundwater is often the only resource available in arid and semi-arid areas. These resources have long been considered protected by the overlying soil layers and the self-cleansing power of the soil1. Due to the intensification of industrial, agricultural and mining activities, and the large number of resulting by-products, many aquifers are experiencing deterioration in water quality. The presence of undesirable substances in that water makes it unfit for consumption as drinking water or for use in irrigation etc.1. The pollutants are of various natures: pathogenic organisms, chemical substances dissolved in water, and organic contaminants, mainly hydrocarbons that form a non-aqueous liquid phase. Pollution, most often of anthropogenic origin, comes from the spilling or spreading of polluting products on the surface of the soil, or their injection into wells or their burial in the soil2. These products migrate until they reach the water table. The aquifer is said to be polluted if the concentrations of these contaminants in the water exceed certain limits by practical standards. In coastal aquifers, marine intrusion is one of the main causes of water quality degradation through salinization. Indeed, coastal areas concentrate most human activities, and aquifers are overexploited. The consequence of marine intrusion is the mixing of aquifer water with seawater, the former becomes saline and wells must be abandoned3. The degradation of water quality frequently concerns a particular zone of the aquifer domain and it is important to specify this zone and to prohibit any exploitation there. The delimitation of polluted areas of an aquifer is first and foremost a matter of public health and environmental protection. It is also a prerequisite for any remediation or clean-up and restoration operation. The processes of groundwater contamination by pollutant-laden water or by marine intrusion are essentially due to convection induced by the difference in density between fresh water and polluted water or seawater. The physical and numerical study of this phenomenon must therefore be carried out by considering the vertical extension of the domain in 2D or even 3D4. However, in the context of a large-scale regional study, it may be useful to, first, carry out a horizontal study in 2D to delineate, in a first iteration, the extension of the contamination of the aquifer and thus target the affected areas. Then, in these areas, a study involving density convection processes, can be performed in 3D or on a vertical section in order to better specify the extent of contamination5,6.

Consider an aquifer adjacent to a polluted area, which may be a polluted lake, a landfill, a sewage discharge, another aquifer contaminated by nitrates, heavy metals or marine intrusion; it would be wise to be able to give instructions for the use of this aquifer by rationing the pumping rates of wells to avoid contamination. Numerical modeling is a useful tool for studying such problem, however, scientists and especially modelers are in many cases confronted with missing data, such as aquifer delineation, hydrodynamic characteristics, upstream and downstream flow conditions, salt or pollutant transport parameters, salt or pollutant concentrations at boundaries. Inverse problem solving is used to identify missing data. Many studies are performed in hydrogeology using these methods. Comprehensive and exhaustive reviews of the use of inverse problems in hydrogeology are available.

1LMHE-ENIT-Université Tunis El Manar, BP 37, 1002 Tunis Le Belvédère, Tunisia. 2LAMSIN-ENIT-Université Tunis El Manar, Tunis, Tunisia. *email: lamia.guellouz@enit.utm.tn
such as those by De Marsily et al.\textsuperscript{7} and Zhou et al.\textsuperscript{8}. Many papers deal with parameter identification\textsuperscript{9–11}, while others focus on determining boundary conditions and positions and flows of wells\textsuperscript{12–14}.

The objective of this work is to develop a numerical model to study and quantify the pollution of an aquifer, exploited by pumping wells, and located upstream of a polluted area. In this case, the direction of the flow, from or towards this zone, is dependent on the pumping rates of the wells. In the case of pollution intrusion, the extent of the pollution should be determined. Since not all boundary conditions of this aquifer are known a priori, the data completion method is used to solve the problem. This method has been applied in several fields, including medical diagnosis, remote sensing, traffic studies, and groundwater flow\textsuperscript{15}. The data completion method has also been used for marine intrusion problems in\textsuperscript{6,15,16}. In\textsuperscript{6}, the authors applied this technique to locate the landsea interface in a coastal aquifer. In\textsuperscript{15}, this method was used by Mansouri et al. to identify the well position and pumping rate from the knowledge of the hydraulic head and flow on a part of the domain boundary. In\textsuperscript{16}, this method was used by Mansouri et al. to identify the well position and pumping rate from the knowledge of the hydraulic head and flow on a part of the domain boundary. In\textsuperscript{15}, Bel Hadj Hassine et al. have reconstructed the fluid flow and the concentration of a solute on a non-accessible part of the boundary for the advection-diffusion equations. In all these studies, only stationary flow problems were considered. In this work, two coupled problems are dealt with: the flow of water in the aquifer and the transport of the pollutant as a solute. Boundary conditions can be imposed pressures, flow rates or solute concentrations. One or more of these conditions may be known on one boundary and missing on another. The inverse algorithm, based on data completion, determines the missing values and thus solves the problem. The novelty of this work lies in the resolution of two coupled problems of flow and transport, which are also non-stationary. The partial differential equations of the two direct problems are solved by the finite element method. The model is tested and used to identify pollution intrusion in an aquifer due to a high pumping rate. The direction and intensity of the exchange flow of the aquifer with the polluted zone are determined and, if pollution intrusion occurs, the concentrations of the pollutant in the aquifer are calculated.

Problem statement

The study and monitoring of an aquifer threatened by pollution intrusion requires knowledge of the geometry of the aquifer, its hydrodynamic characteristics such as transmissivity and porosity, and the diffusion-dispersion parameters of the pollutant propagating as a solute. Upstream and downstream water flows, water heads and pollutant concentrations on the boundaries must be determined. Some values of these physical quantities are measured or determined by previous studies and others must be calculated.

To study the problem of pollution intrusion in an aquifer, and as a primary approach, a theoretical case with a simple geometry and constant flow and transport parameters is considered. The unconfined aquifer is a rectangular domain (Fig. 1), it is bordered by a polluted zone on the lower side $\Gamma_s$ and the upper side $\Gamma_U$ can be the upstream boundary of the aquifer or a line of piezometers or homogeneous flow. A pumping well is located at point $S_k$. The flow regime in the aquifer is the primary factor governing pollution intrusion. Pollution intrusion occurs when the flow of water from the aquifer into the polluted zone decreases, causing the contaminated water to advance into the aquifer.

The mixing zone between contaminated and clean water, where density-induced flow occurs at high concentrations, is called the transition zone. This zone may be wide, narrow, or approximated by a sharp interface, depending on the characteristics of the aquifer and the nature of the pollutant. A large-scale domain is adopted in this work. By comparison, the mixing zone is narrow. The trace pollutant is dissolved in water, so the density effect is not taken into account. The flow in the aquifer is essentially horizontal according to the conception of
underground hydrodynamics under the assumptions of Dupuit\textsuperscript{2,14}. Polluted water and fresh water are considered as one single liquid phase with different concentrations of pollutant as solute carried by the water. The solute transport processes taken into account are convection, diffusion and dispersion. Therefore, two coupled problems are solved: water flow and solute transport in porous media. They are described by unsteady partial differential equations, and need initial conditions and boundary conditions (see Eqs. (1) and (2) below). Under natural conditions, the upstream boundary of the aquifer is \( \Gamma_T \) and the downstream boundary is \( \Gamma_G \). Anthropogenic action: high well pumping rate, leads to pollution intrusion by reversing the direction of flow. The \( \Gamma_G \) boundary is, a priori, a hydrodynamically well defined. The hydraulic head can be measured by a series of piezometers. The flow rate, perpendicular to \( \Gamma_G \), can be determined according to the Dupuit hypothesis, by hydraulic head measurements given by two rows of piezometers separated by a few meters. It can also be the result of previous studies, or it can be a well known aquifer boundary such as the discharge of an upstream aquifer. It is also assumed that this area is beyond the influence of the contaminated water intrusion. On \( \Gamma_S \), the polluted side boundary, the hydraulic head and water flow are unknown and variable. The flow can be either towards the polluted area (discharge case) or towards the aquifer (intrusion situation). This boundary can also be in a mixed situation i.e. partly discharge and partly intrusion. Since the water flow is essentially parallel to \( \Gamma_S \), the well pumping rate, leads to pollution intrusion by reversing the direction of flow. The inaccessible part of the boundary \( \Gamma_S \) is well known at the polluted side, so for the inverse problem (PTI), \( \Gamma_G \) is the prescribed flux and concentration on the corresponding lateral boundary \( \Gamma_S \). For any portion \( i \) of the boundary \( \partial \Omega \), we note \( H^{i}(\Omega) \) the space of elements of \( H^1(\Omega) \) the space of the traces over \( \Gamma \) of elements of \( H^1(\Omega) \). We designate by \( V^1(\Gamma) = H^{0\tau}_{\text{co}}(\Gamma) \) the set of all the restrictions to \( \Gamma \) of the functions of \( H^{1\tau}(\partial \Omega) \) that vanish on \( \partial \Omega \setminus \Gamma \). It’s topological dual space will be noted \( V^{-1}(\Gamma) \).

By the Darcy law, the velocity \( V \) satisfies the relation:

\[
V = - T \text{grad}(h)
\]

where \( h \) is the hydraulic head, obtained as a solution of the flow problem in saturated porous media:

\[
\begin{align*}
\left(\text{PTI}^H\right) \quad \frac{\partial h}{\partial t} - \text{div}(T \nabla h) &= \sum_{k=1}^{2} Q_k \delta(x - S_k) \quad \text{in } \Omega \times [t_0, t_f] \\
\nabla h \cdot n &= \delta_i \quad \text{on } \Gamma_T \times [t_0, t_f] \\
\nabla h \cdot n &= H_i \quad \text{on } \Gamma_G \times [t_0, t_f] \\
h &= h_i \quad \text{on } \Gamma_s \times [t_0, t_f] \\
h(x, y, t_0) &= 0 \quad \text{in } \Omega
\end{align*}
\]

Data completion

For \( \Omega \subset \mathbb{R}^2 \) a lipschitz bounded and connected domain such that \( \partial \Omega \) is composed of three disjoint parts of non-vanishing measure: \( \Gamma_T \) is the polluted side, \( \Gamma_S \) is the lateral sides and \( \Gamma_G \) corresponds to the upstream side. Consider the problem of pollutant transport as a dissolved solute in the aquifer water, at a concentration \( C \):

\[
\begin{align*}
\left(\text{PTI}\right) \quad \frac{\partial C}{\partial t} - \text{div}(D V C - V C) &= \sum_{k=1}^{2} Q_k \delta(x - S_k) \quad \text{in } \Omega \times [t_0, t_f] \\
(D V C) \cdot n &= \Phi_c \quad \text{on } \Gamma_T \times [t_0, t_f] \\
\nabla C \cdot n &= \delta_i \quad \text{on } \Gamma_G \times [t_0, t_f] \\
C &= \Psi_c \quad \text{on } \Gamma_G \times [t_0, t_f] \\
C(x, y, t_0) &= C_0(x, y) \quad \text{in } \Omega
\end{align*}
\]

\( \Phi_c \) and \( \Psi_c \) are the overspecified data corresponding respectively to the prescribed flux and concentration on the accessible part of the boundary \( \Gamma_G \), on which measures are available. \( D \) is the hydrodynamic diffusion-dispersion second-order tensor and \( V \) is the Darcy velocity. We make the assumption that the matrix \( D \) is symmetric positive definite and that \( V \cdot n > 0 \) on \( \Gamma_G \). For \( k = 1, 2, Q_k \) is the well abstraction corresponding to the \( k \)th point source \( S_k \) whose coordinates are \((x_k, y_k)\). \( \delta \) denotes the Dirac distribution.

We denote by \( H^1(\Omega) = \{ u \in L^2(\Omega), \partial u_i/\partial x_j \in L^2(\Omega) \quad \text{for } i, j = 1, 2 \} \) the standard Sobolev space equipped with usual first-order Sobolev norm. For any portion \( i \) of the boundary \( \partial \Omega \), we note \( H^{1\tau}(\Gamma) \) the space of the traces over \( \Gamma \) of elements of \( H^1(\Omega) \). We designate by \( V^1(\Gamma) = H^{0\tau}_{\text{co}}(\Gamma) \) the set of all the restrictions to \( \Gamma \) of the functions of \( H^{1\tau}(\partial \Omega) \) that vanish on \( \partial \Omega \setminus \Gamma \). It’s topological dual space will be noted \( V^{-1}(\Gamma) \).

First, we suppose that the hydraulic head is unknown on the polluted side, so for the inverse problem \( (\text{PTI}^H) \), \( \Gamma_T = \Gamma_G \), and \( \Gamma_S = \Gamma_{\text{poll}} \). For the pollutant transport inverse problem, the concentration \( C \) is well known at the polluted side, which is the \( \Gamma_{\text{poll}} \) for \( (\text{PTI}) \) and unknown at the upstream side. So we will use the reconstructed hydraulic head on the polluted side and the Darcy law to reconstruct concentration on \( \Gamma_G = \Gamma_T \).

In order to reconstruct the missing hydraulic head and concentration on the corresponding \( \Gamma_G \), we use a method developed by Andrieux et al.\textsuperscript{19}. We first introduce a fictious inner boundary \( \Gamma_{\text{poll}} \) excluding the region containing the wells. Thus we obtain a fictious domain \( \Omega_{\text{F}} \) such that \( \partial \Omega_{\text{F}} = \Gamma_N \cup \Gamma_G \cup \Gamma_{\text{F}} \cup \Gamma_F \). In the sequel, we...
note \( \Gamma_H = \Gamma_D \cup \Gamma_F \) the new inaccessible boundary on which we must reconstruct the data. The Cauchy problem consists then in finding \((t, \varphi)\) such that:

\[
\begin{align*}
\frac{\partial C}{\partial t} - \text{div}(D\nabla C - VC) &= 0 \quad \text{in } \Omega_F \times [t_0, t_f] \\
\frac{\partial h}{\partial t} - \text{div}(T\nabla h_i) &= 0 \quad \text{in } \Omega_F \times [t_0, t_f] \\
(D\nabla C) \cdot n &= \Phi_c \quad \text{on } \Gamma_N \times [t_0, t_f] \\
(D\nabla C) \cdot n &= \Psi_c \quad \text{on } \Gamma_D \times [t_0, t_f] \\
(C(x, y, t_0) &= C_0(x, y) \quad \text{in } \Omega_F \\
\end{align*}
\]

(3)

Then, we split each one of the problems (2) and (3) into two mixed and well-posed problems which use simultaneously the overspecified data: the first one uses the Dirichlet condition on \( \Gamma_C \) as given data and called Dirichlet problem \( (P_D^H) \), the second one uses the Neumann condition as given data on \( \Gamma_C \) and consequently called Neumann problem \( (P_N^H) \):

\[
\begin{align*}
\frac{\partial h_1}{\partial t} - \text{div}(T\nabla h_1) &= 0 \quad \text{in } \Omega_F \times [t_0, t_f] \\
\frac{\partial h_2}{\partial t} - \text{div}(T\nabla h_2) &= 0 \quad \text{in } \Omega_F \times [t_0, t_f] \\
(D\nabla C_1) \cdot n &= \lambda \quad \text{on } \Gamma_D \times [t_0, t_f] \\
(C(x, y, t_0) &= C_0 \quad \text{in } \Omega_F \\
\end{align*}
\]

(4)

(5)

We do the same thing for (3):

\[
\begin{align*}
\frac{\partial C_1}{\partial t} - \text{div}(D\nabla C_1 - VC_1) &= 0 \quad \text{in } \Omega_F \times [t_0, t_f] \\
\frac{\partial C_2}{\partial t} - \text{div}(D\nabla C_2 - VC_2) &= 0 \quad \text{in } \Omega_F \times [t_0, t_f] \\
(D\nabla C_1) \cdot n &= \lambda \quad \text{on } \Gamma_D \times [t_0, t_f] \\
(C(x, y, t_0) &= C_0 \quad \text{in } \Omega_F \\
\end{align*}
\]

(6)

(7)

\( C_1 \) and \( C_2 \) are equal only when the pair \((\tau, \lambda)\) meets the real data \((t, \varphi)\) on \( \Gamma_H \).

In order to make the problem "symmetric," we will define some function changes. Since \( D \) is supposed to be symmetric positive definite, we define the vector \( K \) and the positive constant \( \varphi \) such that:

\[ 2DK - \nabla \varphi = 0 \quad \text{and} \quad \varphi = K \cdot \nabla \varphi. \]

Now, for \( X = (x, y) \in \Omega_F \), we introduce new functions\(^{20}\): \( U_1 = e^{-K \cdot X} C_1 \) and \( U_2 = e^{-K \cdot X} C_2 \). These functions are solutions of the systems:

\[
\begin{align*}
\frac{\partial U_1}{\partial t} - \text{div}(D\nabla U_1) + \varphi U_1 &= 0 \quad \text{in } \Omega_F \times [t_0, t_f] \\
(D\nabla U_1) \cdot n + \frac{\varphi}{2} V \cdot U_1 &= 0 \quad \text{on } \Gamma_D \times [t_0, t_f] \\
U_1(x, y, t_0) &= C_0 \quad \text{in } \Omega_F \\
\end{align*}
\]

(8)

\[
\begin{align*}
\frac{\partial U_2}{\partial t} - \text{div}(D\nabla U_2) + \varphi U_2 &= 0 \quad \text{in } \Omega_F \times [t_0, t_f] \\
(D\nabla U_2) \cdot n + \frac{\varphi}{2} V \cdot U_2 &= 0 \quad \text{on } \Gamma_D \times [t_0, t_f] \\
U_2(x, y, t_0) &= C_0 \quad \text{in } \Omega_F \\
\end{align*}
\]

(9)

where we have set
\[
\begin{align*}
\mathbf{V}_e &= e^{-KX}\mathbf{v}_e, \\
\mathbf{V}_\xi &= e^{-KX}\mathbf{v}_\xi, \\
\Lambda &= e^{-KX}\mathbf{\lambda}, \\
\Theta &= e^{-KX}\mathbf{\tau}.
\end{align*}
\]

So, let us define the functional
\[
\mathcal{E}(\Lambda, \Theta) = \frac{1}{2} \int_0^t \int_\Omega \left( ((D^{1/2}\nabla U_1 - D^{1/2}\nabla U_2)^2 + \rho |U_1 - U_2|^2 \right) dx + \frac{1}{4} \int_0^t \int_\Gamma \cdot \nabla U_1 - U_2)^2
\]

**Remark 1**

1. \( U_1 \) and \( U_2 \) are equal only when the pair \((\Lambda, \Theta)\) meets the real data \((e^{-KX}\mathbf{v}, e^{-KX}\mathbf{t})\) on \(\Gamma_H\) which is equivalent to say that \((\mathbf{\lambda}, \mathbf{\tau})\) meets the real data \((\mathbf{\varphi}, \mathbf{t})\) on \(\Gamma_H\).
2. \( \mathcal{E}(\Lambda, \Theta) \) reaches its minimum when \( U_1 = U_2 = U \) with \( U = e^{-KX}C \) and where \( C \) is the unique solution to the data completion problem.
3. The inverse problem (3) is equivalent to the following minimization:
   \[
   (\Pi, \Upsilon) = \arg \min_{(\Lambda, \Theta)} \mathcal{E}(\Lambda, \Theta),
   \]
   with \((\Pi, \Upsilon) = (e^{-KX}\mathbf{v}, e^{-KX}\mathbf{t}), \Theta \in \mathcal{V}(\Gamma_H)\) and \(\Lambda \in \mathcal{V}^{-1}(\Gamma_H)\).
4. \( \mathcal{E} \) is an energy-like error functional that represents the energy gap between the so-called Neumann solution of (9) and the Dirichlet solution of (8) (see [13,21]).

**Theorem 2**

1. The functional \((\Lambda, \Theta) \mapsto \mathcal{E}(\Lambda, \Theta)\) is a positive quadratic functional. It is strictly convex on \(\mathcal{V}^{-1}(\Gamma_H) \times \mathcal{V}(\Gamma_H)\) and consequently has a unique minimum for a compatible data \((\mathbf{V}_e, \mathbf{V}_\xi)\).
2. When \( \mathcal{E} \) reaches its minimum, the solutions \( C_1 \) and \( C_2 \) of the inverse problem (3) verify the first-order optimality conditions:
   \[
   \left\{ \begin{array}{ll}
   C_1 = C_2 + \alpha & \text{on } \Gamma_H \times [t_0, t_f] \\
   (D\nabla C_1 - \frac{1}{2} V C_1) \cdot n = (D\nabla C_2 - \frac{1}{2} V C_2) \cdot n & \text{on } \Gamma_H \times [t_0, t_f]
   \end{array} \right.
   \]
   where \( \alpha \) is a constant.

**Proof**

1. A similar calculus as in [18] shows that the first derivatives of \( \mathcal{E} \) are given by:
   \[
   \frac{\partial \mathcal{E}}{\partial \Lambda} \cdot w = \int_0^t \int_{\Gamma_H} (D\nabla (U_1 - U_2) \cdot n) r^w_1,
   \]
   \[
   \frac{\partial \mathcal{E}}{\partial \Theta} \cdot h = -\int_0^t \int_{\Gamma_H} (D\nabla r^h_2 \cdot n)(U_1 - U_2),
   \]
   where \( r^w_1 \) and \( r^h_2 \) are respectively the solution of
   \begin{equation}
   \left\{ \begin{array}{ll}
   \frac{\partial r^w_1}{\partial t} - \text{div}(D\nabla r^w_1) + \rho r^w_1 = 0 & \text{in } \Omega_F \times [t_0, t_f] \\
   (D\nabla r^w_1 + \frac{1}{2} V \cdot r^w_1) \cdot n = 0 & \text{on } \Gamma_N \times [t_0, t_f] \\
   r^w_1 = 0 & \text{on } \Gamma_C \times [t_0, t_f] \\
   (D\nabla r^w_1 + \frac{1}{2} V \cdot r^w_1) \cdot n = w & \text{on } \Gamma_H \times [t_0, t_f] \\
   r^w_1(t_f) = 0 & \text{in } \Omega_F
   \end{array} \right.
   \end{equation}
   \begin{equation}
   \left\{ \begin{array}{ll}
   \frac{\partial r^h_2}{\partial t} - \text{div}(D\nabla r^h_2) + \rho r^h_2 = 0 & \text{in } \Omega_F \times [t_0, t_f] \\
   (D\nabla r^h_2 + \frac{1}{2} V \cdot r^h_2) \cdot n = 0 & \text{on } \Gamma_N \times [t_0, t_f] \\
   (D\nabla r^h_2 + \frac{1}{2} V \cdot r^h_2) \cdot n = 0 & \text{on } \Gamma_C \times [t_0, t_f] \\
   r^h_2 = h & \text{on } \Gamma_H \times [t_0, t_f] \\
   r^h_2(t_f) = 0 & \text{in } \Omega_F
   \end{array} \right.
   \end{equation}

In the same manner, the second derivatives are
\[
\begin{align*}
\frac{\partial^2 E}{\partial \alpha^2}(w, w) &= \int_{\Omega} \int_{\Gamma} \nabla r_{1}^w \nabla r_{1}^w + \int_{\Omega} \int_{\Gamma} q |r_{1}^w|^2 \\
\frac{\partial^2 E}{\partial \beta^2}(h, h) &= \int_{\Omega} \int_{\Gamma} \nabla r_{2}^h \nabla r_{2}^h + \int_{\Omega} \int_{\Gamma} q |r_{2}^h|^2 + \frac{1}{2} V \cdot n \int_{\Gamma} |r_{2}^h|^2
\end{align*}
\]

which shows the convexity of \(E\). To prove the strict convexity, we use the trace theorem to deduce that we have, for some constant \(c > 0\),

\[
\begin{align*}
\frac{\partial^2 E}{\partial \alpha^2}(w, w) &\geq c \|r_{1}^w\|_{H^1(\Omega_F)}^2 \geq c \|r_{1}^w\|_{V_1(\Gamma_H)}^2 \\
\frac{\partial^2 E}{\partial \beta^2}(h, h) &\geq c \|r_{2}^h\|_{H^1(\Omega_F)}^2 \geq c \|r_{2}^h\|_{V_1(\Gamma_H)}^2 = c \|h\|_{V_1(\Gamma_H)}^2
\end{align*}
\]

2. From Eqs. (12) and (13), we deduce that when \(E\) reaches its minimum, the solutions \(U_1\) and \(U_2\) verify:

\[
\begin{aligned}
U_1 &= U_2 + \alpha \quad \text{on } \Gamma_H, \\
(\nabla U_1) \cdot n &= (\nabla U_2) \cdot n \quad \text{on } \Gamma_H.
\end{aligned}
\]

so by the inverse variable change, we have immediately the conditions (11).

□

Now, we can use the above transmission conditions to set up an iterative procedure among subdomains and based on the preconditioned gradient algorithm:

Algorithm.

1- Initialization: set \(k = 0\) and chosen \(\tau^{(0)}\) and \(\lambda^{(0)}\).

2- The stopping criterion: \(E(\tau^{(k)}, \lambda^{(k)}) \leq \epsilon\) where \(\epsilon\) is a given tolerance level:

(a) solve the problems (6) and (7) using \(\lambda = \lambda^{(k)}\) and \(\tau = \tau^{(k)}\).

(b) compute \(r_1^{(k)}\) and \(r_2^{(k)}\) solutions to two adjoint problems \((A_1)\) and \((A_2)\) defined by:

\[
\begin{aligned}
(A_1) &\quad \left\{ \begin{array}{l}
\frac{\partial r_1^{(k)}}{\partial t} - \text{div}(D\nabla r_1^{(k)} - V r_1^{(k)}) = 0 \\
(D\nabla r_1^{(k)}) \cdot n = 0 \\
r_1^{(k)}(t_0) = 0
\end{array} \right. \quad \text{in } \Omega_F \times [t_0, t_f] \\
&\quad \text{on } \Gamma_N \times [t_0, t_f] \\
&\quad \text{on } \Gamma_e \times [t_0, t_f] \\
&\quad \text{on } \Gamma_H \times [t_0, t_f] \\
&\quad \text{in } \Omega_F
\end{aligned}
\]

\[
\begin{aligned}
(A_2) &\quad \left\{ \begin{array}{l}
\frac{\partial r_2^{(k)}}{\partial t} - \text{div}(D\nabla r_2^{(k)} - V r_2^{(k)}) = 0 \\
(D\nabla r_2^{(k)}) \cdot n = 0 \\
(D\nabla r_2^{(k)} - \frac{1}{2} V r_2^{(k)}) \cdot n = 0 \\
r_2^{(k)}(t_0) = C_1 - C_2
\end{array} \right. \quad \text{in } \Omega_F \times [t_0, t_f] \\
&\quad \text{on } \Gamma_N \times [t_0, t_f] \\
&\quad \text{on } \Gamma_e \times [t_0, t_f] \\
&\quad \text{on } \Gamma_H \times [t_0, t_f] \\
&\quad \text{in } \Omega_F
\end{aligned}
\]

(c) set \(\lambda^{(k+1)} = \lambda^{(k)} - \rho_1 r_1^{(k)}\) and \(\tau^{(k+1)} = \tau^{(k)} - \rho_2 r_2^{(k)}\).

(d) \(k \rightarrow k + 1\).

3- End do.

where \(\rho_1\) and \(\rho_2\) are the relaxation parameters.

**Numerical results**

We consider a rectangular domain with edges \(\ell_1 = 5\) km and \(\ell_2 = 7\) km. The upstream side is the upper edge referred to by \(\Gamma_e = [0, 5] \times [7]\), the polluted side is the lowest one and denoted by \(\Gamma_s = [0, 5] \times [0]\) as depicted in Fig. 1. The lateral edges with zero flux correspond to \(\Gamma_N\). \(\Gamma_H\) is chosen to be the rectangle: \([1.5, 3.5] \times [2, 5]\).

The three computed examples differ by the location and the intensity of the points sources. The transmissivity is taken \(T = 0.1\) m\(^2\)s\(^{-1}\) for the first and the second examples while it is \(T = 0.0001\) m\(^2\)s\(^{-1}\) for the third one. \(D\) is the diagonal matrix such that \(d_{11} = d_{22} = 10^{-6}\). Synthetic data are generated with constant exact values: \(h = 3\)

\[
\begin{aligned}
\frac{\partial^2 E}{\partial \alpha^2}(w, w) &= \int_{\Omega} \int_{\Gamma} \nabla r_{1}^w \nabla r_{1}^w + \int_{\Omega} \int_{\Gamma} q |r_{1}^w|^2 \\
\frac{\partial^2 E}{\partial \beta^2}(h, h) &= \int_{\Omega} \int_{\Gamma} \nabla r_{2}^h \nabla r_{2}^h + \int_{\Omega} \int_{\Gamma} q |r_{2}^h|^2 + \frac{1}{2} V \cdot n \int_{\Gamma} |r_{2}^h|^2
\end{aligned}
\]
on $\Gamma_f$ and $h = 1$ on $\Gamma_f$ for the hydraulic problem corresponding to both first and second examples. For the last one, we took $h = 0$ on $\Gamma_f$. For the transport problem, $C$ is chosen to be constant on each boundary such that: $C = 30$ on $\Gamma_i$, and $C = 1$ on $\Gamma_f$. The stopping criterion is $\varepsilon = 0.015$. We chose $t_0 = 0$ and $t_f = 3$ for first and second examples and $t_f = 2$ for the third one. We initialize with a constant concentration: $C_0 = C(t = 0) = 1$ on $\Omega$ for all cases. All calculations are run under the Freefem Software environment. For time treatment, a finite difference method and precisely the Euler scheme is used. The computation time was for case 1: 417.187 s, for case 2: 211.392 s and for case 3: 238.163 s.

**First example.** This first example referred to in the figures by Case 1 corresponds to two point sources such that $S_1(2, 2.8)$ with $Q_1 = -0.1$ and $S_2(2.8, 3.5)$ with $Q_2 = -0.8$.

Figures 2 and 3 give the reconstructed traces and fluxes for the flow and the transport problem respectively. Figure 4 gives the flow velocity. Since the values are small, they are multiplied by 10 to represent them correctly. In Fig. 11 are shown the isovalues of the pollutant concentrations in the aquifer at the final time.

**Second example.** The second example denoted by Case 2 corresponds to the same coordinates for the point source as in Case 1, but with $Q_1 = -1$ and $Q_2 = -0.02$.

The reconstructed hydraulic head and its flux are given in Fig. 5, while the concentration are depicted in Fig. 6. Figure 7 describes the flow velocity of this second example. In Fig. 12 are shown the isovalues of the pollutant concentrations in the aquifer at the final time.

**Third example.** It’s denoted by Case 3 in the sequel. Here we consider a single point source with location: $S(2.3, 1.2)$ and $Q = -0.0003$. We take $T = 0.0001$, $h = 3$ on $\Gamma_f$, and $h = 0$ on $\Gamma_f$. 

---

**Figure 2.** The reconstructed flow—Case 1.

**Figure 3.** The reconstructed concentration—Case 1.
The reconstructed hydraulic head and the flux are given in Fig. 8, while the concentration is represented in Fig. 9. The flow velocity of this final example is given in Fig. 10.

**Discussion**

Synthetic data are used in the three simulated cases. Boundary conditions and pumping well flows are specifically chosen so that polluted water will invade the aquifer. In all three cases, the hydraulic head and water flow are given on the upstream side while the pollutant concentration is imposed on the other side. Missing data on either boundary are reconstructed by the numerical model and compared to the exact values in Figs. 2, 3 for case 1, Figs. 5, 6 for case 2 and Figs. 8, 9 for case 3. For all cases, the boundary values are correctly identified. In Figs. 4 and 7, the velocity fields are shown for Case 1 and Case 2 respectively. It is pointed out that in both
cases, the velocity vectors converge towards the wells, they cross the domain boundary on the polluted side, showing that there is intrusion of polluted water into the aquifer. Since the pumping fluxes from the wells are larger in case 2, the velocities are therefore larger in this case. The isovalues of the pollutant concentrations in the aquifer at the final time (Figs. 11 and 12) are concentric curves around the well. They show the intrusion of the pollution and its intensification near the well. The third case is specifically simulated with a single off-center well, located near the polluted side, with a low pumping rate, to test the model’s ability to identify a mixed inflow and outflow boundary condition on $\Gamma_s$. Figure 10, shows that the velocity vectors on this boundary are directed toward the aquifer near the well and toward the polluted zone elsewhere. The three simulated cases showed that the numerical model based on the data completion method is able to find the missing boundary conditions for the water flow coupled with solute transport in an aquifer.

Figure 6. The reconstructed concentration—Case 2.

Figure 7. Velocity $\times 10$—Case 2.
**Conclusion**

The main goal of this work was the development of a numerical tool to study the pollution of an aquifer by the intrusion of polluted water from a neighboring aquifer located downstream of the studied one. For this purpose, an inverse procedure based on the data completion method was developed. The direct problem model solves the coupled transient problems of groundwater flow and solute transport. The inverse method was used to reconstruct the missing data at the domain boundaries, thus circumventing the lack of information and exploiting the existence of over-specified data on the other boundaries.

---

**Figure 8.** The reconstructed flow—Case 3.

**Figure 9.** The reconstructed concentration—Case 3.
The numerical model is tested on three examples with synthetic data. In these examples, the aquifers are exploited by pumping wells. To create different situations, well positions and flow rates were varied. The computational results show that the missing boundary conditions are well identified, even in the case of a mixed inlet-outlet boundary. The model is therefore able to answer the questions of the aquifer manager: is there polluted water intrusion? If so, where does it occur? And what is its extent?

As we have proven the validity of the method used and the robustness of the numerical model, we plan to use this promising tool, in future work, on real aquifers to detect the intrusion of polluted water or to locate the intrusion of seawater in coastal aquifers. The challenge would be to deal with complex geometry and variable parameters. Professional code would then be used to solve the direct problem. The FEFLOW, MODFLOW and SUTRA models can be called from user-defined code or from a bash file. SUTRA has already been implemented in an inverse problem solving procedure by Riahi et al. We can therefore consider using them for the direct problem as suggested by Bouhlila and Hariga.

The present work developed tool could also be used to test pollution remediation scenarios by taking advantage of the available data and the low computational cost of the inverse method, thus reducing costly in situ measurement campaigns.

Figure 10. Velocity x 1e4—Case 3.
Figure 11. Pollutant concentrations—Case 1.
Data availability
All data generated or analysed during this study are included in this published article.

Received: 16 February 2022; Accepted: 8 September 2022
Published online: 28 September 2022

References
1. Besbes, M. Hydrogéologie de l'ingénieur, Ed. CPU Tunis (2010).
2. Bear, J. & Cheng, A.H.-D. Modeling Groundwater Flow and Contaminant Transport. Theory and Applications of Transport in Porous Media (Springer Science+Business Media B.V., 2010).
3. Bouzourra, H., Bouhlila, R., Elango, L., Slama, F. & Ouali, N. Characterization of mechanisms and processes of groundwater salinization in irrigated coastal area using statistics. GIS Hydrogeochem. Investig. Environ. Sci. Pollut. Res. 22, 2643–2660 (2015).
4. Misut, P. E. & Voss, C. I. Simulation of seawater intrusion resulting from proposed expanded pumpage in New York City, USA. Dev. Water Sci. 55, 1595–1606 (2004).
5. Sherif, M. & Al-Rashed, M. F. Vertical and horizontal simulation of seawater intrusion in the Nile Delta aquifer. In First International Conference on Saltwater Intrusion and Coastal Aquifer, Monitoring, Modeling, and Management. Essaouira, Morocco, April 23–25 (2001).
6. Hariga, N. T., Baranger, T. N. & Bouhlila, R. Land-sea interface identification and submarine groundwater exchange (SGE) estimation. Comput. Fluids 88, 569–578 (2013).
7. De Marsily, G., Delhomme, J. P., Couderc, A., & Marsh Lavenue, A. Four Decades of Inverse Problem in Hydrogeology, Special Paper of the Geological Society of America 348, 1–17, (2000).
8. Zhou, H., Gomez-Hernandez, I. & Li, I. Inverse methods in hydrogeology: Evolution and recent trends. Adv. Water Resour. 63, 22–37 (2013).
9. Hazra, S. B., Class, H., Helmi, R. & Schulz, V. Forward and inverse problems in modeling of multiphase flow and transport through porous media. Comput. Geosci. 8, 21–47 (2004).
10. Dai, Z. & Samper, J. Inverse problem of multicomponent reactive chemical transport in porous media: Formulation and applications. Water Resour. Res. 40(7), (2004).
11. Guellouz, L., Askri, B., Jaffre, J. & Bouhlila, R. Estimation of the soil hydraulic properties from field data by solving an inverse problem. Sci. Rep. 10, 9359 (2020).
12. Safi, M. & Chekkadi, A. Identification of illegal groundwater pumping in semi-confined aquifers. Hydrol. Sci. J. 55(8), 1348–1356 (2010).
13. Hariga, N. T., Bouhlila, R. & Ben Abda, A. Recovering data in groundwater: Boundary conditions and wells’ positions and fluxes. Comput. Geosci. 15, 637–645 (2011).
14. Bouhlila, R. & Hariga, N. T. Identification of aquifer pollution’s point sources with the reciprocity principle. Sci. Rep. 12, 9968. https://doi.org/10.1038/s41598-022-13795-w (2022).
15. Bel Hadj Hassin, A., Hariga, N. T. & Khayat, F. Data completion problem for the advection-diffusion equation with aquifer points sources. *Math. Methods Appl. Sci.* **44**(2), 1651–1663 (2021).
16. Mansouri, W., Baranger, T. N., Ben Ameur, H. & Hariga, N. T. Identification of injection and extraction wells from overspecified boundary data. *Inverse Probl. Sci. Eng.* **25**(8), 1091–1111 (2017).
17. Besbes, M., Chahed, J. & Hamdane, A. *National Water Security: Case Study of an Arid Country: Tunisia* (Springer, 2019).
18. de Marsily, G. *Quantitative Hydrogeology: Groundwater Hydrology for Engineers* (Academic Press, 1986).
19. Andrieux, S., Baranger, T. N. & Ben Abda, A. Solving Cauchy problems by minimizing an energy-like functional. *Inverse Probl.* **22**, 115–133 (2006).
20. Hamdi, A. Identification of point sources in two-dimensional advection-diffusion-reaction equation: Application to pollution sources in a river. *Station. Case Inverse Probl. Sci. Eng.* **15**(8), 855–870 (2007).
21. Hariga, N. T., Bouhila, R. & Ben Abda, A. Identification of aquifer point sources and partial boundary condition from partial overspecified boundary data. *C. R. Geosci.* **340**, 245–250 (2008).
22. Pironneau, O., Hecht, F., Hyaric, A. & Ohtsuka, K. *FreeFEM*. http://www.freefem.org (2006).
23. Riahi, M. H., BenAmeur, H., Jaffré, J. & Bouhila, R. Refinement indicators for estimating hydrogeologic parameters. *Inverse Probl. Sci. Eng.* https://doi.org/10.1080/17415977.2018.1460366 (2018).

**Author contributions**
L.G. formulated the problem statement and the objectives of solving the pollution intrusion problem. F.K. developed the algorithm for the direct and inverse problem and applied it for the identification of missing boundary data. Both authors contributed to the analysis of the results and the writing of the paper.

**Competing interests**
The authors declare no competing interests.

**Additional information**

**Correspondence** and requests for materials should be addressed to L.G.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022