Mathematical Geosciences manuscript No.
(will be inserted by the editor)

Contaminant Source Identification in Aquifers: A Critical View

J. Jaime Gómez-Hernández · Teng Xu

Received: date / Accepted: date

Abstract Forty years and one hundred fifty-seven papers later, research on contaminant source identification has grown exponentially in numbers but seems to be stalled concerning advancement towards the problem solution and its field application. This paper presents a historical evolution of the subject, highlighting its major advances. It also shows how the subject has grown in sophistication on the solution of the core problem (the source identification), forgetting that, from a practical point of view, such identification is worthless unless it is accompanied by a joint identification of the other uncertain parameters that characterize flow and transport in aquifers.

Keywords simulation-optimization, backward tracking, Bayesian approach, machine learning, surrogate models, heuristic approaches

1 Introduction

The year 2021 will mark the 40$^{\text{th}}$ anniversary of the first work on contaminant source identification in aquifers: the Ph. D. thesis defended by Steven Gorelick at Stanford University (Gorelick 1981). The subject attracted some attention in the following decade. It flourished during the last decade of the previous century, and has grown exponentially during the current century; unfortunately, this growth was not accompanied by the breadth of new ideas and approaches that took place between 1991 and 1997. Figure 1 shows a histogram of the

J. Jaime Gómez-Hernández ✉
Institute of Water and Environmental Engineering, Universitat Politècnica de València, Valencia, Spain
E-mail: jgomez@upv.es

Teng Xu
State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China
E-mail: teng.xu@hhu.edu.cn

number of works published in the field per year and its cumulative version. The figure includes a few papers not precisely about identifying contaminant sources in aquifers but in streamflows, lakes, and water distribution systems, mainly when these papers are based on the findings in aquifer research. As this paper is being written, the total number of works found is 157, but, most likely, that number will have increased by the time the paper is published.

The rate at which papers have been published in the last three years is above ten papers per year; yet, very few discuss applications to real problems.

This paper revisits how the subject has evolved after the pioneering work by Gorelick (1981) pointing out those papers that, in the opinion of the authors, signified an apparent breakthrough in the subject. The paper ends with a discussion of why, after forty years, the subject is not mature enough to find routine applications to real cases and is still far from being applied regularly.

2 The Problem

The problem of identifying a contaminant source in an aquifer using concentration measurements observed downgradient from the point of contamination falls in the realm of inverse problems (Tarantola 2005; Zhou et al. 2014). Consider, first, a forward model

$$\mathbf{d} = \mathbf{G}(\mathbf{m}) \tag{1}$$

where $\mathbf{d}$ is the outcome of the model providing the state of the system, $\mathbf{m}$ represents the model parameters at large, including not only material parameters but also those variables that need to be specified to characterize the system before any prediction is performed, and $\mathbf{G}$ is the function that maps parameters into system states. For example, in an aquifer where groundwater flow is under study, the state of the system is given by the piezometric heads, the model parameters are the hydraulic conductivities and porosities, but also the infiltration rates, boundary conditions, and pumping rates; and the function $\mathbf{G}$ is the groundwater flow equation, or better the numerical model solving the groundwater flow equation on a discretized version of the aquifer.

Consider, now, that several observations of the state of the system are available $\mathbf{d}_{\text{obs}}$; one could attempt to guess the values of the model parameters by inverting (1)

$$\mathbf{m} = \mathbf{G}^{-1}(\mathbf{d}). \tag{2}$$

This inversion is much more challenging to perform than the forward modeling because seldom the inverse model $\mathbf{G}^{-1}$ is explicitly known, nor the number of necessary observations available. In such case, the solution is to use the forward model to try to determine the parameters by means of an optimization or search procedure. During the search, the objective is to find a set of parameters $\mathbf{m}$ that produces state values $\mathbf{G}(\mathbf{m})$ that are as close as possible to the observed ones. Issues that must be considered in solving this problem include taking into account measurement errors—observations $\mathbf{d}_{\text{obs}}$ may be corrupted estimates of the true state values—; and model errors—$\mathbf{G}$ is only a numerical approximation of a system state equation that may not be representing exactly all relevant processes, and therefore, predictions $\mathbf{d}$ may not be exactly of the system state.

Contaminant source identification is an inverse problem where the target parameters to identify are the number and spacetime locations of the contamination events and their strengths. As discussed next, focusing on identifying the parameters characterizing the source results in an interesting and difficult-to-solve problem. Still, it may remain purely academic if realism is not introduced as part of the solution to the general problem.

3 Milestones in the Timeline

3.1 Early Work · The Simulation-Optimization Approach

The Ph.D. thesis by Steven Gorelick (1981) centers on groundwater pollution management problems, one of which is determining the location and strength of a contaminant leaking into an aquifer; making this the first work addressing this problem. The work was later published (Gorelick et al. 1983) and, to the best of the authors' knowledge, is the first paper on the subject.

The problem addressed is identifying the locations and strengths of the leaking portions of a pipe that is in contact with an aquifer where the contaminant disperses. The problem is cast as an optimization problem to minimize an objective function measuring the discrepancy between model-predicted concentrations and observed ones

$$J(\mathbf{m}) = \mathbf{w}\left[|\mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{cal}}|^p\right]^T \tag{3}$$

with

$$\mathbf{d}_{\text{cal}} = \mathbf{H}\mathbf{d} = \mathbf{H}\mathbf{G}(\mathbf{m}) \tag{4}$$

where $\mathbf{d}_{\text{obs}}$ is a vector with the observed concentrations, and $\mathbf{d}_{\text{cal}}$ is a vector with the calculated concentrations at the same locations, which are obtained after applying an observation matrix $\mathbf{H}$ to the model outcome $\mathbf{G}(\mathbf{m})$; $\mathbf{w}$ is a row vector of positive weights, the exponent p is generally 1 or 2, depending on the norm to be minimized, and the upper-script T stands for transpose. In Gorelick's work, he uses two optimization approaches, a linear programming one, in which the exponent is 1, and a least-squares one, in which the exponent is 2. In both cases, the weights are inversely proportional to the magnitude of the observations.

The vector of parameters $\mathbf{m}$, on which $\mathbf{d}$ depends, includes all the parameters needed to run the numerical model $\mathbf{G}$, such as the material parameters describing the aquifer (conductivity, porosity, etc.), the boundary conditions, the external stresses and the parameters describing the source. Not all of these parameters are subject to identification, and, in most papers, many of them are considered known without uncertainty. For example, in the work by Gorelick, all model parameters are known (and homogeneous) except for the intensities at eight potential pipe leaks. Under these setting, application of the principle of superposition yields $\mathbf{d}_{\mathrm{cal}}$ as a linear function of the unknown parameters $\mathbf{m}$; this allows writing the minimization problem as a linear programming one. Each observed concentration in time and space acts as a linear constraint to be satisfied by the parameters. Gorelick et al. (1983) also analyze the results obtained by multiple regression, which amounts to minimizing (3) with an exponent p equal to 2 using a least-squares approach. a This paper sets the scene for the papers to come. Gorelick et al. (1983) had identified a new and interesting inverse problem. He also established a specific way to approach the problem, which is the combination of simulation—to solve the forward model (1)—and optimization—to minimize an objective function like (3). For this reason, the papers using this approach are referred to as simulation-optimization papers.

The Ph.D. thesis by Gorelick (1981) also got the attention of Hwang and Koerner (1983) who looked for an alternative solution to the problem of source identification coupled with a dynamic network design. They use system sensitivity theory (a branch of control theory). Aquifer transport is treated as a dynamic system for which an initial guess of parameters is made, and feedback is obtained after concentrations are observed. The mismatch between predicted and observed concentrations is used to compute a so-called trajectory function that provides a perturbation of the parameters to be added to their last estimate before making the next prediction. The authors demonstrate the method in a two-dimensional synthetic aquifer and announce that a three-dimensional case study would follow, which was never published.

The decade of the 1980s finished with the publication of a research report by Datta et al. (1989) who use the same approach as Gorelick et al. (1983) to solve the problem.

From here on, the text will focus on the papers that, according to the authors, have supposed a significant advancement either in the solution of the core problem or in making the solution closer to its potential application to real cases. These papers are indicated in the timeline shown in Fig. 2. The text will end with a quick discussion and classification of the papers published in these 40 years. Two tables, including all 157 papers, are appended as supplementary material.

3.2 Backward Probability

Bagtzoglou et al. (1991) formulate a probabilistic solution for the problem of source identification based on the stochastic transport theory by Dagan (1982). In a heterogeneous media, solute concentrations resulting from an injection of a contaminant at location $\mathbf{X}_0$ are proportional to the probability that such a particle may be at location $\mathbf{X}$ after some time t. Dagan's theory revolves around trying to find these probability functions. Reversing the concept, one can think of finding the probability that a given particle that has been observed at $\mathbf{X}$ at time t was at $\mathbf{X}_0$ at time zero. When the release time is known, running a backward-in-time particle tracking using the current spatial distribution of the concentrations will yield a map of probabilities. The locations with the highest values would correspond with the source locations. As described, this identification is possible only if the velocity field is perfectly known and if all the sources start at the same known time. The paper leaves some unresolved issues but opens a new avenue for the solution of the source identification that will be explored by several authors in the future.

Bagtzoglou et al. (1992) addressed some of those problems in their next paper, such as not knowing precisely the time of the release or the velocity field, and propose the calculation of location and time probabilities with attached uncertainty.

3.3 Joint Identification of Source and Hydraulic Conductivity

Wagner (1992) is the first author that realizes that assuming that the hydraulic conductivity or the velocity spatial distributions are known is unrealistic and proposes a maximum likelihood parameter estimation following the steps by Carrera (1984) and Carrera and Neuman (1986). The forward problem remains the same, but now the objective function not only depends on the source parameters, but also on other parameters such as hydraulic conductivities, dispersivities, and boundary fluxes, which must be identified, too. The author demonstrates the application of maximum likelihood estimation, which had been utilized successfully for aquifer parameter identification in zoned aquifers, to the simultaneous estimation of material and source parameters. Our main criticism of this work is that the conceptualization of aquifer heterogeneity is very simple. It is limited to two zones with homogeneous flow and transport parameters. It also assumes that the source location is known, with the only source-related unknown being the mass load. In total, there are ten parameters to estimate.

The objective function to minimize in this work is the negative log-likelihood function, which under the assumption of normally distributed errors has an expression very similar to (3). Observations $\mathbf{d}$, in this case, were not limited to concentration values but also included piezometric heads (to help in the hydraulic conductivity identification).

The simultaneous estimation of aquifer and source parameters will reappear in several papers published later, but, almost always, with very simplistic representations of aquifer heterogeneity.

3.4 Time-Varying Injection and Tikhonov Regularization

The next major step was to consider the identification of a continuously time-varying solute injection function. Until the work by Skaggs and Kabala (1994), the identification of contaminant source was either of a constant pulse of finite duration or a series of them. Still, nobody had contemplated the possibility of identifying a pulse that was a continuous function of time. In their one-dimensional seminal paper, Skaggs and Kabala proposed to identify the three-peaked release function represented in Fig. 3; the identification was formulated by discretizing the span during which the release occurred into 100 points. They argue that such identification would be bound to fail due to the ill-posedness of the inversion problem and introduced, for the first time, the idea of regularizing the solution. Regularization implies modifying the objective function (3) by adding a term

$$J(\mathbf{m}) = \mathbf{w}\left[|\mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{cal}}|^{p}\right]^{T} + \alpha^2 \|L(\mathbf{m})\|^2 \tag{5}$$

where L is the regularization function and α is a weighting factor controlling its strength in the objective function. Skaggs and Kabala's regularization is a function of the 100 parameters discretizing the input function penalizing rapid oscillations in time. The authors focus exclusively on identifying the discretized source function, assuming that all other parameters controlling flow and transport in a homogeneous aquifer are known, including the source location. The observations are concentrations sampled in time and space at selected intervals. Two case studies were analyzed, one with exact observations without observation errors and one with inexact observations with measurement errors of varying magnitude. The authors conclude that Tikhonov regularization could be used to solve an inherently ill-posed inverse problem as long as the observation errors are not too large and that the measurements are taken before the plume has dissipated too much.

3.5 Minimum Relative Entropy

The year 1996 saw the publication of two significant contributions to the solution of the source identification problem. One of them is the work by Woodbury and Ulrych (1996), who shifted the focus of the problem from a deterministic one into a stochastic one. The other one is described in the next section.

The parameters $\mathbf{m}$ to be identified are considered as random variables with unknown probability distribution functions (pdfs), and the optimization approach is aimed at determining these pdfs, from which the expected value or the median could be retrieved as the model parameter estimate. Let p be the parameter prior pdf, which could be as uninformative as a uniform distribution between some lower and upper bounds, let q be the pdf of the parameters that are consistent with the observation data. By consistent, it is meant that the expected value of the predicted state at observation locations be equal to the observed values, $E\{\mathbf{HG}(\mathbf{m})\} = \mathbf{d}_{obs}$. Pdf q will result from the minimization of the relative entropy

$$H(q, p) = \int q(\mathbf{x}) \ln \left[\frac{q(\mathbf{x})}{p(\mathbf{x})}\right] d\mathbf{x} \tag{6}$$

subject to several linear constraints that result from the consistency requirement described above. The authors describe in detail how the minimization is performed, retrieve $q(\mathbf{m})$, and compute its expected value, which is compared with the reference injection curve with satisfactory results. The same injection function used by Skaggs and Kabala (1994) is analyzed, and the impact of observation errors is also studied. The location of the source is not subject to identification.

3.6 Heuristic Approaches

The work by Aral and Guan (1996) is the second of the landmark papers of 1996. It is the first paper that uses a heuristic approach to solve the optimization problem. The problem statement is the same one used by Gorelick et al. (1983), that is, the minimization of (3) subject to linearity constraints (one for each observed concentration). These constraints can be easily derived from the solute transport equation when the aquifer is homogeneous and of known parameters. The authors also add the additional constraint that the parameters to identify (the contaminant fluxes into the aquifer) must be positive. The originality of the solution is to depart from standard optimization algorithms and move into the, then new, heuristic algorithms, of which a genetic algorithm was chosen. As with all heuristic algorithms, multiple evaluations of the forward model (1) are needed, which makes the method computationally demanding; as a counterpart, these heuristic algorithms are supposed not to get stuck in local minima and are capable of getting the global minimum for objective functions with potentially many local extremes. Aral and Guan (1996) demonstrate the application of genetic algorithms to identify the contaminant fluxes from six known locations time-varying stepwise in three known time intervals. The aquifer is synthetic, two-dimensional, and of known parameters. Exact and measurement error-corrupted observations are used. The authors conclude that genetic algorithms are a viable alternative.

3.7 Geostatistical/Bayesian Approaches

Following the path by Woodbury and Ulrych (1996), Snodgrass and Kitanidis (1997) also use a stochastic approach for the solution of the identification problem. The authors focus on the solution of the same problem, estimating a contaminant time-varying release function into an aquifer, assuming that the source location and the rest of the parameters describing the aquifer are known. Following a standard geostatistical approach, the parameters $\mathbf{m}$ (which, in this case, are the injection strengths discretized in time over the injection period) are modeled as a random function with a stationary but unknown mean value and a stationary but unknown covariance function of which its shape is known (for instance, it may be an exponential function). There are no observations of the parameters, but there are observations of the concentrations downgradient from the source, which, for conservative solutes, are linearly related to the source parameters. This linearity permits the computation of the conditional expected value and the conditional covariance of the unknown parameters given the observed concentrations.

The geostatistical approach starts by first estimating the parameters of the multivariate random function, which, in this case, are the unknown mean and the unknown parameters of the covariance function (variance and correlation length for the case of an exponential isotropic covariance). Then, the estimation is done maximizing the likelihood of the observations given the structural parameters. Snodgrass and Kitanidis (1997) argue that simultaneous estimation of both mean and covariance parameters results in biased estimates and proceed to maximize the likelihood after filtering out the unknown mean by integrating over all possible mean values. Once the parameters have been estimated, the rest is a standard co-kriging estimation to obtain the conditional (also referred to as posterior) estimate of mean and covariance of the parameters describing the injection function.

Since kriging cannot enforce non-negativity, Snodgrass and Kitanidis (1997) describe an iterative approach to the estimation of a non-linear transform of the input concentrations (what breaks the linearity between parameters to be estimated and observations) that ensures that all concentrations estimates are positive. The method is demonstrated using the benchmark injection function by Skaggs and Kabala (1994) in a one-dimensional aquifer with satisfactory results. An interesting discussion in the paper is the indication that Tikhonov regularization or thin-plate spline interpolation would yield the same results as the geostatistical approach for specific shapes of the covariance of the multivariate random function.

Although not explicitly stated in the paper, this is the first one in which a Bayesian approach is used.

3.8 Jumping into Three Dimensions

In 1998, the first paper addressing contaminant source identification in a three-dimensional domain was published. Woodbury et al. (1998) extend their application of Minimum Relative Entropy (MRE) in one dimension to the reconstruction of a three-dimensional plume source. The source is a rectangular patch of known dimensions, and in order to maintain the linearity between observations and source concentrations, the aquifer is considered homogeneous and with known parameters. An analytical solution of the transport equation is used that relates aquifer concentrations and source values. The benchmark input function of Fig. 3 is used, and the capabilities of the MRE in three dimensions are demonstrated. Case studies using observations with and without errors and the interplay between spatial data and temporal data are analyzed.

The method was also applied to a real case to identify the source of a 1,4-dioxane plume observed at the Gloucester landfill in Ontario, Canada. The underlying model of the aquifer had to adhere to the simplifications used for the derivation of the algorithm; that is, it was modeled as homogeneous with known flow and transport parameters. The authors are pretty satisfied with the results since the parameter uncertainty intervals are smaller than previous estimates.

3.9 Artificial Neural Networks

It is not until 2004 that the first paper that explores the potential of machine learning to identify a contaminant source appears. Singh et al. (2004) and Singh and Datta (2004) publish two very similar papers to demonstrate the use of artificial neural networks to estimate the parameters describing a contamination event and the aquifer properties. Focusing on the joint identification problem, the authors consider that aquifer and source can be characterized with fourteen parameters: one for the isotropic conductivity, one for porosity, two for dispersivity (longitudinal and transversal) and ten for the injection strengths in five years at two locations (injections remain constant within the year). The aquifer is two-dimensional and homogeneous in its parameters and perfectly known in size and shape; the location of the two sources is also known. Using a numerical code, the authors generate 8500 sets of values for the fourteen parameters, which are used to predict concentrations at 40 time intervals at four observation wells. From these sets, 4500 are chosen as training sets and 4000 as testing ones. The authors consider different artificial neural network architectures until they find the one that produces the smaller prediction errors. They follow with a demonstration using data with observation errors and conclude that these models could be used for source identification with a warning: the artificial neural network would have to be retrained for a different case study or if the aquifer system changes in any way.

3.10 Markov Chain Monte Carlo and Surrogate Models

The work by Zeng et al. (2012) marks a new development that goes beyond an incremental contribution. The problem is cast in a probabilistic framework aimed at computing the posterior probability of the parameters (location and strength source) given the observations (concentration measurements) using a Bayesian framework

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{m})p(\mathbf{d}|\mathbf{m})}{p(\mathbf{d})}, \tag{7}$$

where $p(\mathbf{m}|\mathbf{d})$ is the posterior pdf, $p(\mathbf{m})$ is the prior pdf, $p(\mathbf{d}|\mathbf{m})$ is the likelihood, and $p(\mathbf{d})$ can be regarded as a normalizing constant. Then, instead of using the geostatistical approach to determine the posterior pdf, the authors propose two novelties. One to use Markov chain Monte Carlo (McMC) to sample the posterior distribution, and the other one to use a surrogate model for the forward problem (1) (since McMC requires many evaluations of the likelihood function, which, in turn, requires many runs of the forward model). In particular, the McMC algorithm chosen is delayed rejection combined with an adaptive Metropolis sampler as described by Haario et al.

(2006). The surrogate model chosen is a sparse grid-based interpolation using the Smolyak algorithm (Wasilkowski and Wozniakowski 1995), which provides an estimate of the forward model by interpolating the forward model values computed on a sparse grid in parameter space. Let N be the number of parameters in $\mathbf{m}$, a grid of points is defined within the model domain $\{m_{i_1}, m_{i_2}, \ldots, m_{i_n}; i_1 = 1, \ldots, Q_1, i_2 = 1, \ldots, Q_2, \ldots, i_N = 1, \ldots, Q_N\}$, where $\{Q_1, Q_2, \ldots Q_N\}$ are the number of points along each dimension. The forward problem is evaluated at each of these points, and then the forward problem is estimated at any point by interpolating these values using some predefined basis functions

$$\mathbf{G}(\mathbf{m}) \approx \sum_{i=1}^{Q} f_{\mathbf{m_i}}(\mathbf{m})\mathbf{G}(\mathbf{m_i}) \tag{8}$$

where Q is the number of surrounding points to use in the interpolation, and $f_{\mathbf{m_i}}(\mathbf{m})$ are the basis functions. How to select the number of points to use, the grid on which they are defined, and the basis functions is discussed in the paper.

The authors analyze two synthetic two-dimensional case studies, one with five unknown parameters: location coordinates, beginning and ending times, and source strength, and the other one with ten parameters representing the source strength variability in time. Another difference between the two cases used to test the surrogate model is that the first case uses a homogeneous aquifer and the second one a heterogeneous one, although conductivities are not subject to identification and therefore assumed known. Nevertheless, in both cases, the algorithm can retrieve the parameters sought.

3.11 Network Design

Jha and Datta (2014) introduce a component of reality that had only be treated in a very imprecise way by Hwang and Koerner (1983), and mentioned without any demonstration in the review by Amirabdollahian and Datta (2013): that of designing the monitoring network to identify the source at the lowest observation cost possible. Even though the aquifer was still modeled as homogeneous and perfectly known, the authors propose a realistic situation whereby there is not a network of observation locations already in place, but rather, a contaminant is observed in a well during a period. Then, a network of observations is deployed, maximizing the chances of detecting the source locations and magnitudes correctly. The method proposed is a two-stage one; in the first stage, once the contaminant has been monitored during a specific time in the detection well, several potential source locations, which are consistent with the observations, are identified in the aquifer. Then, with this set of potential sources, a dense grid of potential observation locations is designed out of which a small number of points are chosen as the observation network. This network is defined to maximize the possible observed concentrations coming from the potential source locations. Once the network has been defined, concentrations are collected in the newly designed network and used as data to solve the source identification problem. This problem is solved by a simulation-optimization approach in which the objective function (3) is written in terms of a dynamic time warping distance, a distance that coincides with the traditional Euclidean distance when two series of values spanning the same length, with the same number of samples and without missing data are compared. The authors demonstrate the effectiveness of optimal network design for identifying a time-varying contaminant source in their synthetic aquifer.

3.12 Ensemble Kalman Filter and Joint Identification of Source and Hydraulic Conductivity

The ensemble Kalman filter (EnKF) (Evensen 2003) had been used for parameter identification in petroleum engineering and hydrogeology for some time (Aanonsen et al. 2009; Chen and Zhang 2006; Li et al. 2011a, 2012; Xu et al. 2013a,b) but focusing on static parameters such as hydraulic conductivities. The EnKF is an assimilation technique based on gathering observations in time and updating the parameter estimates after each collection step. Comparison of the forward model predictions and the observations allows the correction of the estimates into a newly updated estimate for the next forward prediction. However, when the parameter to be estimated is the location of a contaminant source, an updated location cannot be incorporated into the model to predict in time unless the forward model is restarted from time zero to account for the updated location. This procedure is known as the restart EnKF (r-EnKF). Xu and Gómez-Hernández (2016) demonstrated that the r-EnKF can be used for source identification and went a step further (Xu and Gómez-Hernández 2018) to prove that a channelized heterogeneous hydraulic conductivity spatial distribution could be jointly identified with the contaminant source parameters (location, release time, and source strength).

At last, after many years, a true leap towards the applicability of contaminant source identification algorithms was done, since, for the first time, a complex, realistic spatial distribution of hydraulic conductivity was not assumed known and was subject of identification simultaneously with the source. However, the rest of the parameters defining the aquifer, such as porosity, dispersivity, boundary conditions, and stresses were known.

3.13 Bayesian Model Selection

The paper by Cao et al. (2019) is the last paper that proposes a new paradigm to address the problem of contaminant source identification. In most of the papers published before that use synthetic experiments, the reference data were obtained adopting a specific model for the aquifer (whether deterministic or probabilistic). Then the same model was used for the solution of the identification problem. In a real situation, the uncertainties around the aquifer model are significant, and it is virtually impossible to claim that the aquifer system known. Cao et al. (2019) adopt a probabilistic model to select among a set of potential aquifer models using a Bayesian approach. The plausibility of the approach will depend on the span covered by the alternative models proposed. The authors demonstrate their proposal in two synthetic case studies. One is a two-dimensional aquifer with a zoned spatial distribution for hydraulic conductivity. The other is a three-dimensional experiment in a laboratory column made up of two sands arranged in two continuous blocks of very different shapes and sizes. The different models considered are not so different after all; in both case studies, the models only differ in the size and shape of the zones used to describe the heterogeneity of the hydraulic conductivity, but the paper marks a route of how to incorporate different descriptions of the aquifer system and to identify jointly the model description and the source parameters that best reproduce the observations.

4 But There Are More Papers

In the previous sections, the papers that marked a change in the line of research towards the solution of contaminant source identification have been discussed. However, there are more, all in all, 157 papers have been encountered, and they deserve a short analysis that will help place the whole research field in perspective. Table 1 in the Supplementary Material lists the papers and highlights their main contributions, while Table 2, also in the Supplementary Material, uses the same paper numbering as Table 1 and includes some characteristic features of the papers of interest. More precisely, Table 2 includes the dimensionality of the problem, the type of source, whether the source is time dependent or not (it is marked as time dependent if it is a continuous function of time as in Fig. 3 or a step function that changes according to some stress periods; it is marked as time independent if it is either a pulse or a continuous injection), the type of solution algorithm used to solve the identification problem, the state equation considered with indication of the code used to solve it when available, the type of case study analyzed (it could be synthetic, laboratory or field), the parameters describing the source being identified (the most common parameters are the source locations and the release functions; in some occasions, the locations are chosen out of a set of release candidates or the strength of the source changes stepwise according to predefined stress periods), whether other parameters apart from the ones describing the source are identified (some papers identify flow and transport parameters, too; although in most of them these parameters are homogeneous or piecewise homogeneous within the domain) and, finally, whether hydraulic conductivity was considered as a heterogeneous parameter, and if heterogeneous, whether this heterogeneity was piecewise, that is, variable but homogeneous within well-defined zones, and whether the heterogeneity was known or was subject of identification, too.

Analyzing these attributes for all the papers, we can see an evolution towards applicability that looks more like the upper limb of a logistic curve reaching its asymptote rather than the exponential rise of the number of papers published.

Next, the different attributes will be discussed, stressing the potential applicability of the results to real problems.

4.1 Dimensionality

While the first papers presented applications in two-dimensional aquifers, there is a substantial number of papers going on until today addressing the problem in one dimension. Figure 4 shows the histograms of the papers classified by their dimensionality. It is not until 1998, with the paper by Woodbury et al. (1998) that the first three-dimensional analysis is published. The majority of papers are for two-dimensional aquifers, and only in the last few years the applications in three dimensions have increased.

From a practical point of view, solutions are needed in two or three dimensions. The scale of the problem will mark the need to use a two-dimensional model (regional flow) or a three-dimensional one (local flow).

4.2 Source

The problems addressed by the different authors can be classified as single or multiple sources and as point, areal or volumetric sources. Some authors assume that the source locations are known or that the source locations should be chosen out of a set of possible locations; this situation could be plausible in some occasions when the agent originating the contamination in the aquifer is known; but, in many occasions, this is not the case, and the location must be treated as an unknown to identify. The case of multiple locations where the number and coordinates of the sources have to be jointly identified has not been addressed; always that multiple sources are considered, there are some potential source locations to choose from, transforming a difficult continuous-mixed integer optimization problem into a not much simpler combinatorial one.

The papers for which the type of source is identified as areal consider the shape of the area to identify as known and only seek the release strength, except for Ala and Domenico (1992), Mahinthakumar and Sayeed (2005), Hosseini et al. (2011), Ayvaz (2016) and Zhou and Tartakovsky (2021) who also attempt to find the shape of the areal source. Only two of these consider an unknown generic shape.

Of the papers addressing a volumetric source, all of them assume that the shape is known, except for Mahinthakumar and Sayeed (2006), Mirghani et al. (2009), Aghasi et al. (2013), Jin et al. (2014) and Yeh et al. (2016) who also attempt to identify the shape of the source, most of them using a simple prismatic parameterization.

From a practical point of view, it does not seem feasible (because of its difficulty) to ask for a solution in which the sources are unknown in number, locations, and shapes, but some degree of lack of knowledge regarding these three attributes will always be present, and methods should aim to address all three of them in the most general way possible.

4.3 Time dependency

When the source varies in time, even if it is a single point at a known location, the difficulty of the identification problem increases dramatically, unless the variation is very simple and can be parameterized with a few unknowns (as is the case of a rectangular pulse, or a train of pulses, that only needs the pulse beginning and end times and the pulse concentrations).

A substantial number of papers consider that the source is either an instantaneous injection or a continuous one of constant intensity, in which case the number of parameters to describe it is only two, the (initial) time of the release and its concentration. Adopting this type of release means that there is a good knowledge of what happened, as it could be the case of an illegal overnight dump into an abandoned well or a continuous leakage out of a deposit. These cases are labeled as not being time-dependent.

Another important number of papers assume that the concentration history varies stepwise in time according to several stress periods. The duration of each stress period is known, and during each period, the concentration remains constant. Unless the stress periods are considered relatively short in time, the number of parameters to describe the time dependency is relatively small; adopting this formulation also implies that there is essential knowledge about the history of the release and the time periods during which the release remained constant. In Table 2, care has been taken to indicate when the case study assumes that the source strengths are identified at specified stress periods.

Finally, another group of papers attempts to identify an unknown continuous-time function that describes the release. This group starts with the one-dimensional case by Skaggs and Kabala (1994) for which the location was known, and continuous with papers in higher dimensions and the simultaneous identification of the source location (Todaro et al. 2021).

4.4 Solution Approach

As already said in the section describing the landmark papers, there are three main approaches to address contaminant source identification: The simulation-optimization approach, the backward probability tracking approach, and the probabilistic approach.

Most of the efforts in these forty years since the publishing of the first paper have focused on solving the identification problem under the premise that some concentration observations are available (in space and time), and there is a need to find out the parameters that describe the originating contamination, with little consideration on trying to account for other uncertainties inherent to groundwater flow and mass transport. Many refinements have been proposed concerning the initial papers, with the latest papers making use of the most sophisticated techniques regarding optimization by heuristic approaches, machine learning to build surrogate models, and innovative applications of Markov chain Monte Carlo.

It can be concluded that the identification problem is solved, provided that there is a perfect knowledge of the underlying aquifer in which the contamination has occurred. However, when uncertainties about the parameters describing the aquifer are considered, no approach has been able to get close enough to real conditions to grant its routine application to field cases.

4.5 State Equation

The state equation information included in Table 2 highlights whether flow and transport were solved, or just transport assuming the flow velocity known; it makes reference to the codes used to solve the state equation, when known; and, in the most recent years, whether surrogate models have been used to speed up the multiple evaluations of the forward problem needed by most of the solution algorithms.

4.6 Case Study

Five papers have used laboratory data, 28 papers used field data, and 113 papers used synthetic data. Although the number of papers using field data has increased in the last few years, the corpus of the subject is mainly based on results using synthetic aquifers.

While synthetic aquifers are necessary to test new algorithms and techniques, the subject should be mature enough to prove the latest development in closer-to-field conditions. Besides, most of the papers using field data do not use the most elaborated techniques at the time, but, generally, they make rather simplistic approximations, weakening the contribution of the field case demonstration.

There is a need for more research with field data. A task that on most occasions is hindered by the difficulty to have access to data that can be publicly shared, which may explain the relatively low number of field papers, but that does not explain the even lower number of papers using laboratory data.

4.7 Source Parameters being Identified

It is important to note that not all papers attempt to identify the source location, many assume it is known, and many assume that it could be one of a small set of candidates. The rest of the papers identify the source coordinates, either a point in space or a small set of parameters that identify an areal or volumetric source.

The time distribution of the source strength was already discussed above.

The hardest problem is the simultaneous identification of the number and the locations of multiple sources, and it has not been addressed by anyone yet. Only a very recent paper considers the problem of identifying the location of two sources (the number of sources is, therefore, known) and the parameters describing them (Zhou and Tartakovsky 2021).

4.8 Other Parameters being Identified

In practical terms, the aquifer parameters are never known, and, therefore, they should be subject to identification. Some authors consider that all parameters other than the source ones are known, without entering into any consideration of whether this decision is meaningful or not; some authors argue that they work with previously calibrated aquifer models, not using the additional concentration data to refine the aquifer model calibration further; finally, a few of the authors do perform a simultaneous identification of source and aquifer parameters.

When in Table 2 it is indicated that other parameters are identified, these additional parameters are described in Table 1.

4.9 Hydraulic Conductivity Heterogeneity

Hydraulic conductivity heterogeneity is of paramount importance for the proper prediction of contaminant transport (Capilla et al. 1999; Gómez-Hernández and Wen 1994; Li et al. 2011b). For this reason, it is necessary to include a realistic representation of conductivity if the techniques developed are to be applied in practice. The papers have been classified as not accounting for heterogeneity (N), accounting for heterogeneity using a zonation with constant conductivities within each zone (Z), and accounting for heterogeneous conductivity using a stochastic realization (Y).

However, using a heterogeneous conductivity is not enough to make the analysis realistic. The conductivity field cannot be perfectly known, so an additional set of papers is tagged as accounting for heterogeneity but not fully knowing the hydraulic conductivity spatial distribution (YN). Of these papers, the subset that, in addition, attempts to identify the unknown conductivity field contains the ones closest to applicability. The number of papers meeting these latter conditions, that is, that assume that conductivity is fully heterogeneous in space, unknown (except for a few sampling points) and subject to identification, is only 11. The techniques described in these papers are the ones closer to be applicable in field conditions.

4.10 Additional Information

Figure 5 shows a word cloud with the last names of all authors signing the papers. While some of the last names of the Chinese authors may correspond to several people, it is clear that some authors have made an imprint in the field, with Datta being the leader and the responsible that both India and Australia are in third and fourth positions in the number of papers published by country.

Figure 6 shows a histogram of the number of papers published by the country of the first author institution. The USA is the country that has produced the largest number of papers overall, but if these numbers are broken by year of publication, it is noticeable that China is the leader in the latest years.

5 Conclusions

The productivity in terms of the number of papers published in the subject has grown exponentially in the forty years since the first work. Unfortunately, this exponential growth in numbers does not go in parallel with similar growth in added value. The field seems to be stalled with only minor incremental advances towards a solution that can be applied with reasonable expectations to field cases.

From a practical point of view, it seems unreasonable to attempt to solve a source contamination identification without any prior knowledge about the source itself. The optimal method should identify all parameters at once: the number of contaminant events, their locations, their extent, and their time history; but nobody has tried to do it, and probably nobody will try since it is too complex a problem. Therefore, it must be admitted that some information about the source is available, such as the number of sources, potential locations of them, whether it is punctual or not, and if it is not punctual, some idea about the shape or the duration of the contamination event.

At the same time, from a practical point of view, it seems unreasonable to develop new techniques that do not incorporate the inherent uncertainties involved in groundwater flow and mass transport modeling. Thus, whatever technique that wishes to have a chance to be applied in practice has to incorporate the uncertainty on the parameters describing flow and transport and other variables such as infiltration, pumping, or boundary conditions. Also, these techniques should consider a proper data acquisition since many of the papers assume dense networks of observations already existing before detecting the contaminant.

There is still room for improvement and new papers on the subject, but they should either propose a radically new approach to solving the problem or recognize previous work's limitations regarding its applicability and advance towards it.

Acknowledgements The first author wishes to acknowledge the financial contribution of the Spanish Ministry of Science and Innovation through project number PID2019-109131RB-I00 and the second author acknowledges the financial support from the Fundamental Research Funds for the Central Universities (B200201015) and Jiangsu Specially-Appointed Professor Program from Jiangsu Provincial Department of Education (B19052).

References

Aanonsen S, Nævdal G, Oliver D, Reynolds A, Vallès B (2009) The ensemble kalman filter in reservoir engineering–a review. SPE Journal 14(3):393–412

Aghasi A, Mendoza-Sanchez I, Miller EL, Ramsburg CA, Abriola LM (2013) A geometric approach to joint inversion with applications to contaminant source zone characterization. Inverse Problems 29(11):115014, DOI 10.1088/0266-5611/29/11/115014

Ala NK, Domenico PA (1992) Inverse Analytical Techniques Applied to Coincident Contaminant Distributions at Otis Air Force Base, Massachusetts. Groundwater 30(2):212–218, DOI 10.1111/j.1745-6584.1992.tb01793.x

Amirabdollahian M, Datta B (2013) Identification of Contaminant Source Characteristics and Monitoring Network Design in Groundwater Aquifers: An Overview. Journal of Environmental Protection DOI 10.4236/jep.2013.45a004

Aral MM, Guan J (1996) Genetic Algorithms in Search of Groundwater Pollution Sources. In: Advances in Groundwater Pollution Control and Remediation, Springer Netherlands, Dordrecht, pp 347–369, DOI 10.1007/978-94-009-0205-3_17

Ayvaz MT (2016) A hybrid simulation-optimization approach for solving the areal groundwater pollution source identification problems. Journal of Hydrology 538:161–176, DOI 10.1016/j.jhydrol.2016.04.008

Bagtzoglou AC, Tompson AFB, Dougherty DE (1991) Probabilistic Simulation for Reliable Solute Source Identification in Heterogeneous Porous Media. In: Water Resources Engineering Risk Assessment, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 189–201, DOI 10.1007/978-3-642-76971-9_12

Bagtzoglou AC, Dougherty DE, Tompson AFB (1992) Application of particle methods to reliable identification of groundwater pollution sources. Water Resources Management 6(1):15–23, DOI 10.1007/BF00872184

Cao T, Zeng X, Wu J, Wang D, Sun Y, Zhu X, Lin J, Long Y (2019) Groundwater contaminant source identification via Bayesian model selection and uncertainty quantification. Hydrogeology Journal 27(8):2907–2918, DOI 10.1007/s10040-019-02055-3

Capilla JE, Rodrigo J, Gómez-Hernández JJ (1999) Simulation of non-gaussian transmissivity fields honoring piezometric data and integrating soft and secondary information. Math Geosc 31(7):907–927

Carrera J (1984) Estimation of aquifer parameters under transient and steady-state conditions. PhD thesis, Univ. of Arizona, Dep. of Hydrology and Water Resources

Carrera J, Neuman SP (1986) Estimation of aquifer parameters under transient and steady state conditions, 1, Maximum likelihood method incorporating prior information. Water Resources Research 22(2):199–210

Chen Y, Zhang D (2006) Data assimilation for transient flow in geologic formations via ensemble kalman filter. Advances in Water Resources 29(8):1107–1122

Dagan G (1982) Stochastic modeling of groundwater flow by unconditional and conditional probabilities: 2. the solute transport. Water Resources Research 18(4):835–848

Datta B, Beegle J, Kavvas M, Orlob G (1989) Development of an expert system embedding pattern recognition techniques for pollution source identification. Tech. rep., University of California-Davis

Evensen G (2003) The ensemble kalman filter: Theoretical formulation and practical implementation. Ocean dynamics 53(4):343–367

Gómez-Hernández J, Wen XH (1994) Probabilistic assessment of travel times in groundwater modeling. Stochastic Hydrology and Hydraulics 8(1):19–55

Gorelick SM (1981) Numerical Management Models of Groundwater Pollution. Ph.d., Stanford University

Gorelick SM, Evans B, Remson I (1983) Identifying sources of groundwater pollution: An optimization approach. Water Resources Research 19(3):779–790, DOI 10.1029/WR019i003p00779

Haario H, Laine M, Mira A, Saksman E (2006) Dram: efficient adaptive mcmc. Statistics and computing 16(4):339–354

Hosseini AH, Deutsch CV, Mendoza CA, Biggar KW (2011) Inverse modeling for characterization of uncertainty in transport parameters under uncertainty of source geometry in heterogeneous aquifers. Journal of Hydrology 405(3-4):402–416, DOI 10.1016/j.jhydrol.2011.05.039

Hwang JC, Koerner RM (1983) Groundwater Pollution Source Identification From Limited Monitoring Data. Part 1 - Theory and Feasibility. Journal of Hazardous Materials 8:105–119

Jha MK, Datta B (2014) Linked Simulation-Optimization based Dedicated Monitoring Network Design for Unknown Pollutant Source Identification using Dynamic Time Warping Distance. Water Resources Management 28(12):4161–4182, DOI 10.1007/s11269-014-0737-5

Jin X, Ranjithan RS, Mahinthakumar GK (2014) A Monitoring Network Design Procedure for Three-Dimensional (3D) Groundwater Contaminant Source Identification. Environmental Forensics 15(1):78–96, DOI 10.1080/15275922.2013.873095

Li L, Zhou H, Franssen H, Gómez-Hernández J (2011a) Groundwater flow inverse modeling in non-multigaussian media: performance assessment of the normal-score ensemble kalman filter. Hydrol Earth Syst Sci Discuss 8(4):6749–6788

Li L, Zhou H, Gómez-Hernández JJ (2011b) A comparative study of three-dimensional hydraulic conductivity upscaling at the macro-dispersion experiment (made) site, Columbus Air Force Base, Mississippi (usa). Journal of Hydrology 404(3-4):278–293

Li L, Zhou H, Gómez-Hernández J, Hendricks Franssen H (2012) Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble kalman filter. Journal of Hydrology

Mahinthakumar GK, Sayeed M (2005) Hybrid Genetic Algorithm-Local Search Methods for Solving Groundwater Source Identification Inverse Problems. Journal of Water Resources Planning and Management 131(1):45–57, DOI 10.1061/(asce)0733-9496(2005)131:1(45)

Mahinthakumar GK, Sayeed M (2006) Reconstructing Groundwater Source Release Histories Using Hybrid Optimization Approaches. Environmental Forensics 7(1):45–54, DOI 10.1080/15275920500506774

Mirghani BY, Mahinthakumar KG, Tryby ME, Ranjithan RS, Zechman EM (2009) A parallel evolutionary strategy based simulation-optimization approach for solving groundwater source identification problems. Advances in Water Resources 32(9):1373–1385, DOI 10.1016/j.advwatres.2009.06.001

Singh RM, Datta B (2004) Groundwater pollution source identification and simultaneous parameter estimation using pattern matching by artificial neural network. Environmental Forensics 5(3):143–153, DOI 10.1080/15275920490495873

Singh RM, Datta B, Jain A (2004) Identification of Unknown Groundwater Pollution Sources Using Artificial Neural Networks. Journal of Water Resources Planning and Management 130(6):506–514, DOI 10.1061/(ASCE)0733-9496(2004)130:6(506)

Skaggs TH, Kabala ZJ (1994) Recovering the release history of a groundwater contaminant. Water Resources Research 30(1):71–79, DOI 10.1029/93WR02656

Snodgrass MF, Kitanidis PK (1997) A geostatistical approach to contaminant source identification. Water Resources Research 33(4):537–546, DOI 10.1029/96WR03753

Tarantola A (2005) Inverse problem theory and methods for model parameter estimation. SIAM

Todaro V, D'Oria M, Tanda MG, Gómez-Hernández JJ (2021) Ensemble smoother with multiple data assimilation to simultaneously estimate the source location and the release history of a contaminant spill in an aquifer. Journal of Hydrology 598:126215, DOI 10.1016/j.jhydrol.2021.126215

Wagner BJ (1992) Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modelling. Journal of Hydrology 135(1-4):275–303, DOI 10.1016/0022-1694(92)90092-A

Wasilkowski GW, Wozniakowski H (1995) Explicit cost bounds of algorithms for multivariate tensor product problems. Journal of Complexity 11(1):1–56

Woodbury A, Sudicky E, Ulrych TJ, Ludwig R (1998) Three-dimensional plume source reconstruction using minimum relative entropy inversion. Journal of Contaminant Hydrology 32(1-2):131–158, DOI 10.1016/S0169-7722(97)00088-0

Woodbury AD, Ulrych TJ (1996) Minimum Relative Entropy Inversion: Theory and Application to Recovering the Release History of a Groundwater Contaminant. Water Resources Research 32(9):2671–2681, DOI 10.1029/95WR03818

Xu T, Gómez-Hernández JJ (2016) Joint identification of contaminant source location, initial release time, and initial solute concentration in an aquifer via ensemble Kalman filtering. Water Resources Research 52(8):6587–6595, DOI 10.1002/2016WR019111

Xu T, Gómez-Hernández JJ (2018) Simultaneous identification of a contaminant source and hydraulic conductivity via the restart normal-score ensemble Kalman filter. Advances in Water Resources 112:106–123, DOI 10.1016/j.advwatres.2017.12.011

Xu T, Gómez-Hernández JJ, Zhou H, Li L (2013a) The power of transient piezometric head data in inverse modeling: An application of the localized normal-score EnKF with covariance inflation in a heterogenous bimodal hydraulic conductivity field. Advances in Water Resources 54:100–118, DOI 10.1016/j.advwatres.2013.01.006

Xu T, Jaime Gómez-Hernández J, Li L, Zhou H (2013b) Parallelized ensemble kalman filter for hydraulic conductivity characterization. Computers & Geosciences 52:42–49

Yeh HD, Lin CC, Chen CF (2016) Reconstructing the release history of a groundwater contaminant based on AT123D. Journal of Hydro-Environment Research 13:89–102, DOI 10.1016/j.jher.2015.06.001

Zeng L, Shi L, Zhang D, Wu L (2012) A sparse grid based Bayesian method for contaminant source identification. Advances in Water Resources 37:1–9, DOI 10.1016/j.advwatres.2011.09.011

Zhou H, Gómez-Hernández JJ, Li L (2014) Inverse methods in hydrogeology: evolution and recent trends. Advances in Water Resources 63:22–37

Zhou Z, Tartakovsky DM (2021) Markov chain Monte Carlo with neural network surrogates: Application to contaminant source identification. Stochastic Environmental Research and Risk Assessment 35(3):639–651

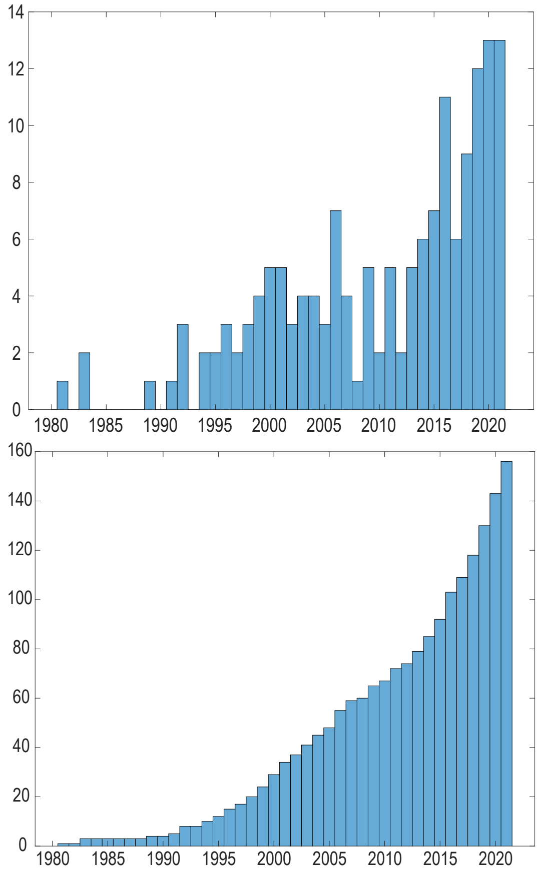

Fig. 1 Histogram and cumulative histogram of the number of papers published in the subject of contaminant source identification. Total number of papers is 157.

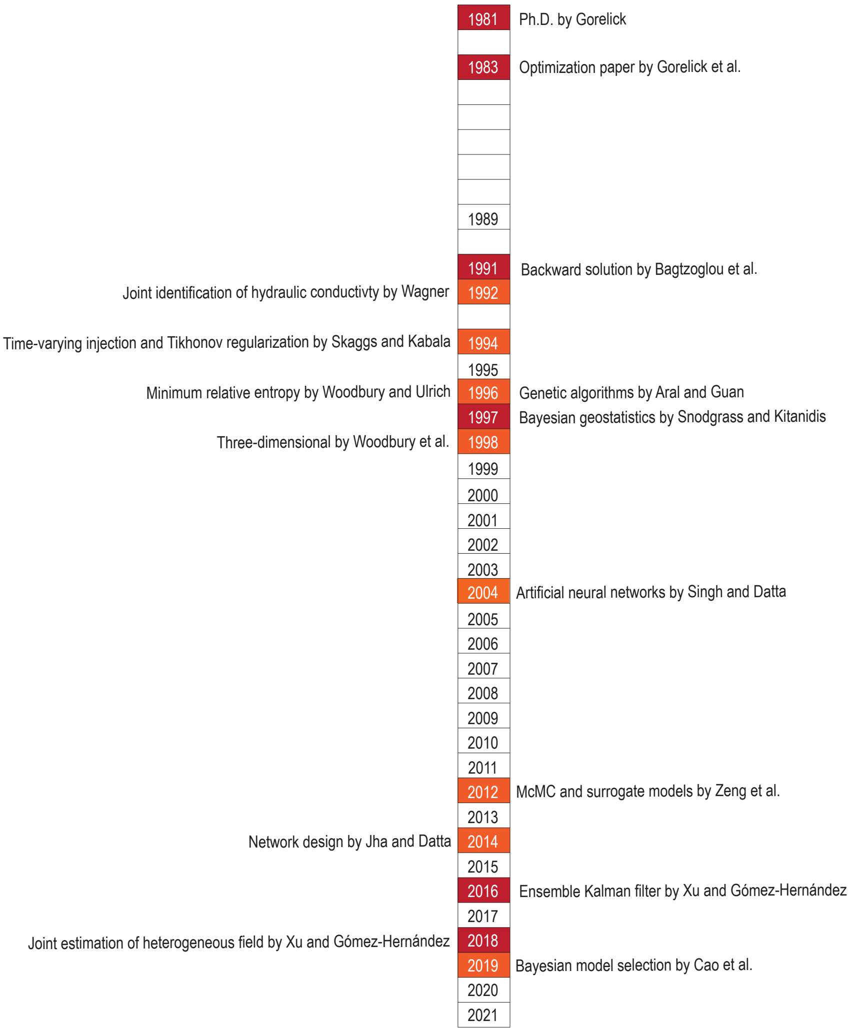

Fig. 2 Timeline with the papers that marked a difference in the solution of the contaminant source identification problem. The unlisted years are those without any published work. The reddish years correspond to major breakthroughs and the orangish ones to minor ones.

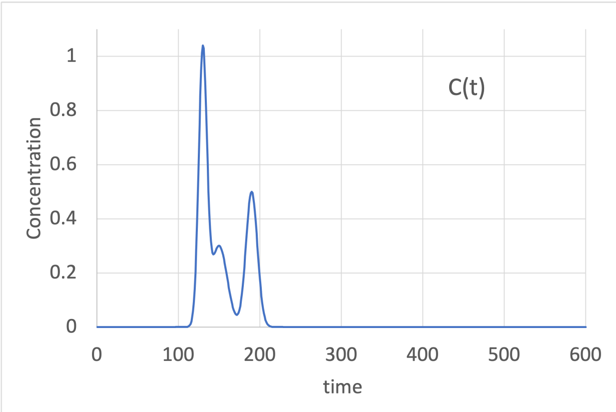

Fig. 3 Time varying pulse injection used by Skaggs and Kabala (1994) and repeatedly used later as a benchmark problem.

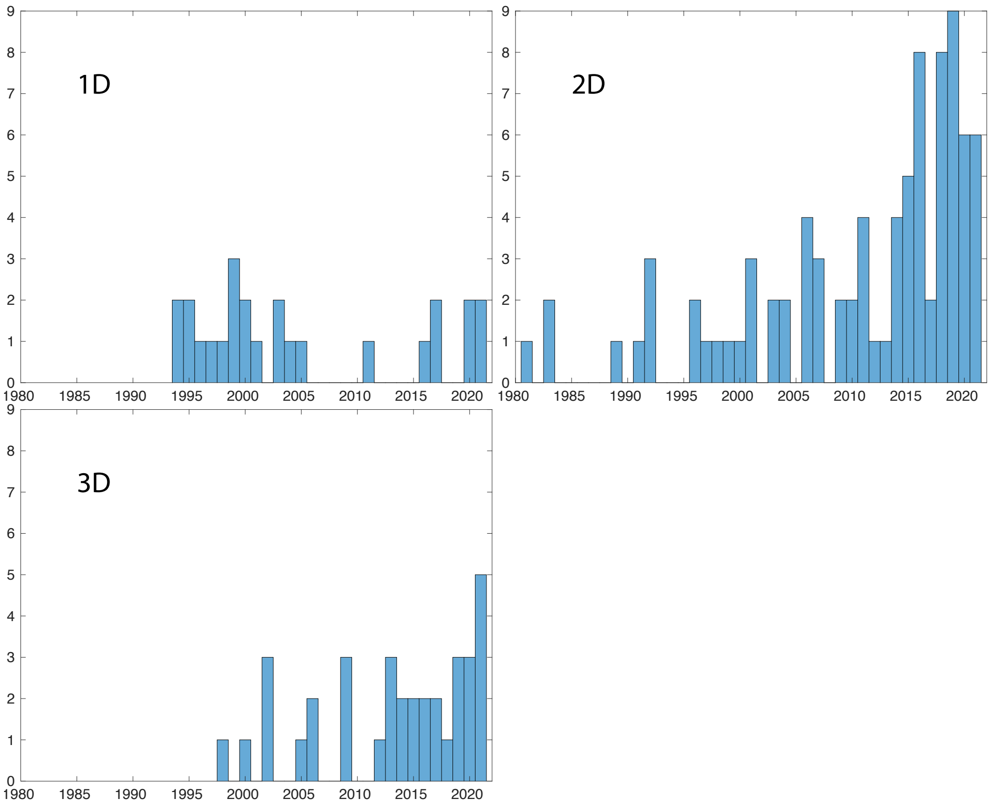

Fig. 4 Histograms of the number of papers classified by the dimensionality of the case studies

Fig. 5 Name cloud of all author's last names signing the papers

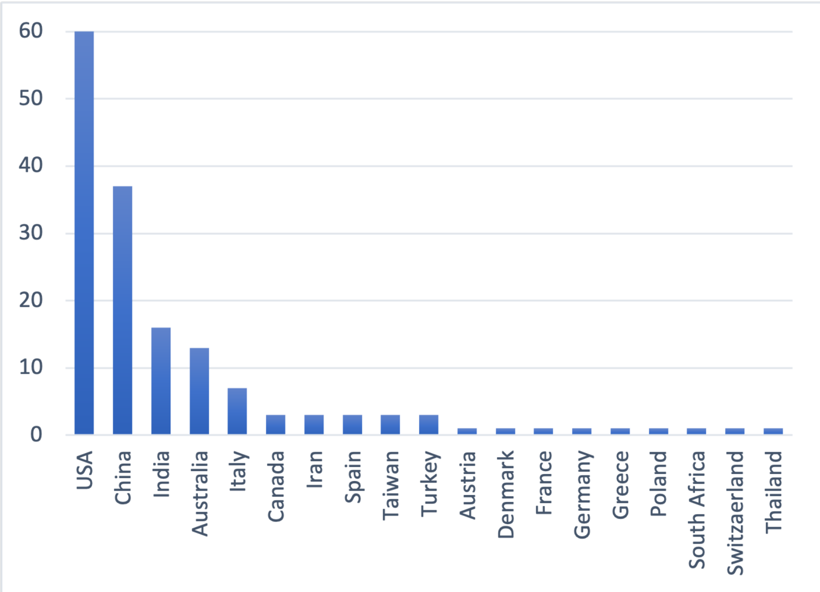

Fig. 6 Papers by country of first-author institution

Supplementary Material

Table 1: Paper highlights

	Reference	Highlights
1	Gorelick (1981)	The work that stated the problem and provided a solution. The details of the work are the same as the ones in the next paper .
2	Gorelick et al. (1983)	The first paper in the subject.
3	Hwang and Koerner (1983)	It could have started the way for a different approach to solve the problem using system sensitivity theory but failed to deliver two other papers on a real case study and on a three-dimensional one. The work also sketches a potential optimal network design.
4	Datta et al. (1989)	Very similar approach to Gorelick's, based on the use of response functions and optimization with the aid of an expert system for statistical pattern recognition.
5	Bagtzoglou et al. (1991)	It uses a random walk solution of the transport equation backward in time that can be interpreted as the source location probability. The authors perform the analysis assuming that the conductivity field is heterogeneous and that the covariance of the underlying random function is known. In this first implementation, it is necessary to know the time when the contaminant entered into the aquifer.
6	Ala and Domenico (1992)	The authors use a regression-like approach to fit observed concentration data to analytically-derived solutions for the transport equation. It is the first application to a field site, the Otis Air Force Base near Cape Cod (USA). Assuming all parameters homogeneous, the technique identifies the lateral extent of the source area, its position, the starting time and the concentration at the source, but also the parameters that control transport in the aquifer, such as dispersivities, decay rate or biodegradation rate.
7	Bagtzoglou et al. (1992)	Same concepts as in the previous paper by Bagtzoglou et al. (1991).
8	Wagner (1992)	For the first time, the parameters defining the source location (highly simplified to four unknowns corresponding to the mass fluxes at two locations and two stress periods) together with the parameters defining the aquifer model (all of them homogeneous in space: two conductivities, longitudinal and transversal dispersivities, porosity and lateral flux entering the aquifer) are estimated simultaneously.
9	Skaggs and Kabala (1994)	This is the first work that attempts to identify a time-dependent release history from measurements taken downgradient.
10	Wilson and Liu (1994)	The authors derived expressions for backward-in-time location and travel time probabilities using a heuristic approach.
11	Macdonald (1995)	The author proposes an approach for the location and strengths in a hypothetical one-dimensional domain in which a temperature pulse diffuses laterally. The impact of measurement errors is analyzed, since for an idealized medium, if the observations are error-free, the inversion is exact.
12	Skaggs and Kabala (1995)	The plume is traced back in time from a given snapshot by solving the diffusion equation with quasi-reversibility in a moving coordinate system. The estimation of the originating plume, not too far in time, from the observations is good, but it deteriorates when it gets closer to the source.

	Reference	Highlights
13	Aral and Guan (1996)	Simulation-optimization paper in which the response function is used to establish linear constraints for the optimization algorithm, which is solved using a genetic algorithm.
14	Sonnenborg et al. (1996)	Inverse modeling in the Vertskoven site (Denmark) containing a waste residue deposit. The geometry of the site and the location of the deposit are known, and the inversion focuses on identifying the flow and transport parameters in a zoned-conductivity aquifer, one of which is the contaminant concentration at the source.
15	Woodbury and Ulrych (1996)	The problem of contaminant source identification shifts from a deterministic inverse problem into a stochastic one.
16	Mahar and Datta (1997)	A standard weighted least-squares optimization using the embedding approach.
17	Snodgrass and Kitanidis (1997)	First paper using a Bayesian approach to solve the problem of source identification. Taking advantage of the response function solution, the problem can be solved using the geostatistical approach.
18	Sidauruk et al. (1998)	Curve fitting by least-squares to analytical solutions. Parameters fitted are location and strength of the source, but also the dispersivity coefficient.
19	Skaggs and Kabala (1998)	The authors perform an analysis of contaminant release function identification using a Monte-Carlo approach to test how the measurement errors hinder the optimization with Tikhonov regularization.
20	Woodbury et al. (1998)	First paper that addresses the problem in three dimensions.
21	Birchwood (1999)	Breakthrough curves and release function are expressed as Fourier series and a causality condition is imposed to determine the Fourier coefficients of the release function .
22	Liu and Ball (1999)	Using an analytical solution of the sorption-diffusion equation in a two-layer system, the authors infer the time function at which PCE has been crossing from an aquitard into an aquifer at the Dover Air Force Base in Delaware, USA.
23	Neupauer and Wilson (1999)	Improvement on the backward-in-time solution by Wilson and Liu (1994) based on the use of the adjoint equation of the advection dispersion equation.
24	Seibert and Stohl (1999)	This paper deals with dispersion in the atmosphere but it addresses the same problem: the identification of a pollution source and its time history.
25	Alapati and Kabala (2000)	The authors claim that there is no need to include a regularization term in the least-squares optimization process.
26	Mahar and Datta (2000)	Another simulation-optimization approach including a performance analysis of what happens if the solute concentration time series have missing data.
27	Morrison (2000)	Review paper of forensic techniques for age dating and source identification.
28	Neupauer et al. (2000)	Tikhonov regularization is less efficient in reproducing a step release function than minimum relative entropy.
29	Sciortino et al. (2000)	Application of least-squares optimization to identify the size and location of a pool of DNAPL being dissolved in a sandbox.
30	Aral et al. (2001)	Introduction of a variant of genetic algorithm combined with an iterative approach to reduce the number of forward model evaluations typically needed by heuristic algorithms such as genetic ones.

	Reference	Highlights
31	Atmadja and Bagtzoglou (2001a)	The authors develop a new approach for the backward-in-time solution of the transport equation. Once the observed plume has been traced back in time, deducing the release function is a much harder problem solved with relative success.
32	Atmadja and Bagtzoglou (2001b)	Very good review paper at the time it was written.
33	Duffy and Brandes (2001)	In a contaminated site in the Midwestern US, 184 observations are taken on 116 chemical species. Principal component analysis help in reducing the dimensionality of the problem and identifying potential sources.
34	Mahar and Datta (2001)	Simultaneous estimation of homogeneous aquifer parameters (conductivity, porosity and dispersivities) and source locations and strengths.
35	Michalak and Kitanidis (2002)	Application of the geostatistical approach to the analysis of 136 measurements taken at the Gloucester landfill in Ontario, Canada. The forward model used is an analytical solution with homogeneous coefficients in a semi-infinite layer of given thickness.
36	Sohn et al. (2002a)	Bayesian approach to determine the room at which an air pollution event started in a three-story building.
37	Sohn et al. (2002b)	Similar to the previous paper applied to a synthetic single-story building.
38	Akçelik et al. (2003)	The authors use a variational finite-element method for source inversion of convective-diffusive transport.
39	Butera and Tanda (2003)	Extension of the geostatistical approach in two dimensions. Demonstration for point, non-point and multiple sources, the locations of which are known.
40	Bagtzoglou and Atmadja (2003)	The paper focuses on heterogeneous conductivity distributions in one dimension and shows that marching-jury works better that quasi-reversibility for these cases.
41	Michalak and Kitanidis (2003)	Application of the geostatistical approach enforcing parameter non-negativity to a diffusion problem from a two-layer aquifer onto an aquitard at the Dover Air Force Base in Delaware, USA.
42	Michalak and Kitanidis (2004b)	Coupling the geostatistical approach and the adjoint state method allows the identification of the spatial distribution of the contaminant at a given instant back in time.
43	Michalak and Kitanidis (2004a)	Application of the geostatistical approach for source contaminant identification at the Dover Air Force Base in Delaware. The problem solved is a diffusion one using two-layers overlaying an aquitard.
44	Singh and Datta (2004)	After training an artificial neural network, homogeneous aquifer parameters, such as conductivity, porosity and dispersivity, as well as the source strengths at two given locations are identified.
45	Singh et al. (2004)	Same approach as previous one.
46	Bagtzoglou and Atmadja (2005)	Best review paper published until then.
47	Boano et al. (2005)	Given the conceptual similarity between source contaminant detection in aquifers and rivers, the geostatistical approach is used to identify the release function by Skaggs and Kabala (1994) in a river contamination event.
48	Mahinthakumar and Sayeed (2005)	The authors perform a comparison of many optimization approaches, concluding that the hybrid ones work best.
49	Mahinthakumar and Sayeed (2006)	Similar to the previous paper, now with a three dimensional parallelepiped source.

	Reference	Highlights
50	Neupauer and Lin (2006)	Extension of the work by Neupauer and Wilson (1999) but now conditioning on measured concentrations. The aquifer is heterogeneous but the transmissivity distribution is perfectly known.
51	Newman et al. (2006)	Laboratory experiment with a DNAPL source zone. The inverse flux plane model coupled with a hybrid simulated annealing-minimum relative entropy is used to identify the source and also to quantify the uncertainty on the estimates.
52	Singh and Datta (2006)	Same as Singh and Datta (2004) and Singh et al. (2004) but different case studies.
53	Sun et al. (2006a)	Model uncertainty is considered thanks to the use of constrained robust least squares optimization.
54	Sun et al. (2006b)	In this paper, in addition to the release function, the source location is also identified.
55	Wang and Zabaras (2006)	The hierarchical Bayesian method is used to reconstruct contaminant history backward in time until the time of an instantaneous release. The authors analyze several cases with homogeneous conductivities and one with heterogeneous, but known, uncorrelated values.
56	Liu and Zhai (2007)	Although in the field of air pollution, the paper includes an extensive review of groundwater literature plus a few references from air pollution.
57	Milnes and Perrochet (2007)	Backtracking a predicted concentration contour line until its point of disappearance. The underlying conductivity field is binary and heterogeneous.
58	Sun (2007)	Conductivity is only known in some statistical sense, the robust geostatistical method incorporates this uncertainty. The conductivity field is not subject to identification, though.
59	Yeh et al. (2007)	Tabu search is used to detect trial source locations, then simulated annealing is used to identify strength and release period. In one of the case studies, conductivity is considered heterogeneous and unknown. Several realizations are generated to conclude that only in 50% of them it was possible to identify the source. No attempt to identify conductivities.
60	Sun (2008)	CONSID, a program that implements nonnegative least squares, constrained robust least squares and robust geostatistical inversion.
61	Datta et al. (2009)	Simultaneous identification of aquifer and source parameters. However, aquifer parameters are homogeneous except for a case in which the aquifer is divided in two homogeneous known zones.
62	Dokou and Pinder (2009)	A stochastic flow and transport model is used to generate multiple realizations of the plume evolution. These realizations are averaged to obtain mean plumes that are then used to decide on the potential source location using an iterative approach.
63	Jin et al. (2009)	An application at the Canadian Force Base Borden site, near Toronto, of a simulation-optimization method to identify, with reasonable accuracy, a contaminant source under field conditions. A heterogeneous conductivity field that had been previously used elsewhere is used here and assumed known.
64	Liu and Zhai (2009)	Another application from a groundwater solution to the field of indoor airborne contaminant source locations.
65	Mirghani et al. (2009)	A simulation-optimization approach with emphasis in the parallelization of the computations. A heterogeneous conductivity field is used but it is assumed known.
66	Ayvaz (2010)	A new heuristic approach comes into play. Conductivity is, at most, zoned, but with known values. Source locations are known.

	Reference	Highlights
67	Cheng and Jia (2010)	Another extension of the adjoint method by Neupauer and Wilson (1999), in this case, to a point contamination in an open-surface water body.
68	Datta et al. (2011)	The authors claim that gradient-based optimization approaches can beat heuristic ones, which need too many evaluations of the forward problem.
69	Hosseini et al. (2011)	The paper focuses in determining the geometry of a DNAPL pool using wells inside and outside the pool. In the process, heterogeneous conductivity fields are calibrated, as well as other homogeneous transport parameters.
70	Jha and Datta (2011)	The authors claim that simulated annealing is better than genetic algorithms for source identification.
71	Li and zhong Mao (2011)	The novelty of the paper is the use of a new way to solve for concentrations using the global space-time multiquadric.
72	Telci and Aral (2011)	A paper dealing with identification of multiple instantaneous spills in a river network from a large set of potential locations. It is demonstrated in the Althamah river in Georgia, USA.
73	Chadalavada et al. (2012)	The novelty of the paper is the use of feedback information obtained from sequentially-designed monitoring networks. An application to a contaminated farmland in South Australia is shown.
74	Zeng et al. (2012)	First paper to use a surrogate model to speed up forward model evaluation.
75	Aghasi et al. (2013)	Multi-objective optimization combined with level set functions to determine a DNAPL pool. Conductivity is heterogeneous and known.
76	Amirabdollahian and Datta (2014)	Good review paper.
77	Butera et al. (2013)	For the first time, the geostatistical approach was applied to the identification of the release function and the source location in a known heterogeneous aquifer.
78	Jha and Datta (2013)	Comparison between genetic algorithm and simulated annealing.
79	Wang and Jin (2013)	A synthetic case based on the Borden site using a heterogeneous known distribution of conductivities.
80	Amirabdollahian and Datta (2014)	Another simulation-optimization approach. Conductivities are interpolated from spare data using inverse-distance weighting.
81	Gzyl et al. (2014)	A combination of integral pumping tests to identify the sources, then the release history is recovered by the geostatistical approach. The method is applied in a chemical plant at southern Poland in the city of Jaworzno in the valley of the Wawolnic river that has been contaminated with lindane. The aquifer model had been calibrated with homogeneous parameters.
82	Jha and Datta (2014)	Multi-step approach involving an intelligent network design.
83	Jin et al. (2014)	The authors focus in determining the best monitoring network for the purposes of source identification.
84	Srivastava and Singh (2014)	An artificial neural network is trained by running many contamination scenarios originating in one or two potential sources.
85	Yeh et al. (2014)	Demonstration of a hybrid heuristic approach. The aquifer is divided in three zones of known parameters.
86	Cupola et al. (2015)	The minimum entropy approach works equally well for source identification in two-dimensional homogeneous and heterogeneous conductivity fields.
87	Gurarslan and Karahan (2015)	Application of the differential evolution algorithm as the best genetic algorithm for source identification.

	Reference	Highlights
88	Jha and Datta (2015)	Adaptive simulated annealing is used for contaminant source identification including surface-subsurface water interaction. The flow model had been previously calibrated with PEST using a zoned conductivity. The method is applied to the upper Macquarie Groundwater Management Area in New South Wales, Australia.
89	Ngamsritrakul et al. (2015)	Simulation-optimization approach.
90	Prakash and Datta (2015)	Application at the Macquarie Groundwater Management Area, the authors claim that the feedback provided by the new monitoring wells drilled as a result of the optimal network design improves the final results.
91	Srivastava and Singh (2015)	Simultaneous identification of aquifer and source parameters. However, aquifer parameters are homogeneous.
92	Zhang et al. (2015)	Coupling experimental design and source identification for the optimal placement of observation wells. In the final case study, the authors show a joint identification problem of heterogeneous conductivities and source parameters. The conductivity field is represented with only five parameters using a Karhunen-Loève expansion.
93	Ayvaz (2016)	Identification of a distributed source of arbitrary shape. The conductivity is heterogeneous but known.
94	Bashi-Azghadi et al. (2016)	Application to the Tehran aquifer in the Tehran refinery region.
95	Borah and Bhattacharjya (2016)	The paper includes a comparison of genetic algorithms and artificial neural networks.
96	Hansen and Vesselinov (2016)	Unsupervised machine optimization to evaluate the most likely location and time of point contamination .
97	Koch and Nowak (2016)	Conditional realizations of conductivity are generated and used to build probabilistic estimates of DNAPL presence.
98	Xu and Gómez-Hernández (2016)	First application of the ensemble Kalman filter for the identification of the space-time coordinates of a contaminant source.
99	Yeh et al. (2016)	Application of simulated annealing to fit some analytical functions to observed concentrations to reconstruct the release function of a contaminant of known location.
100	Zanini and Woodbury (2016)	Application of the method developed by Woodbury and Ferguson (2006) for heat flow inversion to contaminant release function identification.
101	Zhang et al. (2016a)	Application of surrogate models to Markov chain Monte Carlo implemented in a two-stage manner to improve the accuracy of the surrogate model.
102	Zhang et al. (2016b)	Backward tracking with a fractional advection-dispersion model applied to the MADE-2 tracer tests.
103	Zhao et al. (2016)	A kriging surrogate model is used to solve a simulation-optimization problem with known source locations in a conductivity-zoned aquifer.
104	Gu et al. (2017)	Many realizations of conductivity are generated and plumes are computed from potential sources. These plumes are weighted to approximate the observed values. From these weights the most likely source location and strength are determined.
105	Hamdi (2017)	Self-organized maps are used to build the surrogate models. In one of the applications, conductivity is only known at a few locations and it is interpolated by inverse distance weighting over the rest of aquifer.

	Reference	Highlights
106	Long et al. (2017)	A standard application of the geostatistical approach including a thorough sensitivity analysis to different kinds of errors.
107	Onyari and Taigbenu (2017)	Use of the Green element method.
108	Rajeev Gandhi et al. (2017)	Application to the identification of the source of a virus in a three-dimensional unconfined aquifer.
109	Zhang et al. (2017)	Application at the MADE-2 site of backward probabilities computed using a fractional advection dispersion equation.
110	Chen et al. (2018)	Application of the ensemble Kalman filter to identify a source but also the position and length of a plank that was inserted in a laboratory sandbox.
111	Esfahani and Datta (2018)	Adaptive simulated annealing for the identification of source concentrations in a contaminated mine site in Queensland, Australia.
112	Guneshwor et al. (2018)	Application to an industrial site in Gujarat, India. Model is homogeneous and known.
113	Hou and Lu (2018)	A comparison of support vector regression, kernel extreme learning machine and kriging for multiphase flow .
114	Huang et al. (2018)	Very simplistic model to test a new optimization algorithm: the shuffled complex evolution algorithm.
115	Jiang et al. (2018)	A concentration field library is constructed and used as a surrogate during the identification. Conductivities are heterogeneous but known.
116	Stanev et al. (2018)	Application of blind source separation to identify the number of sources. The authors also identify key parameters such as advective velocity and dispersivity, although they are homogeneous.
117	Vesselinov et al. (2018)	Non-negative matrix factorization is applied to synthetic and real data to identify the original source from a series of observations that include different geochemical constituents. Unsupervised machine learning is used. Application at the regional aquifer beneath Los Alamos National Laboratory.
118	Xu and Gómez-Hernández (2018)	First attempt to identify jointly contaminant source and conductivity field without simplifications.
119	Amirabdollahian et al. (2019)	The authors apply the adaptive simulated annealing to the Eastlakes Experimental Site at the Botany Sands aquifer, South Wales, Australia.
120	Ayub et al. (2019)	After presenting a groundwater model of the Duplin county research site in North Carolina, the authors build a synthetic exercise based on this model to demonstrate their application of Markov chain Monte Carlo for the sampling of the posterior distribution of source contaminant concentrations.
121	Cao et al. (2019)	First attempt to bring model uncertainty into play. Although the alternative models analyzed only differ on the geometry of the zonation of the conductivity spatial distribution.
122	Jiao et al. (2019)	A heterogeneous aquifer of known conductivities based on the Texas High Plain is used to test a new inverse method based on local approximation solutions of the transport equation.
123	Li et al. (2019)	Improved Kalman filter coupled with 0-1 mixed-integer nonlinear programming and simulated annealing.
124	Mo et al. (2019)	Joint identification of a heterogeneous conductivity field and the contaminant source parameters. Unfortunately, the authors use a Gaussian covariance function to characterize the spatial heterogeneity of the log-conductivities, which is completely unrealistic but helps in the inversion process.

	Reference	Highlights
125	Vesselinov et al. (2019)	Continuation of paper by Vesselinov et al. (2018) to further analyze the use of blind source separation coupled with non-negative matrix factorization applied at the Los Alamos Laboratories chromium-contaminated site.
126	Xia et al. (2019a)	Another simulation-optimization approach using genetic algorithms.
127	Xia et al. (2019b)	The self-organized map-based surrogate model can solve the identification problem without using a simulation-optimization approach.
128	Xing et al. (2019)	Ensemble of three surrogate models in which each model is weighted according to an adaptive Metropolis-Markov chain Monte Carlo.
129	Yan et al. (2019)	Bayesian approach with a very efficient kriging surrogate model.
130	Zhang et al. (2019)	Bayesian approach coupled with Markov chain Monte Carlo using an improved Metropolis-Hastings algorithm.
131	Chaubey and Srivastava (2020)	Artificial neural networks in one dimension.
132	Colombo et al. (2020)	Interesting application of backward tracking to identify potential PCE sources in Milan, Italy.
133	Essouayed et al. (2020)	Gaussian-Levenberg-Marquardt algorithm using pilot points to identify a heterogeneous conductivity field, homogeneous dispersivity and the source location.
134	Han et al. (2020)	Genetic algorithm applied to a laboratory experiment and then to a gas station in Beijing .
135	Jamshidi et al. (2020)	Simulation-optimization approach in a groundwater-river integrated system.
136	Kang et al. (2020)	Joint inversion of hydraulic heads and self-potential measurements to improve conductivity estimation, followed by the inversion of tracer mean travel times to identify the source.
137	Li et al. (2020)	Use of several surrogate models with extreme learning machine.
138	Lu et al. (2020)	Heuristic search to determine the coordinates of the source and the strength of the release together with the discrete values of a zoned conductivity, and homogeneous porosity and dispersivities.
139	Wang et al. (2020)	Extremely similar to the previous one.
140	Wang and Lu (2020)	Extremely similar to the two previous ones.
141	Zhang et al. (2020)	Very ingenious combination of the ensemble smoother and deep learning for the joint identification of source parameters and non-Gaussian heterogeneous conductivity distributions.
142	Zhao et al. (2020b)	Identification of the contaminant source by the interpretation of concentration time series after an artificially enhanced catchment is generated. Application to a low natural flow velocity in the city of Cangzhou, China.
143	Zhao et al. (2020a)	Application in the valley of the Wawolnica river to identify the source of lindane.
144	Ayaz (2021)	Artificial neural network application.
145	Ayaz et al. (2021)	Genetic algorithm application.
146	Chakraborty and Prakash (2021)	Evolutionary search algorithm application.
147	Dodangeh et al. (2021)	Contaminant source identification in a coastal aquifer using the ensemble Kalman filter and artificial neural networks.
148	He et al. (2021)	Fitting an analytical solution to observed concentration by least-squares. Application to a gas station site in Beijing .
149	Hou et al. (2021)	Comparison of several heuristic algorithms.
150	Jiang et al. (2021)	Simultaneous identification of hydraulic conductivity using pilot points and source strengths for two known contaminant sources.

	Reference	Highlights
151	Liu et al. (2021)	Application of the ensemble smoother to identify source and conductivities. Conductivity fields are built using multipoint geostatistics conditioned at some pilot points that are subject to identification by the ensemble smoother
152	Todaro et al. (2021)	Ensemble smoother applied to a calibrated model of a laboratory experiment
153	Wang et al. (2021a)	Multiphase flow case with joint identification of source strengths and aquifer parameters, although the later are homogeneous within the aquifer. A surrogate model based on an adaptive chaotic particle swarm optimization and extreme learning machine is used.
154	Wang and Zhang (2021)	A case of contamination source detection in a lake where the state equation is the diffusion equation.
155	Wang et al. (2021b)	Combination of heuristic approaches to solve the joint identification of the source strengths plus zone conductivity values
156	Yuan and Liang (2021)	One-dimensional exercise.
157	Zhou and Tartakovsky (2021)	The source is instantaneous and described by two Gaussian pulses, the parameters to identify are those that define the Gaussian bells. The posterior distribution is sampled using Markov chain Monte Carlo using a fast surrogate model based on deep convolutional networks. Conductivity is treated as heterogeneous but known.

Table 2: Paper characteristics

	Reference	D[1]	Source	$f(t)$	Solution approach	State eq.[2]	CS[3]	Source parameters	O.[4]	H.[5]
1	Gorelick (1981)				All items same as next one					
2	Gorelick et al. (1983)	2	Multiple	N	Linear programming and multiple regression. State equation produces linear constraints at the points and times where solute concentrations are measured	Response function	S	Source strengths and locations out of a predetermined set of potential ones	N	N
3	Hwang and Koerner (1983)	2	Point	N	System sensitivity theory with feedback	Transport equation	S	Location and strength	N	N
4	Datta et al. (1989)	2	Point	N	Expert system based on a statistical patter recognition	Response function	S	Source strength and location	N	N
5	Bagtzoglou et al. (1991)	2	Multiple	N	Backward solution of transport equation	Random walk solution of transport equation	S	Most likely location	N	Y
6	Ala and Domenico (1992)	2	Areal	N	Inverse analytical regression-like technique	Transport equation	F	Source extent and location, strength, starting time	Y	N
7	Bagtzoglou et al. (1992)	2	Point	N	Backward solution of transport equation	Random walk	S	Most likely location	N	Y
8	Wagner (1992)	2	Point	N	Non-linear maximum likelihood estimation	Flow and transport solved with SUTRA	S	Mass flux at two potential locations and during two stress periods	Y	Z
9	Skaggs and Kabala (1994)	1	Point	Y	Optimization with Tikhonov regularization	Response function	S	Time-dependent release history	N	N
10	Wilson and Liu (1994)	1	Point	N	backward-in-time location and time probabilities	Response function	S	Most likely location	N	N
11	Macdonald (1995)	1	Pulse	N	Non-linear least-square fitting of a convolution integral	Diffusion equation	S	Location and strength	N	N
12	Skaggs and Kabala (1995)	1	Point	Y	Method of quasi-reversibility to solve the advection diffusion equation backward in time	Backward solution of transport equation	S	Release function	N	N

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
13	Aral and Guan (1996)	2	Multiple	Y	Genetic algorithm	Response function	S	Locations and stress periods out of a collection of candidates, plus strengths	N	N
14	Sonnenborg et al. (1996)	2	Areal	N	Non-linear least-squares optimization	Flow and transport equations	F	Solute concentration at the source	Y	Z
15	Woodbury and Ulrych (1996)	1	Point	Y	Minimum relative entropy estimation	Convolution integral	S	Release function	N	N
16	Mahar and Datta (1997)	2	Multiple	N	Weighted least-squares optimization	Flow and transport equations	S	Source locations out of a potential set plus source strengths	N	N
17	Snodgrass and Kitanidis (1997)	1	Point	Y	Geostatistical approach	Response function	S	Release function	N	N
18	Sidauruk et al. (1998)	2	Point	N	Method of correlation coefficient optimization	Analytical solutions of transport equations	S	Strength and location	Y	N
19	Skaggs and Kabala (1998)	1	Point	Y	Monte-Carlo analysis of the optimization with Tikhonov regularization	Transport equation	S	Release function	N	N
20	Woodbury et al. (1998)	3	Areal	Y	Minimum relative entropy inversion	Analytical solutions of transport equations	S	Release function	N	N
21	Birchwood (1999)	1	Point	Y	Fourier-based inverse method	Transport equation	S	Location and release function	N	N
22	Liu and Ball (1999)	1	Interface	Y	Least-squares minimization with Tikhonov regularization	Diffusion equation with linear sorption	F	Release function	N	Z
23	Neupauer and Wilson (1999)	1	Pulse	N	Adjoint method to solve backward-in-time location and travel probabilities	Adjoint of the advection-dispersion equation	S	Likely locations	N	N
24	Seibert and Stohl (1999)	2	Point	Y	Backward tracking coupled with regularized non-linear optimization	Lagrangian particle dispersion model	F	Release function	N	N
25	Alapati and Kabala (2000)	1	Point	Y	Non-linear least-squares without regularization	Convolution integral	S	Gradual release function	N	N

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
26	Mahar and Datta (2000)	2	Multiple	Y	Non-linear weighted least-squares optimization with the flow and transport equations acting as binding equality constraints	Finite difference solution of flow and transport	S	Magnitude and duration of release plus location out of a number of candidates	N	N
27	Morrison (2000)	-	-	-	Review paper of forensic techniques for age dating and source identification	-	-	-	-	-
28	Neupauer et al. (2000)	1	Point	Y	A comparison between Tikhonov regularization and minimum relative entropy	Transport equation	S	Release function	N	N
29	Sciortino et al. (2000)	3	Volumetric	N	Weighted least-squares optimization	Analytical solutions of transport equations	L	Contaminant pool location and size	N	N
30	Aral et al. (2001)	2	Point	Y	Non-linear optimization using a progressive genetic algorithm. Source location and release histories are defined as explicit unknown variables	Transport equation	S	Release function and source locations	N	N
31	Atmadja and Bagtzoglou (2001a)	1	Point	Y	Marching-Jury Backward Beam Equation method for the solution of the advection-dispersion equation with heterogeneous (by zones) transport parameters	Transport equation solved backward in time	S	Release function	N	N
32	Atmadja and Bagtzoglou (2001b)	-	-	-	This was a very good review paper at the time	-	-	-	-	
33	Duffy and Brandes (2001)	2	Areal	N	Principal component analysis applied to a set of data on organic and inorganic chemicals at a given time. The results help to reduce the dimensionality of the problem and the identification of the sources	No state equation	F	Potential sources	Y	N

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
34	Mahar and Datta (2001)	2	Multiple	Y	Non-linear programming solved with the projected augmented Lagrangian algorithm	Flow and transport equations	S	Source strengths and locations out of a predetermined set of potential ones	Y	N
35	Michalak and Kitanidis (2002)	3	Areal	Y	Bayesian/geostatistical approach enforcing non-negativity of results	Transport equation	F	Release function	N	N
36	Sohn et al. (2002a)	3	Room	N	Bayesian approach	Air flow and pollutant transport in buildings represented as a collection of well-mixed zones	F	Room source	N	N
37	Sohn et al. (2002b)	3	Room	N	Bayesian approach	Same as above	S	Room source	N	N
38	Akçelik et al. (2003)	2	Multiple	Y	Constrained least-squares method with regularization	Transport equation	S	Release function	N	N
39	Bagtzoglou and Atmadja (2003)	1	Point	Y	Marching-jury backward beam equation versus quasi-reversibility	Transport equation	S	Release function	N	Y
40	Butera and Tanda (2003)	2	Multiple	Y	Geostatistical approach extended in two dimensions	Convolution integral	S	Release function	N	N
41	Michalak and Kitanidis (2003)	1	Interface	Y	Bayesian approach based on the geostatistical approach	Diffusion equation	F	Release function	N	Z
42	Michalak and Kitanidis (2004b)	2	Areal	N	Geostatistical approach with the adjoint method	Transport equation	S	Historical contaminant distribution	N	Y
43	Michalak and Kitanidis (2004a)	1	Interface	Y	Geostatistical approach	Diffusion equation	F	Release function	N	Z
44	Singh and Datta (2004)	2	Multiple	Y	Artificial neural network	USGS Konikow code	S	Release function	Y	N
45	Singh et al. (2004)	-	-	-	Same approach as previous one	-	-	-	-	-
46	Bagtzoglou and Atmadja (2005)	-	-	-	Best review paper until then	-	-	-	-	-
47	Boano et al. (2005)	1	Multiple	Y	Geostatistical approach applied to contamination source identification in rivers	Transient storage equation	S	Release function	N	N

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
48	Mahinthakumar and Sayeed (2005)	3	Areal	N	Simulation-optimization approach using a hybrid genetic algorithm with local search	Flow and transport equations	S	Coordinates of the corners of the rectangular source and its concentration	N	N
49	Mahinthakumar and Sayeed (2006)	3	Volumetric	Y	Several hybrid genetic algorithms with local search strategies compared	Transport equation	S	Coordinates of the corners of the parallelepiped source and its concentration	N	N
50	Neupauer and Lin (2006)	2	Point	N	Backward probabilities using the adjoint state method	Adjoint equations conditioned to concentration data	S	Source location	N	Y
51	Newman et al. (2006)	3	Volumetric	N	Combination of simulated annealing and minimum relative entropy	Flux plane model	L	Magnitude and distribution of mass influx from the DNAPL source zones	N	N
52	Singh and Datta (2006)	-	-	-	Same as Singh and Datta (2004) and Singh et al. (2004) but different cases	-	-	-	-	-
53	Sun et al. (2006a)	2	Point	Y	Constrained robust least-squares optimization	MODFLOW and MT3DMS and convolution approach	S	Release function	N	N
54	Sun et al. (2006b)	2	Multiple	Y	Constrained robust least-squares optimization to identify release history and branch and bound for location	Flow and transport equations	S	Release function and source locations	N	N
55	Wang and Zabaras (2006)	2	Point	N	Posterior mean estimate using a Gibbs sampler	Flow and transport equations	S	Location of instantaneous release	N	Y
56	Liu and Zhai (2007)	-	-	-	Although in the field of air pollution, the paper includes an extensive review of groundwater literature	-	-	-	-	-

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
57	Milnes and Perrochet (2007)	2	Point	N	Backtracking using flow reversal with FEFLOW	FEFLOW	S	Location and time	N	Y
58	Sun (2007)	2	Point	Y	Robust geostatistical approach	Transport equation	S	Release function	N	YN
59	Yeh et al. (2007)	2	Multiple	Y	Combination of simulated annealing and tabu search	MODFLOW and MOC	S	Source location, strength and release period	N	YN
60	Sun (2008)	-	-	-	CONSID, a program that implements nonnegative least squares, constrained robust least squares and robust geostatistical inversion	MODFLOW	-	-	-	-
61	Datta et al. (2009)	2	Multiple	Y	Nonlinear optimization with MINOS	SUTRA	S	Source strengths and locations out of a predetermined set of potential ones	Y	Z
62	Dokou and Pinder (2009)	3	Multiple	N	Optimal search strategy including a Monte Carlo flow and transport model and a predetermined set of potential sources	Flow and transport equations	S	Location out of set of potential ones	N	YN
63	Jin et al. (2009)	3	Prism	N	Parallel hybrid optimization with the real genetic algorithm and local search approaches	Transport equation	F	Seven unknowns: the corners of the prism plus the concentration at the source	N	Y
64	Liu and Zhai (2009)	2	Point	N	The adjoint probability method by Neupauer and Wilson (1999) applied to the detection of an air pollution in a building	Contaminant transport equation	S	Room source	N	N
65	Mirghani et al. (2009)	3	Cuboid	N	Simulation-optimization approach coupling numerical modeling and an evolutionary search algorithm.	Transport equation	S	Coordinates of the corners of the parallelepiped source and its concentration	N	Y

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
66	Ayvaz (2010)	2	Multiple	Y	Simulation-optimization using heuristic harmony search	Transport equation	S	Locations and release histories for a total of 12 parameters	N	Z
67	Cheng and Jia (2010)	2	Point	N	Backward location probability density function	Backward probability density function model in an open water body	S	Source location	N	N
68	Datta et al. (2011)	2	Multiple	Y	Classical optimization-simulation approach	SUTRA	S	Source strengths and locations out of a predetermined set of potential ones	N	Z
69	Hosseini et al. (2011)	2	Areal	N	Sequential self-calibration coupled with distance functions to determine the size and shape of DNAPL source.	Flow and transport equation	S	Shape of the DNAPL pool	Y	YN
70	Jha and Datta (2011)	2	Multiple	N	Simulated annealing for a global heuristic search optimization	MODFLOW and MT3DMS	S	Strength, location and timing	N	N
71	Li and zhong Mao (2011)	2	Areal	Y	Least-squares radial basis collocation method based on the global space-time multiquadric	GST-MQ collocation method	S	Release function	N	N
72	Telci and Aral (2011)	1	Multiple	N	Classification routine to associate observations with candidate source locations. Applied in a river network.	Storm water management model, SWMM	F	Source locations and timings	N	N
73	Chadalavada et al. (2012)	3	Multiple	Y	Simulation-optimization approach	MODFLOW and MT3DMS	F	Locations out of three potential ones, plus strengths	N	N
74	Zeng et al. (2012)	2	Point	Y	Bayesian method based on adaptive sparse grid interpolation	Surrogate model	S	Locations out of potential candidates, plus release function	N	N
75	Aghasi et al. (2013)	3	Volumetric	N	Joint inversion of geophysical data and concentration data to delimit a DNAPL pool	Multiphase flow and electrical resistance tomography	S	DNAPL pool and intensity	N	Y

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
76	Amirabdollahian and Datta (2013)	-	-	-	Review paper on source identification, network monitoring design, joint use of them, and identification of distributed sources	-	-	-	-	-
77	Butera et al. (2013)	2	Point	Y	Geostatistical approach	Convolution integral	S	Release function and source location	N	Y
78	Jha and Datta (2013)	3	Multiple	Y	Simulation-optimization	MT3DMS	S	Strengths at two known locations	N	N
79	Wang and Jin (2013)	3	Point	N	Bayesian approach coupled with Markov chain Monte Carlo using the Metropolis-Hastings algorithm	Flow and transport equations	S	Location coordinates and strength	N	Y
80	Amirabdollahian and Datta (2014)	2	Multiple	Y	Adaptive simulated annealing and fuzzy logic	MODFLOW and MT3DMS	S	Intensities at three locations	N	Y
81	Gzyl et al. (2014)	2	Areal	Y	Geostatistical approach	MODFLOW and MT3D	F	Release function	N	N
82	Jha and Datta (2014)	2	Point	Y	Simulation-optimization approach using a dynamic time warping distance	Flow and transport equations	S	Release function	N	N
83	Jin et al. (2014)	3	Cuboid	N	Network design for groundwater contaminant source identification	Flow and transport equations	S	Location and concentration	N	Y
84	Srivastava and Singh (2014)	2	Point	Y	Artificial neural network	MODFLOW-GWT	S	Location (out of two potential sources), strength and duration	N	N
85	Yeh et al. (2014)	3	Point	Y	Simulated annealing plus tabu search plus ordinal optimization plus roulette wheel.	MF2K-GWT	S	Location and release function	N	Z
86	Cupola et al. (2015)	2	Point	Y	A comparison between the simultaneous release function and location identification method by Butera et al. (2013) and the backward probability model using the adjoint method	Convolution integral	L	Location and release function	N	N

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
87	Gurarslan and Karahan (2015)	2	Multiple	Y	Differential evolution optimization	MODFLOW and MT3DMS	S	Source strengths and locations out of a predetermined set of potential ones	N	N
88	Jha and Datta (2015)	3	Areal	Y	Adaptive simulated annealing	MODFLOW and MT3DMS	F	Location, starting time and duration	N	Z
89	Ngamsritrakul et al. (2015)	2	Point	N	Non-linear optimization as proposed by Aral et al. (2001)	MODFLOW and MT3DMS	S	Location	N	N
90	Prakash and Datta (2015)	3	Areal	Y	Simulated annealing coupled with optimal monitoring network design	GMS	F	Location and starting time	N	Z
91	Srivastava and Singh (2015)	2	Point	Y	Artificial neural network	MOC	S	Source location	Y	N
92	Zhang et al. (2015)	2	Point	Y	Full Bayesian approach for optimal sampling well location and source parameter identification using Markov chain Monte Carlo	Flow and transport equations	S	Location, starting time, ending time and strength	Y	YN
93	Ayvaz (2016)	2	Areal	N	Simulation-optimization using a hybrid optimization where a binary genetic algorithm and a generalized gradient method are used	MODFLOW and MT3DMS	S	Shape of areal source and input concentration	N	Z
94	Bashi-Azghadi et al. (2016)	2	Multiple	N	Regret-based optimization model to minimize the number of monitoring wells. Bayesian network trained to identify pollutant source	MODFLOW and MT3DMS	F	Location	N	N
95	Borah and Bhattacharjya (2016)	2	Multiple	Y	Simulation-optimization approach using a new genetic algorithm	GMS	S	Strengths	N	N
96	Hansen and Vesselinov (2016)	2	Point	N	Monte-Carlo analysis to run multiple forward models to generate simulated data which are later used to identify the source by an inverse procedure using a simplified model	Transport equation	S	Location and time	N	Y

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
97	Koch and Nowak (2016)	3	Volumetric	N	Inverse Bayesian methodology for the joint inversion of source geometry and aquifer parameters	Galerkin finite element code	S	Probability of DNAPL presence	Y	YN
98	Xu and Gómez-Hernández (2016)	2	Point	Y	Ensemble Kalman filter used for the identification of a contaminant source	Transport equation	S	Coordinates of source, starting and ending time, and strength	N	Y
99	Yeh et al. (2016)	3	Volumetric	Y	Function fitting by simulated annealing	Analytical solutions of transport equations	S	Release function plus some geometric parameters of source	N	N
100	Zanini and Woodbury (2016)	2	Point	Y	Empirical Bayesian method combined with Akaike's Bayesian Information Criterion	Convolution integral	S	Release function	N	N
101	Zhang et al. (2016a)	2	Point	Y	Bayesian formulation for experimental design	Surrogate model integrated with Markov chain Monte Carlo	S	Source locations and strengths	Y	YN
102	Zhang et al. (2016b)	1	Point	N	Backward tracking using a fractional advection dispersion model	Fractional advection dispersion model	F	Source location	N	N
103	Zhao et al. (2016)	2	Multiple	Y	Simulation-optimization approach using a surrogate model	Kriging surrogate model	S	Strengths per stress period	N	Z
104	Gu et al. (2017)	2	Multiple	N	Monte Carlo approach	Transport equation	S	Number, location and strength of sources	N	Y
105	Hazrati Y. (2017)	3	Multiple	N	Surrogate models based on self-organized maps	MODFLOW and MT3DMS	S	Strengths at two known locations	N	Y
106	Long et al. (2017)	1	Point	Y	Geostatistical approach	Convolution integral	S	Release function	N	N
107	Onyari and Taigbenu (2017)	2	Point	Y	Green element method to solve inverse contaminant transport problems	Green element method	S	Release function	N	N
108	Rajeev Gandhi et al. (2017)	3	Multiple	Y	Simulated annealing coupled with analytical solutions	MODFLOW and MT3DMS	S	Wastewater release at potential locations	N	N

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
109	Zhang et al. (2017)	1	Point	N	Backward probabilities computed assuming a fractional advection-dispersion equation	Fractional advection-dispersion model	F	Location	N	N
110	Chen et al. (2018)	2	Point	N	Ensemble Kalman filter used for the identification of a contaminant source and the geometry of the system	MODFLOW and MT3DMS	L	Location, starting time, ending time and strength, and geometry of system	N	N
111	Esfahani and Datta (2018)	3	Multiple	N	Fractal singularity-based multi-objective monitoring networks	Genetic programming-based surrogate model for reactive transport	F	Source concentrations at known locations	N	Z
112	Guneshwor et al. (2018)	2	Multiple	Y	Simulation-optimization approach using particle swarm optimization	Mesh-free simulator using radial point collocation method	F	Source strengths at potential locations	N	N
113	Hou and Lu (2018)	2	Multiple	N	Comparative study of surrogate models	UTCHEM	S	Potential sources	N	N
114	Huang et al. (2018)	2	Multiple	N	Simulation-optimization using a shuffled complex evolution algorithm	MODFLOW and MT3DMS	S	Sources out of a potential set plus strengths	N	N
115	Jiang et al. (2018)	2	Multiple	N	Network design using Kalman filter to improve source identification	MODFLOW and MT3DMS	S	Five potential sources of unknown strengths	N	Y
116	Stanev et al. (2018)	2	Multiple	Y	Machine learning combined with Green's function inverse method	Green's function	S	Number, location and strength of sources accounting for different components	Y	N
117	Vesselinov et al. (2018)	2	Point	N	Blind source separation coupled with non-negative matrix factorization	Linear mixing of unknown signals	F	Origin of mixed waters observed	N	N
118	Xu and Gómez-Hernández (2018)	2	Point	Y	Simultaneous estimation of conductivity spatial distribution and source parameters using the ensemble Kalman filter	MODFLOW and MT3DMS	S	Location, strength	Y	YN

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
119	Amirabdollahian et al. (2019)	3	Multiple	N	Simulation-optimization approach using adaptive simulated annealing	MODFLOW and MT3DMS	F	Candidate locations out a set of potential sources	N	N
120	Ayub et al. (2019)	2	Multiple	N	Bayesian approach coupled with Markov chain Monte Carlo	Flow and transport equations	S	Locations out of potential candidates, plus release function	N	Y
121	Cao et al. (2019)	3	Multiple	Y	Bayesian model selection	MODFLOW and MT3DMS	S	Source strengths and locations out of a predetermined set of potential ones	N	Z
122	Jiao et al. (2019)	2	Point	Y	Local approximation solution of concentrations	MT3DMS	S	Release function and source locations	N	Y
123	Li et al. (2019)	2	Multiple	N	Simulation-optimization model with a surrogate model and Kalman filtering combined with a mixed-integer nonlinear programming	Kriging surrogate model	S	Location and release function	N	Y
124	Mo et al. (2019)	2	Point	Y	Joint identification using deep autoregressive neural networks	Surrogate model	S	Location and strengths	Y	YN
125	Vesselinov et al. (2019)	3	Multiple	Y	Unsupervised machine learning based on nonnegative tensor factorization of the original geochemical componentes	Unsupervised machine learning	F	Source locations	N	N
126	Xia et al. (2019a)	2	Multiple	Y	Simulation-optimization approach using a genetic algorithm tuned using Taguchi experimental design	MODFLOW and MT3DMS	S	Source strengths for given stress periods	N	Z
127	Xia et al. (2019b)	2	Multiple	Y	Optimal self-organized map-based surrogate model	Surrogate model	S	Strengths per stress period	N	Y
128	Xing et al. (2019)	2	Multiple	Y	Simulation-optimization solved with a genetic algorithm	Ensemble of surrogate models	S	Strengths per stress period	N	Z
129	Yan et al. (2019)	2	Multiple	Y	Bayesian approach coupled with Markov chain Monte Carlo using the Metropolis-Hastings algorithm	Kriging surrogate model	S	Strengths per stress period	N	Z

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
130	Zhang et al. (2019)	2	Point	N	Bayesian approach coupled with Markov chain Monte Carlo using an improved Metropolis-Hastings algorithm	Analytical solution	S	Strength, location and timing	N	N
131	Chaubey and Srivastava (2020)	1	Point	N	Artificial neural network	Analytical solution	S	Location and strength	N	N
132	Colombo et al. (2020)	3	Point	Y	Null-space Monte Carlo backward particle tracking	MODFLOW and MODPATH	F	Source location	N	Y
133	Essouayed et al. (2020)	2	Point	N	Standard non-linear optimization including pilot points	MODFLOW and MT3DMS	S	Source location	Y	YN
134	Han et al. (2020)	1	Point	N	Genetic algorithm	Analytical solution	F	Strength, location and timing	N	N
135	Jamshidi et al. (2020)	2	Multiple	Y	Standard gradient descent approach using Newton-Raphson	Convolution integral	S	Release function and source location	N	Z
136	Kang et al. (2020)	3	Volumetric	N	Principal component geostatistical approach using temporal moments of concentration instead of individual concentrations	COMSOL Multiphysics	S	Initial DNAPL saturation	Y	YN
137	Li et al. (2020)	2	Multiple	Y	Simulation-optimization using hybrid particle swarm optimization and extreme learning machine	Several surrogate models	S	Location out of set of potential ones	N	Z
138	Lu et al. (2020)	3	Point	Y	Parallel heuristic search based on a Bayesian approach	Surrogate model combining Gaussian process, kernel extreme learning machine and support vector machine	S	Coordinates of source, strength	Y	Z
139	Wang et al. (2020)	-	-	-	Extremely similar to the previous one	-	-	-	-	
140	Wang and Lu (2020)	-	-	-	Extremely similar to the two previous ones	-	-	-	-	-
141	Zhang et al. (2020)	2	Point	Y	Ensemble smoother modified with deep learning	MODFLOW and MT3DMS	S	Coordinates of source, strength	Y	YN

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
142	Zhao et al. (2020b)	2	Multiple	N	Artificially enhanced catchment	Interpretation of concentration time series	F	Location	N	N
143	Zhao et al. (2020a)	2	Multiple	N	Simulation-optimization using four different heuristic approaches	Surrogate model using kernel-based extreme machine learning	F	Strengths per stress period	N	Z
144	Ayaz (2021)	3	Point	Y	Artificial neural network	Transport equation	S	Release function	N	N
145	Ayaz et al. (2021)	2	Line	N	Constrained non-linear optimization with genetic algorithm	Transport equation	S	Source location, strength and release period	N	N
146	Chakraborty and Prakash (2021)	3	Multiple	Y	Simulation-optimization coupling numerical modeling and evolutionary search algorithm	Flow and transport equation	S	Source locations and strengths	N	N
147	Dodangeh et al. (2021)	3	Point	Y	Ensemble Kalman filter	Surrogate model	S	Source coordinates and initial concentration	N	N
148	He et al. (2021)	1	Point	N	Least squares	Analytical solution	F	Source location and strength	N	N
149	Hou et al. (2021)	3	Point	N	Homotopy-based hyper-heuristic approach	Multiphase surrogate model	S	Source location and strength	Y	N
150	Jiang et al. (2021)	2	Multiple	Y	Construction of a large table of inputs and outputs	Surrogate model based on self-organized maps	S	Source strengths for given stress periods	Y	YN
151	Liu et al. (2021)	2	Point	N	Ensemble smoother	MODFLOW and MT3DMS	S	Location and strength	Y	YN
152	Todaro et al. (2021)	2	Point	Y	Ensemble smoother	MODFLOW and MT3DMS	L	Location and release function	N	Y
153	Wang et al. (2021a)	3	Point	Y	Iterative updating heuristic search strategy	Surrogate multiphase flow	S	Strengths per stress period	Y	N
154	Wang and Zhang (2021)	2	Point	Y	Standard gradient descent	Diffusion equation	S	Location and strength	N	N
155	Wang et al. (2021b)	2	Point	Y	Differential evolution and tabu search optimization	Ensemble of surrogate models	S	Strengths per stress period	Y	Z

	Reference	D	Source	$f(t)$	Solution approach	State eq.	CS	Source parameters	O.	H.
156	Yuan and Liang (2021)	1	Point	Y	Simulation-optimization approach using genetic algorithms	Transport equation	S	Release function	N	N
157	Zhou and Tartakovsky (2021)	2	Areal	N	Bayesian approach coupled with Markov chain Monte Carlo	Surrogate model using a deep convolutional neural network	S	Location and Gaussian spread of initial concentrations	N	Y

[1] Dimensionality, 1: One-dimensional, 2: Two-dimensional, 3: Three-dimensional

[2] Acronyms used in the table: SUTRA, the flow and transport model by Voss (1984); MODFLOW, the modular groundwater flow code by McDonald and Harbaugh (1984); MT3DMS, the mass transport code by Zheng and Wang (1999); GMS, the Groundwater Modeling System by Aquaveo; MOC, the transport model developed by Konikow and Bredehoeft (1978); MF2K-GWT, implementation of MOC in MODFLOW, see, for instance, Konikow and Hornberger (2006); UTCHEM, the University of Texas multiphase simulator code; MODPATH, the particle tracking code by Pollock (1994); COMSOL Multiphysics, a general purpose simulation software by COMSOL; EPANET, flow and transport in water distribution networks by Rossman et al. (2000)

[3] Case Study, S: Synthetic, L: Laboratory, F: Field

[4] Other non-source parameters also identified

[5] Aquifer heterogeneity, N: Homogeneous, Z: Zonation, Y: Heterogeneous and unknown, YN: Heterogeneous but known

References

Aanonsen S, Nævdal G, Oliver D, Reynolds A, Vallès B (2009) The ensemble kalman filter in reservoir engineering–a review. SPE Journal 14(3):393–412

Aghasi A, Mendoza-Sanchez I, Miller EL, Ramsburg CA, Abriola LM (2013) A geometric approach to joint inversion with applications to contaminant source zone characterization. Inverse Problems 29(11):115014, DOI 10.1088/0266-5611/29/11/115014, URL https://iopscience.iop.org/article/10.1088/0266-5611/29/11/115014

Akçelik V, Biros G, Ghattas O, Long KR, Van Bloemen Waanders B (2003) A variational finite element method for source inversion for convective-diffusive transport. Finite Elements in Analysis and Design 39(8):683–705, DOI 10.1016/S0168-874X(03)00054-4

Ala NK, Domenico PA (1992) Inverse Analytical Techniques Applied to Coincident Contaminant Distributions at Otis Air Force Base, Massachusetts. Groundwater 30(2):212–218, DOI 10.1111/j.1745-6584.1992.tb01793.x

Alapati S, Kabala ZJ (2000) Recovering the release history of a groundwater contaminant using a non-linear least-squares method. Hydrological Processes 14(6):1003–1016, DOI 10.1002/(SICI)1099-1085(20000430)14:6⟨1003::AID-HYP981⟩3.0.CO;2-W

Amirabdollahian M, Datta B (2013) Identification of Contaminant Source Characteristics and Monitoring Network Design in Groundwater Aquifers: An Overview. Journal of Environmental Protection DOI 10.4236/jep.2013.45a004

Amirabdollahian M, Datta B (2014) Identification of pollutant source characteristics under uncertainty in contaminated water resources systems using adaptive simulated anealing and fuzzy logic. International Journal of GEOMATE 6(1):757–762, DOI 10.21660/2014.11.3258

Amirabdollahian M, Datta B, Beck PH (2019) Application of a link simulation optimization model utilizing quantification of hydrogeologic uncertainty to characterize unknown groundwater contaminant sources. Modeling Earth Systems and Environment 5(1):119–131, DOI 10.1007/s40808-018-0522-1, URL http://link.springer.com/10.1007/s40808-018-0522-1

Aral MM, Guan J (1996) Genetic Algorithms in Search of Groundwater Pollution Sources. In: Advances in Groundwater Pollution Control and Remediation, Springer Netherlands, Dordrecht, pp 347–369, DOI 10.1007/978-94-009-0205-3_17, URL http://link.springer.com/10.1007/978-94-009-0205-3_17

Aral MM, Guan J, Maslia ML (2001) Identification of Contaminant Source Location and Release History in Aquifers. Journal of Hydrologic Engineering 6(3):225–234, DOI 10.1061/(asce)1084-0699(2001)6:3(225)

Atmadja J, Bagtzoglou AC (2001a) Pollution source identification in heterogeneous porous media. Water Resources Research 37(8):2113–2125, DOI 10.1029/2001WR000223

Atmadja J, Bagtzoglou AC (2001b) State of the art report on mathematical methods for groundwater pollution source identification. Environmental Forensics 2(3):205–214, DOI 10.1006/enfo.2001.0055

Ayaz M (2021) Estimation of release history of groundwater pollution source using ANN model. Modeling Earth Systems and Environment (0123456789), DOI 10.1007/s40808-021-01142-3, URL https://doi.org/10.1007/s40808-021-01142-3http://link.springer.com/10.1007/s40808-021-01142-3

Ayaz M, Ansari SA, Singh OK (2021) Detection of pollutant source in groundwater using hybrid optimization model. International Journal of Energy and Water Resources DOI 10.1007/s42108-021-00118-4, URL http://link.springer.com/10.1007/s42108-021-00118-4

Ayub R, Messier KP, Serre ML, Mahinthakumar K (2019) Non-point source evaluation of groundwater nitrate contamination from agriculture under geologic uncertainty. Stochastic Environmental Research and Risk Assessment pp 1–18, DOI 10.1007/s00477-019-01669-z, URL http://link.springer.com/10.1007/s00477-019-01669-z

Ayvaz MT (2010) A linked simulation-optimization model for solving the unknown groundwater pollution source identification problems. Journal of Contaminant Hydrology DOI 10.1016/j.jconhyd.2010.06.004

Ayvaz MT (2016) A hybrid simulation-optimization approach for solving the areal groundwater pollution source identification problems. Journal of Hydrology 538:161–176, DOI 10.1016/j.jhydrol.2016.04.008, URL https://www.sciencedirect.com/science/article/pii/S0022169416301895

Bagtzoglou AC, Atmadja J (2003) Marching-jury backward beam equation and quasi-reversibility methods for hydrologic inversion: Application to contaminant plume spatial distribution recovery. Water Resources Research 39(2)

Bagtzoglou AC, Atmadja J (2005) Mathematical Methods for Hydrologic Inversion: The Case of Pollution Source Identification. Water Pollution 5:65–96, DOI 10.1007/b11442, URL http://www.springerlink.com/index/10.1007/b11442http://link.springer.com/10.1007/b11442

Bagtzoglou AC, Tompson AFB, Dougherty DE (1991) Probabilistic Simulation for Reliable Solute Source Identification in Heterogeneous Porous Media. In: Water Resources Engineering Risk Assessment, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 189–201, DOI 10.1007/978-3-642-76971-9_12, URL http://link.springer.com/10.1007/978-3-642-76971-9_12

Bagtzoglou AC, Dougherty DE, Tompson AFB (1992) Application of particle methods to reliable identification of groundwater pollution sources. Water Resources Management 6(1):15–23, DOI 10.1007/BF00872184

Bashi-Azghadi SN, Kerachian R, Bazargan-Lari MR, Nikoo MR (2016) Pollution Source Identification in Groundwater Systems: Application of Regret Theory and Bayesian Networks. Iranian Journal of Science and Technology, Transactions of Civil Engineering 40(3):241–249, DOI 10.1007/s40996-016-0022-3, URL http://link.springer.com/10.1007/s40996-016-0022-3

Birchwood R (1999) Identifying the location and release characteristics of a groundwater pollution source using spectral analysis. In: 19th Annual American Geophysical Union Hydrology Days Conference, pp 37–50

Boano F, Revelli R, Ridolfi L (2005) Source identification in river pollution problems: A geostatistical approach. Water Resources Research DOI 10.1029/2004WR003754

Borah T, Bhattacharjya RK (2016) Development of an Improved Pollution Source Identification Model Using Numerical and ANN Based Simulation-Optimization Model. Water Resources Management 30(14):5163–5176, DOI 10.1007/s11269-016-1476-6, URL http://link.springer.com/10.1007/s11269-016-1476-6

Butera I, Tanda MG (2003) A geostatistical approach to recover the release history of groundwater pollutants. Water Resources Research 39(12), DOI 10.1029/2003WR002314, URL http://doi.wiley.com/10.1029/2003WR002314

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

Cao T, Zeng X, Wu J, Wang D, Sun Y, Zhu X, Lin J, Long Y (2019) Groundwater contaminant source identification via Bayesian model selection and uncertainty quantification. Hydrogeology Journal 27(8):2907–2918, DOI 10.1007/s10040-019-02055-3, URL http://link.springer.com/10.1007/s10040-019-02055-3

Capilla JE, Rodrigo J, Gómez-Hernández JJ (1999) Simulation of non-gaussian transmissivity fields honoring piezometric data and integrating soft and secondary information. Math Geosc 31(7):907–927

Carrera J (1984) Estimation of aquifer parameters under transient and steady-state conditions. PhD thesis, Univ. of Arizona, Dep. of Hydrology and Water Resources

Carrera J, Neuman SP (1986) Estimation of aquifer parameters under transient and steady state conditions, 1, Maximum likelihood method incorporating prior information. Water Resources Research 22(2):199–210

Chadalavada S, Datta B, Naidu R (2012) Optimal Identification of Groundwater Pollution Sources Using Feedback Monitoring Information: A Case Study. Environmental Forensics 13(2):140–153, DOI 10.1080/15275922.2012.676147, URL http://www.tandfonline.com/doi/abs/10.1080/15275922.2012.676147

Chakraborty A, Prakash O (2021) CHARACTERIZATION OF GROUNDWATER POLLUTION SOURCES BY KRIGING BASED LINKED SIMULATION

OPTIMIZATION. International Journal of GEOMATE 20(81):79–85, DOI 10.21660/2021.81.6209, URL https://www.geomatejournal.com/sites/default/files/articles/79-85-6209-Anirban-May-2021-81.pdf

Chaubey J, Srivastava R (2020) Simultaneous identification of groundwater pollution source location and release concentration using Artificial Neural Network. Environmental Forensics 0(0):1–8, DOI 10.1080/15275922.2020.1850566, URL https://doi.org/10.1080/15275922.2020.1850566

Chen Y, Zhang D (2006) Data assimilation for transient flow in geologic formations via ensemble kalman filter. Advances in Water Resources 29(8):1107–1122

Chen Z, Gómez-Hernández JJ, Xu T, Zanini A (2018) Joint identification of contaminant source and aquifer geometry in a sandbox experiment with the restart ensemble Kalman filter. Journal of Hydrology 564:1074–1084, DOI 10.1016/j.jhydrol.2018.07.073, URL https://linkinghub.elsevier.com/retrieve/pii/S0022169418305870

Cheng WP, Jia Y (2010) Identification of contaminant point source in surface waters based on backward location probability density function method. Advances in Water Resources 33(4):397–410, DOI 10.1016/j.advwatres.2010.01.004, URL http://dx.doi.org/10.1016/j.advwatres.2010.01.004

Colombo L, Alberti L, Mazzon P, Antelmi M (2020) Null-Space Monte Carlo Particle Backtracking to Identify Groundwater Tetrachloroethylene Sources. Front Environ Sci 8:142, DOI 10.3389/fenvs.2020.00142, URL www.frontiersin.org

Cupola F, Tanda MG, Zanini A (2015) Laboratory sandbox validation of pollutant source location methods. Stochastic Environmental Research and Risk Assessment 29(1):169–182, DOI 10.1007/s00477-014-0869-4

Dagan G (1982) Stochastic modeling of groundwater flow by unconditional and conditional probabilities: 2. the solute transport. Water Resources Research 18(4):835–848

Datta B, Beegle J, Kavvas M, Orlob G (1989) Development of an expert system embedding pattern recognition techniques for pollution source identification. Tech. rep., University of California-Davis

Datta B, Chakrabarty D, Dhar A (2009) Simultaneous identification of unknown groundwater pollution sources and estimation of aquifer parameters. Journal of Hydrology 376(1-2):48–57, DOI 10.1016/j.jhydrol.2009.07.014

Datta B, Chakrabarty D, Dhar A (2011) Identification of unknown groundwater pollution sources using classical optimization with linked simulation. Journal of Hydro-Environment Research 5(1):25–36, DOI 10.1016/j.jher.2010.08.004, URL http://dx.doi.org/10.1016/j.jher.2010.08.004

Dodangeh A, Rajabi MM, Fahs M (2021) Simultaneous identification of contaminant source location, initial contaminant concentration, horizontal and vertical hydraulic conductivity in a three-dimensional coastal aquifer via ensemble Kalman filter. In: EGU General Assembly 2021, online, 19–30 Apr 2021, p 10890, DOI 10.5194/egusphere-egu21-10890

Dokou Z, Pinder GF (2009) Optimal search strategy for the definition of a DNAPL source. Journal of Hydrology 376(3-4):542–556, DOI 10.1016/j.jhydrol.2009.07.062, URL http://dx.doi.org/10.1016/j.jhydrol.2009.07.062

Duffy CJ, Brandes D (2001) Dimension reduction and source identification for multispecies groundwater contamination. Journal of Contaminant Hydrology DOI 10.1016/S0169-7722(00)00169-8

Esfahani HK, Datta B (2018) Fractal Singularity–Based Multiobjective Monitoring Networks for Reactive Species Contaminant Source Characterization. Journal of Water Resources Planning and Management 144(6):04018021, DOI 10.1061/(ASCE)WR.1943-5452.0000880, URL http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000880

Essouayed E, Verardo E, Pryet A, Chassagne R, Atteia O (2020) An iterative strategy for contaminant source localisation using GLMA optimization and Data Worth on two synthetic 2D Aquifers. Journal of Contaminant Hydrology 228:103554, DOI 10.1016/j.jconhyd.2019.103554, URL https://linkinghub.elsevier.com/retrieve/pii/S0169772219301251

Evensen G (2003) The ensemble kalman filter: Theoretical formulation and practical implementation. Ocean dynamics 53(4):343–367

Gómez-Hernández J, Wen XH (1994) Probabilistic assessment of travel times in groundwater modeling. Stochastic Hydrology and Hydraulics 8(1):19–55

Gorelick SM (1981) Numerical Management Models of Groundwater Pollution. Ph.d., Stanford University

Gorelick SM, Evans B, Remson I (1983) Identifying sources of groundwater pollution: An optimization approach. Water Resources Research 19(3):779–790, DOI 10.1029/WR019i003p00779, URL http://doi.wiley.com/10.1029/WR019i003p00779

Gu W, Lu W, Zhao Y, Ouyang Q, Xiao C (2017) Identification of groundwater pollution sources based on a modified plume comparison method. Water Science and Technology: Water Supply 17(1):188–197, DOI 10.2166/ws.2016.122

Guneshwor L, Eldho TI, Vinod Kumar A (2018) Identification of Groundwater Contamination Sources Using Meshfree RPCM Simulation and Particle Swarm Optimization. Water Resources Management 32(4):1517–1538, DOI 10.1007/s11269-017-1885-1, URL http://link.springer.com/10.1007/s11269-017-1885-1

Gurarslan G, Karahan H (2015) Solving inverse problems of groundwater-pollution-source identification using a differential evolution algorithm. Hydrogeology Journal 23(6):1109–1119, DOI 10.1007/s10040-015-1256-z, URL http://link.springer.com/10.1007/s10040-015-1256-z

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

Haario H, Laine M, Mira A, Saksman E (2006) Dram: efficient adaptive mcmc. Statistics and computing 16(4):339–354

Hamdi A (2017) Detection and identification of multiple unknown time-dependent point sources occurring in 1D evolution transport equations. Inverse Problems in Science and Engineering 25(4):532–554, DOI 10.1080/17415977.2016.1172224, URL https://www.tandfonline.com/doi/full/10.1080/17415977.2016.1172224

Han K, Zuo R, Ni P, Xue Z, Xu D, Wang J, Zhang D (2020) Application of a genetic algorithm to groundwater pollution source identification. Journal of Hydrology p 125343, DOI 10.1016/J.JHYDROL.2020.125343, URL https://www.sciencedirect.com/science/article/pii/S0022169420308039#b0105

Hansen SK, Vesselinov VV (2016) Contaminant point source localization error estimates as functions of data quantity and model quality. Journal of Contaminant Hydrology 193:74–85, DOI 10.1016/j.jconhyd.2016.09.003, URL https://linkinghub.elsevier.com/retrieve/pii/S0169772216301875

Hazrati Y S (2017) Self-organigizing map based surrogate models for contaminant source identification under parameter uncertainty. International Journal of GEOMATE 13(36), DOI 10.21660/2017.36.2750, URL http://www.geomatejournal.com/sites/default/files/articles/11-18-2750-Shahrbanoo-Aug-2017-36-a1.pdf

He Z, Zuo R, Zhang D, Ni P, Han K, Xue Z, Wang J, Xu D (2021) A least squares method for identification of unknown groundwater pollution source. Hydrology Research pp 1–11, DOI 10.2166/nh.2021.088

Hosseini AH, Deutsch CV, Mendoza CA, Biggar KW (2011) Inverse modeling for characterization of uncertainty in transport parameters under uncertainty of source geometry in heterogeneous aquifers. Journal of Hydrology 405(3-4):402–416, DOI 10.1016/j.jhydrol.2011.05.039, URL http://dx.doi.org/10.1016/j.jhydrol.2011.05.039

Hou Z, Lu W (2018) Comparative study of surrogate models for groundwater contamination source identification at DNAPL-contaminated sites. Hydrogeology Journal 26(3):923–932, DOI 10.1007/s10040-017-1690-1, URL http://link.springer.com/10.1007/s10040-017-1690-1

Hou Z, Lao W, Wang Y, Lu W (2021) Homotopy-based hyper-heuristic searching approach for reciprocal feedback inversion of groundwater contamination source and aquifer parameters. Applied Soft Computing 104:107191, DOI 10.1016/j.asoc.2021.107191, URL https://doi.org/10.1016/j.econlet.2020.109008https://linkinghub.elsevier.com/retrieve/pii/S1568494621001149

Huang L, Wang L, Zhang Y, Xing L, Hao Q, Xiao Y, Yang L, Zhu H (2018) Identification of groundwater pollution sources by a SCE-UA algorithm-based simulation/optimization model. Water (Switzerland) 10(2):193, DOI 10.3390/w10020193, URL http://www.mdpi.com/2073-4441/10/2/193

Hwang JC, Koerner RM (1983) Groundwater Pollution Source Identification From Limited Monitoring Data. Part 1 - Theory and Feasibility. Journal of Hazardous Materials 8:105–119

Jamshidi A, Samani JMV, Samani HMV, Zanini A, Tanda MG, Mazaheri M (2020) Solving Inverse Problems of Unknown Contaminant Source in Groundwater-River Integrated Systems Using a Surrogate Transport Model Based Optimization. Water 12(9):2415, DOI 10.3390/w12092415, URL https://www.mdpi.com/2073-4441/12/9/2415

Jha M, Datta B (2013) Three-Dimensional Groundwater Contamination Source Identification Using Adaptive Simulated Annealing. Journal of Hydrologic Engineering 18(3):307–317, DOI 10.1061/(ASCE)HE.1943-5584.0000624, URL http://ascelibrary.org/doi/10.1061/%28ASCE%29HE.1943-5584.0000624

Jha M, Datta B (2015) Application of Unknown Groundwater Pollution Source Release History Estimation Methodology to Distributed Sources Incorporating Surface-Groundwater Interactions. Environmental Forensics 16(2):143–162, DOI 10.1080/15275922.2015.1023385, URL http://www.tandfonline.com/doi/full/10.1080/15275922.2015.1023385

Jha MK, Datta B (2011) Simulated annealing based simulation-optimization approach for identification of unknown contaminant sources in groundwater aquifers. Desalination and Water Treatment 32(1-3):79–85, DOI 10.5004/dwt.2011.2681, URL http://www.tandfonline.com/doi/abs/10.5004/dwt.2011.2681

Jha MK, Datta B (2014) Linked Simulation-Optimization based Dedicated Monitoring Network Design for Unknown Pollutant Source Identification using Dynamic Time Warping Distance. Water Resources Management 28(12):4161–4182, DOI 10.1007/s11269-014-0737-5, URL http://link.springer.com/10.1007/s11269-014-0737-5

Jiang S, Fan J, Xia X, Li X, Zhang R (2018) An effective Kalman filter-based method for groundwater pollution source identification and plume morphology characterization. Water (Switzerland) 10(8):1063, DOI 10.3390/w10081063, URL http://www.mdpi.com/2073-4441/10/8/1063

Jiang S, Liu J, Xia X, Wang Z, Cheng L, Li X (2021) Simultaneous identification of contaminant sources and hydraulic conductivity field by combining geostatistics method with self-organizing maps algorithm. Journal of Contaminant Hydrology 241:103815, DOI 10.1016/j.jconhyd.2021.103815, URL https://linkinghub.elsevier.com/retrieve/pii/S0169772221000541

Jiao J, Zhang Y, Wang L (2019) A new inverse method for contaminant source identification under unknown solute transport boundary conditions. Journal of Hydrology 577:123911, DOI 10.1016/j.jhydrol.2019.123911, URL https://linkinghub.elsevier.com/retrieve/pii/S0022169419306316

Jin X, Mahinthakumar GK, Zechman EM, Ranjithan RS (2009) A genetic algorithm-based procedure for 3D source identification at the Borden emplacement site. Journal of Hydroinformatics 11(1):51–64, DOI 10.2166/hydro.2009.002, URL https://linkinghub.elsevier.com/retrieve/pii/S0022169409004053https://iwaponline.com/jh/article/11/1/51/30998/A-genetic-algorithmbased-procedure-for-3D-source

Jin X, Ranjithan RS, Mahinthakumar GK (2014) A Monitoring Network Design Procedure for Three-Dimensional (3D) Groundwater Contaminant Source Identification. Environmental Forensics 15(1):78–96, DOI 10.1080/15275922.2013.873095, URL http://www.tandfonline.com/doi/abs/10.1080/15275922.2013.873095

Kang X, Kokkinaki A, Kitanidis P, Shi X, Revil A, Lee J, Soueid Ahmed A, Wu J (2020) Improved characterization of DNAPL source zones via sequential hydrogeophysical inversion of hydraulichead, selfpotential and partitioningtracer data. Water Resources Research pp 1–26, DOI 10.1029/2020WR027627

Koch J, Nowak W (2016) Identification of contaminant source architectures-A statistical inversion that emulates multiphase physics in a computationally practicable manner. Water Resources Research 52(2):1009–1025, DOI 10.1002/2015WR017894, URL http://doi.wiley.com/10.1002/2015WR017894

Konikow LF, Bredehoeft JD (1978) Computer model of two-dimensional solute transport and dispersion in ground water, vol 7. US Government Printing Office Washington, DC

Konikow LF, Hornberger GZ (2006) Use of the multi-node well (mnw) package when simulating solute transport with the modflow ground-water transport process. Tech. rep., US Geological Survey

Li J, Lu W, Wang H, Fan Y (2019) Identification of groundwater contamination sources using a statistical algorithm based on an improved Kalman filter and simulation optimization. Hydrogeology Journal 27(8), DOI 10.1007/s10040-019-02030-y

Li J, Lu W, Wang H, Fan Y, Chang Z (2020) Groundwater contamination source identification based on a hybrid particle swarm optimization-extreme learning machine. Journal of Hydrology 584(February):124657, DOI 10.1016/j.jhydrol.2020.124657, URL https://doi.org/10.1016/j.jhydrol.2020.124657

Li L, Zhou H, Franssen H, Gómez-Hernández J (2011a) Groundwater flow inverse modeling in non-multigaussian media: performance assessment of the normal-score ensemble kalman filter. Hydrol Earth Syst Sci Discuss 8(4):6749–6788

Li L, Zhou H, Gómez-Hernández JJ (2011b) A comparative study of three-dimensional hydraulic conductivity upscaling at the macro-dispersion experiment (made) site, columbus air force base, mississippi (usa). Journal of Hydrology 404(3-4):278–293

Li L, Zhou H, Gómez-Hernández J, Hendricks Franssen H (2012) Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble kalman filter. Journal of Hydrology

Li Z, zhong Mao X (2011) Global multiquadric collocation method for groundwater contaminant source identification. Environmental Modelling and Software DOI 10.1016/j.envsoft.2011.07.010

Liu C, Ball WP (1999) Application of inverse methods to contaminant source identification from aquitard diffusion profiles at Dover AFB, Delaware. Water Resources Research 35(7):1975–1985, DOI 10.1029/1999WR900092

Liu Jb, Jiang Sm, Zhou Nq, Cai Y, Cheng L, Wang Zy (2021) Groundwater contaminant source identification based on. Journal of Groundwater Science and Engineering 9(1):73–82, DOI 10.19637/j.cnki.2305-7068.2021.01.007

Liu X, Zhai Z (2007) Inverse modeling methods for indoor airborne pollutant tracking: Literature review and fundamentals. Indoor Air 17(6):419–438, DOI 10.1111/j.1600-0668.2007.00497.x

Liu X, Zhai Z (2009) Prompt tracking of indoor airborne contaminant source location with probability-based inverse multi-zone modeling. Building and Environment 44(6):1135–1143, DOI 10.1016/j.buildenv.2008.08.004, URL http://dx.doi.org/10.1016/j.buildenv.2008.08.004

Long YQ, Cui TT, Li W, Yang ZP, Gai YW (2017) Sensitivity analysis of geostatistical approach to recover pollution source release history in groundwater Sensitivity analysis of geostatistical approach to recover pollu-tion source release history in groundwater. IOP Conf Ser: Earth Environ Sci 81:12105, DOI 10.1088/1755-1315/81/1/012105, URL https://iopscience.iop.org/article/10.1088/1755-1315/81/1/012105/pdf

Lu W, Wang H, Li J (2020) Parallel heuristic search strategy based on a Bayesian approach for simultaneous recognition of contaminant sources and aquifer parameters at DNAPL-contaminated sites. Environmental Science and Pollution Research

Macdonald JR (1995) Solution of an "impossible" diffusion-inversion problem. Computers in Physics 9(5):546, DOI 10.1063/1.168551, URL http://scitation.aip.org/content/aip/journal/cip/9/5/10.1063/1.168551

Mahar PS, Datta B (1997) Optimal Monitoring Network and Ground-Water–Pollution Source Identification. Journal of Water Resources Planning and Management 123(4):199–207, DOI 10.1061/(ASCE)0733-9496(1997)123:4(199), URL http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9496%281997%29123%3A4%28199%29

Mahar PS, Datta B (2000) Identification of pollution sources in transient groundwatersystems. Water Resources Management 14(3):209–227, DOI 10.1023/A:1026527901213

Mahar PS, Datta B (2001) Optimal Identification of Ground-Water Pollution Sources and Parameter Estimation. Journal of Water Resources Planning and Management 127(1):20–29, DOI 10.1061/(asce)0733-9496(2001)127:1(20)

Mahinthakumar GK, Sayeed M (2005) Hybrid Genetic Algorithm-Local Search Methods for Solving Groundwater Source Identification Inverse Problems. Journal of Water Resources Planning and Management 131(1):45–57, DOI 10.1061/(asce)0733-9496(2005)

131:1(45)

Mahinthakumar GK, Sayeed M (2006) Reconstructing Groundwater Source Release Histories Using Hybrid Optimization Approaches. Environmental Forensics 7(1):45–54, DOI 10.1080/15275920500506774, URL http://www.tandfonline.com/doi/abs/10.1080/15275920500506774

McDonald MG, Harbaugh AW (1984) A modular three-dimensional finite-difference groundwater flow model. Open-file report 83-875, U.S. Geological Survey

Michalak AM, Kitanidis PK (2002) Application of Bayesian inference methods to inverse modelling for contaminants source identification at Gloucester Landfill, Canada. In: Computational methods in water resources - Proceedings of the XIVth International Conference on Computational Methods in Water Resources (CMWR XIV) - Volume 2, Elsevier, pp 1259–1266

Michalak AM, Kitanidis PK (2003) A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification. Water Resources Research 39(2):1–14, DOI 10.1029/2002WR001480

Michalak AM, Kitanidis PK (2004a) Application of geostatistical inverse modeling to contaminant source identification at Dover AFB, Delaware. Journal of Hydraulic Research 42(EXTRA ISSUE):9–18, DOI 10.1080/00221680409500042

Michalak AM, Kitanidis PK (2004b) Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling. Water Resources Research 40(8), DOI 10.1029/2004WR003214

Milnes E, Perrochet P (2007) Simultaneous identification of a single pollution point-source location and contamination time under known flow field conditions. Advances in Water Resources 30(12):2439–2446, DOI 10.1016/j.advwatres.2007.05.013

Mirghani BY, Mahinthakumar KG, Tryby ME, Ranjithan RS, Zechman EM (2009) A parallel evolutionary strategy based simulation-optimization approach for solving groundwater source identification problems. Advances in Water Resources 32(9):1373–1385, DOI 10.1016/j.advwatres.2009.06.001, URL http://dx.doi.org/10.1016/j.advwatres.2009.06.001

Mo S, Zabaras N, Shi X, Wu J (2019) Deep Autoregressive Neural Networks for High-Dimensional Inverse Problems in Groundwater Contaminant Source Identification. Water Resources Research DOI 10.1029/2018WR024638, 1812.09444

Morrison RD (2000) Application of Forensic Techniques for Age Dating and Source Identification in Environmental Litigation. Environmental Forensics 1(3):131–153, DOI 10.1006/enfo.2000.0015, URL http://www.tandfonline.com/doi/abs/10.1006/enfo.2000.0015

Neupauer RM, Lin R (2006) Identifying sources of a conservative groundwater contaminant using backward probabilities conditioned on measured concentrations. Water Resources Research 42(3):1–13, DOI 10.1029/2005WR004115

Neupauer RM, Wilson JL (1999) Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant. Water Resources Research 35(11):3389–3398, DOI 10.1029/1999WR900190

Neupauer RM, Borchers B, Wilson JL (2000) Comparison of inverse methods for reconstructing the release history of a groundwater contamination source. Water Resources Research 36(9):2469–2475, DOI 10.1029/2000WR900176

Newman MA, Hatfield K, Hayworth J, Rao PSC, Stauffer T (2006) Inverse Characterization of NAPL Source Zones. Environmental Science & Technology 40(19):6044–6050, DOI 10.1021/es060437s, URL https://pubs.acs.org/doi/10.1021/es060437s

Ngamsritrakul T, Ruangrassamee P, Putthividhya A (2015) Backward Probabilistic Model for Contaminant Source Identification in Thailand. In: World Environmental and Water Resources Congress 2015, American Society of Civil Engineers, Reston, VA, pp 475–485, DOI 10.1061/9780784479162.043, URL http://ascelibrary.org/doi/10.1061/9780784479162.043

Onyari E, Taigbenu A (2017) Inverse Green element evaluation of source strength and concentration in groundwater contaminant transport. Journal of Hydroinformatics 19(1):81–96, DOI 10.2166/hydro.2016.028, URL https://iwaponline.com/jh/article/19/1/81/3526/Inverse-Green-element-evaluation-of-source

Pollock DW (1994) User's guide for modpath/modpath-plot, version 3: A particle tracking post-processing package for modflow, the us: Geological survey finite-difference groundwater flow model

Prakash O, Datta B (2015) Optimal characterization of pollutant sources in contaminated aquifers by integrating sequential-monitoring-network design and source identification: methodology and an application in Australia. Hydrogeology Journal 23(6):1089–1107, DOI 10.1007/s10040-015-1292-8, URL http://link.springer.com/10.1007/s10040-015-1292-8

Rajeev Gandhi BG, Bhattacharjya RK, Satish MG (2017) Simulation–Optimization-Based Virus Source Identification Model for 3D Unconfined Aquifer Considering Source Locations and Number as Variable. Journal of Hazardous, Toxic, and Radioactive Waste 21(2):04016019, DOI 10.1061/(ASCE)HZ.2153-5515.0000334, URL http://ascelibrary.org/doi/10.1061/%28ASCE%29HZ.2153-5515.0000334

Rossman LA, et al. (2000) Epanet 2 users manual, us environmental protection agency. Water Supply and Water Resources Division, National Risk Management Research Laboratory, Cincinnati, OH 45268

Sciortino A, Harmon TC, Yeh WW (2000) Inverse modeling for locating dense nonaqueous pools in groundwater under steady flow conditions. Water Resources Research 36(7):1723–1735, DOI 10.1029/2000WR900047

Seibert P, Stohl A (1999) Inverse modelling of the ETEX-1 release with a Lagrangian particle model. In: Third GLOREAM Workshop

Sidauruk P, Cheng AH, Ouazar D (1998) Ground water contaminant source and transport parameter identification by correlation coefficient optimization. DOI 10.1111/j.1745-6584.1998.tb01085.x

Singh RM, Datta B (2004) Groundwater pollution source identification and simultaneous parameter estimation using pattern matching by artificial neural network. Environmental Forensics 5(3):143–153, DOI 10.1080/15275920490495873

Singh RM, Datta B (2006) Identification of Groundwater Pollution Sources Using GA-based Linked Simulation Optimization Model. Journal of Hydrologic Engineering 11(2):101–109, DOI 10.1061/(ASCE)1084-0699(2006)11:2(101), URL http://ascelibrary.org/doi/10.1061/%28ASCE%291084-0699%282006%2911%3A2%28101%29

Singh RM, Datta B, Jain A (2004) Identification of Unknown Groundwater Pollution Sources Using Artificial Neural Networks. Journal of Water Resources Planning and Management 130(6):506–514, DOI 10.1061/(ASCE)0733-9496(2004)130:6(506), URL http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9496%282004%29130%3A6%28506%29

Skaggs TH, Kabala ZJ (1994) Recovering the release history of a groundwater contaminant. Water Resources Research 30(1):71–79, DOI 10.1029/93WR02656, URL http://doi.wiley.com/10.1029/93WR02656

Skaggs TH, Kabala ZJ (1995) Recovering the History of a Groundwater Contaminant Plume: Method of Quasi-Reversibility. Water Resources Research 31(11):2669–2673, DOI 10.1029/95WR02383, URL http://doi.wiley.com/10.1029/95WR02383

Skaggs TH, Kabala ZJ (1998) Limitations in recovering the history of a groundwater contaminant plume. Journal of Contaminant Hydrology 33(3-4):347–359, DOI 10.1016/S0169-7722(98)00078-3

Snodgrass MF, Kitanidis PK (1997) A geostatistical approach to contaminant source identification. Water Resources Research 33(4):537–546, DOI 10.1029/96WR03753

Sohn M, Reynolds P, Gadgil A, Sextro R (2002a) Rapidly locating sources and predicting contaminant dispersion in buildings. In: Proceedings ofthe Indoor Air 2002 Conference, vol 2002, pp 211–216, URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Rapidly+locating+sources+and+predicting+contaminant+dispersion+in+buildings#0

Sohn MD, Reynolds P, Singh N, Gadgil AJ (2002b) Rapidly locating and characterizing pollutant releases in buildings. Journal of the Air and Waste Management Association 52(12):1422–1432, DOI 10.1080/10473289.2002.10470869

Sonnenborg TO, Engesgaard P, Rosbjerg D (1996) Contaminant transport at a waste residue deposit. 1. Inverse flow and nonreactive transport modeling. Water Resources Research 32(4):925–938, DOI 10.1029/95WR03823

Srivastava D, Singh RM (2014) Breakthrough Curves Characterization and Identification of an Unknown Pollution Source in Groundwater System Using an Artificial Neural Network (ANN). Environmental Forensics 15(2):175–189, DOI 10.1080/15275922.2014.890142, URL http://www.tandfonline.com/doi/abs/10.1080/15275922.2014.890142

Srivastava D, Singh RM (2015) Groundwater System Modeling for Simultaneous Identification of Pollution Sources and Parameters with Uncertainty Characterization. Water Resources Management 29(13):4607–4627, DOI 10.1007/s11269-015-1078-8, URL http://link.springer.com/10.1007/s11269-015-1078-8

Stanev VG, Iliev FL, Hansen S, Vesselinov VV, Alexandrov BS (2018) Identification of release sources in advection–diffusion system by machine learning combined with Green's function inverse method. Applied Mathematical Modelling 60:64–76, DOI 10.1016/J.APM.2018.03.006, URL https://www.sciencedirect.com/science/article/pii/S0307904X18301227

Sun AY (2007) A robust geostatistical approach to contaminant source identification. Water Resources Research 43(2):1–12, DOI 10.1029/2006WR005106

Sun AY (2008) CONSID: A toolbox for contaminant source identification. Ground Water DOI 10.1111/j.1745-6584.2008.00434.x

Sun AY, Painter SL, Wittmeyer GW (2006a) A constrained robust least squares approach for contaminant release history identification. Water Resources Research 42(4):1–13, DOI 10.1029/2005WR004312

Sun AY, Painter SL, Wittmeyer GW (2006b) A robust approach for iterative contaminant source location and release history recovery. Journal of Contaminant Hydrology 88(3-4):181–196, DOI 10.1016/j.jconhyd.2006.06.006

Tarantola A (2005) Inverse problem theory and methods for model parameter estimation. SIAM

Telci IT, Aral MM (2011) Contaminant Source Location Identification in River Networks Using Water Quality Monitoring Systems for Exposure Analysis. Water Quality, Exposure and Health 2(3-4):205–218, DOI 10.1007/s12403-011-0039-6

Todaro V, D'Oria M, Tanda MG, Gómez-Hernández JJ (2021) Ensemble smoother with multiple data assimilation to simultaneously estimate the source location and the release history of a contaminant spill in an aquifer. Journal of Hydrology 598:126215, DOI 10.1016/j.jhydrol.2021.126215, URL https://linkinghub.elsevier.com/retrieve/pii/S0022169421002626

Vesselinov VV, Alexandrov BS, O'Malley D (2018) Contaminant source identification using semi-supervised machine learning. Journal of Contaminant Hydrology 212:134–142, DOI 10.1016/J.JCONHYD.2017.11.002, URL https://www.sciencedirect.com/science/article/pii/S0169772217301201

Vesselinov VV, Alexandrov BS, O'Malley D (2019) Nonnegative tensor factorization for contaminant source identification. Journal of Contaminant Hydrology 220:66–97, DOI 10.1016/j.jconhyd.2018.11.010, URL https://linkinghub.elsevier.com/retrieve/pii/S0169772218300809

Voss CI (1984) Sutra (saturated-unsaturated transport). a finite-element simulation model for saturated-unsaturated, fluid-density-dependent ground-water flow with energy transport or chemically-reactive single-species solute transport. Tech. rep., Geological Survey Reston Va Water Resources Div

Wagner BJ (1992) Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modelling. Journal of Hydrology 135(1-4):275–303, DOI 10.1016/0022-1694(92)90092-A

Wang H, Jin X (2013) Characterization of groundwater contaminant source using Bayesian method. Stochastic Environmental Research and Risk Assessment 27(4):867–876, DOI 10.1007/s00477-012-0622-9

Wang H, Lu W (2020) Recognizing groundwater DNAPL contaminant source and aquifer parameters using parallel heuristic search strategy based on Bayesian approach. Stochastic Environmental Research and Risk Assessment DOI 10.1007/s00477-020-01909-7, URL http://link.springer.com/10.1007/s00477-020-01909-7

Wang H, Lu W, Chang Z, Li J (2020) Heuristic search strategy based on probabilistic and geostatistical simulation approach for simultaneous identification of groundwater contaminant source and simulation model parameters. Stochastic Environmental Research and Risk Assessment 1, DOI 10.1007/s00477-020-01804-1, URL https://doi.org/10.1007/s00477-020-01804-1

Wang H, Lu W, Chang Z (2021a) An iterative updating heuristic search strategy for groundwater contamination source identification based on an ACPSO–ELM surrogate system. Stochastic Environmental Research and Risk Assessment 2, DOI 10.1007/s00477-021-01994-2, URL https://doi.org/10.1007/s00477-021-01994-2http://link.springer.com/10.1007/s00477-021-01994-2

Wang H, Lu W, Chang Z (2021b) Simultaneous identification of groundwater contamination source and aquifer parameters with a new weighted–average wavelet variable–threshold denoising method. Environmental Science and Pollution Research DOI 10.1007/s11356-021-12959-x, URL http://link.springer.com/10.1007/s11356-021-12959-x

Wang J, Zabaras N (2006) A Markov random field model of contamination source identification in porous media flow. International Journal of Heat and Mass Transfer 49(5-6):939–950, DOI 10.1016/j.ijheatmasstransfer.2005.09.016

Wang X, Zhang D (2021) Inverse identification of multiple contaminant sources and pollution states using future spatial-temporal information. Journal of Cleaner Production 289:125790, DOI 10.1016/j.jclepro.2021.125790, URL https://doi.org/10.1016/j.jclepro.2021.125790https://linkinghub.elsevier.com/retrieve/pii/S095965262100010X

Wasilkowski GW, Wozniakowski H (1995) Explicit cost bounds of algorithms for multivariate tensor product problems. Journal of Complexity 11(1):1–56

Wilson JL, Liu J (1994) Backward tracking to find the source of pollution. Water Manage Risk Remediation 1:181–199

Woodbury A, Sudicky E, Ulrych TJ, Ludwig R (1998) Three-dimensional plume source reconstruction using minimum relative entropy inversion. Journal of Contaminant Hydrology 32(1-2):131–158, DOI 10.1016/S0169-7722(97)00088-0

Woodbury AD, Ferguson G (2006) Ground surface paleotemperature reconstruction using information measures and empirical bayes. Geophysical research letters 33(6)

Woodbury AD, Ulrych TJ (1996) Minimum Relative Entropy Inversion: Theory and Application to Recovering the Release History of a Groundwater Contaminant. Water Resources Research 32(9):2671–2681, DOI 10.1029/95WR03818, URL http://doi.wiley.com/10.1029/95WR03818

Xia X, Jiang S, Zhou N, Li X, Wang L (2019a) Genetic algorithm hyper-parameter optimization using taguchi design for groundwater pollution source identification. Water Science and Technology: Water Supply 19(1), DOI 10.2166/ws.2018.059

Xia X, Zhou N, Wang L, Li X, Jiang S (2019b) Identification of transient contaminant sources in aquifers through a surrogate model based on a modified self-organizing-maps algorithm. Hydrogeology Journal 27(7):2535–2550, DOI 10.1007/s10040-019-02003-1

Xing Z, Qu R, Zhao Y, Fu Q, Ji Y, Lu W (2019) Identifying the release history of a groundwater contaminant source based on an ensemble surrogate model. Journal of Hydrology 572, DOI 10.1016/j.jhydrol.2019.03.020

Xu T, Gómez-Hernández JJ (2016) Joint identification of contaminant source location, initial release time, and initial solute concentration in an aquifer via ensemble Kalman filtering. Water Resources Research 52(8):6587–6595, DOI 10.1002/2016WR019111, URL http://doi.wiley.com/10.1002/2016WR019111

Xu T, Gómez-Hernández JJ (2018) Simultaneous identification of a contaminant source and hydraulic conductivity via the restart normal-score ensemble Kalman filter. Advances in Water Resources 112:106–123, DOI 10.1016/j.advwatres.2017.12.011, URL https://linkinghub.elsevier.com/retrieve/pii/S030917081730756X

Xu T, Gómez-Hernández JJ, Zhou H, Li L (2013a) The power of transient piezometric head data in inverse modeling: An application of the localized normal-score EnKF with covariance inflation in a heterogenous bimodal hydraulic conductivity field. Advances in Water Resources 54:100–118, DOI 10.1016/j.advwatres.2013.01.006, URL http://dx.doi.org/10.1016/j.advwatres.2013.01.006

Xu T, Jaime Gómez-Hernández J, Li L, Zhou H (2013b) Parallelized ensemble kalman filter for hydraulic conductivity characterization. Computers & Geosciences 52:42–49

Yan X, Dong W, An Y, Lu W (2019) A Bayesian-based integrated approach for identifying groundwater contamination sources. Journal of Hydrology 579, DOI 10.1016/j.jhydrol.

2019.124160
Yeh HD, Chang TH, Lin YC (2007) Groundwater contaminant source identification by a hybrid heuristic approach. Water Resources Research 43(9):1–16, DOI 10.1029/2005WR004731
Yeh HD, Lin CC, Yang BJ (2014) Applying hybrid heuristic approach to identify contaminant source information in transient groundwater flow systems. Mathematical Problems in Engineering 2014:1–14, DOI 10.1155/2014/369369
Yeh HD, Lin CC, Chen CF (2016) Reconstructing the release history of a groundwater contaminant based on AT123D. Journal of Hydro-Environment Research 13:89–102, DOI 10.1016/j.jher.2015.06.001, URL https://linkinghub.elsevier.com/retrieve/pii/S1570644315000441
Yuan Y, Liang D (2021) Optimization of Identifying Point Pollution Sources for the Convection-Diffusion-Reaction Equations. Advances in Applied Mathematics and Mechanics 13(1):1–17, DOI 10.4208/aamm.OA-2019-0121
Zanini A, Woodbury AD (2016) Contaminant source reconstruction by empirical Bayes and Akaike's Bayesian Information Criterion. Journal of Contaminant Hydrology 185-186:74–86, DOI 10.1016/j.jconhyd.2016.01.006, URL https://www.sciencedirect.com/science/article/pii/S0169772216300067
Zeng L, Shi L, Zhang D, Wu L (2012) A sparse grid based Bayesian method for contaminant source identification. Advances in Water Resources 37:1–9, DOI 10.1016/j.advwatres.2011.09.011, URL http://dx.doi.org/10.1016/j.advwatres.2011.09.011
Zhang J, Zeng L, Chen C, Chen D, Wu L (2015) Efficient Bayesian experimental design for contaminant source identification. Water Resources Research 51(1):576–598, DOI 10.1002/2014WR015740, URL http://doi.wiley.com/10.1002/2014WR015740
Zhang J, Li W, Zeng L, Wu L (2016a) An adaptive Gaussian process-based method for efficient Bayesian experimental design in groundwater contaminant source identification problems. Water Resources Research 52(8):5971–5984, DOI 10.1002/2016WR018598, URL https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016WR018598
Zhang J, Zheng Q, Wu L, Zeng L (2020) Using Deep Learning to Improve Ensemble Smoother: Applications to Subsurface Characterization. Water Resources Research 56(12):1–15, DOI 10.1029/2020WR027399, 2002.09100
Zhang S, Liu H, Qiang J, Gao H, Galar D, Lin J (2019) Optimization of Well Position and Sampling Frequency for Groundwater Monitoring and Inverse Identification of Contamination Source Conditions Using Bayes' Theorem. Computer Modeling in Engineering & Sciences 119(2):373–394, DOI 10.32604/cmes.2019.03825, URL http://www.techscience.com/CMES/v119n2/29791
Zhang Y, Meerschaert MM, Neupauer RM (2016b) Backward fractional advection dispersion model for contaminant source prediction. Water Resources Research 52(4):2462–2473, DOI 10.1002/2015WR018515, URL http://doi.wiley.com/10.1002/2015WR018515
Zhang Y, Sun H, Lu B, Garrard R, Neupauer RM (2017) Identify source location and release time for pollutants undergoing super-diffusion and decay: Parameter analysis and model evaluation. Advances in Water Resources 107:517–524, DOI 10.1016/j.advwatres.2017.05.017, URL https://linkinghub.elsevier.com/retrieve/pii/S0309170817301628
Zhao H, Zhang J, Chen H, Wang L, Yang Z (2020a) Localization of Groundwater Contaminant Sources Using Artificially Enhanced Catchment. Water 12(7):1949, DOI 10.3390/w12071949, URL https://www.mdpi.com/2073-4441/12/7/1949
Zhao Y, Lu W, Xiao C (2016) A Kriging surrogate model coupled in simulation-optimization approach for identifying release history of groundwater sources. Journal of Contaminant Hydrology 185-186:51–60, DOI 10.1016/j.jconhyd.2016.01.004, URL https://www.sciencedirect.com/science/article/pii/S0169772216300043
Zhao Y, Qu R, Xing Z, Lu W (2020b) Identifying groundwater contaminant sources based on a KELM surrogate model together with four heuristic optimization algorithms. Advances in Water Resources 138:103540, DOI 10.1016/j.advwatres.2020.103540, URL https://linkinghub.elsevier.com/retrieve/pii/S0309170819309698
Zheng C, Wang PP (1999) MT3DMS: A Modular Three-Dimensional Multispecies Transport Model (December):219
Zhou H, Gómez-Hernández JJ, Li L (2014) Inverse methods in hydrogeology: evolution and recent trends. Advances in Water Resources 63:22–37

Zhou Z, Tartakovsky DM (2021) Markov chain monte carlo with neural network surrogates: Application to contaminant source identification. Stochastic Environmental Research and Risk Assessment 35(3):639–651