Contaminant source reconstruction by empirical Bayes and Akaike's Bayesian Information Criterion

Contaminant source reconstruction by empirical Bayes and Akaike's Bayesian Information Criterion

Andrea Zanini, Department of Civil Environmental Land Management Engineering and Architecture, University of Parma, Parma Italy

Allan D. Woodbury*, Department of Civil Engineering, University of Manitoba, Winnipeg, Manitoba, Canada

*now Visiting Professor, Department of Earth and Environmental Sciences, University of British Columbia Okanagan, Kelowna, BC, Canada

Corresponding author: A. Zanini, Department of Civil Environmental Land Management Engineering and Architecture, University of Parma, Parco Area delle Scienze 181/A 43124 Parma, Italy (andrea.zanini@unipr.it).

Abstract

The objective of the paper is to present an empirical Bayesian method combined with Akaike's Bayesian Information Criterion (ABIC) to estimate the contaminant release history of a source in groundwater starting from few concentration measurements in space and/or in time. From the Bayesian point of view, the ABIC considers prior information on the unknown function, such as the prior distribution (assumed Gaussian) and the covariance function. The unknown statistical quantities, such as the noise variance and the covariance function parameters, are computed through the process; moreover the method quantifies also the estimation error through the confidence intervals. The methodology was successfully tested on three test cases: the classic Skaggs and Kabala release function, three sharp releases (both cases regard the transport in a one-dimensional homogenous medium) and data collected from laboratory equipment that consists of a two-dimensional homogeneous unconfined aquifer. The performances of the method were tested with

two different covariance functions (Gaussian and exponential) and also with large measurement error. The obtained results were discussed and compared to the geostatistical approach of Kitanidis (1995).

Highlights

- A new approach to estimate the contaminant release history is presented
- The procedure is very robust and efficient
- The method has been tested on complex cases

Keywords

Akaike's Bayesian Information Criterion; Groundwater; Pollutant transport; Release History; Inverse Problems.

1 Introduction

Contaminant release history identification has received considerable attention in the literature over the past several decades. Although a number of reasonable approaches have been developed during this time no panacea has yet emerged. This is in part due to its ill-posed nature, and frequently, either the data stream is of insufficient length, contains missing data points, or is inaccurate. The reader is referred to Atmadja and Batzoglou (2001), Michalak and Kitandis (2004), Sun et al. (2006) or Cupola et al. (2015) for extensive reviews of this specific problem in groundwater hydrology.

Interest in this area continues because it is a good representative of an inverse problem in hydrology. Since mathematical inversion is a cornerstone-problem in geophysics, the impact of any successful works will be high. Inverse theory, in it's truest sense, is different from standard parameter estimation problems in statistics in that the unknowns sought are functions and not a small set of numbers (Parker, 1977; Tarantola, 1985; Ulrych and Sacchi, 2005). This means that in principle, there are an infinite number of variables sought. A variety of approaches exist and there

are two main avenues to take. One of these deals with the ideal case of an infinite amount of exact data and the unknowns sought are continuous functions. For example, a Fourier transform and its inverse. This is the realm of the applied mathematician and these approaches tend to be analytic or quasi-analytic in nature. Analytic techniques are sensitive to the way data are collected and to noise present. Nevertheless these approaches are useful for their results concerning uniqueness, stability and so on (see also Tarantola, 2005, Ch 5, functional space inversion). The other main avenue relates to the practical problems encountered in the geophysical sciences where the model is "parameterized" into a finite set of parameters and involves the collection of incomplete and noisy data. One could on purpose propose a small number of structures such that more data than unknowns are present. It is in this area where the vast majority of efforts in groundwater are concentrated and a variety of approaches are possible. The more computationally demanding, and perhaps interesting problems are those in which the parameterization is done so that a high degree of resolution is possible, if one is willing to tolerate the ambiguity of the result (e.g. Woodbury and Ulrych, 2000; Painter et al., 2007). Stability in the presence of noise is always an issue, as is uniqueness which is difficult to prove. The technique that permits unique and stable inverse solutions by introducing prior information is called regularization. The widely used Levenberg-Marquardt method imposes "smoothness" to the model. This is essentially the basis for the well known PEST wrap-around code. In fact, Tikhonov showed that once an ill-posed problem becomes properly regularized it becomes stable. For these reasons, parameterized inverse problems are stabilized by weighting with error terms and are regularized to achieve some measure of uniqueness under one norm, or a variety of norms. The validity of the regularization terms becomes apparent, and perhaps justified when the inverse problem is approached from Bayesian or maximum entropy perspectives (Ulrych and Sacchi, 2005).

Specifically, in our review of the literature on this subject (see the above references) suggests that improvements are needed in terms of a reliable procedure, one that is easy to implement, with only few hyperparameters to estimate, and is able to evaluate confidence intervals. For these reason the

purpose of this work is to propose an empirical Bayesian approach combined to the Akaike's Bayesian Information Criterion (ABIC) to estimate the contaminant release history, and to demonstrate its effectiveness.

This work estimates the temporal contaminant release history of a point source with the following simplifying assumptions: the solute is conservative, it is a 1-D or 2-D problem, the source location is known, the flow is uniform and steady, and the transport parameters are known at each point of the domain. These assumptions are necessary for the development and testing of the current methodology. Further, the release concentration is uncertain and its probability density function is assumed multivariate Gaussian. Specifically, we adopt a probabilistic approach to the inversion, assume a Gaussian likelihood and Gaussian prior to the problem, and seek the solution that minimizes Akaike's Bayesian Information Criterion, the ABIC. We propose an important extension to the algorithm that constrains solutions to only positive models and we test the method out on three test cases: the classic Skaggs and Kabala (1994) source, a "midnight dump" example that consists of three delta-like sources and lastly a laboratory experimental dataset, consisting of two measurement points spatially but with synoptic observations, obtained from a laboratory equipment, that reproduces the response of a 2-D unconfined aquifer.

2. Theory

2.1 Contaminant Transport in Groundwater

The following equation (1) describes the transport process in an aquifer reacting to the injection of a non-sorbing, non-reactive solute at a point source (Bear and Verruijt, 1987):

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

where: φ [-] is the effective porosity, $\mathbf{u}(\mathbf{x},t)$ [LT^{-1}] is the effective velocity at location $\mathbf{x}$ and time t [T], $\mathbf{D}(\mathbf{x})$ [L^2T^{-1}] is the dispersion tensor, $C(\mathbf{x},t)$ [ML^{-3}] is the concentration, $m(\mathbf{x_0},t)=c_{in}(t)\cdot q_{in}(\mathbf{x_0},t)$ [MT^{-1}] is the amount of pollutant per unit time injected into the aquifer through the source located

at $\mathbf{x_0}$, $c_{in}(t)$ [ML^{-3}] is the concentration of the released contaminant at time t and $q_{in}(\mathbf{x_0},t)$ [L^3T^{-1}] is the injection flow rate.

Equation (1), considering uniform porosity, can be rewritten as:

$$\varphi \frac{\partial C(\mathbf{x},t)}{\partial \mathrm{t}} = \nabla \cdot [\varphi \mathbf{D}(\mathbf{x}) \nabla C(\mathbf{x},t)] - \nabla \cdot [\varphi \mathbf{u}(\mathbf{x},t) C(\mathbf{x},t)] + m(\mathbf{x_0},t)\delta(\mathbf{x}-\mathbf{x_0}) \tag{2}$$

The solution of equation (2) when associated with the initial and boundary conditions: C($\mathbf{x}$,0) = 0; C(∞,t) = 0, is given by the convolution integral:

$$C(\mathbf{x},t) = \int_0^t m(\mathbf{x_0},\tau) g(\mathbf{x}, t-\tau) d\tau \tag{3}$$

where $g(\mathbf{x},t\text{-}\tau)$ [L^{-3}] is the Kernel function that describes the effects at $\mathbf{x}$ at time t [T] by an impulse injection occurring at $\mathbf{x_0}$ at time τ.

Under simple flow conditions (such as homogeneous, isotropic, absence of withdrawal or recharge) the Kernel functions can be determined analytically, for instance for 1-D flow

$$g(x,t-\tau) = \frac{x}{2\sqrt{\pi D(t-\tau)^3}} \exp\left[-\frac{(x - v(t-\tau))^2}{4D(t-\tau)} \right] \tag{4}$$

In non-uniform flow field it is necessary to employ numerical approaches, such as the Stepwise Input Function procedure methodology developed by Butera et al. (2006; 2013), that is a numerical strategy for Kernel functions calculation. The time derivative of equation (3), considering a constant and known input function $m(\mathbf{x}_0,t) = F_0 \cdot H(t)$, where $H(t)$ [-] is the Heaviside step function and $F_0 = c_0 \cdot q_{in}(\mathbf{x_0},t)$ [MT^{-1}] is the amount of pollutant per unit time injected into the aquifer with constant and known concentration c_0, results in:

$$g(\mathbf{x},t) = \frac{1}{F_0} \frac{\partial C(\mathbf{x},t)}{\partial t} = \frac{1}{c_0 q_{in}(\mathbf{x}_0,t)} \frac{\partial C(\mathbf{x},t)}{\partial t} \quad t > 0 \tag{5}$$

Equation (5) shows that it is possible to compute the Kernel functions at a generic point $\mathbf{x}$ by processing the concentration history (breakthrough curve) at the same location due to a stepwise tracer injection at $\mathbf{x_0}$. The application of equation (5) under field conditions is rarely possible but it can be coupled with a numerical flow and transport model that can easily simulate the effect of a

pollutant injection on an aquifer and can calculate the response (breakthrough curve) at each monitoring point.

Note that a continuous inverse problem can be converted into a discrete one by making an assumption that the model $m(\tau)$ can be represented by a finite number of M coefficients, that is:

$$m(\tau) = \sum_{j=1}^{M} m_j \Phi_j(\tau) \quad (0 < \tau < t) \tag{6}$$

One commonly used assumption is that the model is constant over a certain sub-region and in this case Φ is unity inside the subregion and zero outside of it. Other choices of $\Phi_j(\tau)$ are possible and an alternative is the assumption that the Φ_j are linear basis functions; that is, the model is assumed to vary linearly between point estimates. For whatever choice is made the forward problem becomes:

$$d_i = \sum_{j=1}^{M} G_{ij} m_j \tag{7}$$

where the discrete kernel is

$$G_{ij} = \int_0^t g_i(\tau) \Phi_j(\tau) d\tau \tag{8}$$

In a matrix form

$$\mathbf{d} = \mathbf{Gm} \tag{9}$$

In a typical linear inverse problem the observed data $\mathbf{d}^*$ (N×1) are related to a set of model parameters $\mathbf{m}$ (M×1) that are partially hidden by noise $\boldsymbol{\upsilon}$ (N×1), and $M >> N$, so that:

$$\mathbf{d}^* = \mathbf{Gm} + \boldsymbol{\upsilon} \tag{10}$$

The equation (3), thanks to the linearity of the problem, could be written as equation (10) and the matrix $\mathbf{G}$ is represented through the following:

$$\mathbf{G} = \Delta t \cdot \begin{bmatrix} g(\mathbf{x}_1, T - t_1) & \dots & g(\mathbf{x}_1, T - t_M) \\ g(\mathbf{x}_2, T - t_1) & \dots & g(\mathbf{x}_2, T - t_M) \\ \dots & \dots & \dots \\ g(\mathbf{x}_N, T - t_1) & \dots & g(\mathbf{x}_N, T - t_M) \end{bmatrix} \tag{11}$$

where $g(\mathbf{x}_i,t)$ is the kernel function computed at $\mathbf{x}_i$ (observation location) at time t, T is the sampling time and Δt is the time interval.

2.2. Empirical Bayes and Hyperparameter Estimation by ABIC

In this work we follow a Bayesian paradigm to construct an inverse solution to the contaminant source problem as noted above. The Bayesian philosophy for inverse problems in groundwater is detailed in Woodbury (2007), amongst other sources, and is briefly restated below.

If the conditional pdf of $\mathbf{d}^*$ given $\mathbf{m}$ and prior information I, is given by $p(\mathbf{d}^*|\mathbf{m}, I)$, Bayes' rule dictates that:

$$p(\mathbf{m}|\mathbf{d}^*, I) = \frac{p(\mathbf{d}^*|\mathbf{m}, I)p(\mathbf{m}|I)}{\int p(\mathbf{d}^*|\mathbf{m}, I)p(\mathbf{m}|I)d\mathbf{m}} \tag{12}$$

In (12) $p(\mathbf{m}|I)$ is the prior probability density (pdf) of the model parameters, given prior information, I, and $p(\mathbf{d}^*|\mathbf{m}, I)$, is the likelihood of observing $\mathbf{d}^*$ given the model parameters and the prior information. The left hand side of (12) is the posterior probability density and the term in the denominator is a normalizing constant which is important in this work, as it represents the pdf of observing data with the uncertainty in the model parameters marginalized out of consideration. This discussion will no doubt be familiar to the reader but it is appropriate here to repeat this for the sake of completeness and understanding of the notation used.

Given the prior pdf and new information in the form of sample data, we compute expected values of the posterior pdf, which in this paradigm is taken as the "answer". Here we also assume the observations have errors as described by matrix $\mathbf{C}_d$. If the errors in the data and the prior information on the model parameters are described by the Gaussian hypothesis (covariance $\mathbf{C}_m$ and mean $\mathbf{s}$) then the posterior probability in model space is also Gaussian. The latter assumption on the model space is a very common one to be made in the hydrologic sciences. According to many previous works (such as: Kitanidis and Vomvoris (1983); Carrera and Neuman (1986); Kitanidis (1995); Snodgrass and Kitanidis (1997); Woodbury and Ulrych (2000); Ulrych et al. (2001)) the

prior probability distribution of the unknowns is assumed as Gaussian. In assigning prior probabilities, the principle of indifference it is assumed, adopting probabilistic models with maximum entropy; for instance, if the mean and variance are known, the Gaussian is the maximum-entropy assignment of probabilities. The first two moments of this pdf are given by Tarantola (1987), see also Woodbury (2011):

$$\langle \mathbf{m} \rangle = \mathbf{s} + \mathbf{C}_m \mathbf{G}^T \left(\mathbf{G}\mathbf{C}_m \mathbf{G}^T + \mathbf{C}_d \right)^{-1} \left(\mathbf{d}^* - \mathbf{G}\mathbf{s} \right) \tag{13}$$

$$\mathbf{C}_q = \mathbf{C}_m - \mathbf{C}_m \mathbf{G}^T \left(\mathbf{G}\mathbf{C}_m \mathbf{G}^T + \mathbf{C}_d \right)^{-1} \mathbf{G}\mathbf{C}_m \tag{14}$$

where $\langle \mathbf{m} \rangle$ and $\mathbf{C}_q$ are the expected value and covariance of the posterior pdf, respectively. These results are equally valid for the continuous inverse; the only difference between them is that probability densities are defined for discrete spaces and not for continuous ones.

$\mathbf{C}_m = \sigma_M^2 \exp\left(-\left| t_i - t_j \right| / \lambda \right)$ is typically assumed to have an exponential autocovariance (see Beck et al., 1992), where t_i-t_j is the separation distance (time units), σ_M^2 the variance and λ the integral scale. The premise of the empirical Bayesian approach is that the prior probability is considered flexible (Ulrych et al., 2001). This means that the noise C_d and matrix $\mathbf{C}_m$ with the statistical hyperparameters, such as the mean **s**, the variance σ_M^2 and the integral scale λ may not be known initially, and are determined from information contained in the input data. In this way, the prior pdf is used constrain the solution to a form that is considered appropriate. For example constraining ln(K), the hydraulic conductivity, to follow a Gaussian distribution even though we may not know its mean and variance.

Note the denominator of (12) represents the pdf of observing data, with the uncertainty in the model parameters marginalized out. In the empirical Bayes approach the denominator depends on any hyperparameters in the prior, for example:

$$\int p\left(\mathbf{d}^* \middle| \mathbf{m}, I\right) p\left(\mathbf{m} \middle| I\right) d\mathbf{m} = p\left(\mathbf{d}^* \middle| I\right) = p\left(\mathbf{d}^* \middle| \sigma_M^2, \sigma_d^2, \lambda\right) \tag{15}$$

(if, say three hyperparameters are present). For example, suppose there are two alternative assumptions related to the prior information, such that $p(\mathbf{d}^*|I_1) \geq p(\mathbf{d}^*|I_2)$; we would want to select I_1 over I_2 as the "best" candidate for the prior information. As will be shown in the sections below there are different ways that choice could be made.

The particular approach that we adopt is based on the work of Akaike (1980), specifically Akaike's Information Criterion (AIC) and the Akaike's Bayesian Information Criterion (ABIC). A detailed derivation of the AIC may be found in Matsuoka and Ulrych (1986). The AIC is based on the Kullback-Liebler information measure (Shibata, 2002). The minimum AIC, $\mathrm{AIC}|_{\min}$, is the best trade-off between errors in parameter estimation and errors in fitting of the model. Denoting the estimated model by $\hat{\mathbf{m}}$, the form of the measure is (Ulrych et al., 2001)

$$\mathrm{AIC} = -2\ln\left\{p\left(\mathbf{d}^* \mid \hat{\mathrm{m}}, I\right)\right\}_{\max} + 2k \tag{16}$$

In (16) the first term on the RHS is essentially the likelihood function, which for normally distributed errors is s_M^2 is the residual sum of squares, and the second term is twice k is the number of parameters independently adjusted for the maximization of the likelihood (Akaike, 1974); most of the time k is equal to the dimension of $\hat{\mathbf{m}}$. The minimum of the AIC allows the computation of the appropriate number of parameters, a particularly difficult task in many geophysical problems.

The ABIC is similar in form to the AIC and is computed in terms of the Bayesian likelihood (Ulrych et al., 2001); it is appropriate when using Bayes' rule and is defined in equation (17)

$$\mathrm{ABIC} = -2\ln\left\{p\left(\mathbf{d}^*|I\right)\right\} + 2N_h \tag{17}$$

N_h is the number of hyperparameters evaluated at the minimum value of the ABIC.

For the discrete linear inverse problem with Gaussian priors and likelihood, Mitsuhata (2004) after some manipulation showed that:

$$p\left(\mathbf{d}^*|I\right) = (2\pi)^{-N/2}\left\|\mathbf{C}_{dp}\right\|^{-1/2}\exp\left\{-\frac{1}{2}\left(\mathbf{d}^* - \mathbf{Gs}\right)^T\mathbf{C}_{dp}^{-1}\left(\mathbf{d}^* - \mathbf{Gs}\right)\right\} \tag{18}$$

Where $\mathbf{C}_{dp} = \mathbf{GC}_m\mathbf{G}^T + \mathbf{C}_d$.

The AIC is appropriate when there is no prior information on the model parameters, while the ABIC when prior information is considered in the form of a prior probability. Moreover, the AIC uses the quality of the optimum estimation, obtained under the model employed in the stage of deriving the criterion, which is treated in the form of an estimate of the probability density function; whereas the ABIC treats the overall approximation capability for the unknowns of the entire models (Ulrych et al., 2001; Shibata, 2002; Woodbury and Ferguson, 2006).

For the simulations in this paper, we assume three principle hyperparameters are of interest. The first is σ_d^2, the variance of the noise in the observed data, such that $\mathbf{C}_d = \sigma_d^2 \mathbf{I}$. Second, there is an unknown scale parameter for the correlation matrix; i.e., $\mathbf{C}_c \sigma_M^2 = \mathbf{C}_m$. Third, embedded within the correlation matrix $\mathbf{C}_c$ is another hyperparameter, λ the integral scale.

According to the work of Mitsuhata (2004) and Woodbury and Ferguson (2006), combining equation (17) with (18) the ABIC is:

$$\text{ABIC} = N \ln(2\pi) + \ln \| \mathbf{C}_{dp} \| + (\mathbf{d}^* - \mathbf{Gs})^T \mathbf{C}_{dp}^{-1} (\mathbf{d}^* - \mathbf{Gs}) + 2N_h \qquad (19)$$

assuming three hyperparameters, $2N_h = 6$. In the above, terms two and three both depend on σ_d^2 and σ_M^2.

The procedure for determining the hyperparameters through the ABIC and to estimate $\langle \mathbf{m} \rangle$ is as follows:

1. input a starting value for σ_d^2, λ and σ_M^2, form $\mathbf{C}_m$;
2. form $\mathbf{C}_{dp}$, decompose, compute the natural log of its determinant;
3. form $(\mathbf{d}^* - \mathbf{Gs})^T \mathbf{C}_{dp}^{-1} (\mathbf{d}^* - \mathbf{Gs})$ (the χ^2 misfit) and minimize the ABIC for σ_M^2 and λ holding σ_d^2 fixed;
4. compute $\langle \mathbf{m} \rangle$ and estimate σ_d^2 from the χ^2 misfit between computed and observed data;

5. if the improvement in the objective function, computed between two successive iterations as $2 \cdot \left|\mathrm{ABIC}_i - \mathrm{ABIC}_{i-1}\right| / \left[\left|\mathrm{ABIC}_i\right| + \left|\mathrm{ABIC}_{i-1}\right|\right]$, converges to a stipulated tolerance then step out of the loop, otherwise go back to step 2;
6. the iterations are complete and then compute the final model, covariance, the ABIC, and any other pertinent measures.

Our procedure accounts for the uncertainty in σ_M^2 in any one iteration by including its variability in terms two and three in (19). It is important to remark that the determinant of the matrix $\mathbf{C}_{dp}$ for under-determined inverse problems can be very small and consequently its natural logarithmic is a significant negative number and therefore it has an important role in the computation of ABIC. In this work, we simultaneously estimated σ_M^2 and λ holding σ_d^2 fixed during the iteration.

This method is very effective and efficient (Kennedy et al., 2000; Hendry and Woodbury, 2007) but does not constrain the release **m** to only positive values, which is an issue with respect to concentrations whose values must only be positive. This issue is not of concern with respect to entropy based solutions (see Woodbury and Ulrych, 1998) as the solutions in MRE are naturally constrained as part of the theory. Non-positive models though, in Bayesian and other inversion techniques may suffer from side lobe "ringing" and this is often noted in geophysical problems. A solution can be obtained by working in a transformed space and solving the equations iteratively. Box and Cox (1964) and Snodgrass and Kitanidis (1997), with the aim at avoiding negative concentrations suggested the use of a power transformation of the unknown variable **m**. The new unknown function becomes:

$$\tilde{\mathbf{m}} = \alpha\left(\mathbf{m}^{1/\alpha} - 1\right) \tag{20}$$

where α is a positive number and it is chosen as small as possible while ensuring $\tilde{m}_i > -\alpha$. For $\tilde{m}_i < -\alpha$ imaginary results from equation (20) are possible. This transformation is general: it

includes the unmodified **m** ($\alpha = 1$), the square root of **m** ($\alpha = 2$), and in the limit for large value of α it reduces to a logarithm transformation.

Considering the variable transformation of equation (20), the equation (10) becomes:

$$\mathbf{d}^* = \tilde{\mathbf{g}}(\tilde{\mathbf{m}}) + \nu \tag{21}$$

Where $\tilde{\mathbf{g}}(\tilde{\mathbf{m}}) = \mathbf{g}\left[\left(\frac{\tilde{\mathbf{m}} + \alpha}{\alpha}\right)^{\alpha}\right]$.

In this case the kernel function $\tilde{\mathbf{g}}(\tilde{\mathbf{m}})$ is not linear with the transformed unknown $\tilde{\mathbf{m}}$. The solution is reached iteratively starting from an initial estimate of the unknowns $\tilde{\mathbf{m}}_0$ and making the derivative of $\tilde{\mathbf{g}}$ with respect to $\tilde{\mathbf{m}}$ at $\tilde{\mathbf{m}}_0$: $\tilde{\mathbf{G}}_0 = \left.\frac{\partial \tilde{\mathbf{g}}}{\partial \tilde{\mathbf{m}}}\right|_{\tilde{\mathbf{m}}=\tilde{\mathbf{m}}_0}$.

The procedure for determining the hyperparameters through the ABIC and to estimate $\langle \mathbf{m} \rangle$, given a value of $\tilde{\mathbf{s}}$ and α, is:

1. input the starting value for σ_d^2, λ, σ_M^2 and $\tilde{\mathbf{m}}$;
2. compute the transformed Kernel function:

$$\tilde{\mathbf{G}} = \Delta t \cdot \begin{bmatrix} \left(\frac{\tilde{m}(t_1)+\alpha}{\alpha}\right)^{\alpha-1} g(\mathbf{x}_1, T-t_1) & \dots & \left(\frac{\tilde{m}(t_M)+\alpha}{\alpha}\right)^{\alpha-1} g(\mathbf{x}_1, T-t_M) \\ \left(\frac{\tilde{m}(t_1)+\alpha}{\alpha}\right)^{\alpha-1} g(\mathbf{x}_2, T-t_1) & \dots & \left(\frac{\tilde{m}(t_M)+\alpha}{\alpha}\right)^{\alpha-1} g(\mathbf{x}_2, T-t_M) \\ \dots & \dots & \dots \\ \left(\frac{\tilde{m}(t_1)+\alpha}{\alpha}\right)^{\alpha-1} g(\mathbf{x}_N, T-t_1) & \dots & \left(\frac{\tilde{m}(t_M)+\alpha}{\alpha}\right)^{\alpha-1} g(\mathbf{x}_N, T-t_M) \end{bmatrix} \tag{22}$$

3. compute $\tilde{\mathbf{g}}(\tilde{\mathbf{m}}) = \int_0^t \left(\frac{\tilde{\mathbf{m}} + \alpha}{\alpha}\right)^{\alpha} g(\mathbf{x}, t-\tau) d\tau$;
4. compute $\mathbf{d}_l^* = \mathbf{d}^* - \tilde{\mathbf{g}}(\tilde{\mathbf{m}}) + \tilde{\mathbf{G}}\tilde{\mathbf{s}}$;
5. form $\mathbf{C}_m$;
6. form $\mathbf{C}_{dp}$, decompose, compute the natural log of its determinant;

7. form $\left(\mathbf{d}_l^* - \tilde{\mathbf{G}}\tilde{\mathbf{s}}\right)^T \mathbf{C}_{dp}^{-1}\left(\mathbf{d}_l^* - \tilde{\mathbf{G}}\tilde{\mathbf{s}}\right)$ (the χ^2 misfit) and minimize the ABIC for σ_M^2 and λ holding σ_d^2 fixed;
8. compute $\langle\tilde{\mathbf{m}}\rangle = \tilde{\mathbf{s}} + \mathbf{C}_m\tilde{\mathbf{G}}^T\left(\tilde{\mathbf{G}}\mathbf{C}_m\tilde{\mathbf{G}}^T + \mathbf{C}_d\right)^{-1}\left(\mathbf{d}_l^* - \tilde{\mathbf{G}}\tilde{\mathbf{s}}\right)$ and $\langle\mathbf{m}\rangle = \left(\frac{\langle\tilde{\mathbf{m}}\rangle + \alpha}{\alpha}\right)^{\alpha}$ and estimate σ_d^2 from the misfit between computed and observed data;
9. if the improvement in the objective function, computed between two successive iterations as $2\cdot\left|\mathrm{ABIC}_i - \mathrm{ABIC}_{i-1}\right| / \left[\left|\mathrm{ABIC}_i\right| + \left|\mathrm{ABIC}_{i-1}\right|\right]$, converges to a stipulated tolerance then step out of the loop, otherwise go back to step 2;
10. the iterations are completed and then compute the final model, covariance, the ABIC, resolution, and so on.

Once that the iterative procedure is completed, it is possible to estimate the covariance as $\tilde{\mathbf{C}}_q = \mathbf{C}_m - \mathbf{C}_m\tilde{\mathbf{G}}^T\left(\tilde{\mathbf{G}}\mathbf{C}_m\tilde{\mathbf{G}}^T + \mathbf{C}_d\right)^{-1}\tilde{\mathbf{G}}\mathbf{C}_m$ and the 95% confidence interval as: $\tilde{\mathbf{C}}_{\text{intervals}} = \tilde{\mathbf{m}} \pm 1.96\sqrt{\tilde{\mathbf{C}}_{q_{ii}}}$, where $\tilde{\mathbf{C}}_{q_{ii}}$ are the diagonal elements of $\tilde{\mathbf{C}}_q$. The final step is the transformation of the confidence intervals as: $\text{confidence intervals} = \left(\frac{\tilde{\mathbf{C}}_{\text{intervals}} + \alpha}{\alpha}\right)^{\alpha}$.

The reader should also be aware of the conundrum that this poses to us, at least philosophically. In empirical Bayes we start out with an assumption about the likelihood function (Gaussian) and the model itself (Gaussian). Following the algorithm, the solution is also Gaussian and mean values, confidence limits and so on follow naturally and immediately. However, if we follow the transformation algorithm proposed above we now assume that the transformed variable $\tilde{\mathbf{m}}$ is Gaussian (not the actual model, **m)** with perhaps little justification. Practically speaking this poses really no issues and aids greatly in suppressing side lobes in the solution. More on these issues will written in sections below.

2.3 Incorporation of prior information from observed data

The inverse procedure requires initial values of the hyperparameters in order to start the estimation process. Moreover, some general information, with regards to the phenomenon in study, is needed. The starting values for the hyperparameters should be selected preferring a simple, flat solution from which complexity in the form of release variability could be added by means the proposed approach when it is supported by the data. The initial estimate of the hyperparameters, see Butera and Tanda (2001), can be carried out assuming as initial variance a value that is greater than the variance of the observed values: $\sigma_M^2 > \sigma_{Observations}^2$ and as initial correlation length λ [T] a value that is smaller than the one of the observed values $\lambda_{Observations}$ [L] divided by the mean effective velocity [L/T]: $\lambda < \frac{\lambda_{Observations}}{v}$. The measurement error σ_d^2 can also be inferred from sampling methods and techniques.

The proposed approach assumes the prior mean of the unknowns. This enforces the convergence of the procedure. An approximate idea on the mean could be obtained from a simple forward modeling of the phenomenon and a further weak calibration. Knowing the source location, the flow field and transport parameters, it is possible to simulate the injection into the aquifer of a flux with constant and known concentration. Considering a few (<5) injections with concentrations that present different orders of magnitude, the mean value could be chosen as the injected concentration that presents the minimum variance between the observed and computed values at the monitoring points.

3. Study cases

Three cases were considered to test and validate the methodology. The first is a widely regarded study case initially proposed by Skaggs and Kabala (1994) and successively adopted by several researchers (Woodbury and Ulrych, 1996; Snodgrass and Kitanidis, 1997; Butera and Tanda, 2003; Sun, 2007; Butera et al., 2013). The second study case, which we refer to as a “midnight dump” presents the same forward model of the first but considers a very simple but difficult release function that could be summarized in three uncorrelated and sharp releases (this case essentially

examines the ability to reconstruct a delta-like source). The third case uses data collected under controlled condition in a laboratory framework. All the study cases assumed the source location is known.

3.1 CASE 1 - Skaggs and Kabala release function

The problem consists of 1-D steady state flow and it can be reproduced through equations (3) and (4). The velocity v = 1 m/d, the dispersion coefficient D = 1 m^2/d and the sampling time is T = 300 d. The adopted release function was the one initially proposed by Skaggs and Kabala (1994):

$$c_{in}(\mathbf{x}_0,t) = \exp\left(\frac{-\left(\frac{t}{\Delta t}-130\right)^2}{50}\right) + 0.3\cdot\exp\left(\frac{-\left(\frac{t}{\Delta t}-150\right)^2}{200}\right) + 0.5\cdot\exp\left(\frac{-\left(\frac{t}{\Delta t}-190\right)^2}{98}\right) \tag{23}$$

Since $q_{in}(\mathbf{x}_0,t)$ is of unit value, the identification of the release history $\mathbf{m}$ is equivalent to the identification of the concentration history $c_{in}(\mathbf{x}_0,t)$. The same observations (N = 25) and release history discretization (M = 300) of the work of Snodgrass and Kitanidis (1997) were considered in order to allow the reader a comparison to other Bayesian approaches (Figure 1). According to previous works the results are reported dimensionless.

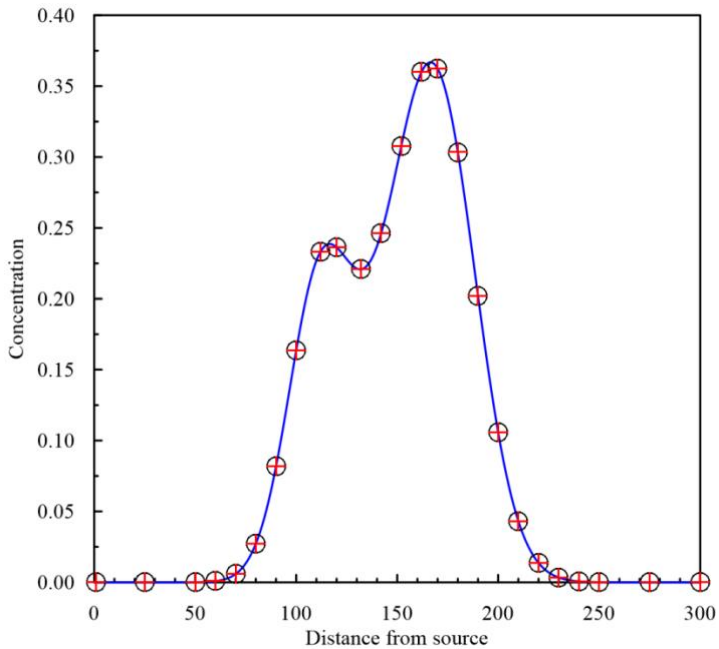

Figure 1. CASE 1 – The blue line is the concentration observed after 300 time units, the circles denote the measurement locations used in the inverse procedure and the red crosses are the estimated concentration by means of the constrained case. The observed concentrations were not corrupted by errors.

At first the release history (Figure 2) was recovered considering an exponential covariance function $\mathbf{C}_m = \sigma_M^2 \exp\left(-\left|t_i - t_j\right| / \lambda\right)$, data without measurement errors and through the unconstrained case. Figure 2 shows the true release history, the best estimate and its 95% confidence intervals. The recovered release history is well recovered, but negative concentration values and confidence limits were estimated and appeared as side-lobes. In order to avoid these negative values the constrained approach was applied.

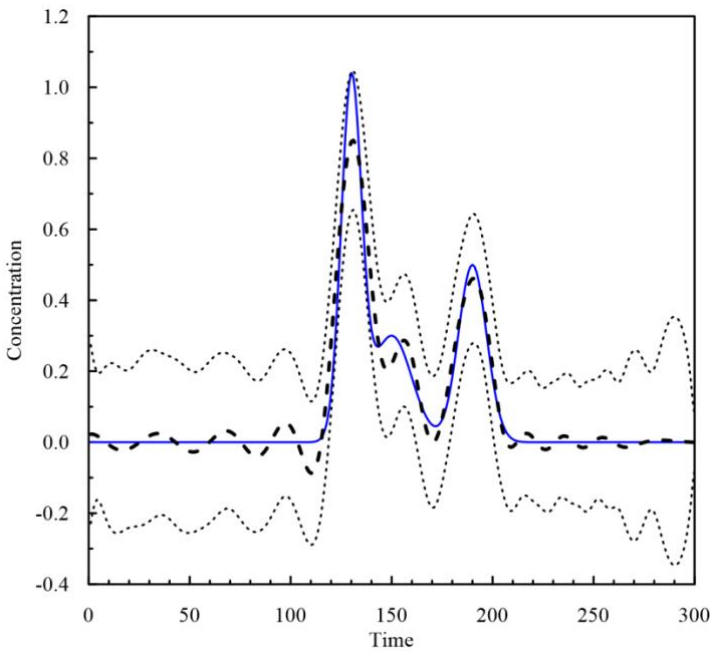

Figure 2. CASE 1 – Unconstrained ABIC with exponential covariance function: the true solution (solid blue line), best estimate (dashed thick line), and 5-95% confidence interval (dotted lines).

Figure 3 shows that the release history, with the same assumptions and dataset of the previous test, is very well recovered through the proposed approach; moreover Figure 1 reports the good agreement between the observed and computed concentrations. Figure 4 shows the results obtained using the dataset of the previous examples but considering a Gaussian covariance function: $\mathbf{C}_m = \sigma_M^2 \exp\left(-\left(t_i - t_j\right)^2 / \lambda^2\right)$. It is clear that the release function is very well recovered and presents

a 95% confidence interval much smaller than the case with an exponential covariance function. A comparison between the results obtained under different covariance functions could be carried out comparing the final values of the ABIC considering the same initial values, α and convergence criteria. The estimate due to the Gaussian covariance function yields an ABIC = -347.88, while using the exponential covariance the ABIC = -367.00. It is clear that, in this case, the exponential covariance function identifies a solution that presents a marginally lower value of ABIC. Any observer analyzing a generic problem without knowing the true source release can identify the most appropriate covariance function choosing the one that generates the minimum ABIC value.

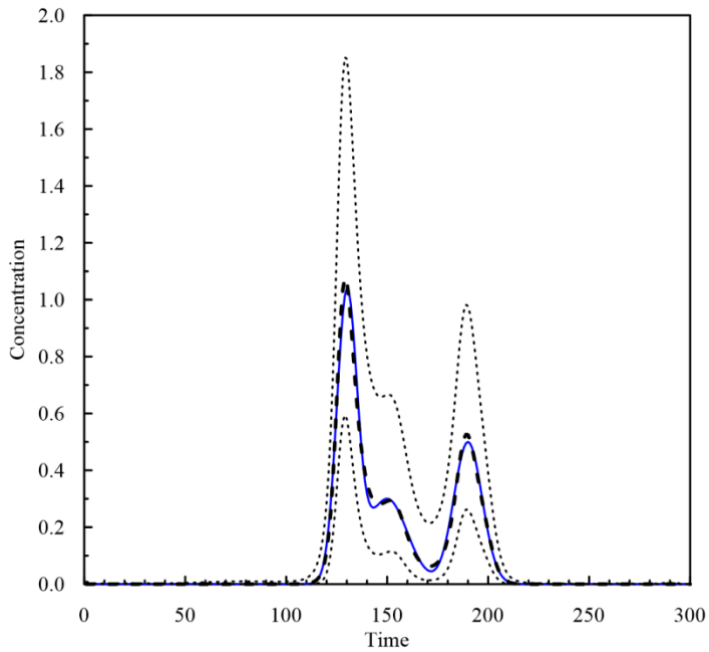

Figure 3. CASE 1 – Constrained ABIC with exponential covariance function: the true solution (solid blue line), best estimate (dashed thick line), and 5-95% confidence interval (dotted lines).

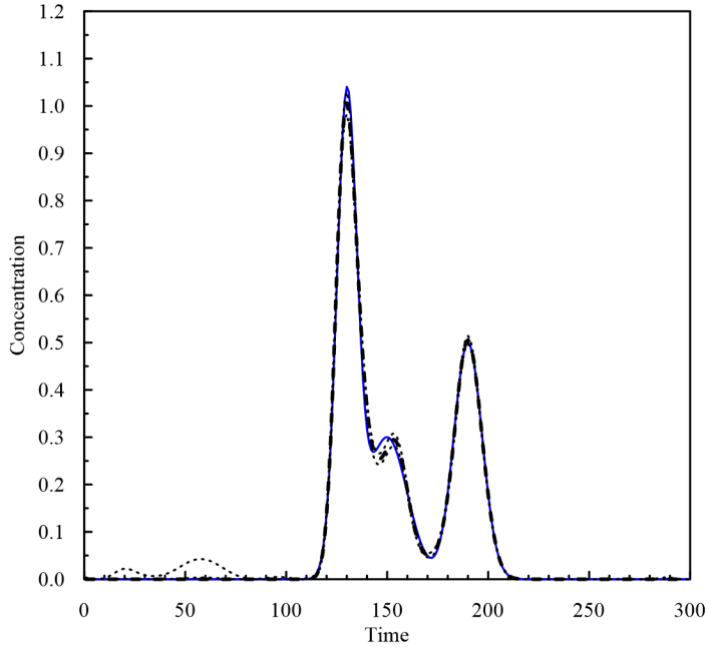

Figure 4. CASE 1 – Constrained ABIC with Gaussian covariance function: the true solution (solid blue line), best estimate (dashed thick line), and 5-95% confidence interval (dotted lines).

Finally, the observed concentrations were corrupted by a random normally distributed error with $\sigma_d^2 = 10^{-3}$. Figure 5 shows that the release history is quite well recovered but only the two main peaks are detected and they are underestimated. Considering the quality of the dataset, the performance of the procedure was considered satisfactory and the "true" signal is contained within the upper and lower confidence limits.

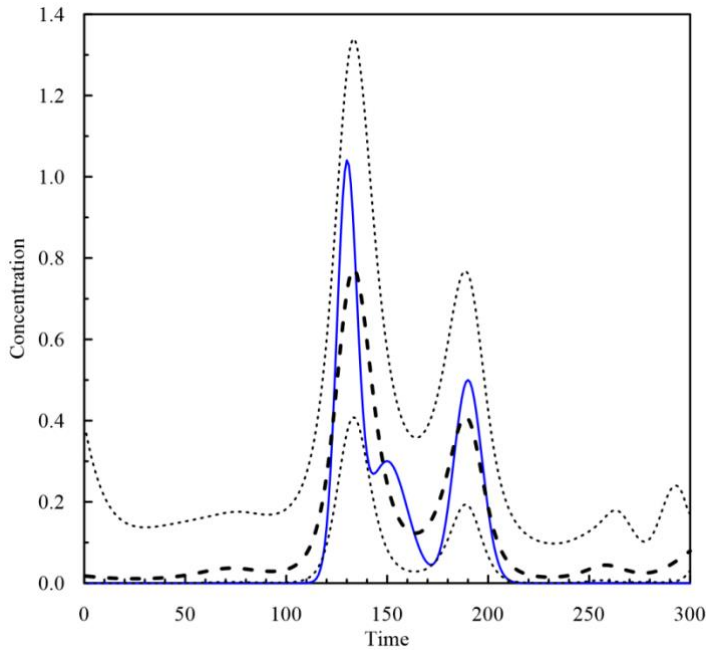

Figure 5. CASE 1 – Constrained ABIC with exponential covariance function, observations added with normally distributed errors ($\sigma^2 = 10^{-3}$): the true solution (solid blue line), best estimate (dashed thick line), and 5-95% confidence interval (dotted lines).

3.2 CASE 2 - Midnight dump

The second case used the same forward model of the previous one but with the release function of Eq. (24).

$$c_{in}(\mathbf{x}_0,t) = \exp\left(\frac{-\left(\frac{t}{\Delta t}-129.75\right)^2}{1}\right) + 0.3\cdot\exp\left(\frac{-\left(\frac{t}{\Delta t}-150\right)^2}{1}\right) + 0.5\cdot\exp\left(\frac{-\left(\frac{t}{\Delta t}-189.75\right)^2}{1}\right) \qquad (24)$$

In this case 30 observations equal spaced (from x = 10 to 300 m with Δx = 10 m) were considered in the inverse procedure (see Figure 6) and the release history was discretized in 400 unknowns with Δt = 0.75 days.

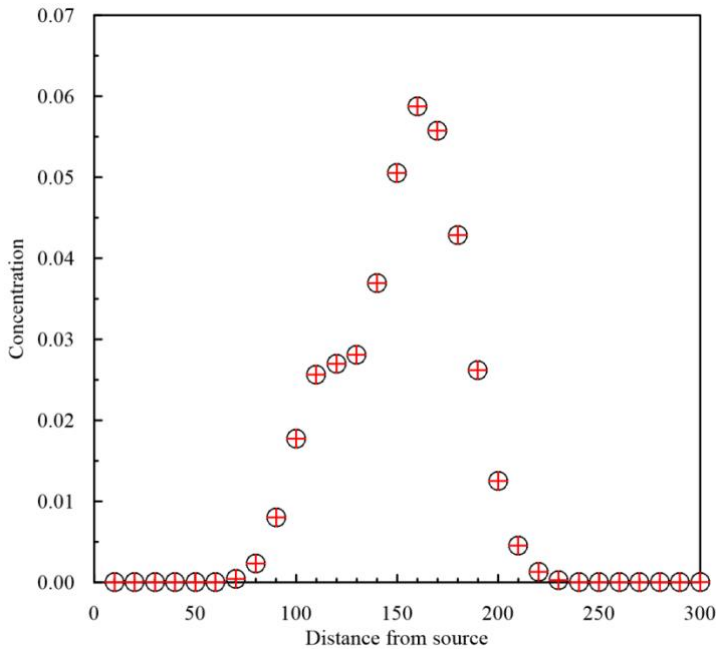

Figure 6. CASE 2 –Concentration after 300 time units, the circles denote the observed concentrations used in the inverse procedure and the red crosses are the estimated concentration by means of the constrained case. The observed concentrations were not corrupted by errors.

This case is very difficult because it presents three uncorrelated peaks that could cause a convergence problem by way of a large step in the newton algorithm we use. This means the new estimate $\langle \mathbf{m}_i \rangle$ could have a greater ABIC value than the previous one $\langle \mathbf{m}_{i-1} \rangle$. In order to avoid this problem a relaxation factor was adopted:

$$\langle \mathbf{m}_i \rangle = \langle \mathbf{m}_i \rangle \cdot (1-\delta) + \langle \mathbf{m}_{i-1} \rangle \cdot \delta \tag{25}$$

where δ is a scalar that defines the point along a line. In this case δ = 0.9 was considered appropriate with the aim at varying the estimate from one iteration to the next one very slowly and avoid oscillations. This implies a large number of iterations and to speed up the computation process the value of δ could be estimated at each iteration, as suggested by Zanini and Kitanidis (2009) for the geostatistical approach and choosing the one that minimizes the ABIC value. Figure 7 shows that the true release history is very well recovered and the observations are optimally reproduced (Figure 6). That indicates that the procedure, amended with the relaxation factor, is efficient also in complex cases. Also in this case, the observed concentrations were corrupted through a random normally distributed error with $\sigma_d^2 = 10^{-9}$. Figure 8 shows that the approach underestimated the release concentrations, but identified the time of the peaks quite well.

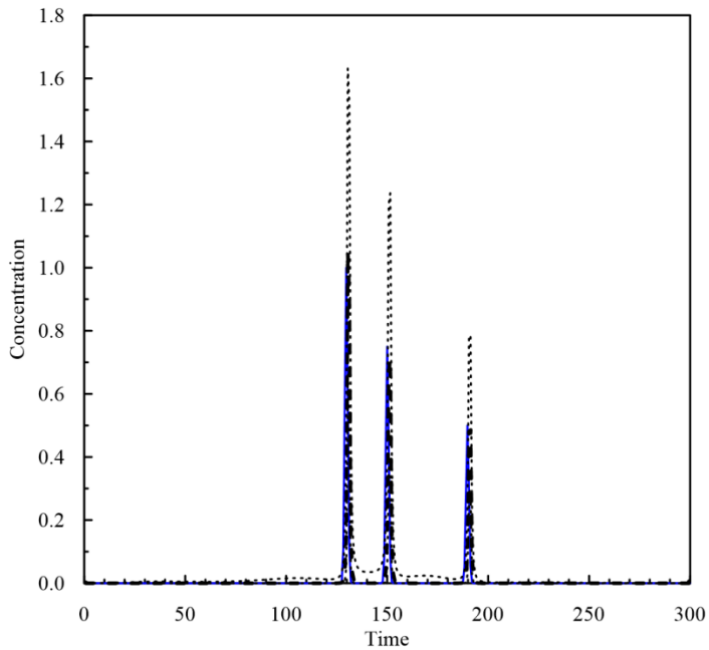

Figure 7. CASE 2 – Constrained ABIC with exponential covariance function: the true solution (solid blue line), best estimate (dashed thick line), and 5-95% confidence interval (dotted lines).

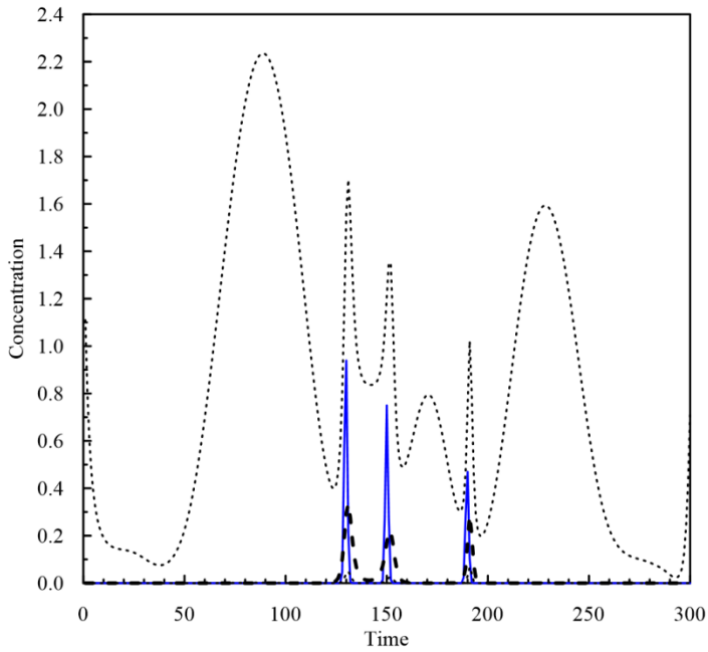

Figure 8. CASE 2 – Constrained ABIC with exponential covariance function, observations added with normally distributed errors ($\sigma^2 = 10^{-9}$): the true solution (solid blue line), best estimate (dashed thick line), and 5-95% confidence interval (dotted lines).

3.3 CASE 3- Laboratory dataset

The procedure was finally tested using data collected through a laboratory device (sandbox) following the work of Cupola et al. (2015). This is an important comparison to note in that the kernel matrix **G** here is developed from a numerical procedure, see Eq. (5), and not an analytic one. In this way all of these procedures here can be used for any inverse problem provided that the matrix **G** is provided.

The sandbox reproduces an unconfined aquifer governed by two constant head levels (upstream and downstream). The porous medium consists of glass beads with diameter in the range between 0.75 and 1 mm with a mean hydraulic conductivity of 0.652 cm/s and porosity of 0.37. A tracer solution (fluorescein sodium salt) with variable mass rate was injected through an injector positioned in the upstream part of the sandbox. It is important to notice that the experiments carried out by Citarella

et al. (2015) and Cupola et al. (2015) were performed under simple conditions; in fact, the porous medium, can represent a non-sorbing coarse sand, and the tracer, may be regarded as conservative under neutral or moderately basic pH (Smart and Laidlaw, 1977). All variables, such as upstream and downstream levels, injected discharge, temperature, background discharge, start and ending of injection, were acquired by means of a data acquisition system. The luminosity at each point of the sandbox was recorded by a digital camera and then converted in concentration through an imaging technique (for more details see Citarella et al., 2015). The average background flow rate was measured as 25 mL/s. The injector was located at the coordinates x_{inj}= 14.25 cm, z_{inj} = 32.75 cm and it was as wide as the central chamber. The test had a time length of 2,200 s. The injection started at time t_{start} = 310 s and finished at t_{end} = 1,800 s considering a constant concentration of 20 mg/L and a variable injection rate (from 0 to about 3 mL/s); consequently the mass rate varied in a range between 0 and approximately 60 μg/s. Through a photographic survey, the tracer concentration was estimated at the monitoring points P1 and P2 (see Figure 3 of Cupola et al. (2015)) with coordinates x_{P1}= 41.25 cm, z_{P1} = 30.71 cm and x_{P2}= 60.25 cm, z_{P2} = 32.18 cm. At such point the concentration was estimated every 5 s for the whole duration of the test, but only 32 (see Figure 9) for each monitoring point were considered in the inverse procedure. The contaminant release history, in terms of mass rate, presented three peaks of different magnitude (Figure 10). As mentioned, this case considers an injection of a pollutant in a 2-D aquifer that does not have an analytical solution. For this reason the Kernel function was computed by means of a numerical flow and transport model (Citarella et al, 2015; Cupola et al, 2015) that was set up from the data collected at the experimental device.

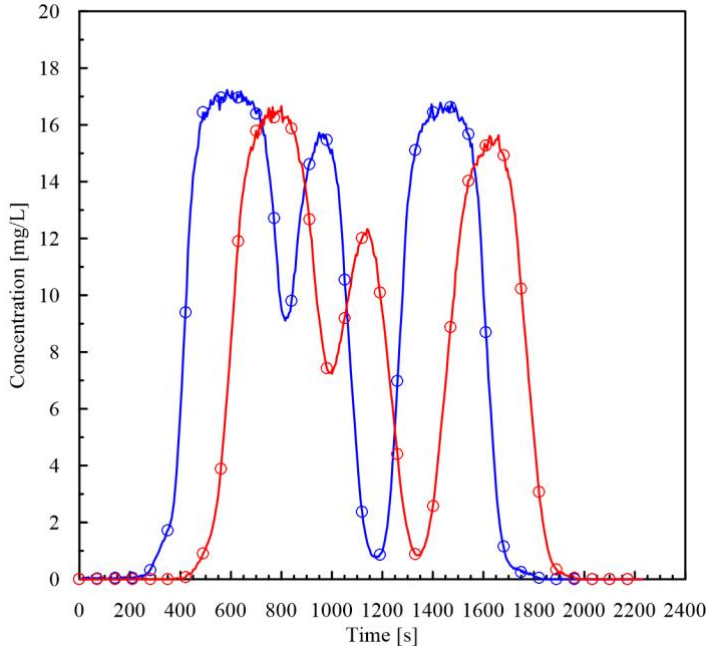

Figure 9. CASE 3 – The blue and red lines are respectively the observed concentration at the monitoring point P1 and P2 and the circles denote the measurement locations. Time 0 s represents the time at which the injection starts (t_{start}).

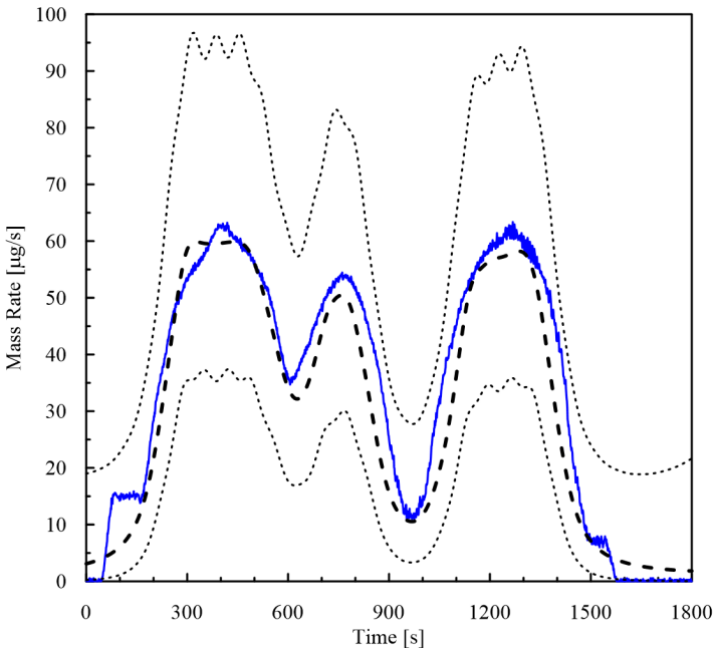

Figure 10. CASE 3 – Constrained ABIC with exponential covariance function: the true solution (solid blue line), best estimate (dashed thick line), and 5-95% confidence interval (dotted lines). Time 0 s represents the time at which the injection starts (t_{start}).

The first tests estimated a very good agreement between the computed and observed concentrations and consequently a very small σ_d^2; this caused the underestimation of the confidence intervals and

implied that the actual release history was not included in the confidence interval. An issue could be the choice of the covariance function and its parameters. Considering Figure 10, we have an actual release history that at early (100 s) and late (1500 s) time, it is constant for a short period. This implies long correlation, but in the middle of the two constant periods, the release history goes up in time and then back down. In order to study this problem, we used the actual release history with the Kernel matrix **G** to compute the observations (in order to obtain noise free data), then we corrupted the data with a measurement error and we estimated the release history. The inverse procedure was very efficient (the true release history was fully included in the 95% confidence interval) and able to reproduce the two constant release periods. This analysis highlights that there are other errors, than just strictly the measurement ones, which we should keep into account. In particular in this case we had had errors due to: image acquisition, calibration, numerical approximations and so on (see Cupola et al. 2015).

In a problem of this kind, it is worthwhile to estimate the measurement error with a minimum boundary, which is obtained through data analysis. According to Cupola et al. (2015) we choose as minimum value $\sigma_d = 3\,\mathrm{mg/L}$. With this assumption, the true release history (Figure 10) is included in the 95% confidence interval and it was well estimated through the procedure except the “steps”, after about 100 s and at about 1500 s, which were difficult to reproduce due to the large measurement error.

4. Comparison to other Bayesian approaches

Another efficient method is the Bayesian extension to the geostatistical approach (GA) proposed by Kitanidis (1995) and applied to recovery the release history by Snodgrass and Kitanidis (1997). Both the approaches (ABIC and GA) start from Bayes' theorem (Eq. (12)), are based on the assumption that the prior and the posterior probability functions are Gaussian, and can be summarized in two main steps: 1) estimation of the hyperparameters; 2) estimation of the unknown vector, **m**.

The first step, for both methodologies, is achieved maximizing the probability of the measurements called also "predictive distribution" $p\left(\mathbf{d}^{*}\middle|I\right)$, for ABIC by means of Eq. (18) and for GA through Eq. (26), but starting from different prior information and hypotheses on the estimation of the unknown vector, **m** (see for more details Mitsuhata (2004) and Kitanidis (1995)).

$$p\left(\mathbf{d}^{*}\middle|I\right) \propto \left\|\mathbf{\Sigma}\right\|^{-1/2}\left\|\mathbf{X}^{T}\mathbf{G}^{T}\mathbf{\Sigma}^{-1}\mathbf{G}\mathbf{X}\right\|^{-1/2}\exp\left[-\frac{1}{2}\mathbf{d}^{*T}\mathbf{\Xi}^{-1}\mathbf{d}^{*}\right] \tag{26}$$

Where $\mathbf{\Sigma} = \mathbf{G}\mathbf{C}_{m}\mathbf{G}^{T} + \mathbf{C}_{d}$ and $\mathbf{\Xi} = \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{G}\mathbf{X}\left(\mathbf{X}^{T}\mathbf{G}^{T}\mathbf{\Sigma}^{-1}\mathbf{G}\mathbf{X}\right)^{-1}\mathbf{X}^{T}\mathbf{G}^{T}\mathbf{\Sigma}^{-1}$. In GA the unknown **m** is viewed as being the sum of a deterministic part and a stochastic part. It is characterized by a prior mean E[**m**] = **s** = **Xβ** and a covariance matrix $\mathbf{C}_m$ = E[(**m**- **Xβ**)(**m**- **Xβ**)^T], where **X** is a known matrix and **β** are unknown drift coefficients. Here, embedded into the prior covariance is a vector **θ** which may contain the same hyperparameters as those in ABIC. The main assumption here is in the nature of the prior model **s**, which is a linear function of a set of drift parameters. In GA, essentially **s** is unknown at the prior stage but not in ABIC.

The maximization of Eq. (26) could be simplified (Kitanidis, 1995) to the minimization of the following objective function respects to the hyperparameters:

$$\mathrm{L} = \frac{1}{2}\ln\left\|\mathbf{\Sigma}\right\| + \frac{1}{2}\ln\left\|\mathbf{X}^{T}\mathbf{G}^{T}\mathbf{\Sigma}^{-1}\mathbf{G}\mathbf{X}\right\| + \frac{1}{2}\mathbf{d}^{*T}\mathbf{\Xi}^{-1}\mathbf{d}^{*} \tag{27}$$

This is a non-linear estimation problem and minimization is achieved numerically, through a Gauss-Newton iterative process, by taking the derivatives of L with respect to the hyperparameters and setting them to zero.

The second step is carried out for ABIC by estimating the first moment of the posterior pdf through Eq. (13). For a linear inverse problem this step is entirely linear. The overall-iterative procedure is completed when the improvement of the normalized objective function is lower than a tolerance value. GA carries out the second step of maximizing the posterior probability

$p(\mathbf{m}|\mathbf{d}^*) \propto \|\mathbf{C}_d\|^{-1/2} \|\mathbf{Q}\|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{d}^* - \mathbf{Gm})^T \mathbf{C}_d^{-1}(\mathbf{d}^* - \mathbf{Gm})\right] \cdot \exp\left[-\frac{1}{2}(\mathbf{m} - \mathbf{X\beta})^T \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{X\beta})\right]$, which could be simplified by minimizing the following objective function (see Kitanidis, 1995):

$$\Psi = (\mathbf{d}^* - \mathbf{Gm})^T \mathbf{C}_d^{-1} (\mathbf{d}^* - \mathbf{Gm}) + \mathbf{m}^T \mathbf{Hm} \tag{28}$$

where $\mathbf{H} = \mathbf{C}_m^{-1} - \mathbf{C}_m^{-1}\mathbf{s}(\mathbf{s}^T\mathbf{C}_m^{-1}\mathbf{s})^{-1}\mathbf{s}^T\mathbf{C}_m^{-1}$. The unknown vector **m** is finally estimated through a co-Kriging system when, in the iterative algorithm, the improvement in the objective function is lower than a tolerance value, see Eqs. (14-19) of Kitanidis (1995).

Analyzing the ABIC objective function we find that the first and the last term of Eq. (19) are constant and the second may be small compared to the others. The core of the function to be minimized is $(\mathbf{d}^* - \mathbf{Gs})^T \mathbf{C}_{dp}^{-1}(\mathbf{d}^* - \mathbf{Gs})$ which depends on the covariance function $\mathbf{C}_m$, observations $\mathbf{d}^*$, observational error σ_d^2, forward model **G** and the mean of the prior pdf **s**. The objective function of the GA depends on two parts: the first $(\mathbf{d}^* - \mathbf{Gm})^T \mathbf{C}_d^{-1}(\mathbf{d}^* - \mathbf{Gm})$ represents the misfit and is based on observations $\mathbf{d}^*$, observations error σ_d^2, forward model **G** and unknown vector **m**; and the second term $\mathbf{m}^T\mathbf{Hm}$ that represents the misfit which in turn is based on the unknown vector **m**, the covariance function $\mathbf{C}_m$ and the mean of the unknown (in this case) vector **s**.

5. Conclusions

In this work we have outlined an empirical Bayesian approach to the inverse problem for the release history of a groundwater contaminant. This work utilizes Akaike's Bayesian Information Criterion to estimate all hyperparameters within the Bayesian framework. The applied method, initially developed for heat flow inversion by Woodbury and Ferguson (2006), is for the first time applied to contaminant release history estimation, and it has subsequently been improved as noted below:

1. To solve for only positive model unknowns by transforming parameters space with a power transformation and this has the effect of turning a purely linear inverse to a non-linear one;

2. A relaxation factor is introduced in order to keep into account strongly non-linear release functions and to ensure convergence. With this improvement, the ABIC objective function monotonically decreases from one iteration to the next;

3. Confidence bands for the inverse model are produced even though results show skewed distribution for the posterior pdf;

4. The simultaneous estimation of the covariance function parameters (the variance and the integral scale) and of the measurement error is carried out;

5. To keep into account flow fields which are not described through analytical solutions.

We thoroughly tested and verified the algorithm and produced very positive results in all three test cases: a classic benchmark, a very difficult three spike example and finally, actual data from a controlled experiment. Results show a complete suppression of side lobes, which is an improvement over previous results. The method is based on the hypothesis that both the prior and posterior pdfs are Gaussian and requires only the assumption of the mean of the unknown function, while the parameters of the covariance function and the error on measurements are estimated during the iterative process. As shown by the test cases the method is robust and efficient, and managed to deal with very complex cases. One important advantage of the proposed procedure is the small number of the hyperparameters that have to be estimated: the covariance function parameters and, if needed the measurement error. Moreover the mean of the unknowns is independent from the prior mean. At the same time the procedure is very robust, converges to a minimum (Mitsuhata, 2004) and it is relatively simple to implement. Finally a general approach to estimates the initial values of the hyperparameters is provided.

As was demonstrated in CASE 3, the ABIC method is applicable to field cases by using a flow and transport numerical model of the test site. In fact, the kernel matrix **G** can be computed from a numerical procedure starting from laboratory/field dataset, and of course, could include more general conditions such as heterogeneity. Moreover, following the work of Butera et al. (2013), the procedure can be used also to simultaneously identify the source location and its release history.

Finally we show the linkage between ABIC and the geostatistical approach, first pioneered by Kitanidis (1995), and how the methods differ in philosophy.

Acknowledgments

The authors would like to dedicate this work to Tadeusz J. Ulrych who passed away in August, 2014 and left a rich legacy of mentorship, teaching, and research. We warmly thank the Editor and two anonymous Reviewers for their helpful comments. All the data are accessible by contacting the corresponding author.

REFERENCES

Akaike, H., 1974. A new look at the statistical model identification, IEEE Transactions on Automatic Control, Vol AC-19 (6).

Akaike, H., 1980. Likelihood and Bayes procedure. In: Bayesian Statistics (eds J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith), pp. 141–166. University Press, Valencia.

Atmadja, A., Bagtzoglou, A.C., 2001. State of the Art Report on Mathematical Methods for Groundwater Pollution Source Identification, Environmental Forensics, 2, 205-214.

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

Beck, A.E., Shen, P.Y., Betrami, H., Mareschal, J.-C., Safanda, J., Sebagenzi, M.N., Vasseur, G., Wang., K., 1992. A comparison of five different analyses in the interpretation of five borehole temperature data sets, Global and Planetary Change, 98, 101-112.

Box, G.E.P., Cox, D.R., 1964. An Analysis of Transformations, Journal of the Royal Statistical Society. Series B (Methodological), 26(2), 211-252.

Butera, I., and M. G. Tanda (2001), L'approccio geostatistico per la ricostruzoine della storia di rilascio di inquinanti in falda, L'Acqua, 4, 27-36 (in Italian).

Butera, I., Tanda M.G., 2003. A geostatistical approach to recover the release history of groundwater pollutants, Water Resour Research, 39(12), 1372.

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

Butera, I., Tanda, M. G., 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.

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

Citarella, D., Cupola, F., Tanda, M.G., Zanini A., 2015. Evaluation of dispersivity coefficients by means of a laboratory image analysis, Journal of Contaminant Hydrology, Volume 172, January 2015, Pages 10-23, ISSN 0169-7722, http://dx.doi.org/10.1016/j.jconhyd.2014.11.001.

Cupola, F., Tanda, M.G., Zanini, A., 2015. Laboratory sandbox validation of pollutant source location methods. Stochastic Environmental Research and Risk Assessment, Volume 29, issue 1, 169-182, doi: 10.1007/s00477-014-0869-4.

Hendry, M.J., Woodbury, A.D., 2007. Clay Aquitards as Archives of Holocene Paleoclimate: δ18O and Thermal Profiling, Ground Water, 45(6), 683-691.

Kennedy, P.L., Woodbury, A.D., Wang, K., 3000. Minimum relative entropy: theory and application to surface temperature reconstruction from borehole temperature measurements, Geop. Res. Letters, 27(19), 3081-3085.

Kitanidis, P.K., 1995. Quasi-Linear Geostatistical Theory for Inversing, Water Resour. Res., 31(10), 2411–2419, doi:10.1029/95WR01945.

Kitanidis, P.K., Vomvoris, E.G., 1983. A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, Water Resour. Res., 19(3), 677–690, doi:10.1029/WR019i003p00677.

Matsuoka, T., Ulrych, T.J., 1986. Information-theory measures with application to model identification, IEEE Trans Acoust Speech Sign, 34(3), 511-517.

Michalak, A.M., Kitanidis, P.K., 2004. Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling, Water Resources Research, 40(8), W08302.

Mitsuhata, Y., 2004. Adjustment of regularization in ill-posed linear inverse problems by the empirical Bayes approach, Geophysical Prospecting, 52(3), 213-239.

Painter, S., Woodbury A.D., Jiang, Y., 2007. Transmissivity Estimation for Highly Heterogeneous Aquifers: Comparison of Three Methods Applied to the Edwards Aquifer, Hydrogeology Journal, 15(2), 315-331.

Parker, R.L., 1977. Understanding inverse theory, Ann. Rev. Earth Planet. Sci., 5, 35-64.

Shibata, R., 2002. Information Criteria for Statistical Model Selection, Electronics and Communications in Japan (Part III: Fundamental Electronic Science), 85(4), 32-38.

Skaggs, T.H., Kabala, Z.J., 1994. Recovering the release history of a groundwater contaminant, Water Resources Research, 30(1), 71-79.

Smart, P.L., Laidlaw, I.M.S., 1977. An evaluation of some fluorescent dyes for water tracing. Water Resources Research,13(1), 15-33.

Snodgrass, M.F., Kitanidis, P.K., 1997. A geostatistical approach to contaminant source identification, Water Resources Research, 33(4), 537-546.

Sun, A.Y., 2007. A robust geostatistical approach to contaminant source identification, Water Resources Research, 43(2), W02418.

Sun, A.Y., Painter, S.L., Wittmeyer, G.W., 2006. A constrained robust least squares approach for contaminant release history identification, Water Resour Research, 42, W04414.

Ulrych, T.J., Sacchi, M.D., 2005. Information-based Inversion and Processing with Applications. ISBN: 0-08-044721-X, 404 pages, Elsevier.

Tarantola, A., 1987. Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation, 613 pages, Elsevier.

Ulrych, T.J., Sacchi, M.D., Woodbury, A.D., 2001. A Bayes tour of inversion: a tutorial, Geophysics, 66(1), 55-69.

Woodbury, A.D., 2007. Combining Thermal and Hydraulic Data for Hydrologic Estimation, Chapter 13, in "Data Integration in Subsurface Hydrology" (D. Hyndman editor), American Geophysical Union Monograph Series, 253 p., ISBN:978-0-87590-437-5.

Woodbury, A.D., 2011. Minimum relative entropy, Bayes and Kapur, Geophys. J. Int. (2011) 185 (1): 181-189 doi:10.1111/j.1365-246X.2011.04932.x

Woodbury, A.D., Ferguson, G., 2006. Ground surface paleotemperature reconstruction using information measures and empirical Bayes, Geophysical Research Letters, 33(6), L06702.

Woodbury, A.D., Ulrych, T.J., 1996. Minimum relative entropy inversion: Theory and application to recovering the release history of a groundwater contaminant, Water Resources Research, 32(9), 2671-2681.

Woodbury, A.D., Ulrych, T.J., 1998. Minimum relative entropy and probabilistic inversion in groundwater hydrology, Stochastic Hydrology and Hydraulics, 12(5), 317-358.

Woodbury, A.D., Ulrych, T.J., 2000. A full-Bayesian approach to the groundwater inverse problem for steady state flow, Water Resources Research, 36(8), 2081-2093.

Zanini, A., Kitanidis, P.K., 2009. Geostatistical Inversing for Large-Contrast Transmissivity Fields, Stochastic Environmental Research and Risk Assessment, 23, 565-577.