### RESEARCH ARTICLE

# Smart pilot points using reversible-jump Markov-chain Monte Carlo

**S. Jiménez[1], G. Mariethoz[2], R. Brauchler[3], and P. Bayer[1]**

[1]Department of Earth Sciences, Federal Institute of Technology, ETH Zurich, Zurich, Switzerland, [2]Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland, [3]AF-Consult Switzerland Ltd, Baden, Switzerland

**Abstract** Pilot points are typical means for calibration of highly parameterized numerical models. We propose a novel procedure based on estimating not only the pilot point values, but also their number and suitable locations. This is accomplished by a trans-dimensional Bayesian inversion procedure that also allows for uncertainty quantification. The utilized algorithm, reversible-jump Markov-Chain Monte Carlo (RJ-MCMC), is computationally demanding and this challenges the application for model calibration. We present a solution for fast, approximate simulation through the application of a Bayesian inversion. A fast path-finding algorithm is used to estimate tracer travel times instead of doing a full transport simulation. This approach extracts the information from measured breakthrough curves, which is crucial for the reconstruction of aquifer heterogeneity. As a result, the "smart pilot points" can be tuned during thousands of rapid model evaluations. This is demonstrated for both a synthetic and a field application. For the selected synthetic layered aquifer, two different hydrofacies are reconstructed. For the field investigation, multiple fluorescent tracers were injected in different well screens in a shallow alluvial aquifer and monitored in a tomographic source-receiver configuration. With the new inversion procedure, a sand layer was identified and reconstructed with a high spatial resolution in 3-D. The sand layer was successfully validated through additional slug tests at the site. The promising results encourage further applications in hydrogeological model calibration, especially for cases with simulation of transport.

## 1. Introduction

Numerical model calibration occupies a central role in hydrogeology and related disciplines where complex environmental systems need to be characterized, such as reservoir modeling or contaminant behavior prediction. A rich bibliography exists on techniques for concerted adjustment of numerous parameters or parameter fields, such as heterogeneous spatial distribution of hydraulic conductivity [*Carrera and Neuman*, 1986; *Sun*, 1994; *Dai and Samper*, 2004; *Ye et al.*, 2005; *Yang et al.*, 2014]. Manual calibration (trial and error) is probably the most simplistic way but it can prove highly inefficient. One of the most popular approaches is the zonation method, where the model domain is divided into a small number of uniform geological subdomains, and their properties are optimized. An efficient technique for zonal calibration is automatic deterministic calibration based on Newton's method [*Thiery*, 1994]; however, it does not allow for uncertainty quantification. In contrast, approaches using a maximum likelihood based formulation of the inverse problem [*Carrera and Neuman*, 1986] provide meaningful uncertainty quantifications. A comprehensive review of most relevant geostatistical inversion approaches can be found in *Zimmerman et al.* [1998], *Harp et al.* [2008], *Zhou et al.* [2014], or *Kitanidis* [2015].

Among the existing concepts, the pilot points method has been widely used for the calibration of groundwater models [e.g., *Marsily et al.*, 1984; *LaVenue and Pickens*, 1992; *Doherty*, 2003; *Tonkin et al.*, 2007]. Instead of formulating an estimation procedure whereby the value of each model cell value is considered as an independent unknown, only a limited number of values at predefined points are adjusted. The space in-between these so-called pilot points is filled by geostatistical interpolation techniques. This is often sufficient for obtaining a satisfactory model fit due to a small amount of available data in comparison to a high model parameterization.

One main criticism of the pilot points approach is that no general rules exist regarding the definition of the number and locations of the pilot points. Number and locations have to be fixed a priori, and the success of

inversion will strongly depends on this initial choice [*Doherty et al.*, 2010]. *Alcolea et al.* [2006] explain the related difficulties as follows: "On the one hand, a large number will result in substantial computational times and instabilities during the inversion. On the other hand, a small number of pilot points will hamper the resolution at which the heterogeneity is resolved."

*RamaRao et al.* [1995] suggest adding pilot points "where their potential for minimizing the objective function of model misfit is the highest." However, according to *Cooley* [2000], this strategy encourages overparametrization. For Gaussian distributed fields, *Gómez-Hernández et al.* [2003] propose, as a general rule, a spacing of 1 or 2 pilot points per correlation length, but *Kowalsky et al.* [2004] find a spacing of 2–3 pilot points per correlation length suitable. *Wen et al.* [2002] present a sequential self-calibration method that makes use of piezometric head and tracer data with a synthetic example where subsets of pilot points are selected whose locations change every 3–4 iterations. In general, a too large number of pilot points (i.e., overparameterization) may result in instabilities during the inversion. To compensate this, regularization (e.g., Thikonov) is commonly employed [*Fienen et al.*, 2009; *Doherty*, 2003, *Yang et al.*, 2012]. Subspace regularization is based on singular value decomposition (SVD) of the sensitivity matrix where pilot points are categorized as identifiable (i.e., used to minimize the objective function) or nonidentifiable (i.e., excluded from the inversion). By using only identifiable subsets, the problem dimension is decreased. Thus, inversion is accelerated and spurious effects from insensitive pilot points are avoided [*Tonkin and Doherty*, 2005]. *Yang et al.* [2012] present a subspace regularization technique using a stochastic SVD in combination with parameter rejection methods. It constructs an intensity field to optimally locate pilot points. However, in all of these refined techniques, an initial static set of pilot points still has to be defined, which strongly condition the inversion results.

In this paper, we propose a novel approach based on learning not only the pilot point values, but also their number and suitable locations during a stochastic inversion procedure, allowing for uncertainty quantification. Since the appropriate number of pilot points is not known a priori, an estimation algorithm is needed that can handle variable numbers of decision parameters. To do this, we use the reversible-jump Markov-chain Monte Carlo (RJ-MCMC) method [*Green*, 1995; *Hastie and Green*, 2012]. This so-called "smart pilot points" method follows a parsimonious principle and automatically searches for a minimum number of points in order to circumvent overparameterization. The inversion simultaneously formulates pilot point values and their coordinates as decision parameters, and thus minimizes the bias related to the initial settings.

RJ-MCMC is a more general formulation of the classic Metropolis-Hastings algorithm [*Metropolis et al.*, 1953; *Hastings*, 1970]. It belongs to the family of trans-dimensional algorithms, which search not only for optimal decision parameter values, but they also optimize their number. Trans-dimensional algorithms have been used in reservoir geology [e.g., *Gallagher et al.*, 2009; *Sambridge et al.*, 2006; *Elsheikh et al.*, 2012] and in geophysics [e.g., *Bodin and Sambridge*, 2009; *Ray and Key*, 2012]. *Bodin and Sambridge* [2009] for instance propose a seismic tomography inversion based on a variable number of Voronoi cells with flexible geometry. The recent work by *Dettmer et al.* [2014] deals with the discretization of faults using a trans-dimensional, self-adapting and irregular grid.

Our new inversion technique is tested using groundwater tracer tomography, similar to hydraulic tomography [*Vargas-Guzmán and Yeh*, 2002; *Cardiff et al.*, 2013; *Berg and Illman*, 2015], which is the combined inversion of several tracer signals. The multiple simultaneously or sequentially activated tracer sources and receivers are arranged in a setting similar to geophysical tomography. Sources denote injection points in different wells or depths, and receivers are sample locations, such as depth-dependent well screens. By means of multiple source-receiver combinations, a given aquifer cross-section or volume can be examined and transport-relevant structures in the subsurface can be located. Due to the relatively demanding field implementation and the challenge of combined tracer data inversion, only a few studies on tracer tomography have been presented so far [*Brauchler et al.*, 2013; *Illman et al.*, 2010]. Variants include applications with solute tracers [*Vasco and Datta-Gupta*, 1999; *Jiménez et al.*, 2015], thermal tracer tomography [*Klepikova et al.*, 2014], and combinations with geophysical methods such as electrical resistivity tomography (ERT) [*Singha and Gorelick*, 2006] and ground penetrating radar (GPR) [*Dorn et al.*, 2012].

In the following, the principle of model parameterization using smart pilot points and the theory of RJ-MCMC is introduced. The strategy for applying smart pilot points to tracer tomography is then presented with a fast transport model proxy for acceleration of the computation. The combined use of smart pilot

points and proxy is first demonstrated for a simple synthetic test case. Second, results from the inversion of tracer tomography field experiments in a shallow alluvial aquifer are presented and validated with multilevel slug tests.

## 2. Methodology

### 2.1. Model Parametrization

The principle of the pilot points method is to:

1. Propose a hydrological model **m** composed of $n_{pp}$ pilot points, each pilot point defined by a set of coordinates and a parameter value such as a hydrofacies indicator.
2. Transfer the pilot points information over the model domain using interpolation. In this manner, an entire spatial field consisting of $q$ unknowns ($q$ being the number of grid cells in a hydrological model **m**) can be parameterized using only $n_{pp} \times 3$, the factor 3 standing for the three adjustable properties of each point (in 2-dimensions): the coordinate $x$, the coordinate $y$, and the parameter value. Since $n_{pp} \times 3 \ll q$, the pilot points are a sparse representation of a hydrological model **m**.
3. Compute the offset between calculated and measured data and return to (1) until a stopping criterion is met.

The main challenge with such sparse representation is that one has to find the correct level of sparsity to be able to explain the data. This is accomplished by using RJ-MCMC, which is a Bayesian method where the parameters values and the number of parameters are estimated simultaneously without introducing unnecessary detail [*Sambridge et al.*, 2013].

### 2.2. RJ-MCMC

The goal of Bayesian inference is to derive the posterior probability based on prior beliefs and a likelihood function. Thus, in the Bayesian framework, the solution to an inverse problem is given as the posterior probability, defined as the conjunction of a prior density and a likelihood by the classical Bayes' theorem (for convenience, the terminology employed in this section has been adopted from *Hastie and Green* [2012]):

$$p(\theta|\mathbf{Y}) = \frac{p(\mathbf{Y}|\theta)p(\theta)}{p(\mathbf{Y})} \tag{1}$$

where $p(\theta|Y)$ is the posterior probability density, which defines the parameter vector values informed by the data **Y**; $p(Y|\theta)$ is the likelihood, representing the probability of observing data **Y** within the given parameter vector $\theta$. In addition, $p(\theta)$ is the prior probability indicating the probability of the parameter vector $\theta$ before data **Y** is observed, and $p(\mathbf{Y})$ is the probability of observing the data, which is commonly an intractable integral. In order to sample the posterior without calculating the denominator in equation (1), we employ a modified Markov-chain Monte Carlo (MCMC) approach based on a Metropolis-Hastings algorithm [*Mosegaard and Tarantola*, 1995]. The steps of such an algorithm are as follows:

1. Select an initial parameter vector $\theta$.
2. Iterate until convergence:
   1. Generate a new set of parameters $\theta^*$ by perturbing $\theta$ based on previous state and according to a given symmetric proposal density.
   2. Compute the likelihood p($\mathbf{Y}|\theta^*$) based on a forward model run.
   3. Accept or reject $\theta^*$ as the next state in the chain, with a probability of acceptance min(1, $a$), $a$ being the acceptance ratio (described below). If accepted, update $\theta$ as $\theta^*$.
3. The set of accepted states, after removal of an initial burn-in phase, is then an ensemble representative of the posterior.

Changes or perturbations in the parameter vector done in point 2a (in our case pilot values and positions) are termed transitions. In the classical Metropolis-Hastings MCMC algorithm, the acceptance criterion is defined by the ratio of likelihoods $a$ = p(**Y**| $\theta^*$)/p(**Y**| $\theta$). This implies that the number of parameters n is constant.

Additionally, the RJ-MCMC algorithm allows for a varying number of parameters. In doing so, the number of pilot points are adjusted automatically in accordance with available data. In RJ-MCMC, samples are generated from a union of spaces of different dimensions such as $X = \cup_{i \in I}(\{i\} \times X_k)$, where $X$ is the sampling

state space and $i$ is a dimensionality variable. To track both dimensions and parameters values, each state $x$ is composed of a parameter vector $\theta_i$ and a model indicator variable $i$, which determines the dimension of the parameter vector. Therefore, RJ-MCMC algorithm allows for two types of transitions:

1. Across-dimensions: transition between two states with different number of parameters $(i, \theta_i) \sim P(i, \theta_i | \mathbf{Y})$.
2. Within-dimensions: transition between two states with same number of parameters $(\theta_i) \sim P(\theta_i | i, \mathbf{Y})$.

When performing a given transition, independently of whether the number of parameters changes or not, both the forward transition and the corresponding reverse transition need to be computed to preserve symmetry of the chain. The transition from a given state $x = (i, \theta_i)$ to a new state $x^* = (i^*, \theta_{i^*}^*)$ consists of two steps:

1. First, draw $r$ random numbers $u$ from a known joint distribution $g$. $u$ is an auxiliary dimension matching variable, which is used to define the locations of new pilot points or select which pilot points should be removed, when the number of pilot points is changed.
2. Second, map to a new state $x^*$ using a deterministic function $h$ such that $(u^*, \theta_{i^*}^*) = h(u, \theta_i)$. It is important to note that for transitions involving states with different dimensions, a correction by a Jacobian determinant is needed in order to make both states comparable. It is out of the scope of this paper to discuss RJ-MCMC details in depth, a complete reference could be found in *Green* [1995, 2003] and *Hastie and Green* [2012].

In the current implementation, we make use of so-called birth and death transitions to increase or decrease the number of dimensions. In other words, we are able to change the number of pilot points at each iteration. This is accomplished by modifying the acceptance ratio $a(x | x^*)$ to accommodate dimension changes in the model, resulting in the Metropolis-Hastings-Green acceptance ratio [*Green*, 1995]:

$$a(x | x^*) = \frac{p(\mathbf{Y} | \theta_{i^*}^*)}{p(\mathbf{Y} | \theta_i)} \frac{J_t}{J_{t^*}} \frac{g_t(x | x^*)}{g_{t^*}(x^* | x)} \left| \frac{\partial(x^* | u^*)}{\partial(x | u)} \right| \tag{2}$$

Equation (2) synthesizes the RJ-MCMC algorithm, with each term expressing a probability. Each term is briefly introduced here and described in detail in the next sections. $p$ expresses the likelihood of a parameters vector, which requires running a forward model that results in values $\mathbf{Y}_{comp}$ that can be comparted to observations $\mathbf{Y}_{obs}$. It is defined as

$$p(\mathbf{Y} | x) = 10^{\left( \frac{-\sum (\mathbf{Y}_{comp} - \mathbf{Y}_{obs})^2}{2\sigma^2} \right)} \tag{3}$$

where $\sigma^2$ is the estimated variance, representing both model and field measurement errors. $J$ is the probability of a given transition type (birth, death, or parameters update). $g$ is the proposal distribution used to update parameters and is state-dependent, meaning that different probability density functions are employed according to the current transition $t$.

The last term in the acceptance ratio (equation (2)) is a Jacobian determinant $\left| \frac{\partial(x^* | u^*)}{\partial(x | u)} \right|$, and as described briefly earlier is used to scale changes derived from the multidimensional transitions. The Jacobian reduces to one for transitions that do not change dimensions, see *Hastie and Green* [2012] for more details. Likewise, for transitions between models, the Jacobian simplifies to:

$$\left| \frac{\partial(x^* | u^*)}{\partial(x | u)} \right| = \begin{pmatrix} \frac{\partial \theta_1^*}{\partial \theta_1} & \cdots & \frac{\partial u}{\partial \theta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \theta_n^*}{\partial u} & \cdots & \frac{\partial u}{\partial u} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} = 1 \tag{4}$$

where $\theta n$ is the nth parameter of the pilot point vector and $u$ is an auxiliary dimension-matching variable. The role of this variable is to act as a placeholder and to allow dimension matching when a pilot point is added or subtracted.

### 2.3. Transition Types and Transition Probability Ratio
We can distinguish between two different transition types, those within dimensions and those across dimensions. Transitions within dimensions do not change the number of pilot points and they represent

updates of hydrofacies parameter values such as hydraulic conductivity and pilot point coordinates. Transitions across dimensions change the number of pilot points. On one hand, the so-called birth transition adds a new pilot point at a random position inside the domain. On the other hand, the so called death transition removes a preexisting pilot point.

To propose a new set of model parameters, first a transition type must be chosen. All available transition types have an equal probability and sampled from a uniform distribution (i.e., the probability of choosing a given transition is equal to $1/r$, where r is the number of possible transitions types):

1. Parameter update $\quad \dfrac{J_t}{J_{t^*}} = \dfrac{r}{r}$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (5)

2. Birth or death transition $\left(\text{maximum} > n_{pp} > 1\right) \quad \dfrac{J_t}{J_{t^*}} = \dfrac{r}{r}$ $\qquad\qquad$ (6)

3. Death transition $\left(n_{pp} = \text{maximum}\right) \quad \dfrac{J_t}{J_{t^*}} = \dfrac{r-1}{r}$ $\qquad\qquad\qquad$ (7)

4. Birth transition $\left(n_{pp} = 1\right) \quad \dfrac{J_t}{J_{t^*}} = \dfrac{r-1}{r}$ $\qquad\qquad\qquad\qquad$ (8)

For transitions where the number of pilot points is kept constant, the transition probability ratio update cancels out (equation (5)). This occurs because the same transition is needed to attain the previous model state. For example, if a new pilot point position is proposed using a coordinates transition, the inverse transition is required (i.e., removing the pilot point) to return to the previous position. The same transition probability ratio (equation (6)) is used for other across-dimensions transitions, where the number of pilot points is within minimum and maximum bounds.

For across-dimensions transitions, the number of available transitions and derived probabilities may change when the minimum or maximum number of pilot points is reached (equations (7) and (8)). The minimum number of pilot points allowed is 1 and the maximum is equal to the number of cells or elements in the model domain. When the number of pilot points is equal to 1, the death transition cannot be chosen. Conversely when the number of pilot points is at a maximum, the birth transition is prohibited.

### 2.4. Proposal Distributions

For transitions within dimensions, new hydraulic parameter values and pilot point coordinates are proposed by sampling from a Gaussian probability density function (pdf), with a mean equal to the parameter values of the previous iteration (i.e., random walk Metropolis sampler). To avoid parameters without physical meaning (e.g., negative hydraulic conductivity), constraints are necessary when proposing new parameter values. Therefore, the proposal probability density function is truncated. This distortion is corrected in the proposal distribution ratio (equation (2)) and reversibility is maintained. For use of discrete variables, such as in sequential indicator simulation, the hydrofacies indicator can be updated by uniformly sampling an integer in the range [0 $n_f$], where $n_f$ is the total number of hydrofacies. For the sake of simplicity, the model domain is considered periodic, with lower/upper and left/right boundaries that are connected. This means that when an update in the coordinate results in a pilot point going over the model boundary, it wraps up at the opposite boundary. Thus, there is no need to truncate the Gaussian pdf when proposing new coordinates.

For facilitating transitions across dimensions, a transformation function is required to map two different sets of pilot points having different dimensions. This function makes use of the auxiliary dimension-matching variable $u$, drawn from a proposal distribution. In the current implementation, the role of this function is to assign a location for the new pilot point in case of a birth transition. Likewise, it will decide which of the existing pilot points is to be deleted in case of a death transition. Additionally, the function needs to be invertible, differentiable and bijective to guarantee reversibility [*Yanan and Scott*, 2011].

Although many transformation functions would be valid, the choice of a well-behaved function is desirable. To ensure a successful inversion, the transformation function is prone to propose high likelihood values (i.e., close to the target posterior probability distribution). A poor choice of the transformation function may
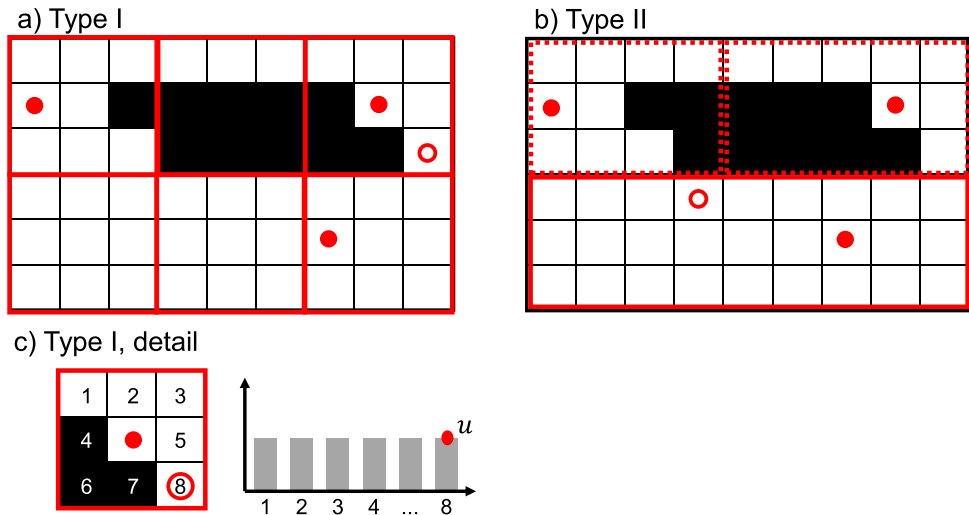
**Figure 1.** Examples of birth transition (a) Type I transition with three regions (red squares), three pilot points (red solid dots), and a new proposed pilot point (red empty dots). (b) Type II with a selected region and a proposed location for a new pilot point. (c) Detail of a selected region, cell identifiers, and the resulting probability mass function to choose the new pilot point cell.

result in long computational times and ultimate failure of the inversion. Two types of birth transition are proposed, types I and II (Figure 1.).

### 2.4.1. Birth Transition Type I

When birth type I transition is performed, a new pilot point is added into the domain in areas were high heterogeneity is suspected (Figure 1.a). The underlying reasoning is that high parametrization (i.e., more pilot points) is capable of characterize areas of high variability better. Birth type I is implemented as follows:

1. The model domain is divided into regular regions and the degree of heterogeneity is computed. For instance, for a model with two discrete hydrofacies (i.e., zones of constant hydraulic properties), heterogeneity may be computed as:

$$\text{degree of heterogeneity} = abs\left(1 - \frac{2*c}{N}\right) \qquad (9)$$

where $c$ is the number of cells with a given hydrofacies within the region, and $N$ is the number of cells (i.e., possible pilot point allocation positions) in the region.

2. The region with a higher degree of heterogeneity is selected.
3. A random number is drawn from a uniform proposal $u \sim U[1, n_{reg} - n_{pp\text{-}region}]$, to decide in which cell the new pilot point will be assigned; $n_{reg}$ is the number of cells in a region, while $n_{pp\text{-}region}$ is the number of pilot point in the chosen region.

An example for birth type I is provided on Figure 1.c. It depicts a region composed of a total of nine cells where a new pilot point is selected (solid red dot). In this case, the transition consists of drawing a candidate cell in the interval uniformly [1 8] (the central cell already has a preexisting pilot point). The new pilot point is symbolized by a red empty dot at cell number 8.

### 2.4.2. Birth Transition Type II

The goal of birth transition type II is to increase parametrization in areas with low density of pilot points (Figure 1b). The procedure is as following:

1. Perform a rectangular tessellation using each pilot point as center.
2. The region with the largest surface is selected.
3. A random number, $u \sim U[1, n_{reg} - n_{pp\text{-}region}]$, is drawn from a uniform proposal to select the position of the new pilot point (empty red dot).

Figure 1b depicts an example with three pilot points (solid red dots) and their corresponding regions (dashed and solid red lines). The region with a solid red line is selected because it is the one with the largest surface. As a result, a new pilot point is created at a random position within the region (red empty dot).
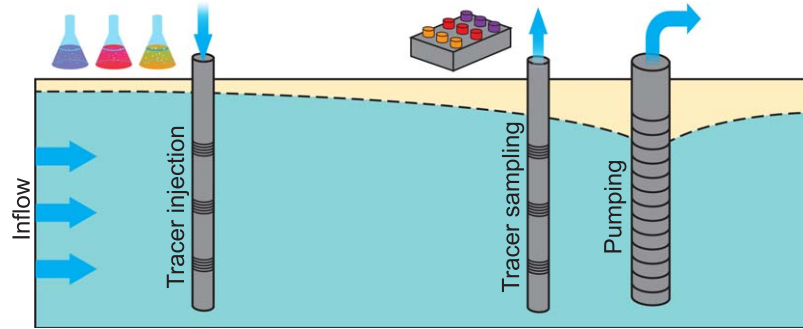
**Figure 2.** Model setup for synthetic and field tracer tests. Tracers are injected in the left borehole and recovered in the right borehole at three different depths. Pumping from a fully screened well creates a steady state flow regime.

Death transitions are implemented as the reverse of both birth types. In a death transition, a pilot point is selected using a uniform proposal u ∼ U[1, n$_{pp}$] and then deleted.

In summary, the proposal distribution ratio for a birth and death transition is presented in the following equations (10) and (11).

$$\frac{g(\mathbf{x}|\mathbf{x}^*)}{g(\mathbf{x}^*|\mathbf{x})} = \frac{n_{pp}+1}{(N-n_{pp})} \tag{10}$$

$$\frac{g(\mathbf{x}|\mathbf{x}^*)}{g(\mathbf{x}^*|\mathbf{x})} = \frac{(N-n_{pp}-1)}{n_{pp}} \tag{11}$$

## 3. Application to Tracer Tomography

### 3.1. Implementation

We choose tracer tomography to demonstrate the smart pilot points inversion, considering two settings: (1) a synthetic case, and (2) a field example. Both cases share the same basic setup in 2 dimensions: the hydraulic gradient is controlled via constant pumping from a single borehole and thus, forced-gradient conditions are imposed. Three tracers are injected simultaneously in a multichamber borehole (sources) and recovered at another downstream borehole at discrete levels (Figure 2). This way, at each level (receiver), tracer breakthrough curves (BTCs) are obtained. The challenge for the inversion procedure is the combined inversion of all measured tracer concentrations for reconstructing the heterogeneous hydraulic parameter distribution of the aquifer.

The implementation of tracer tomography is depicted in Figure 3. First, a given initial pilot point set is updated: coordinates, hydraulic parameters, and possibly the number of pilot points are modified (Figure 3a). Second, based on the proposed pilot point set, sequential indicator simulation is used to assign values to each cell or element in the domain (Figure 3b). Third, a steady state model is computed to derive the velocity field (Figure 3c). Fourth, based on the velocity field, travel times between sources and receivers are computed using a transport proxy (Figure 3d). Fifth, the pilot point set is accepted or rejected according to the Metropolis-Hastings-Green ratio (Figure 3e). A final ensemble is obtained after I iterations, (Figure 3f).

The application of smart pilot points, as most stochastic inversions, requires a large number of iterations, typically on the order of thousands. Therefore, full transient transport model testing during each iteration would yield a computationally intensive procedure. Instead, a more efficient travel time calculation is suggested. For standard pilot point inversion, model surrogates or proxies are often used. For instance, *Burrows and Doherty* [2014] used two models, one with high accuracy and a second approximated model with a modified grid and advective transport to deal with fast implicit finite differences instead of a time variation diminishing (TVD) scheme. *Ciriello et al.* [2015] used polynomial chaos expansion theory to construct transport model proxies. We adopt this idea and use a novel transport proxy based on steady state flow modeling (Appendix A). The proxy is only used to estimate the travel times between sources and receivers, without inspecting the BTCs in more detail.
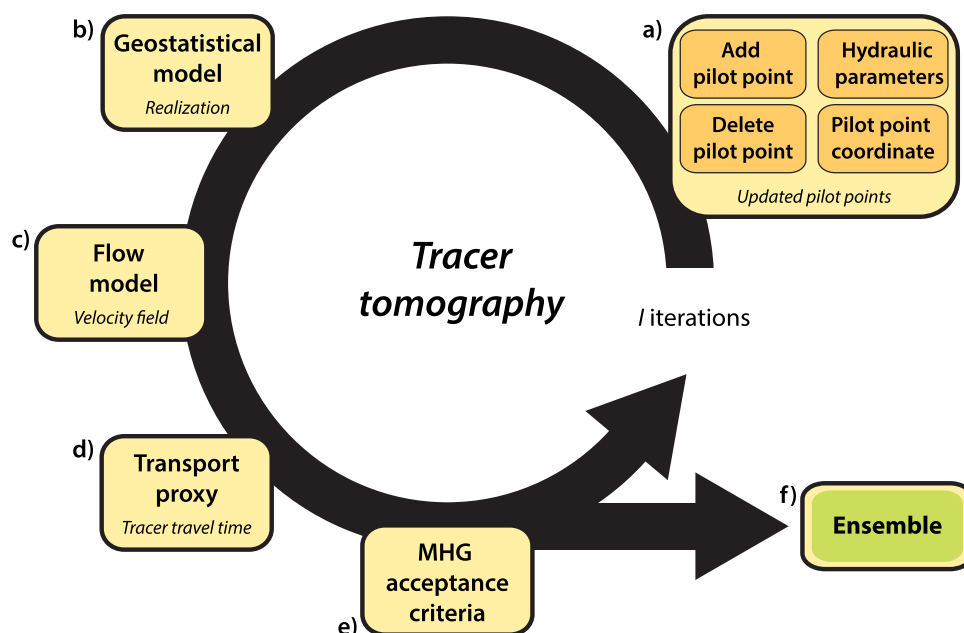
**Figure 3.** Smart pilot points conceptual flowchart for implementation with tracer tomography and transport proxy.

We propose a transport proxy that employs a pathfinding algorithm to estimate the minimum cost path between sources and receivers. The selected method is a variant of the established "A star" (A*) algorithm [*Hart et al.*, 1968], similar to Dijkstra's or Greedy Best-First-Search algorithm. It is implemented in MatlabBGL [*Gleich*, 2008], which is also employed in our study. A* uses a cost function to search between possible paths, composed by two terms: (1) the known distance from the source to the current cell, and (2) a heuristic estimate (e.g., the Manhattan distance). Figure 4 illustrates the individual steps, which are:

1. Computation of a steady state flow model for a suggested hydraulic parameter distribution in order to derive the velocity field of the domain (Figure 4a).
2. The advective velocity is derived from the steady state flow model. The dispersion component is considered proportional to advection (Figure 4b).
3. A weighted graph (i.e., a network) is constructed connecting all adjacent cells based on advection and dispersion values (Figure 4c).
4. Use of the A star search (i.e., A*) algorithm to rapidly find the fastest path between all source and receiver combinations (Figure 4d). Given that the travel time between cells is known, conversion of the fastest paths to travel times is trivial.

### 3.2. Synthetic Case
### 3.2.1. Overview
In a first step, we investigate whether a simple layered structure can be reconstructed. In the second step, we examine in more detail the automatic adjustment of pilot points and the algorithm's robustness by comparing repeated inversion runs.

The same tomographic configuration (Figure 2) is used in the synthetic case as in the subsequent field application. It consists of three sources, three receivers, and a fully screened pumping well that is permanently operated to impose a forced hydraulic gradient.
### 3.2.2. Setup
Figure 5a depicts the two-dimensional (2-D) cross section (5 m width and 2.5 m depth) between source and receiver wells. The reference field consists of a relatively low conductivity background sand matrix (facies 1, $K = 0.001$ m/s). Within the sand matrix, a layer of high conductivity gravel (facies 2, $K = 0.1$ m/s), as depicted in Figure 5a, is embedded. This is implemented in Modflow2005 [*Harbaugh*, 2005] as a uniform finite-differences model, with 25 rows and 50 columns. No-flow conditions are set at the top and bottom of the domain. Along the left side, a specified head boundary with an arbitrarily high value of 100 m is imposed.
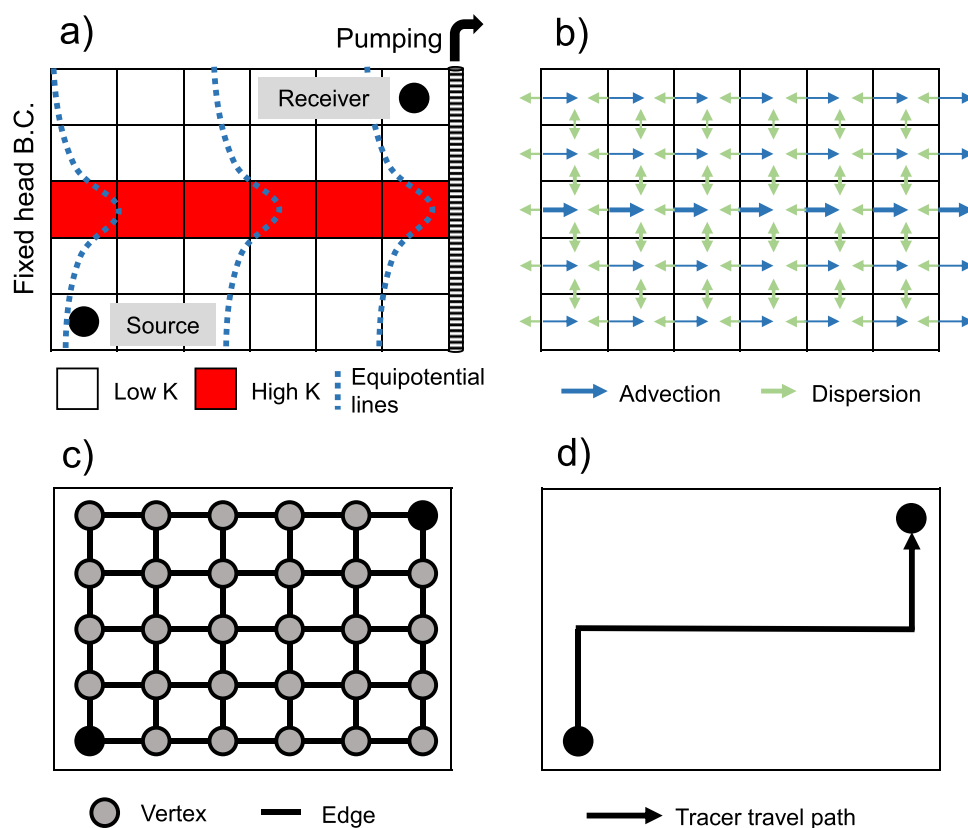
**Figure 4.** First travel time computation: (a) flow model setup and steady state solution; (b) advection and dispersion velocities derived from a steady state flow model; (c) weighted network constructed based on computed velocities between cells; and (d) resulting fastest path between source and receiver based on the pathfinding algorithm (A star).

The well at the right boundary is operated with a constant extraction rate of 2.5 L/s, which establishes a forced gradient towards the right side. With this setup, using a geometric multigrid solver [*Wilson and Naff*, 2004], confined and steady-state flow conditions are simulated. Tracer injection and observation points are located in wells, both at 0.55 m distance from the left and right model boundaries at a depth of 0.65, 1.25, and 1.85 m.

In this synthetic case, we know a priori that only two hydrofacies are present (facies 1 and 2), which have known hydraulic conductivities. In the profile between sources and receivers, 30 pilot points are initially distributed on a regular grid. Each pilot point is initialized with a facies label (i.e., 1 or 2), which may be switched when this transition is selected during the RJ-MCMC. An initial homogeneous aquifer is considered, therefore all pilot points are initialized as facies 1. Pilot point positions are updated using a Gaussian pdf with mean centered on the current coordinates and standard deviation of 0.4 m (i.e., random walk Metropolis sampler). Choosing an optimal standard deviation is a crucial factor. Methods such as adaptive Metropolis [*Haario et al.*, 2001] provide the means to update pdf parameters. However, due the simplicity of the position update, standard deviation was chosen manually after running several chains. Interpolation of the *K*-field between pilot points is performed by sequential indicator simulation using SGeMS software [*Remy*, 2004].

For both scenarios, a chain length of 30,000 iterations is used and the first 9000 iterations are discarded in order to minimize the effects of initial parametrization on the posterior inference (i.e., burn-in period). Furthermore, only every fourth iteration is saved in order to decrease autocorrelation (i.e., chain thinning).

The computed smart pilot points ensemble is compared to the one provided by rejection sampling to validate the layered structure reconstruction and determine whether smart pilot points have converged and explored the entire space of the parameters. Rejection sampling is mostly infeasible due to the enormous number of model runs required; however for a fast model and by parallelization, it is possible to provide an
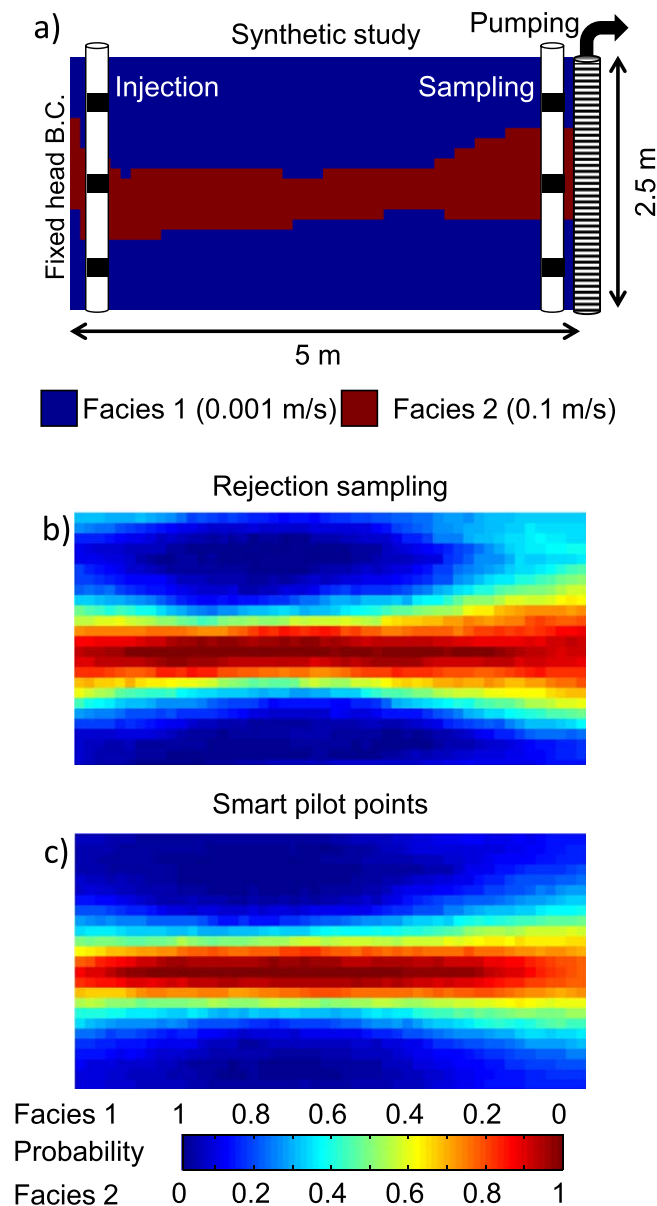
**Figure 5.** (a) Synthetic study with proposed hydrofacies distribution. (b) Probability of facies based on rejection sampling ensemble. (c) Based on smart pilot points.

ensemble. The main attraction of rejection sampling is that it provides an unbiased sampling of the posterior, which can be used as a reference. For rejection sampling in this instance, three million unconditional realizations are generated in parallel using a workstation (Xeon processor, 64 Gb Ram). The mean computation time per iteration is 0.7 s, including geostatistical interpolation, steady state flow model and transport proxy computation. Realizations are accepted [*Mariethoz and Caers*, 2014] if $p \leq \frac{L(r)}{S_L}$, and in this case, the realization is added to the ensemble. Where $p$ is a uniform random number between 0 and 1, $L(r)$ is the likelihood of a given realization $r$, and $S_L$ is the supremum of the likelihood and is set to 1.

### 3.2.3. Results

Figure 5 presents the results of the ensemble obtained from rejection sampling (Figure 5b) and using smart pilot points (Figure 5c). Both ensembles are composed of a subset of 1000 realizations in order to be comparable.

Both ensembles capture the approximate distribution of hydrofacies and are in agreement with the presence of a continuous layer having a probability close to 100% for facies 2. Nevertheless, rejection sampling

presents more variance than smart pilot points on both left and right sections, indicating that smart pilot points tends to slightly underestimate the uncertainty.

The acceptance rate of smart pilot points is 8% in this setup and 0.003% for rejection sampling.

The ability of the method to adjust the problem dimensionality is shown in Figure 6. It provides insight into the trans-dimensional capacity of smart pilot points. Convergence is achieved at approximately 2000 iterations, as shown in the residual trace plot (Figure 6a). The number of pilot points converges at around 10,000 iterations with a mean of 13 pilot points.

### 3.3. Field Implementation
### 3.3.1. Overview
Field tests were conducted at the Widen site in the vicinity of the river Thur in north-east Switzerland. Here, a shallow aquifer has been investigated in previous studies [*Vogt et al.*, 2010; *Doetsch*, 2011; *Schneider et al.*, 2011; *Linde et al.*, 2011; *Schirmer et al.*, 2014], made of highly conductive alluvial sediments with a variable thickness of around 5–10 m. These sediments are overlain by 2 m of alluvial sandy loam [*Vogt et al.*, 2010]. At the bottom, a lacustrine clay of considerable thickness (>30 m) forms an aquitard. At the Widen site, the aquifer has a thickness of approximately 8 m and can be described as a semiconfined gravel aquifer with intercalated sand lenses. The aim here is to characterize the sandy layers that are present in-between the gravel bodies. For this, a tomographic setup is chosen for tracer injection, and the recorded BTCs are inverted by means of smart pilot points. Additionally, slug tests were conducted for validation of the inverted results.

### 3.3.2. Setup
The Widen field site is equipped with 14 wells of 0.0508 m (2 inches) radius, 4 wells of 0.1143 m (4.5 inches), and 4 multichamber wells with seven independent channels and a depth of 10 m apart from each other (Figure 7). We selected the vertical profile between the multichamber boreholes MC2 and MC4 for analysis. For establishing a forced hydraulic gradient with a stable flow field, a surface pump (Grundfos Pomona PO23) is operated at borehole C3 with a pumping rate of 4.5 L/s. For the depth-dependent testing and monitoring, we utilized a multichamber system [*Einarson and Cherry*, 2002]. This system consists of a pipe with seven continuous separate chambers (ID = 0.014 m), which are arranged in a honeycomb shape. In each individual chamber a 0.08 m long opening, covered with a sand filter was cut. At three levels in MC2, different tracers were injected simultaneously, sulforhodamine B in the upper (channel 1 from the top), uranine in the middle (channel 3), and rhodamine in the lower chamber (channel 6). The tracer concentrations were sampled at the same levels (i.e., channels) downstream in MC4. A total of 600 samples (i.e., 200 per channel) were collected and analyzed using a LS-50B Luminescence Spectrometer.

At the Widen site, 240 slug tests were performed at wells P07, P11, and P13. A Campbell Scientific CR3000 Micrologger recorded the signals at 50 Hz with a packer interval of 20 cm. *Springer and Gelhar's* [1991] analytical solution was used to analyze the recorded signals, because this solution takes into account oscillatory water levels, which are commonly observed in highly conductive aquifers.

The flow is simulated in 3-D by means of a uniform finite differences model with 50 rows, 50 columns, and 60 layers. The natural hydraulic gradient is considered negligible and is not simulated. Specified head boundary conditions are imposed at the perimeter of the domain and confined and steady state conditions are simulated using Modflow2005 [*Harbaugh*, 2005].

An initial regular grid of 25 pilot points was placed on a 2-D slide between MC2 and MC4 (Figure 7) with a spacing of 0.8 m and with 1 m separation from the two adjacent domain boundaries. Pilot points are positioned exclusively in this profile since no additional observations are available from further locations. Lateral extrapolation generates a 3-D realization and flow simulations are also performed in 3-D. For calibration with smart pilot points, following the procedure as depicted in Figure 3, 90,000 flow model runs were used. Here, the first 13,000 iterations were discarded (burn-in), and for thinning, only one of four models was saved.

The model was initialized with homogeneous distribution of gravel, and all pilot points having the same initial values. From previous investigations, we know that sand hydrofacies exist locally and for each iteration, RJ-MCMC could choose either gravel or sand facies for each pilot point. The initial values for $K$ of gravel facies were set to $10^{-2}$ m/s, and $10^{-5}$ m/s for sand, as suggested for such hydrofacies in the bibliography
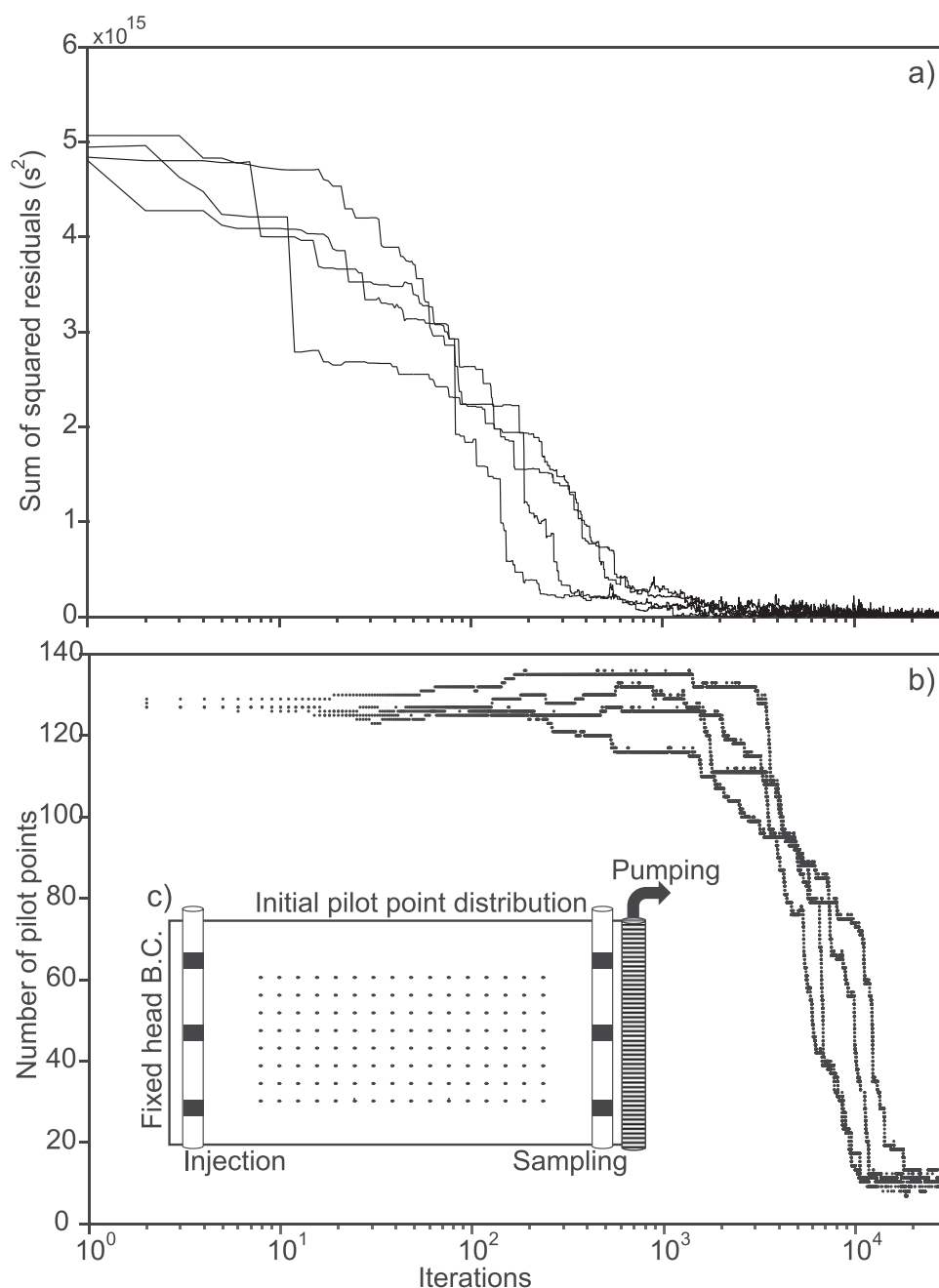
**Figure 6.** (a) Trace plot of the sum of squared residuals between observed and computed travel times. (b) Trace plot of the number of pilot points. (c) Pilot points initial distribution.

[*Fetter*, 2001]. New *K* values for each facies were proposed using a Gaussian pdf centered at the decadal logarithm of the current *K* value, with a standard deviation of 0.3. The sampler is truncated at 0 in order to generate only physically valid values.

### 3.3.3. Results

Figure 8 summarizes the key results from the field campaign and shows evidence of the presence of heterogeneity in the aquifer. First, we focus on the tracer experiments carried out between multichamber wells MC2 and MC4 (Figure 8a). Tracer injected in chamber 1 of MC2 has notably different breakthrough curves at observation chambers of MC4 (Figure 8b and 8c). On the one hand, chamber number 1 (Figure 8b) displays a well-defined breakthrough signal with a mean arrival at 8 min. On the other hand, chamber 2
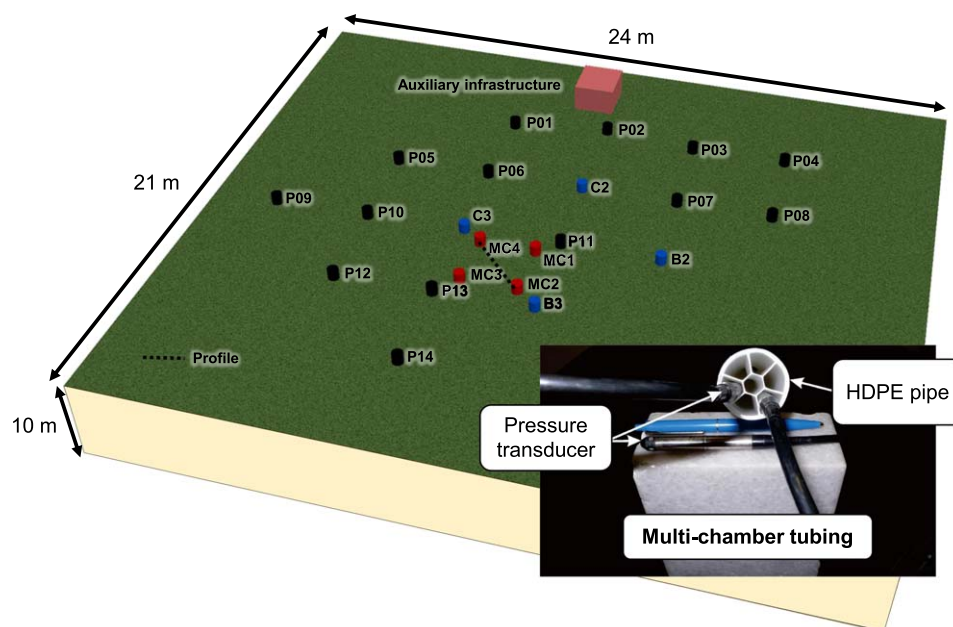
**Figure 7.** Widen site instrumentation layout: 2 inches boreholes are shown in black, 4.5 inches boreholes in blue and multichamber boreholes are depicted in red. Additionally, a detail photo of the multichamber tubing is provided.

(Figure 8c) shows no concentrations above the detection level. A similar behavior is observed in independent hydraulic tests (slug tests) performed at borehole PO7 (Figure 8d), located in the vicinity of boreholes MC2 and MC4. At approximately the same elevation as chamber 1 of borehole MC4, a slug test signal (Figure 8d) was obtained, indicating a relatively high conductivity ($3 \times 10^{-3}$ m/s) as derived from the oscillatory behavior. [*Springer and Gelhar*, 1991]. Likewise, at approximately the same elevation as chamber 2 of well MC4, another slug signal was recorded corresponding to a lower value of $K$ ($5 \times 10^{-4}$ m/s) (Figure 8e). These observations indicate that material of low $K$ is located at approximately 390 m.a.s.l. Another indication for this low $K$ layer is provided by a sieve analysis, which was carried out with the core samples retrieved when drilling the multi-chamber boreholes MC1–MC4. The sieve analysis followed SN670 810c specifications and used 20, 6.3, 2, 0.63, 0.5, 0.25, 0.125, and 0.063 mm mesh sizes. It reveals that test facies 1 is related to sandy material and facies 2 to gravel.
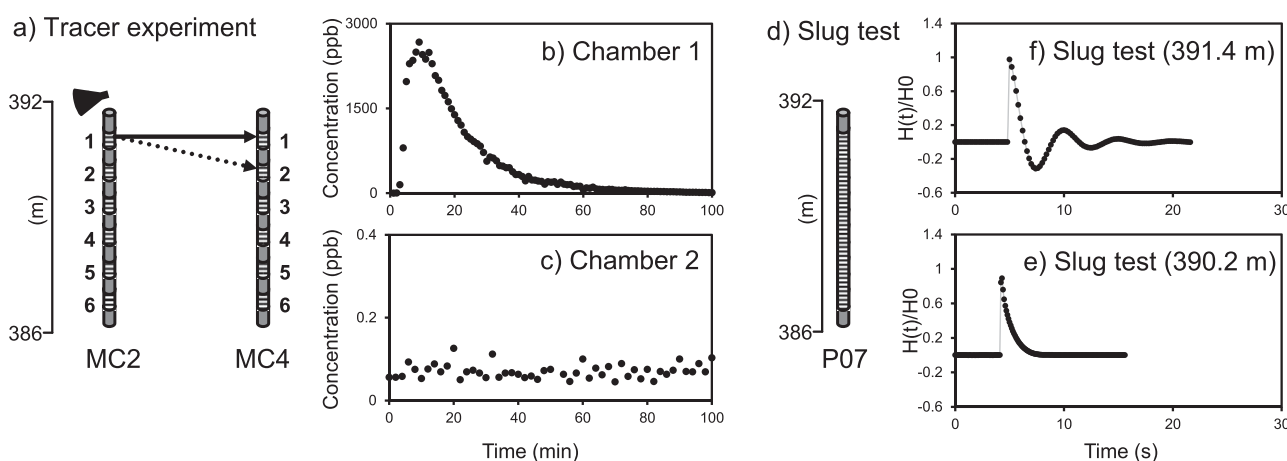


**Figure 8.** Tracer and slug tests field data. (a) Tracer experiment setup for the data presented in this figure; Sulforhodamine B was injected in chamber 1 of MC2 borehole, depicted as an Erlenmeyer flask, and registered in chamber 1 and 2. (b) BTC registered in chamber 1. (c) BTC registered in chamber 2. (d) Borehole P07. (e) Slug test corresponding to 390.2 m.a.s.l. (f) Slug test corresponding to 389.0 m.a.s.l.
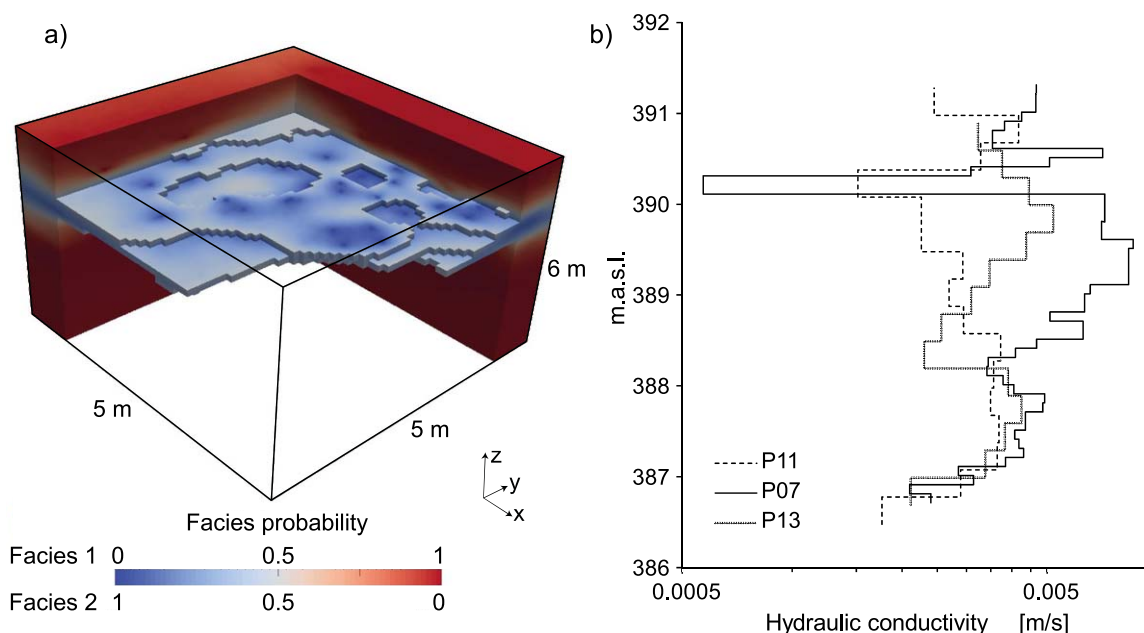
**Figure 9.** (a) Ensemble average resulting of the smart pilot points inversion. (b) Hydraulic conductivity profiles independently obtained from slug tests.

The resulting ensemble average of the smart pilot point method is plotted in Figure 9a, showing a continuous layer corresponding to facies 1. It shows a lower $K$ in comparison to the surrounding facies 2. This facies distribution is in line with the qualitative interpretation of the tracer and slug test signals shown in Figure 8. For more detail, the slug test derived profiles are depicted in Figure 9b. They indicate a low $K$ layer at the same elevation (ca. 390 m a.s.l.) for boreholes P07 and P11, but it is not detected at borehole P13. As a result, even if no slug tests were taken exactly in the profile examined by tomography, the consistent findings support a valid inversion by smart pilot points. In such alluvial aquifers, heterogeneous layers or troughs are common and they often have a variable width and are laterally continuous [*Bayer et al.*, 2011].

A detailed plot of the convergence of the smart pilot points based calibration is shown in Figure 10a. After around 5000 iterations, a significant decrease of the computed sum of squared residuals, which quantify the discrepancy between computed and observed travel times, is achieved. Based on visual inspection, the first 13,000 iterations were discarded (burn-in) and for thinning, only one of four models was saved. For the obtained ensemble of realizations from the 13,000th to the 90,000th iteration, a mean number of 23 pilot points was derived, ranging from 13 to 34 for the individual realizations (i.e., standard deviation of 3). The corresponding histogram is depicted in Figure 10b. Figure 10c shows the evolution of $K$ values for both hydrofacies. $K$ values converge to a mean value of $1.6 \times 10^{-3}$ and $3 \times 10^{-4}$ m/s for gravel and sand, respectively, with a variance of $2 \times 10^{-3}$ m/s. There is a discrepancy with the values obtained by slug test, possibly associated with the fact that slug tests measure a smaller portion of the aquifer [*Karasaki et al.*, 1988; *Beckie and Harvey*, 2002]. Other sources of discrepancy are wellbore skin effects [*Sageev*, 1986, *Yang and Gates*, 1997] and in highly conductive material the existence of positive or low-conductivity skin should be taken into consideration [*Cardiff et al.*, 2011]. Additionally, discrepancies have been observed between theoretical water level response and field data from a pressure transducer when using common analysis models for highly permeable formations [*Zurbuchen et al.*, 2002]. Nevertheless, both methods provide the same order of magnitude for both hydrofacies.

The mean number of only 23 pilot points reflects the low complexity of heterogeneous structures identified by tracer testing. This highlights the tendency of the algorithm to avoid overparameterization. This number is even very low when compared to related applications at other sites. For instance, *Bohling and Butler* [2010] used 177 pilot points to asses transmissivity and storage using hydraulic test
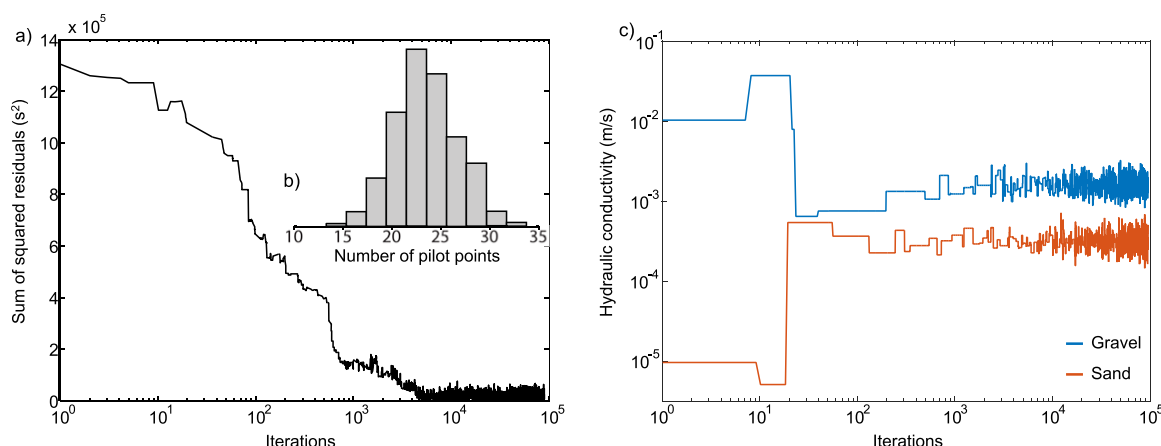
**Figure 10.** Field case results for smart pilot points application: (a) Trace plot of the sum of squared residuals between observed and computed travel times; (b) Histogram for the number of pilot points.

information derived from pumping tests at six different wells (five partially penetrating and one fully penetrating).

An important issue is that for obtaining a 3-D model of the Widen site, only the information between two boreholes (MC2 and MC4) is used. Thus, the few pilot points are only located in the vertical profile between both boreholes and the surrounding aquifer structures are extrapolated based on the imposed geostatistical model (sequential indicator simulation) [*Deutsch and Journel*, 1998; *Deutsch*, 2002]. Note that the choice of the geostatistical model will have an impact on the structures reproduced [*Linde et al.*, 2015]. This means that reliability of the 3-D model declines away from the profile. This is a limitation and can be overcome by the use of additional data points. For instance, *Jiménez et al.* [2013] uses several combinations of sources and receivers to invert pressure signals using pilot points in a 3-D model. Yet calibration with 3-D data and multiple profiles is computationally more demanding. While the use of the presented pathfinding algorithm to compute travel times is highly attractive for 2-D profiles such as in our case study, 3-D application requires higher computing times. These could increase even exponentially with respect to the number of model cells or elements. A possible way around this is to divide the model domain into sub-regions, and this way, limit the number of cells considered for each individual source-receiver combination.

Standard implementations of pilot points apply inter/extrapolation methods such as kriging, using pilot points as conditional data to constrain the geostatistical estimations or simulations. Smart pilot points can be combined with geostatistical simulation such as sequential Gaussian or indicator simulation and multiple-points statistics. This easily facilitates the inclusion of prior geological knowledge as soft information in order to reduce the uncertainty in the calibrated model (e.g., as training images for multipoint statistics). With additional information, a geostatistical model will yield simulated heterogeneous structures closer to the actual structures, and this will reduce the number of smart pilot points needed. However, further investigation is needed for exploring favorable interpolation techniques and geostatistical models in combination with smart pilot points.

## 4. Conclusions

Pilot points are attractive for numerical model calibration because there is rarely sufficient data to estimate unique values for each individual model cell. Only a given number of points is calibrated and the space in-between is interpolated. However, there is one fundamental shortcoming: it is not known where these points should be ideally placed and how many. Available recommendations for selecting ideal locations and amount of points only provide general advice and leave a broad range of freedom for the modeler. This can lead to a crucial bias for the calibration, since the pilot points condition the solution, which may in turn obscure the true uncertainty associated with the model prediction.

We propose to include the search for optimal pilot point numbers and positions in the calibration process. Through a trans-dimensional Bayesian search algorithm (RJ MCMC), an automatic implementation is accomplished. As shown with the two application cases, this yields a minimized number of optimally positioned points. This is appealing since optimized points yield parsimonious models with a resolution that complies with the available data and hence counteracts overparameterization. Favorable locations will naturally have a higher density in regions of observed data.

Applying such smart pilot points results in a much more challenging search procedure when dealing with thousands of candidate model evaluations. This renders the procedure infeasible for computationally demanding numerical models. A solution to this problem is demonstrated in this paper. Instead of full numerical model simulations, simplified variants or proxies may be used. Often, only a specific detail of a full forward simulation is processed during inversion. Such "low-fidelity" [*Ginsbourger et al.*, 2013] modeling of solute transport is used in order to employ a pathfinding algorithm to approximate possible hydraulic connections between monitored wells. This innovative approach proves to be extremely fast and suitable in comparison with numerical methods for supporting the combined inversion of multiple solute tracers. On the whole, a new methodology for tracer tomography is introduced and validated with independent data.

## Appendix A: Transport Proxy Comparison

A comparison in the proposed transport proxy based on the A* algorithm is provided. It consists of a comparison between the A* algorithm and particle tracking method implemented using Modpath 4. The model (Figure 11a) used consists of an injection at a discrete location of 1 $m^3$/d and three observation points downstream. The heterogeneity is the same as proposed for the synthetic model. The injector serves as a source, while the observation points act as receivers. To compute travel times based on particle tracking, one particle is located at each of the three receivers and backward tracking is employed. In Figure 11b, the particles trajectories between source and receivers are shown. In general, the results (Figure 11c) show an acceptable agreement between both methods. Nevertheless, the pathfinding algorithm produces slightly faster travel times than particle tracking. This is to be expected since the pathfinding algorithm provides the global fastest path between sources and receivers, while particle tracking employs a semi analytical scheme that tracks the particle from one cell to the next until the source is reached.
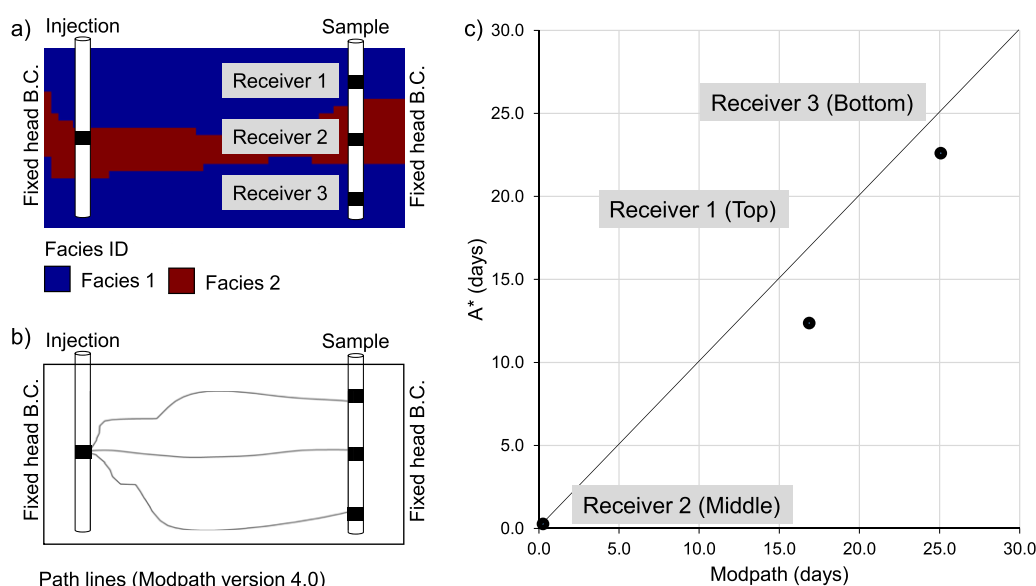


**Figure 11.** Validation of the proposed transport proxy. (a) Model containing an injection and sampling boreholes, (b) particles trajectories, and (c) comparison between the arrival times of the particle tracking (Modpath) and the pathfinding algorithm (A*).

## References

Alcolea, A., J. Carrera, and A. Medina (2006), Pilot points method incorporating prior information for solving the groundwater flow inverse problem, *Adv. Water Resour.*, *29*(11), 1678–1689.

Bayer, P., P. Huggenberger, P. Renard, and A. Comunian (2011), Three-dimensional high resolution fluvio-glacial aquifer analog: Part 1: Field study, *J. Hydrol.*, *405*(1–2), 1–9.

Beckie, R., and C. F. Harvey (2002), What does a slug test measure: An investigation of instrument response and the effects of heterogeneity, *Water Resour. Res.*, *38*(12), 1290, doi:10.1029/2001WR001072.

Berg, S. J., and W. A. Illman (2015), Comparison of hydraulic tomography with traditional methods at a highly heterogeneous site, *Groundwater*, *53*(1), 71–89.

Bodin, T., and M. Sambridge (2009), Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, *178*(3), 1411–1436.

Bohling, G. C., and J. J. Butler (2010), Inherent limitations of hydraulic tomography, *Ground Water*, *48*(6), 809–824.

Brauchler, R., G. Böhm, C. Leven, P. Dietrich, and M. Sauter (2013), A laboratory study of tracer tomography, *Hydrogeol. J.*, *21*(6), 1265–1274.

Burrows, W., and J. Doherty (2014), Efficient calibration/uncertainty analysis using paired complex/surrogate models, *Groundwater*, *53*(4), 531–541, doi:10.1111/gwat.12257

Cardiff, M., W. Barrash, M. Thoma, and B. Malama (2011), Information content of slug tests for estimating hydraulic properties in realistic, high-conductivity aquifer scenarios, *J. Hydrol.*, *403*(1–2), 66–82.

Cardiff, M., W. Barrash, and P. K. Kitanidis (2013), Hydraulic conductivity imaging from 3-D transient hydraulic tomography at several pumping/observation densities, *Water Resour. Res.*, *49*, 7311–7326, doi:10.1002/wrcr.20519.

Carrera, J., and S. P. Neuman (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.

Ciriello, V., Y. Edery, A. Guadagnini, and B. Berkowitz (2015), Multimodel framework for characterization of transport in porous media, *Water Resour. Res.*, *51*, 3384–3402, doi:10.1002/2015WR017047.

Cooley, R. L. (2000), An analysis of the pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, *Water Resour. Res.*, *36*(4), 1159–1163.

Dai, Z., and J. Samper (2004), Inverse problem of multicomponent reactive chemical transport in porous media: Formulation and applications, *Water Resour. Res.*, *40*, W07407, doi:10.1029/2004WR003248.

Dettmer, J., R. Benavente, P. R. Cummins, and M. Sambridge (2014), Trans-dimensional finite-fault inversion, *Geophys. J. Int.*, *199*(2), 735–751.

Deutsch, C. V. (2002), *Geostatistics Reservoir Modeling*, Oxford Univ. Press, Oxford, U. K.

Deutsch, C. V., and A. G. Journel (1998), *GSLIB: Geostatistical Software Library and User's Guide*, 2nd ed., Oxford Univ. Press, Oxford, U. K.

Doetsch, J. A. (2011), Joint and constrained inversion of geophysical data for improved imaging of aquifer structure and processes, ETH Zurich dissertation number: 19762, doi:10.3929/ethz-a-006689501.

Doherty, J. (2003), Ground water model calibration using pilot points and regularization, *Ground Water*, *41*(2), 170–177.

Doherty, J., M. Fienen, and R. Hunt (2010), Approaches to highly parameterized inversion: Pilot-point theory, guidelines, and research directions, *U.S. Geol. Surv. Sci. Invest. Rep.*, *2010–5168*, 36 pp.

Dorn, C., N. Linde, T. Le Borgne, O. Bour, and M. Klepikova (2012), Inferring transport characteristics in a fractured rock aquifer by combining single-hole ground-penetrating radar reflection monitoring and tracer test data, *Water Resour. Res.*, *48*, W11521, doi:10.1029/2011WR011739.

Einarson, M. D., and J. A. Cherry (2002), A new multilevel ground water monitoring system using multichannel tubing, *Ground Water Monit. Rem.*, *22*(4), 52–65.

Elsheikh, A. H., M. D. Jackson, and T. C. Laforce (2012), Bayesian reservoir history matching considering model and parameter uncertainties, *Math. Geosci.*, *44*(5), 515–543.

Fetter, C. W. (2001), *Applied Hydrogeology*, Prentice Hall, N. J.

Fienen, M. N., C. T. Muffels, and R. J. Hunt (2009), On constraining pilot point calibration with regularization in PEST, *Ground Water*, *47*(6), 835–844.

Gallagher, K., K. Charvin, S. Nielsen, M. Sambridge, and J. Stephenson (2009), Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems, *Mar. Pet. Geol.*, *26*(4), 525–535.

Ginsbourger, D., B. Rosspopoff, G. Pirot, N. Durrande, and P. Renard (2013), Distance-based kriging relying on proxy simulations for inverse conditioning, *Adv. Water Resour.*, *52*, 275–291.

Gleich, D. F. (2008), *MatlabBGL: A MATLAB Package for Working With Graphs*. [Available at http://dgleich.github.io/matlab-bgl/.]

Gómez-Hernández, J. J., H. J. W. M. Franssen Hendricks, and A. Sahuquillo (2003), Stochastic conditional inverse modeling of subsurface mass transport: A brief review and the self-calibrating method, *Stochastic Environ. Res. Risk Assess.*, *17*(5), 319–328.

Green, P. J. (1995), Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination, *Biometrika*, *82*(4), 711–732.

Green, P. J. (2003), *Trans-dimensional Markov chain Monte Carlo*, pp. 179–198, Oxford University Press.

Haario, H., E. Saksman, and J. Tamminen (2001), An Adaptive Metropolis algorithm, *Bernoulli*, *7*(2), 223–242.

Harbaugh, A. W. (2005), MODFLOW-2005, The U.S. Geological Survey modular ground-water model: The Ground-Water Flow Process, *U.S. Geol. Surv. Tech. Methods, 6-A16*, variously p.

Harp D. et al. (2008), Aquifer structure identification using stochastic inversion, *Geophys. Res. Lett.*, *35*, L08404, doi:10.1029/2008GL033585.

Hart, P. E., N. J. Nilsson, and B. Raphael (1968), A formal basis for the heuristic determination of minimum cost paths, *IEEE Trans. Syst. Sci. Cybern.*, *4*(2), 100–107.

Hastie, D. I., and P. J. Green (2012), Model choice using reversible jump Markov chain Monte Carlo, *Stat. Neerl.*, *66*(3), 309–338.

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*(1), 97–109.

Illman, W. A., S. J. Berg, X. Liu, and A. Massi (2010), Hydraulic/partitioning tracer tomography for DNAPL source zone characterization: Small-scale sandbox experiments, *Environ. Sci. Technol.*, *44*(22), 8609–8614.

Jiménez, S., R. Brauchler, and P. Bayer (2013), A new sequential procedure for hydraulic tomographic inversion, *Adv. Water Resour.*, *62*, 59–70.

Jiménez, S., R. Brauchler, R. Hu, L. Hu, S. Schmidt, T. Ptak, and P. Bayer (2015), Prediction of solute transport in a heterogeneous aquifer utilizing hydraulic conductivity and specific storage tomograms, *Water Resour. Res.*, *51*, 5504–5520, doi:10.1002/wrcr.21101.

Karasaki, K., J. C. S. Long, and P. A. Witherspoon (1988), Analytical models of slug tests, *Water Resour. Res.*, *24*(1), 115–126.

Kitanidis, P. K. (2015), Persistent questions of heterogeneity, uncertainty, and scale in subsurface flow and transport, *Water Resour. Res.*, *51*, 5888–5904, doi:10.1002/2015WR017639.

Klepikova, M. V., T. Le Borgne, O. Bour, K. Gallagher, R. Hochreutener, and N. Lavenant (2014), Passive temperature tomography experiments to characterize transmissivity and connectivity of preferential flow paths in fractured media, *J. Hydrol.*, *512*, 549–562.

Kowalsky, M. B., S. Finsterle, and Y. Rubin (2004), Estimating flow parameter distributions using ground-penetrating radar and hydrological measurements during transient flow in the vadose zone, *Adv. Water Resour.*, *27*(6), 583–599.

LaVenue, A. M., and J. F. Pickens (1992), Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model, *Water Resour. Res.*, *28*(6), 1543–1569.

Linde, N., J. Doetsch, D. Jougnot, O. Genoni, Y. Dürst, B. J. Minsley, T. Vogt, N. Pasquale, and J. Luster (2011), Self-potential investigations of a gravel bar in a restored river corridor, *Hydrol. Earth Syst. Sci.*, *15*(3), 729–742.

Linde, N., P. Renard, T. Mukerji, and J. Caers (2015), Geological realism in hydrogeological and geophysical inverse modeling: A review, *Adv. Water Resour.*, *86*, 86–101.

Mariethoz, G., and J. Caers (2014), Multiple-point Geostatistics: Stochastic Modeling with Training Images, 376 pp., Wiley-Blackwell, N. J.

Marsily, D., G. Lavedau, M. Boucher, and G. Fasanino (1984), Interpretation of interference test in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model, in *Geostatistics for Natural Resources Characterization, Proceedings of the NATO Advanced Study Institute*, edited by G. Verly, et al., Dordrecht, Holland, Springer.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*(6), 1087–1092.

Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, *100*(B7), 12,431–12,447.

RamaRao, B. S., A. M. LaVenue, G. De Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments, *Water Resour. Res.*, *31*(3), 475–493.

Ray, A., and K. Key (2012), Bayesian inversion of marine CSEM data with a trans-dimensional self-parametrizing algorithm, *Geophys. J. Int.*, *191*(3), 1135–1151.

Remy, N. (2004), The Stanford geostatistical earth modeling software (SGeMS): A tool for new algorithm development, in *Proceedings of the 7th Annual Geostatistics Congress*, pp. 865–871, Banff, Alberta, Canada, Springer.

Sageev, A. (1986), Slug test analysis, *Water Resour. Res.*, *22*(8), 1323–1333.

Sambridge, M., K. Gallagher, A. Jackson, and P. Rickwood (2006), Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.*, *167*(2), 528–542.

Sambridge, M., T. Bodin, K. Gallagher, and H. Tkalčić (2013), Transdimensional inference in the geosciences, *Phil. Trans. R. Soc. A.*, *371*(1984), doi:10.1098/rsta.2011.0547.

Schirmer, M., et al. (2014), Morphological, hydrological, biogeochemical and ecological changes and challenges in river restoration – the Thur River case study, *Hydrol. Earth Syst. Sci.*, *18*(6), 2449–2462.

Schneider, P., T. Vogt, M. Schirmer, J. Doetsch, N. Linde, N. Pasquale, P. Perona, and O. A. Cirpka (2011), Towards improved instrumentation for assessing river-groundwater interactions in a restored river corridor, *Hydrol. Earth Syst. Sci.*, *15*(8), 2531–2549.

Singha, K., and S. Gorelick (2006), Effects of spatially variable resolution on field-scale estimates of tracer concentration from electrical inversions using Archie's law, *Geophysics*, *71*(3), G83–G91.

Springer, R. K., and L. W. Gelhar (1991), Characterization of large-scale aquifer heterogeneity in glacial outwash by analysis of slug tests with oscillatory response, Cape Cod, Massachusetts, *U.S. Geol. Surv. Water Res. Invest. Rep.*, *91-4034*, pp. 36–40.

Sun, N. Z. (1994), *Inverse Problems in Groundwater Modeling*, 364 pp, Kluwer Acad., Netherlands.

Thiery, D. (1994), Calibration of groundwater models by optimization of parameters in homogeneous geological zones, in *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, edited by K. Hipel, pp. 69–82, Springer, Netherlands.

Tonkin, M. J., and J. Doherty (2005), A hybrid regularized inversion methodology for highly parameterized environmental models, *Water Resour. Res.*, *41*, W10412, doi:10.1029/2005WR003995.

Tonkin, M., J. Doherty, and C. Moore (2007), Efficient nonlinear predictive error variance for highly parameterized models, *Water Resour. Res.*, *43*, W07429, doi:10.1029/2006WR005348.

Vargas-Guzmán, J. A., and T. C. J. Yeh (2002), The successive linear estimator: A revisit, *Adv. Water Resour.*, *25*(7), 773–781.

Vasco, D. W., and A. Datta-Gupta (1999), Asymptotic solutions for solute transport: A formalism for tracer tomography, *Water Resour. Res.*, *35*(1), 1–16.

Vogt, T., E. Hoehn, P. Schneider, A. Freund, M. Schirmer, and O. A. Cirpka (2010), Fluctuations of electrical conductivity as a natural tracer for bank filtration in a losing stream, *Adv. Water Resour.*, *33*(11), 1296–1308.

Wen, X. H., C. V. Deutsch, and A. S. Cullick (2002), Construction of geostatistical aquifer models integrating dynamic flow and tracer data using inverse technique, *J. Hydrol.*, *255*(1–4), 151–168.

Wilson, J. D., and R. L. Naff (2004), MODFLOW2000: The U.S. Geological Survey Modular Ground Water Model: GMG Linear Equation Solver Package Documentation, *U.S. Geol. Surv. Open File Rep.*, *2004-1261*, 53 pp.

Yanan, F., and A. S. Scott (2011), Reversible Jump MCMC, in *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, and G. Jones, Chapman and Hall, U. K.

Yang, C., Z. Dai, K. Romanak, S. Hovorka and R. Trevino (2014) Inverse modeling of water-rock-CO2 batch experiments: Implications for potential impacts on groundwater resources at carbon sequestration sites, *Environ. Sci. Technol.*, *48*(5), 2798–2806.

Yang, Y., M. Over, and Y. Rubin (2012), Strategic placement of localization devices (such as pilot points and anchors) in inverse modeling schemes, *Water Resour. Res.*, *48*, W08519, doi:10.1029/2012WR011864.

Yang, Y. J., and T. M. Gates (1997), Wellbore skin effect in slug-test data analysis for low-permeability geologic materials, *Ground Water*, *35*(6), 931–937.

Ye, M., S. P. Neuman, P. D. Meyer, and K. F. Pohlmann (2005), Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff, *Water Resour. Res.*, *41*, W12429, doi:10.1029/2005WR004260.

Zhou, H., J. J. Gómez-Hernández, and L. Li (2014), Inverse methods in hydrogeology: Evolution and recent trends, *Adv. Water Resour.*, *63*, 22–37.

Zimmerman, D. A., et al. (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resour. Res.*, *34*(6), 1373–1413.

Zurbuchen, B. R., V. A. Zlotnik, and J. J. Butler (2002), Dynamic interpretation of slug tests in highly permeable aquifers, *Water Resour. Res.*, *38*(3), 1025, doi:10.1029/2001WR000354.