MEASUREMENT NETWORK DESIGN INCLUDING GROUNDWATER AGE OBSERVATIONS

GIJS M.C.M. JANSSEN[1], JOHAN R. VALSTAR[2] AND SJOERD E.A.T.M. VAN DER ZEE[3]

[1] *Department of Soil Quality, Wageningen University, Wageningen, The Netherlands*
[2] *TNO Built Environment and Geosciences – National Geological Survey, Utrecht, The Netherlands*
[3] *Department of Ecohydrology, Wageningen University, Wageningen, The Netherlands*

ABSTRACT

Groundwater age observations have found increasing application in the characterization of groundwater systems. No algorithms are available, however, to optimally design sampling strategies including this observation type. We propose a first-order methodology to incorporate groundwater age observations in measurement network design directed at contaminant breakthrough uncertainty minimization in short residence time systems. We use the representer approach to the inverse problem to calculate linearized (cross-)covariances (given by the representers) between potential measurements and the goal variables of which we want to reduce the uncertainty. These goal variables are the groundwater age at the control plane and the breakthrough locations of the contaminant. Because the travel time in short residence time systems is not Gaussian distributed and therefore does not comply with the Bayesian framework of the representer approach, we transform the travel time to its natural logarithm and derive representer formulations accordingly. In a synthetic numerical example, the representer approach formulated as such is shown to yield good first-order predictions of all relevant state variances if the variance of the natural logarithm of the transmissivity is less than 3.0. Knowing that the calculated representers are reliable, they can be used to predict, at first order, the posterior breakthrough variance belonging to a candidate network before measurements are actually taken. A Genetic Algorithm is used to efficiently search, among all candidate networks, for a near-optimal one. We show that, in our numerical example, an age observation network outperforms (in terms of breakthrough uncertainty reduction) equally sized head measurement networks and transmissivity measurement networks even if the age observations are highly uncertain.

1. INTRODUCTION

Observation network design (or data worth analysis), whether it is done manually (trial-and-error, based on expert knowledge) or using more or less sophisticated mathematical guidelines, is an inherent part of any soil and groundwater investigation, as it naturally evolves from the measurement campaign's purpose: obtaining the necessary information given the limited resources. Accordingly, this field of research receives constant attention in the literature. The design strategies reported in the literature generally seek the optimal placement and/or sampling times for observations of heads, concentrations, parameters, or a combination of these three. In the past two decades, groundwater age has increasingly found application as

another type of observation to constrain flow models with. When accurate, travel time observations in general, and therefore also age observations, can be more informative than head and conductivity observations, as the sensitivity of heads to parameters is usually limited and the spatial correlation range between travel time and conductivity is often larger than the correlation range of the conductivity itself [*Harvey and Gorelick*, 1995; *Sheets et al.*, 1998; *Stute and Schlosser*, 2000].

Manual calibration of flow models using tracer derived age observations has been performed by, for example, *Reilly et al.* (1994), *Sheets et al.* (1998) and *Izbicki et al.* (2004). Systematic, mathematical approaches to parameter inference from age data (or data on tracer arrival time, which is conceptually equivalent) are given by *Harvey and Gorelick* (1995), *Portniaguine and Solomon* (1998), *Woodbury and Rubin* (2000), *Cirpka and Kitanidis* (2001), and *Feyen et al.* (2003). However, age observations have never been incorporated in the design of optimal measurement strategies. The purpose of the present study is therefore to propose an algorithm that optimally configures measurement networks including age observations.

2. THEORY

2.1 Outline of the first-order design method.

In this paper, we want to reduce breakthrough time uncertainty. If the control plane over which the breakthrough is predicted receives a significant inflow that does not originate from the contaminant source zone, the contaminant breakthrough time probability distribution $p(\tau_{BT})$ is a function of the travel time probability of the entire inflow and the contaminant breakthrough location probability:

$$p(\tau_{BT}) = \int_{x^{CP}=A}^{x^{CP}=B} p\left(t; x^{CP}\right) p\left(x^{CP} \in \chi\right) dx^{CP} \tag{1}$$

Or in discretized form:

$$p(\tau_{BT}) \approx \sum_{x^{CP}=A}^{x^{CP}=B} \frac{1}{2}\left[p\left(t; x^{CP}\right) p\left(x^{CP} \in \chi\right) + p\left(t; x^{CP} + \Delta x^{CP}\right) p\left(x^{CP} + \Delta x^{CP} \in \chi\right)\right] \Delta x^{CP} \tag{2}$$

in which (assuming that t and χ are Gaussian distributed)

$$p\left(\tau; x^{CP}\right) \approx N\left(t; \sigma^2_{t\left(x^{CP}\right)}, \mu_{t\left(x^{CP}\right)}\right) \tag{3}$$

$$p\left(x^{CP} \in \chi\right) \approx \sum_{x^{SZ}=C}^{x^{SZ}=D} \frac{1}{2}\left[N\left(x^{CP}; \sigma^2_{\chi\left(x^{SZ}\right)}, \mu_{\chi\left(x^{SZ}\right)}\right) + N\left(x^{CP}; \sigma^2_{\chi\left(x^{SZ}+\Delta x^{SZ}\right)}, \mu_{\chi\left(x^{SZ}+\Delta x^{SZ}\right)}\right)\right] \Delta x^{SZ} \tag{4}$$

In Eqs. (1-4), t = travel time or groundwater age, $p(t;x^{CP})$ is the marginal travel time probability evaluated at x^{CP}, x^{CP} = a location at the control plane, $\chi(x^{SZ})$ is the breakthrough location of a particle that originated from x^{SZ}, x^{SZ} is a location in the contaminant source zone, $p\left(x^{CP} \in \chi\right)$ is the probability that x^{CP} belongs to the contaminant breakthrough zone χ (χ is the collection of all breakthrough locations χ), A and B are the spatial x limits of the control plane, C and D are the spatial x limits of the contaminant source zone, and $N(p_1;p_2,p_3)$ represents the probability of p_1 according to the normal distribution parameterized with variance p_2 and mean p_3. $\sigma^2_{t\left(x^{CP}\right)}$ = the travel time or age variance at location x^{CP}. $\mu_{t\left(x^{CP}\right)}$ = the mean arrival

time or age at x^{CP}. $\sigma^2_{\chi(x^{SZ})}$ and $\mu_{\chi(x^{SZ})}$ = the breakthrough location variance and mean (along the x axis) of a particle originating from x^{SZ}, respectively.

From Eqs. (1-4) it follows that to evaluate the performance of different measurement network designs, the influences of the observable variables at their potential sampling locations on the prediction of the travel time $t(x^{CP})$ and the breakthrough locations $\chi(x^{SZ})$ have to be known. This makes $t(x^{CP})$ and $\chi(x^{SZ})$ our goal variables, and they will be called as such throughout the remainder of this paper.

The recently proposed representer-based inverse method [*Valstar et al.*, 2004] provides an efficient way for calculating these influences. The representers calculated within this method are equivalent to the linearized (cross-)covariances between the observable variable at the potential sampling location and the variables for which the representers are defined. As such, a travel time representer and a breakthrough location representer provide a first-order estimation of the prior (cross-)covariances of the observable variables and the goal variables $t(x^{CP})$ and $\chi(x^{SZ})$, respectively. These (cross-)covariances can subsequently be used to approximate the posterior covariances of the goal variables before the measurements are actually taken.

For short travel distances, the travel time probability density function will show a significant skewness. Therefore, for short residence time systems, it is necessary to use the representer of an approximately Gaussian transform of the travel time to comply with the Bayesian framework. Here we use the natural logarithm of the travel time. Thus, we have to formulate the representer-based inverse algorithm such that it produces representers for ln t.

The derivation of the ln travel time representer (and the breakthrough location representer) is too elaborate to fully repeat here, also because it would require presenting the mathematical basics of the representer method. The idea, however, is that particle locations and particle travel times are treated as states, which have to obey their own respective balance equations: according to the travel time equation, the travel time of a particle after a certain travel step simply must be equal to the total time of all travel steps taken so far. According to the particle location equation, the breakthrough location of a particle must be equal to its starting position plus its positional changes during all travel steps taken to reach the control plane. By the Lagrange Method, the state equations (flow, travel time and particle breakthrough location equation) are multiplied by unknown Lagrange multipliers and added to a generalized least squares objective function that penalizes deviations of model predictions from measurement values and deviations of parameter estimates from prior information on these parameters. The objective function is minimized if the gradients of the objective function with respect to the parameters, the states and the adjoint variables are zero. Forcing this on the objective function yields a system of coupled Euler-Lagrange equations [*Valstar et al.*, 2004]. In order to decouple the Euler-Lagrange equations, the parameter, state and adjoint variables are expressed as an expansion in a set of basis functions called representers. By inserting the representer definitions in the Euler-Lagrange equations, explicit expressions for all representers (and their coefficients) can be obtained. The adjoint head, adjoint breakthrough location, adjoint ln travel time, parameter and state representers are calculated, respectively, in an iterative procedure in which the unknowns in the representer expressions are replaced by their estimates of the previous iteration. In the first iteration, they are replaced by their prior estimates. In our first-order design method, we stop after this first iteration, because subsequent iterations would involve updating the parameters and states. This requires actual

measurement values, which are not available prior to sampling. The fact that we do not perform updates also means that the representer coefficients do not have to be calculated.

To be able to expand the untransformed travel times in a series of influence functions of ln t observations , the adjoint t representer and the t representer have to be divided by the derivative $dt/d(\ln t) = t$, the value of which is approximated by its prior estimate. This yields the adjoint ln t representer and the ln t representer.

Without giving the derivations we state here that the calculation of the ln t representer involves summing weighted travel time sensitivities with respect to both the heads and the parameters for all travel time steps. These sensitivity terms can be computed by tracking of particles traveling from the locations at which travel time (cross-)covariances are required (either potential measurement locations or pseudo measurement locations at the control plane) backward to the inflow zone and calculating the time step sensitivities of every travel step using the ADV2 package of MODFLOW 2000 [*Anderman and Hill*, 2001]. The calculation of the breakthrough location representer involves summing weighted particle displacement sensitivities with respect to both the heads and the parameters for all travel steps. These sensitivity terms are computed by forward tracking of particles traveling from the pseudo measurement locations (the locations where the breakthrough location (cross-)covariances are required) in the contaminant source zone to the control plane and evaluating the displacement sensitivities of every travel step, again using the ADV2 package of MODFLOW 2000.

2.2 Computation of posterior breakthrough time probability.

Having obtained the prior variances of the goal variances and the (cross-)covariances between the measurements and the goal variables using this first-order method, a first-order approximation can be made of the posterior variances $P^{posterior}_{\ln t\left(x^{CP}\right)}$ and $P^{posterior}_{\chi\left(x^{SZ}\right)}$ of the goal variables $t\left(x^{CP}\right)$ and $\chi\left(x^{SZ}\right)$ (see also *Valstar et al.* (2004), Eqs. (33-35)). $P^{posterior}_{\ln t\left(x^{CP}\right)}$ and $P^{posterior}_{\chi\left(x^{SZ}\right)}$ can subsequently be filled in for $\sigma^2_{t\left(x^{CP}\right)}$ and $\sigma^2_{\chi\left(x^{SZ}\right)}$ in Eqs. (3) and (4), respectively (note that t in Eq. (1-3) is replaced by its natural logarithm). $\mu_{t\left(x^{CP}\right)}$ and $\mu_{\chi\left(x^{SZ}\right)}$ in Eqs. (3) and (4) can be replaced by $E(\ln t(x^{CP}))$ and $E(\chi(x^{SZ}))$, respectively, which can be obtained by a multi-realization approach or, in case the parameters in the prior realization of the stochastic field are set at their prior mean, by running a simulation using this prior realization. Now all necessary information is available to compute $P(\ln \tau_{BT})$ according to Eq. (2).

The algorithm presented above provides an efficient way for calculating, at first order, the expected posterior breakthrough variance of the contaminant for every candidate observation network design. A heuristic search method, such as simulated annealing or a Genetic Algorithm can be used to efficiently search, among all possible networks, for a near-optimal design that minimizes $P(\ln \tau_{BT})$. In this study a Genetic Algorithm was used.

3. NUMERICAL EXAMPLE

We will now apply the developed theory to the specific example of breakthrough prediction uncertainty reduction in confining layers. Figure 1 gives a schematic representation of the numerical model we used in all our calculations. It represents a two-dimensional cross-section of a confining layer (y_2-y_3) that protects Aquifer 2 (y_3-y_4) from contaminants released into Aquifer 1 (y_0-y_2). Both Aquifer 1 and Aquifer 2 are modeled as homogeneous deposits

with a hydraulic conductivity (K) of 3.0 and 0.6 m/d, respectively. In the shaded centre part (x_5-x_{10},y_2-y_3) of the confining layer (see Figure 1) K is assumed to be unknown and randomly distributed, with a natural logarithm of the geometric mean of K (ln K_G) of -3.0 ln(m/d), and horizontal and vertical correlation ranges of ln K (= Y) of 75 m and 25 m, respectively. The variance of ln K, σ_Y^2, is 2.0 unless stated otherwise. Outside the shaded centre part (x_0-x_5 and x_{10}-x_{13}) the confining layer is modeled as a homogeneous deposit with K equal to the K_G applied for the centre part.

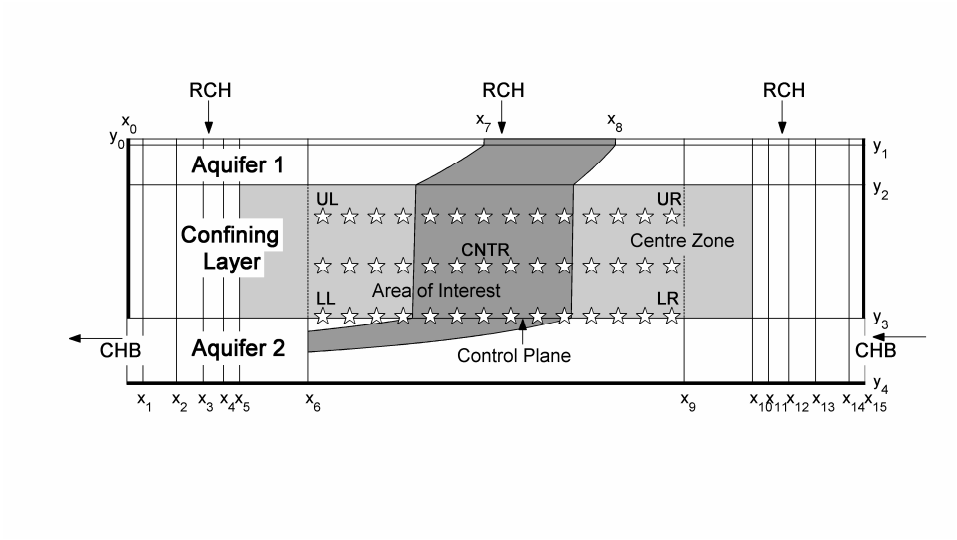

FIGURE 1. Schematic overview of the numerical grid used in the calculations, showing also the prior contaminant flow. RCH = recharge. CHB = constant head boundary. x_0 = 0m, x_1 = 2.0m, x_2 = 2502.0m, x_3 = 2752.0m, x_4 = 2777.0m, x_5 = 2779.5m, x_6 = 2879.5m, x_7 = 3009.5m, x_8 = 3109.5m, x_9 = 3159.5m, x_{10} = 3259.5m, x_{11} = 3262.0m, x_{12} = 3287.0m, x_{13} = 3537.0m, x_{14} = 6037.0m, x_{15} = 6039.0m. y_0 = 0m, y_1 = -0.1m, y_2 = -2.2m, y_3 = -22.2m, y_4 = -32.2m. Bold lines indicate no flow boundaries. Potential measurement locations labelled UR, UL, LR, LL and CNTR are the locations where the head variances are evaluated in Figure 3. Arrows indicate flow directions.

Additional to the discretization of the flow model as shown in Figure 1, the area between x_5 and x_{10} is discretized into 240 equally sized (Δx = 2.0 m) columns, and the confining layer (y_2-y_3) is discretized into 40 equally sized rows (Δy = 0.5 m). A steady-state head distribution is obtained by assigning recharge (250 mm/y) to the top of every Aquifer 1 cell and by imposing a constant head of 0.0 m and 15.0 m in the utmost left and utmost right cell of Aquifer 2, respectively.

The control plane at which breakthrough is evaluated is located at y_3 along the bottom of the area of interest (x_6-x_9, y_2-y_3). The area of interests consists of 140 columns. The purpose of all areas outside the area of interest is solely to reduce the impact of boundary conditions on the flow in the area of interest.

In the top layer of Aquifer 1, a source of a conservative contaminant is assumed. For illustration, Figure 1 shows the contaminant plume that results from this source zone in

combination with the flow model described above, when the confining layer is modeled as a homogeneous medium with ln K_G = -3.0 ln(m/d).

The stars in the area of interest indicate 42 potential measurement locations. All measurements suffer from measurement errors which are assumed to be Gaussian distributed and uncorrelated. The measurement error variances are taken to be 0.001 m^2 and 0.001 $(m^2/d)^2$ for *h* and *T* (transmissivity) measurements, respectively. For *t* measurements, the observation error standard deviation is taken as a percentage of the expected observed (untransformed) age at the sampled location in the prior realization of the confining layer, in which all stochastic parameters are set at their prior means (ln *K* = -3.0 ln(m/d)). This percentage is varied between different calculations.

4. RESULTS

4.1 Validity of the first-order design method

As our inverse method is defined in a Bayesian framework, it is important that all unknown parameters and dependent variables are Gaussian distributed. It is well known that, if *Y* is Gaussian distributed, then, at first order, so are the hydraulic head and the particle displacement [*Dagan*, 1989]. Particle travel times, however, can be significantly skewed if the number of correlation scales travelled is limited (see Section 2). Luckily, the natural logarithm of arrival time is approximately Gaussian if σ_Y^2 is not too large (evidence not shown), at least for the range of σ_Y^2 values for which the linear theory applied here can be assumed valid (see below).

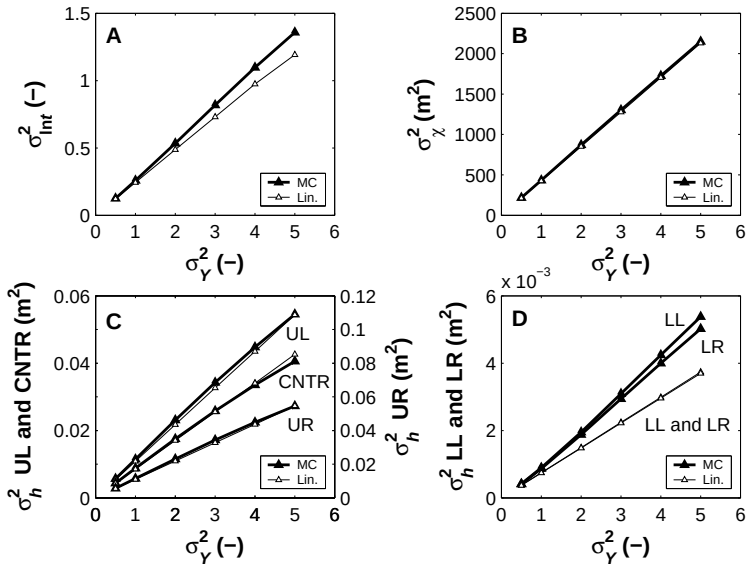

FIGURE 2. Comparison of linearized *t* (a), χ (b) and *h* (c and d) variance predictions as given by the representer approach with variances obtained from a Monte Carlo series, as function of σ_Y^2. In (c and d), the labels UL, LL, UR, LR and CNTR correspond with the labels in Figure 1 and indicate the locations in the domain at which the variances were evaluated.

The question regarding the applicability of the linear theory for larger σ_Y^2 is addressed in Figure 2. Figure 2 compares the variances of the state variables as calculated by the representer approach (which is based on linearizations and therefore in principle only valid for small σ_Y^2) with the variances of these variables computed with a Monte Carlo approach. Figure 2a shows the results for the lognormal transform of the breakthrough time, ln τ_{BT}, of a particle originating from the centre of the contaminant source zone. Up to σ_Y^2 = 3.0 the difference between linearization and Monte Carlo results is less than 10%, suggesting that the first-order approximation in combination with the Gaussian assumption for ln *t* gives fairly good ln *t* variance predictions for σ_Y^2 = 3.0 and lower.

Figure 2b compares Monte Carlo variances with linearized variances of the particle breakthrough location. The correspondence is excellent for the entire range investigated. The same holds for the head variance in the upper part of the domain (Figure 2c). Only for the head variance in the lower part, the first-order results deteriorate rather quickly as σ_Y^2 increases. Nevertheless, considering the small absolute value of the underestimation of the head variance (note the factor 10 difference between the y-axes of Figure 2c and 2d), the error made is not expected to greatly influence contaminant breakthrough uncertainty predictions.

Based on the results given in Figure 2, we estimate that for the synthetic example given in Figure 1, the outcome of our first-order design strategy as outlined in Section 2 is reliable (posterior prediction uncertainty) up to a variance of *Y* of 3.0.

4.2 Performance of age observation networks compared with *T* and *h* networks

Figure 3 shows the posterior breakthrough variance of (near) optimal designs as a function of network size (in number of selected measurements), for observation networks containing only one measurement type. It can be seen that the uncertainty reduction that can be realized with age observations is larger than can be achieved with the other measurement types (with reasonable network sizes), if the travel time measurements are of a reasonable reliability.

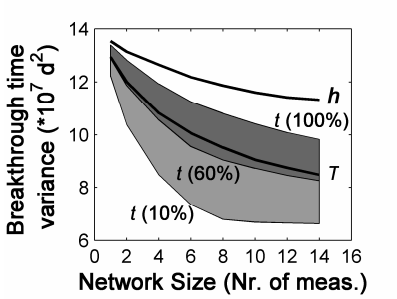

FIGURE 3. Breakthrough time uncertainty reduction achieved by networks consisting of only one observation type, as a function of network size. The percentages given refer to different levels of the age observation error standard deviation.

5. CONCLUSIONS AND DISCUSSION

We proposed a first-order methodology that incorporates age observations into measurement network design for minimizing contaminant breakthrough uncertainty in short residence time systems. The natural logarithm of travel time is approximately Gaussian distributed for systems of low to medium heterogeneity. Building on this knowledge, we derived an expression for the linearized (cross-)covariance between states and parameters on one hand and the natural logarithm of travel time on the other. These (cross-)covariances are necessary for the calculation of posterior breakthrough time uncertainty given a certain set of measurements. In a synthetic example of contaminant breakthrough in a confining layer, the expression was shown to approximate the prior variance of the natural logarithm of travel time well as long as the variance of the natural logarithm of the conductivity was smaller than 3.0. Age observations, if of a reasonable quality, were shown to be more valuable for the reduction of breakthrough time uncertainty than head and conductivity measurements.

REFERENCES

Anderman, E. R., and M. C. Hill (2001), MODFLOW-2000, the U.S. Geological Survey modular ground-water model - documentation of the advective-transport observation (ADV2) package, version 2, t, open-file report 01-54, U.S. Geological Survey, Denver, CO.

Cirpka, O. A., and P. K. Kitanidis (2001), Sensitivity of temporal moments calculated by the adjoint-state method and joint inversing of head and tracer data, *Advances in Water Resources*, 24, 89-103.

Dagan, G. (1989), *Flow and Transport in Porous Formations*, Springer-Verlag, Berlin.

Feyen, L., J. J. Gomez-Hernandez, P. J. Ribeiro Jr., K. J. Beven, and F. De Smedt (2003), A Bayesian approach to stochastic capture zone delineation incorporating arrival times, conductivity measurements, and hydraulic head observations, *Water Resources Research*, 39, 1126, doi:1110.1029/2002WR001544.

Harvey, C. F., and S. M. Gorelick (1995), Mapping hydraulic conductivity: sequential conditioning with measurements of solute arrival time, hydraulic head, and local conductivity, *Water Resources Research*, 31, 1615-1626.

Izbicki, J. A., C. L. Stamos, and P. Martin (2004), Comparison of ground-water flow model particle tracking results and isotopic data in the Mojave River ground-water basin, southern California, USA, *Journal of Hydrology*, 292, 30-47.

Portniaguine, O., and D. K. Solomon (1998), Parameter estimation using groundwater age and head data, Cape Cod, Massachusetts, *Water Resources Research*, 34, 637-645.

Reilly, T. E., L. N. Plummer, P. J. Phillips, and E. Busenberg (1994), The use of simulation and multiple environmental tracers to quantify groundwater flow in a shallow aquifer, *Water Resources Research*, 30, 421-433.

Sheets, R. A., E. S. Bair, and G. L. Rowe (1998), Use of $^{3}H/^{3}He$ ages to evaluate and improve groundwater flow models in a complex buried-valley aquifer, *Water Resources Research*, 34, 1077-1089.

Stute, M., and P. Schlosser (2000), Tritium/^{3}Helium measurements as calibration tools in groundwater transport modelling, paper presented at Tracers and modelling in hydrogeology (Proceedings of the TraM'2000 conference held at Liege, Belgium, May 2000),

Valstar, J. R., D. B. McLaughlin, C. B. M. Stroet, and F. C. van Geer (2004), A representer-based inverse method for groundwater flow and transport applications, *Water Resources Research*, 40, W05116, doi:10.1029/2003WR002922.

Woodbury, A. D., and Y. Rubin (2000), A full-Bayesian approach to parameter inference from tracer travel moments and investigation of scale effects at the Cape Cod experimental site, *Water Resources Research*, 36, 159-171.