ground water

Estimating Flow Using Tracers and Hydraulics in Synthetic Heterogeneous Aquifers

by Marie Larocque[1], Peter G. Cook[2], Klaus Haaken[3], and Craig T. Simmons[4]

Abstract

Regional ground water flow is most usually estimated using Darcy's law, with hydraulic conductivities estimated from pumping tests, but can also be estimated using ground water residence times derived from radioactive tracers. The two methods agree reasonably well in relatively homogeneous aquifers but it is not clear which is likely to produce more reliable estimates of ground water flow rates in heterogeneous systems. The aim of this paper is to compare bias and uncertainty of tracer and hydraulic approaches to assess ground water flow in heterogeneous aquifers. Synthetic two-dimensional aquifers with different levels of heterogeneity (correlation lengths, variances) are used to simulate ground water flow, pumping tests, and transport of radioactive tracers. Results show that bias and uncertainty of flow rates increase with the variance of the hydraulic conductivity for both methods. The bias resulting from the nonlinearity of the concentration–time relationship can be reduced by choosing a tracer with a decay rate similar to the mean ground water residence time. The bias on flow rates estimated from pumping tests is reduced when performing long duration tests. The uncertainty on ground water flow is minimized when the sampling volume is large compared to the correlation length. For tracers, the uncertainty is related to the ratio of correlation length to the distance between sampling wells. For pumping tests, it is related to the ratio of correlation length to the pumping test's radius of influence. In regional systems, it may be easier to minimize this ratio for tracers than for pumping tests.

Introduction

Regional ground water flow rates are most usually estimated using Darcy's law, with hydraulic conductivities estimated from pumping tests. Less frequently, flow rates are estimated using ground water residence times along a flow line derived from analysis of radioactive tracers. A number of studies have demonstrated relatively good agreement between regional fluxes estimated using radioactive tracers and pumping tests in simple aquifer systems (e.g., Pearson and White 1967; Drury et al. 1984). In heterogeneous aquifers, however, both become more uncertain. Zuber et al. (2008), for example, found relatively poor agreement between the two methods in a karst aquifer. Dann et al. (2008) also found very different results from artificial tracers and pumping tests in a channelized aquifer. The reasons why the agreement between the two methods is worse in heterogeneous systems have never been demonstrated. Intuitively, however, it seems reasonable that since both methods sample different portions of the aquifer, the likelihood that the estimated flow rates will be the same and will be representative of the mean flow rate through the system decreases as the level of heterogeneity increases. An important question arises: which method, tracers or hydraulics, is therefore likely

[1]Corresponding author: Centre de recherche pour l'Étude et la Simulation du Climat à l'Échelle Régionale Département des sciences de la Terre et de l'atmosphère (ESCER)-Université du Québec à Montréal, Montréal, Québec, Canada; 514-987-3000 ext. 1515; fax 514-987-7749; larocque.marie@uqam.ca

[2]CSIRO—Land and Water, Private Bag 2, Glen Osmond, Adelaide, SA 5064, Australia

[3]Steinmann Institute for Geology, Mineralogy and Palaeontology, University of Bonn, Nussallee 8, D-53115 Bonn, Germany

[4]School of Chemistry, Physics and Earth Sciences, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia

Received September 2008, accepted May 2009.

doi: 10.1111/j.1745-6584.2009.00595.x

to produce more accurate and reliable estimates of mean regional ground water flow rates in highly heterogeneous systems?

The most common methods used to estimate ground water age rely either on radioactive decay of a tracer whose input concentration has been relatively constant over time (e.g., ^{14}C, ^{36}Cl), on measurement of parent and daughter isotopes for radioactive tracers with uncertain input concentrations ($^{3}H/^{3}He$), or on measurement of compounds that have a variable, but well-known input history (e.g., CFCs, SF_6). It is important to distinguish between advective age (or flowtube age) and tracer age. Advective age is the time elapsed along a streamline since recharge took place. It is the age required for calculation of flow velocities. Tracer age is the mean travel time of tracer molecules between a recharge location and a sampling point. It is affected by diffusion and dispersion, and therefore different tracers can provide different ages depending on their input function, diffusion coefficient, and in some cases rate of radioactive decay (e.g., LaBolle et al. 2006, 2008). Tracer age will correspond to the advective age only in the presence of a closed system and absence of water mixing and will be biased in heterogeneous aquifers where mixing is pronounced (Castro and Gobelet 2005; Weissmann et al. 2002). Radioactive tracers are useful for estimating ground water flow rates to the extent that tracer ages approximate advective ages. Theoretical studies have shown that the difference between advective and tracer ages will usually be less than 10% in homogeneous one-dimensional flow fields (Ekwurzel et al. 1994). Weissmann et al. (2002) showed that for highly heterogeneous systems, the difference can reach 100%. However, the relationship between these ages and the level of heterogeneity has never been systematically analyzed.

A number of studies have examined hydraulic conductivities in heterogeneous flow fields. These studies have found that pumping tests provide valuable information about the aquifer even in heterogeneous media. Leven and Dietrich (2006) have shown that the Theis method (Theis 1935) can be used to assess the spatial arrangement of estimated hydraulic parameters in the area surrounding a single well pumping test. Others have demonstrated that the Cooper-Jacob method (Cooper and Jacob 1946) can provide valuable information in heterogeneous aquifers and that estimated transmissivities on multiple observation points converge to a single value which corresponds to the effective transmissivity of the aquifer (e.g., Sanchèz-Villa et al. 1996, 1999; Meier et al. 1998). This transmissivity produces flow rates that replicate the mean behavior of the system (Gomez-Hernandez and Gorelick 1989). According to Schad and Teutsch (1994), drawdown is the result of a scaling up of hydraulic properties over an increasing averaging volume during the temporal evolution of the depression cone. To provide effective parameters in a heterogeneous aquifer, pumping tests must therefore be sufficiently long to sample a significant portion of the flow field. This comes in contradiction with the common use of short-term pumping tests as an inexpensive and straightforward option for characterization studies.

Some authors have suggested that one advantage of ground water age is that this method integrates spatial variability in hydraulic conductivities over long distances, therefore reflecting the regional aquifer heterogeneity more realistically than pumping tests (Phillips et al. 1989; Zuber et al. 2008). However, this claim has never been tested. The aim of this paper is to compare radioactive tracers and hydraulic approaches to assess ground water flow in heterogeneous aquifers. To our knowledge no previous study has systematically compared the effects of heterogeneity on flow rates estimated using tracer and hydraulic methods within a single analysis. To investigate the heterogeneity conditions in which one method is more or less accurate and reliable than the other in a framework that is both quantitative and comparative, it is necessary to perform tests on many aquifers. Although this can be done in field studies, the easiest and most straightforward approach is with numerical simulation of tracer transport and pumping tests within synthetic heterogeneous aquifers. In this study, two-dimensional (2D) synthetic porous media aquifers with different levels of heterogeneity are developed using a numerical model. Both tracer analysis and pumping tests are performed and ground water flow rates, bias, and uncertainty are assessed for a variety of heterogeneous aquifer conditions considered here as demonstration examples. Our intention is not to quantify absolute bias and uncertainty, but rather to demonstrate in the simplest possible terms the relative performance of tracers and pumping tests and provide cause and effect understanding. Absolute answers (mean and variance of the ground water flow rates) will be dependent on the spatial structure of the underlying hydraulic conductivity field, a proper assessment of which would require a much more exhaustive study.

Methodology

Simulated Domain

Ground water flow and transport are simulated using the HydroGeoSphere model (Therrien et al. 2006). The model uses the control volume finite element method to discretize the ground water flow and advection–dispersion transport equations. Horizontal flow occurs in a 2D sedimentary confined aquifer measuring 4000 by 1250 m and 25 m thick (Figure 1). Confined conditions were used to ensure horizontal flow directions. The choice of a 2D model was made to limit computation time. Although a three-dimensional (3D) model would have provided insight on the impact of vertical heterogeneity, it would also have created a computational intensity beyond the scope of this study. 2D simulations are demonstrative of behavior that is expected to be observed in 3D as well. To limit the influence of boundary conditions on the flow and transport predictions, a 500-m homogeneous buffer zone with the mean hydraulic conductivity (K) of the heterogeneous distribution was used at the upgradient and downgradient ends of the flow domain. Elsewhere, K varies

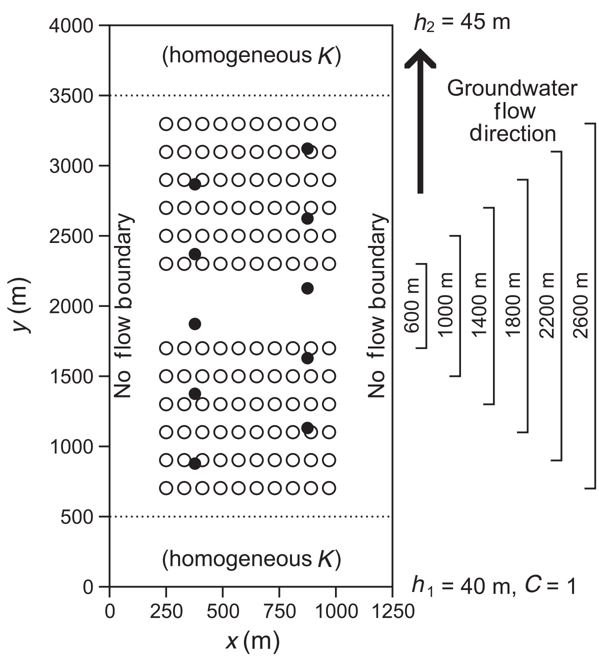


Figure 1. Model setup and location of pumping wells (black circles) and tracer observation wells (white circles). Flow boundary conditions are constant heads on bottom ($h_1 =$ 50 m) and top ($h_2 = 45$ m) and no flow on left and right of the simulated domain. Transport boundary condition is $C = 1$ on the bottom boundary.

spatially according to a defined geostatistical distribution that is described in the following text. Constant head values are prescribed along the bottom ($h_1 = 50$ m) and top ($h_2 = 45$ m) boundaries of the buffer zone, resulting in a hydraulic gradient of 0.00125 from bottom to top. No flow boundaries are specified at the outer edge of the flow domain along the right and left sides of the domain. Horizontally, the domain is divided into 5×5 m cells, giving a total of 200,000 cells (because the confined aquifer is 2D only, there is only one cell in the vertical direction). Automated time steps satisfying the Courant number criterion were used for the transient state simulation.

The Sequential Gaussian Simulation (SGSIM) method of the Geostatistical Software Library (GSLIB) (Deutsch and Journel 1998) with a spherical variogram was used to generate random, isotropic, lognormally distributed, 2D hydraulic conductivity fields for the simulated domain. It is important to underline the fact that the K fields generated with this approach are stationary and isotropic whereas real systems are likely to be both nonstationary and anisotropic. In particular, the connectivity of high conductivity pathways generated by this model may not reflect the connectivity of field aquifers. Nevertheless, the model is useful to assess the effects of heterogeneity on flow rates estimated using tracer analysis and pumping tests. K fields were generated with the SGSIM method for isotropic correlation lengths ($A_x = A_y = A$) of 25, 100, and 250 m. In their review of the literature, Hoeksema and Kitanidis (1985) report field scale correlation lengths on the order of 1 km and Delhomme (1979) reports correlation lengths reaching 10 km. It was impractical to simulate such large correlation lengths in our model while keeping cell size relatively small for pumping test analyses. The mean of the natural logarithm of hydraulic conductivity is 0 for all K fields, and so the geometric mean of K values is 1 m/d. To represent increasing levels of heterogeneity, K fields with ln(K) variances (σ^2) of 0.3, 1, 2.5, and 4 were used. With these variances, 99% of the K values in the flow field are within 1.1, 2.3, 3.6, and 4.4 orders of magnitude from the mean value. These σ values are representative of field conditions reported in the literature ranging from relatively homogeneous aquifers (e.g., LeBlanc et al. [1991] report $\sigma^2 = 0.26$ for the Cape Cod site) to highly heterogeneous aquifers (Barlebo et al. [2004] report $\sigma^2 = 4.5$ for the MADE site). Five realizations of the random correlated hydraulic conductivity field were generated for each A and σ^2 combination. Flow and transport was simulated within each K field. The increasing spatial structure of K with increasing A and the increasing variability of K with increasing σ^2 are clearly visible in Figure 2 that illustrates one K field for each A and σ^2 combination.

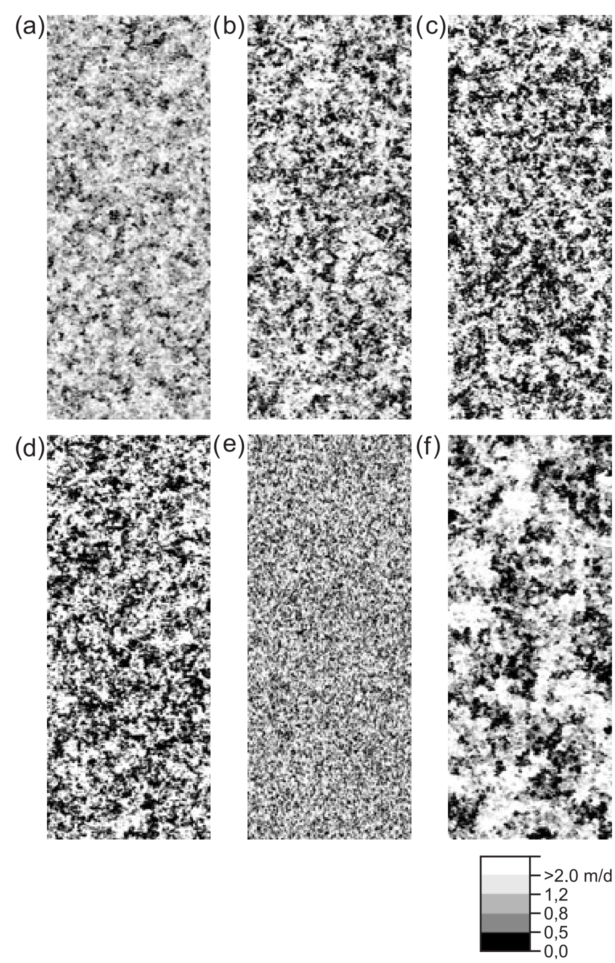


Figure 2. Examples of heterogeneous K field with different variances (σ^2) and correlation lengths (A). (a) $\sigma^2 = 0.3$ and $A = 100$ m, (b) $\sigma^2 = 1.0$ and $A = 100$ m, (c) $\sigma^2 = 2.5$ and $A = 100$ m, (d) $\sigma^2 = 4.0$ and $A = 100$ m, (e) $\sigma^2 = 1.0$ and $A = 25$ m, and (f) $\sigma^2 = 1.0$ and $A =$ 250 m.

Other parameters used in the model include porosity ($n = 0.05$) and specific storage ($S_s = 0.0001$/m). Dispersivity (α) is 0.5 m, that is, a value similar to reported dispersivities at field sites with flow distances similar to the scale of our cells (cf. Gelhar et al. 1992). Macrodispersion is mainly driven by the heterogeneous structure of the flow field. These parameters do not vary spatially. Initial heads for the transient state simulation are those from a steady-state simulation with no pumping. Initial concentrations are set to zero everywhere.

Tracer Analysis

For each K field, 120 observation wells are introduced into the heterogeneous aquifer for tracer analysis. These form 60 upgradient-downgradient well pairs, which would be located on a flow line in the homogeneous case. The observation wells are separated by 600, 1000, 1400, 1800, 2200, and 2600 m (Figure 1). Theoretical radioactive tracers with unit concentration were applied as a constant concentration source across the entire model face at the upgradient boundary condition for all times. As ground water age is usually a nonlinear function of tracer concentration, there is a systematic underestimation of travel time resulting from the application of the radioactive decay equation to a mixture of ground water of different ages (see Park et al. 2002; Cook and Solomon 1997). To verify the impact of this error in the simulated aquifers, theoretical tracers with half-lives of 10, 50, 200, and 1000 years were released in the simulated heterogeneous aquifers. Transport was simulated under steady-state flow conditions for a period that was long enough to ensure that a constant concentration was achieved at all observation wells. Steady tracer concentrations were reported for each observation well. The travel time from the upgradient input boundary condition to each sampling well, also equivalent to the apparent ground water age at this well, is calculated using Equation 1, which is commonly used in studies with radioactive tracers.

$$t = \frac{\ln(C_0) - \ln(C_w)}{\frac{\ln\,(2)}{t_{1/2}}} \tag{1}$$

where t = travel time from the upgradient input boundary condition to each sampling well (T), C_0 = initial concentration at the upgradient limit (1 kg/m^3), C_w = concentration in the sampled well (mL^{-3}), and $t_{1/2}$ = tracer half-life (10, 50, 200, and 1000 years).

The flow rate using tracer age data is estimated using Equation 2 (cf. Figure 3a).

$$Q_T = \left(\frac{\Delta y}{t_d - t_u}\right) Lb\,\phi \tag{2}$$

where Q_T = ground water flow rate estimated using radioactive tracers (L^3/T), Δy = distance between the upgradient and downgradient wells (L), t_d = travel time to the downgradient well from the upgradient boundary (T), t_u = travel time to the upgradient well from the upgradient boundary (T), L = aquifer width (1250 m), b = aquifer thickness (25 m), and ϕ = porosity (0.05).

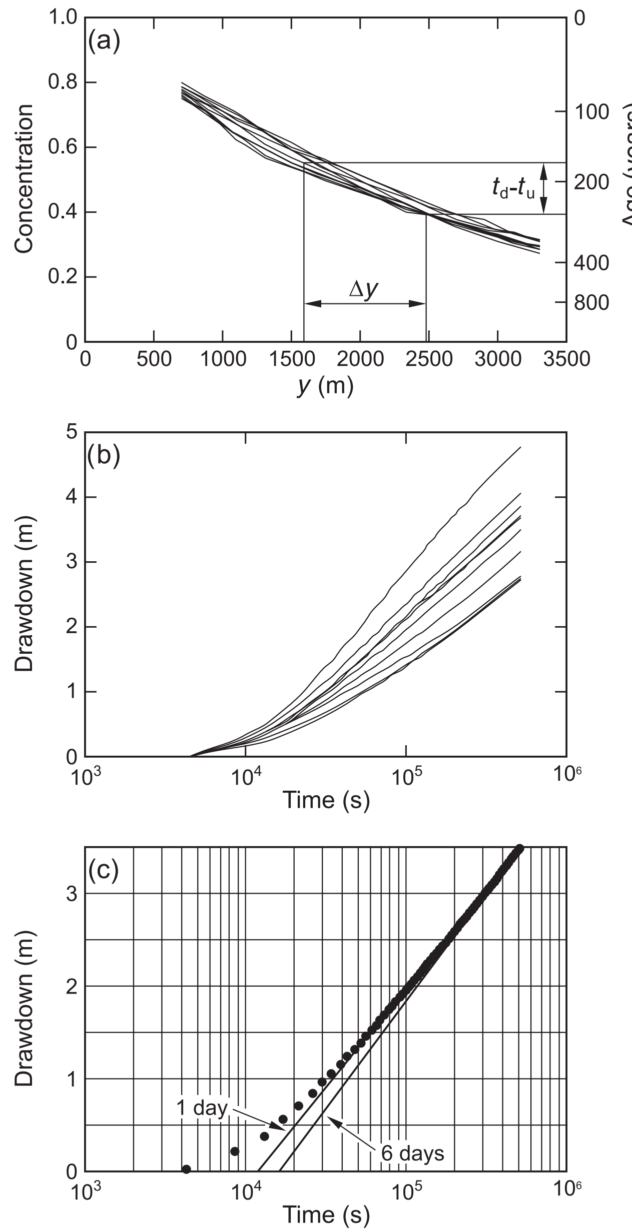


Figure 3. Results from one K field for (a) the evolution of tracer concentrations and ground water ages as a function of distance in the flow direction for 10 equally spaced distances along the x-axis (t_d = travel time to the downgradient well; t_u = travel time to the upgradient well), (b) drawdown as a function of time for 10 pumping tests in the flow domain, and (c) impact of pumping test duration on the rate of drawdown. Tracer half-life is 200 years and the K field is generated for a correlation length (A) of 100 m and a variance (σ^2) of 1.

The use of 10 well pairs for each well separation in the five K fields provides a total of 50 estimates of Q_T for each combination of σ^2 and A. Using multiple well pairs reduces the number of K fields required to get statistically meaningful results. A similar approach has been used, for example, by Sanchèz-Villa et al. (1999), who interpreted pumping test results for all simulation nodes surrounding a pumping well on a single geostatistically simulated K field.

Results are expressed in terms of the mean flow rate (Q_T) and its coefficient of variation CV(Q_T), that is, the standard deviation of the flow rate for the ensemble of realizations divided by the mean flow rate, to reflect flow rate uncertainty. It should be noted that our results do not consider the impact of measurement error in tracer concentrations associated with sampling and analytical procedures on calculated travel times and flow rates.

Pumping Tests

For each K field, 10 fully penetrating pumping wells were introduced in the model (Figure 1). The choice of position ensures that they are separated by a minimum of 500 m, that their radii of influence do not overlap, and that they are not influenced by the model boundaries. Once steady-state flow was established through the model to simulate background ground water flow (with no pumping), each simulation was run in transient state for 6 d, a duration longer than most pumping tests performed for characterization purposes. Pumping wells were activated one by simulation at a rate of 300 m^3/d, which is a representative pumping rate. Drawdown was recorded approximately 50 m from the pumping well (measured at the model node closest to 50 m along a line joining the pumping well and the center of the model domain). The pumping tests were interpreted with the Cooper-Jacob method and transmissivity was estimated by fitting a straight line through the late drawdown data and calculating the slope (Figure 3b). Since in a heterogeneous aquifer, results from pumping test data-time vary in both time and space (see Schad and Teutsch 1994), parameter estimates based on late times will better represent effective values due to an increasing volume of influence (Leven and Dietrich 2006). They are therefore more appropriate for the estimation of the mean ground water flow. This methodology is valid if the influence of the homogeneous transmissivity zone and of lateral no flow boundaries can be neglected. In this study, drawdown after 6 d of pumping was negligible at all model boundaries. The simulation was repeated for each of the 10 pumping wells within the same heterogeneous K field. This entire procedure was then repeated for each of the different heterogeneous realizations (varying both A and σ^2).

Flow rate in the aquifer is calculated hydraulically using Darcy's law:

$$Q_H = KLbi \tag{3}$$

where Q_H = ground water flow rate calculated with the hydraulic conductivity obtained from pumping test (L^3/T), K = hydraulic conductivity estimated from pumping test (L/T), L = aquifer width (1250 m), b = aquifer thickness (25 m), and i = hydraulic gradient (0.00125 m/m).

Similar to the tracer analysis, using 10 pumping wells for each of the five K fields provides a total of 50 estimates of Q_H for each combination of σ^2 and A. Results are expressed in terms of the mean and CV of Q_H.

In a heterogeneous flow field with a regional gradient, the drawdown cone surrounding a pumping well will be irregular and skewed downgradient. The radius of influence of a similar pumping test in homogeneous media with a hydraulic conductivity equal to the geometric mean and no regional gradient nevertheless provides a useful reference with which to compare the relative size of the sampled domain for the different pumping tests. The radius of influence of a well in a homogeneous aquifer is calculated as follows (De Marsily 1986):

$$R = 1.5\sqrt{Kbt/S} \tag{4}$$

where R = radius of influence (L), K = hydraulic conductivity estimated from pumping test (L/T), b = aquifer thickness (25 m), t = time (T), and S = storage coefficient (0.0025).

R keeps increasing as t increases, reflecting the fact that in an infinite aquifer a pumping test will always be in transient state. Using the geometric mean of hydraulic conductivities of 1 m/d, R varies from 150 m for the 1-d pumping test to 367 m after 6 d of pumping. The results do not consider the impact of head measurement error on calculated flow rates, but these errors are generally negligible (generally <1 cm).

Results

Bias in Flow Estimates

Methods to estimate ground water flow will be biased if the expected value of the derived flow rates is different from the true mean flow. In this section, we examine possible bias by comparing mean flow rates estimated from tracers and pumping tests with the total fluid flux through the model domain.

For comparison purposes, simulations were first performed for a homogeneous aquifer with a hydraulic conductivity of 1 m/d, that is, equal to the geometric mean of K values. In these conditions, the simulated fluid flux through the model domain (Q_M) is 39.1 m^3/d. The travel time at pore velocity assuming piston flow from the lower to the upper boundary (Equation 2) is 438 years. Hydraulic conductivities estimated from the 10 pumping tests in this flow field are all equal to 1 m/d after 6 d of pumping, and Q_H is estimated to be 39.1 m^3/d, that is, equal to Q_M. Using a tracer with a 200 years half-life, Q_T was estimated to be 39.4 m^3/d, that is, 1% larger than Q_M and Q_H. The slightly higher flow rate obtained from the tracer analysis can be explained by longitudinal tracer dispersion within the homogeneous aquifer, which is known to lead to overestimation of the concentration of a radioactive tracer (see Castro and Gobelet 2005). Because this reduces the apparent ground water age, dispersion will lead to an interpreted increase in the estimated flow rate. This effect is small in the case of a homogeneous aquifer.

We now turn to cases that consider heterogeneous distributions. With σ^2 ranging from 0.3 to 4.0, values for

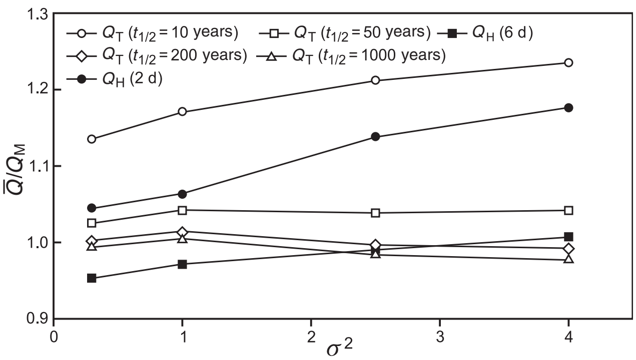


Figure 4. Flow rate ratios $\overline{Q}/Q_M$ estimated from radioactive tracers with different half-lives ($t_{1/2}$) and from pumping tests with different durations as a function of K field variance (σ^2). Well pairs are separated by 2200 m, tracer half-life is 200 years, and K fields are generated for a correlation length (A) of 100 m.

Q_M range from 37.2 to 40.4 m^3/d. Using Darcy's law, Q_M and the applied head gradient provide an equivalent K between 0.95 and 1.03 m/d. These values are very close to the geometric mean of hydraulic conductivities, as expected for a 2D infinite isotropic multi-Gaussian field (Matheron 1967). Figure 4 illustrates the ratios between estimated mean flow rates ($\overline{Q_H}$ and $\overline{Q_T}$) and Q_M for heterogeneous fields as a function of K field variance for the different tracers and two pumping test durations. The $\overline{Q}/Q_M$ ratio will be close to one in conditions where hydraulic methods and tracers provide reliable estimates of the average flow rate in the aquifer system.

With the 200 and 1000 year half-life tracers, the $\overline{Q_T}/Q_M$ ratio is close to one; that is, the mean flow rates determined with these tracers are similar to the simulated fluid flux through the model domain. When tracer half-life decreases to 50 and 10 years, the $\overline{Q_T}/Q_M$ ratio increases markedly, showing that there is a significant bias when using a tracer with a half-life much shorter than the actual travel time within the domain. This bias is due to the nonlinearity of the concentration–time relationship, which causes a mixed ground water sample to have an apparent age that is less than the mean age of the components. This bias will therefore lead to an overestimation of flow rates estimated using Equation 2. Results show that the bias increases as the tracer half-life becomes much shorter than the ground water residence time (when $t_{1/2}$ decreases from 200 to 50 and to 10 years). Our results also show that the bias increases with increasing σ^2 and therefore that the magnitude of this effect is related to the degree of heterogeneity and to the macrodispersion in the aquifer. This cumulative uncertainty can result in a significant error on estimated flow rates. Figure 4 shows that if ground water ages derived from a tracer with a half-life of 10 years are used in this study, the estimated flow rates can be biased by up to 23% for σ^2 equal to 4. In a field scale application, sampling and analytical errors could become a significant issue if $t_{1/2}$ is much shorter (background concentrations may be approached) or much larger (and upgradient and downgradient concentrations are indistinguishable) than the ground water residence time.

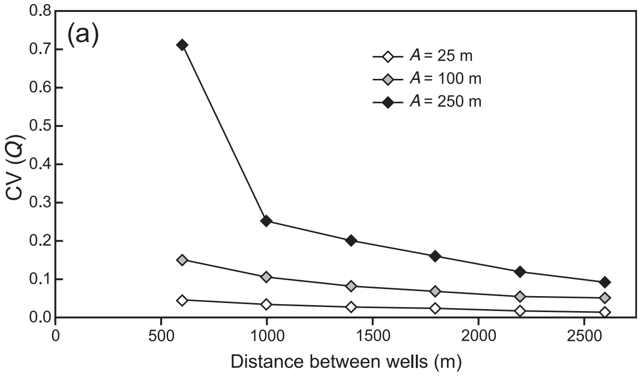
The time-drawdown curves from the pumping test with the different scenarios of aquifer heterogeneity show that when drawdown is plotted against a log scale of time, slopes of the straight line (i.e., the drawdown rates) vary with time (e.g., Figure 3c), as expected when different heterogeneities are encountered (Leven and Dietrich 2006). Each slope can be considered representative of the progressive sampling of different portions of the aquifer as the drawdown cone grows (Schad and Teutsch 1994). Our results show that the $\overline{Q_H}/Q_M$ ratios decrease when pumping test duration increases, and after 6 d are close to one for all variances. The effective hydraulic conductivity has therefore been sampled and the radius of influence is large enough to integrate heterogeneities over a representative portion of the aquifer for all variances. With the 2-d pumping tests, the pumping duration bias is evident for the most heterogeneous K fields reaching 18% for σ^2 equal to 4. We hypothesize that the ratios larger than one can be explained by high K zones having on average a larger influence on drawdown than low K zones. It is possible that using a nonstationary random K field might not have yielded the averaging noted at late times for pumping tests because in these fields interconnected pathways of high K zones may exist at lengths greater than would be predicted by a random stationary field.

Uncertainty of Ground Water Flow Estimates

Uncertainty refers here to the scatter around the mean flow rate and is illustrated with the coefficient of variation of flow rates CV(Q). In a field study, uncertainty can be reduced by increasing the sampling density of the flow field, but it cannot be eliminated.

From hereon, the 200 year half-life tracer is adopted because this half-life is closest to the residence time of water in the aquifer (438 years). Figure 5 shows the effect of well separation, pumping test duration, and correlation length on CV(Q) estimated with tracers and pumping tests. Results show that as the wells get farther apart, the tracer is less likely to sample only high or low conductivity zones, and so the estimated flow rate becomes less variable (Figure 5a). Similarly, as pumping duration increases, the radius of influence of the pumping test increases, the investigated flow field becomes less dependent on local conditions, and CV(Q_H) thus decreases (Figure 5b). The decrease in variance of hydraulic conductivity with increased duration of pumping tests has previously been shown (e.g., Lachassagne et al. 1989; Leven and Dietrich 2006), but the influence of well separation on the uncertainty of flow rates estimated with tracer ages has never been demonstrated before.

The variability of Q also increases with increasing correlation length for both tracers and pumping tests. For example, when the sampling wells are 2600 m apart, CV(Q_T) is 6.2 times higher with $A = 250$ m than with



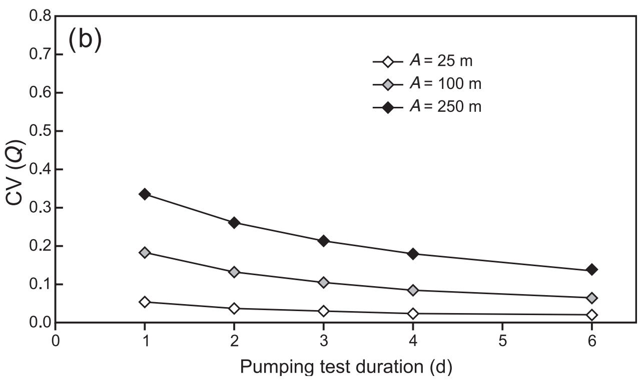


Figure 5. Flow rate variability CV(Q) as a function of correlation length (A) against (a) distance between the sampling wells when using tracers and (b) pumping test duration. Tracer half-life ($t_{1/2}$) is 200 years and K fields are generated for a variance (σ^2) of 1.

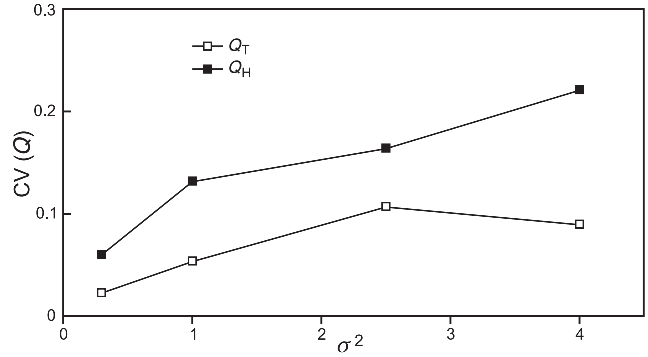


Figure 6. Flow rate variability CV(Q) as a function of σ^2 when using data from tracers (Q_T) and pumping tests (Q_H). Well pairs are separated by 2200 m, tracer half-life ($t_{1/2}$) is 200 years, pumping tests are 2 d long, and K fields are generated for a correlation length (A) of 100 m.

$A = 25$ m. For a 6-d pumping test, CV(Q_H) is 7.2 times higher with $A = 250$ m than with $A = 25$ m. For tracers, CV(Q_T) increases significantly when $A = 250$ m and the distance between wells is smaller than 1000 m. From that point, the tracer is apparently not traveling through a sufficient number of correlation lengths to provide a reliable flow rate estimate that is representative of the larger scale system. The results show that with our model setup, CV(Q_H) estimated from a 6-d pumping test is the same as CV(Q_T) estimated from apparent ages on wells separated by approximately 2000 m for all three A values. If the paired wells used in tracer approaches are closer than this, the 6-d pumping test method is more likely to accurately reflect the true mean ground water flow rate. Conversely, Figure 5 shows that if the wells are farther apart, then the tracer method will be more accurate than a 6-d pumping test. It is important to note that relatively small correlation lengths were used in this study. Larger A values are likely to emphasize even more the need to use well pairs separated by a large distance and long duration pumping tests.

Figure 6 illustrates how σ^2 influences flow rate variability. For all values of σ^2, the average CV(Q) estimated using pumping test data are larger than those estimated from tracer analysis, although this is a function of the well separation and pumping test duration plotted in the figure (results illustrated in Figure 6 are for wells 2200 m apart and 2-d pumping tests). Although the number of points is limited, CV(Q_H) and CV(Q_T) apparently increase in a similar manner in the range of tested σ^2. Neither method appears markedly more reliable than the other in heterogeneous conditions.

In the investigated range of σ^2, CV(Q_T) ranges from 0.02 to 0.11 while CV(Q_H) ranges from 0.06 to 0.22. The CVs of ln(K) range from 0.55 ($\sigma^2 = 0.3$) to 2 ($\sigma^2 = 4$). The uncertainty of predicted flow rates is therefore much smaller than the inherent variability in the heterogeneous flow field itself. These results are probably determined in part by the small correlation lengths used in this work. For very large correlation lengths, CV(Q) should approach the CV of the underlying distribution of hydraulic conductivity, that is, $\sqrt{e^{\sigma^2} - 1}$ for a lognormally distributed K field. For $\sigma^2 = 1$, the maximum CV(Q) corresponding to a large correlation length would thus be 1.31. It can therefore be assumed that with the possibility of larger correlation lengths encountered in the field (cf. the review of Hoeksema and Kitanidis 1985), the variability of flow rates estimated from both methods would be much larger than calculated here.

Figure 7 shows the CV estimated from hydraulic and tracer approaches in terms of the ratio of the correlation length to the well separation ($A/\Delta x$) or twice the radius of influence calculated for pumping test durations of 1, 2, 3, 4, and 6 d using the geometric mean of the K fields ($A/2R$). This scaling is intended to remove the sampling volume effect from the analysis of flow rate variability. For both approaches, the relatively good match between curves with different correlation lengths indicates that CV(Q) is largely determined by these ratios. For a given ratio, Figure 7 shows that pumping tests give smaller values of CV(Q) than the tracers tests. It also shows that CV(Q_T) and CV(Q_H) increase with increasing values of each of the ratios, that is, as the distance between well pairs or the radius of influence of a pumping well increases. It therefore appears reasonable to seek the smallest $A/\Delta x$ or $A/2R$ combinations in order to minimize CV(Q).

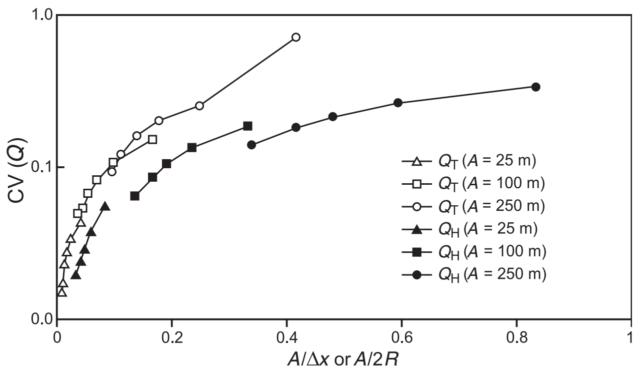


Figure 7. Flow rate variability CV(Q) as a function of the ratio between correlation length (A) and distance between wells (Δx) for radioactive tracers and radius of influence ($2R$) for pumping tests. Tracer half-life ($t_{1/2}$) is 200 years and K fields are generated for a variance (σ^2) of 1.

Discussion

Choice of Tracer

The tracers used in this study, with half-lives of 10, 50, 200, and 1000 years, are theoretical and fictitious. These different half-lives were chosen to demonstrate the effect of this parameter on estimated flow rates. In practice, only tracers with half-lives of 140 (^{32}Si), 269 (^{39}Ar), 5730 (^{14}C), 229,000 (^{81}Kr), and 301,000 years (^{36}Cl) are suitable for ground water dating in the manner described in this paper (Cook and Herczeg 2000). Of course, the use of a number of these tracers for ground water dating is limited by other processes, including sampling and analytical requirements and nonconservative behavior in some aquifers. Ground water dating is also possible with anthropogenic tracers with variable input concentrations (^{3}H/^{3}He, CFCs, SF_6). Findings from this study also have relevance for these tracers.

Results confirm that it is important to match the timescale of the tracer to the residence time of the flow system. Due to the nonlinearity of the concentration–time relationship, tracer ages overestimate the mean ground water age and the degree of overestimation becomes more significant if the tracer half-life is much shorter than the travel time between wells. On the other hand, if $t_{1/2}$ is too long, then negligible radioactive decay will occur and measurement errors will become significant. Results from this study are consistent with those of Weissmann et al. (2002), who found that the extent to which CFC-12 age underestimated ground water residence time increased with the mean residence time. However, the 100% bias reported by these authors is much larger than the one found in the current study. The 2D simulations performed here apparently underestimate the bias (see 2D vs. 3D discussion in the following text). It is also likely that the abrupt boundaries between lithological units in Weissmann et al. (2002) would lead to sharp age gradients, while in our model the smoother multi-Gaussian K fields lead to reduced age gradients that limit the effects of dispersion. Nevertheless, our simulations show that bias is strongly related to degree of heterogeneity and choice of tracer.

Comparison of Methods

Figure 4 has shown that in our model configuration, the flow estimates will have little bias if a tracer with a half-life adapted to the ground water residence time is used and if the pumping test is of sufficiently long duration. The uncertainty of ground water flow rates estimated using tracers and pumping tests will be a function of well separation and pumping test duration, respectively (Figure 5). As these increase, the ground water flow rates determined from tracer analysis or from pumping tests become less sensitive to local conditions and more representative of the aquifer as a whole. Results of simulations have also shown that for a given well separation and pumping test duration, the uncertainty in ground water flow rates estimated with both methods increases as the level of the heterogeneity (A and σ^2) increases (Figure 6). It is possible that using a much larger number of K fields might have led to a clearer distinction of CV(Q) with the two methods as heterogeneity increases. Figure 7 shows that for a given level of heterogeneity, uncertainty in flow rates estimated using the two approaches decreases as the ratio of the correlation length to the well separation ($A/\Delta x$) and the ratio of the correlation length to twice the radius of influence of the pumping test ($A/2R$) decreases.

The accurate estimation of flow rates therefore depends on our ability to conduct measurements of ground water age gradient or hydraulic conductivity at small values of $A/\Delta x$ and $A/2R$. Given that we have no control over the level of heterogeneity or the correlation length of the K field, the accuracy of the flow rate estimates will be determined by the distance between two measurement wells or by the possibility to conduct long pumping tests. In regional aquifer systems, it will generally be easier to maximize well separation than to increase pumping test duration. Consider, for example, a heterogeneous aquifer that is 50 km long, with a mean transmissivity of 100 m^2/d, a correlation length of 1000 m, and a storage coefficient of 0.01. Using these parameters to calculate the radius of influence (Equation 4), and based on the hypothesis that Figure 7 is valid for aquifers with different hydrogeological properties, we find that a pumping test would need to be conducted for 308 d (i.e., provide a radius of influence of 2630 m) to get a CV(Q_H) equal to 0.1. This pumping duration is obviously extremely long and never or rarely performed in the field. On the other hand, a CV(Q_T) of 0.1 would be easily obtained if sampling wells separated by 10 km are available. In this example, if the ground water velocity is 1 m/year, then the ground water age would increase by 1000 years for every kilometer. Given simple geochemistry, ^{14}C may provide reliable ages for water between 1 and 20 km. At shorter distances, there would be insufficient radioactive decay for reliable age dating (and the $A/\Delta x$ ratio would be large) while at greater distances the bias induced by dispersion is likely to become significant. Also, at much greater distances, tracer concentrations may approach background levels. The availability of wells in appropriate locations then becomes an issue.

Although no absolute values of well pair separations or pumping test durations can be recommended, field experiments should be designed to minimize the $A/\Delta x$ and $A/2R$ ratios. Whichever method achieves the lower ratio in a heterogeneous aquifer is likely to be more accurate. However, to quantitatively determine the accuracy of either method, field scale correlation lengths must be known. As this information is rarely available, both methods should be used jointly and implemented in such a way that they have the largest possible sampling volume. Decreased uncertainty of both hydraulic and tracer methods can also be obtained by calculating ground water ages based on many wells and by performing multiple pumping tests, and this is always to be recommended.

Model Limitations

The variability of flow rates estimated from both ground water ages and pumping tests in this study is relatively low, even with the highly heterogeneous aquifer. There are many reasons for this. Practical limitations on the number of cells in the model domain have limited the size of the correlation length (25 to 250 m) that did not reach some observed field values. Our results suggest that increasing the correlation length to larger field scale values would dramatically increase flow rate variability. The multi-Gaussian K fields (from SGSIM) used in this work are known to have low connectivity (e.g., Gomez-Hernandez and Wen 1998) and may not be representative of field conditions where connected flow channels are known to occur (e.g., Proce et al. 2004; Knudby and Carrera 2006). Using K fields with stronger connectivity might have led to larger flow rate uncertainty. The multi-Gaussian K fields are also stationary, a condition that is often not met in real aquifers. As mentioned earlier, the long-term pumping averaging behavior observed with pumping tests may not be visible in nonstationary K fields because of interconnected pathways of high K zones over long distances. Nonstationarity could also influence results from tracer analysis as tracers may also follow these connected pathways over longer distances than in a stationary K field. When stratigraphic heterogeneity varies within a deposit, the segmentation of stratigraphic successions into different units modeled separately can be used to generate a nonstationary K field, as employed by Weissmann and Fogg (1999) with the transition probability geostatistical approach.

During preliminary work, simulation tests were performed with tracers to determine if the uncertainty stabilizes when the number of K fields increases from 1 to 10. It was found that $CV(Q_T)$ reaches a plateau when five or more K fields are used. Based on the hypothesis that uncertainty converges monotonically to a plateau, using five K fields seemed a reasonable approach for this demonstration study. However, it is possible that using a much larger number of K fields might have increased flow rate uncertainty.

The synthetic aquifers used in this study represent horizontal 2D flow in confined conditions. A 3D heterogeneous flow field would lead to more tracer dispersion and hence increased bias and uncertainty in derived estimates of ground water flow (see Varni and Carrera 1998). 3D heterogeneity would also induce larger variability in K estimated from pumping tests and would therefore increase uncertainty in flow estimates. This is especially true for wells that do not fully penetrate the aquifer, which is often encountered in field studies. Anisotropy in a 2D horizontal flow aquifer is expected to impact mostly flows derived from pumping tests. Indeed, the skewness of the radius of influence induced by 2D anisotropy limits the possibility for pumping tests to include heterogeneities over a representative aquifer volume, thereby increasing flow rate uncertainty. Anisotropy in a 3D confined aquifer, corresponding for example to horizontal layering due to depositional processes, could induce large vertical variations in flow rates (cf. Dann et al. 2008), thereby increasing the uncertainty of flow estimates with tracer ages. This could also have an impact on pumping tests performed on wells that do not fully penetrate the aquifer. These are important topics that should be tested in subsequent work with synthetic aquifers.

Conclusion

This study presents the first systematic and integrated assessment of bias and uncertainty associated with the estimation of ground water flow rates using tracers and pumping tests in heterogeneous aquifer systems. In particular, the relative magnitude of bias and uncertainty of ground water flow rates from the two methods has been assessed in terms of the structure (correlation length) of the heterogeneous flow field and magnitude (variance) of the heterogeneity.

We acknowledge that the results are to some extent case specific and model dependent. Future work should examine the impact of spatial dimensionality (2D or 3D) and anisotropy as well as the effect of stationary/nonstationary K fields on the results obtained. This work nevertheless provides a demonstration of the field conditions that can lead to a reduction of bias and uncertainty with both methods. It has been shown that bias is a function of tracer half-life, pumping test duration, and heterogeneity of the flow field. Bias can be reduced by choosing a tracer with a rate of radioactive decay similar to the mean ground water residence time and by performing long-term pumping tests. Uncertainty increases with heterogeneity but is relatively small and comparable for both methods. Results from this study do not show that one method is more reliable than the other in heterogeneous conditions. However, field characterization experiments should be designed to minimize the $A/\Delta x$ and $A/2R$ ratios, that is, in such a way that they cover a maximum number of correlation lengths within their sampling volume. In practice, for regional flow systems, this will often be much easier with tracers than with pumping tests, but whichever method more satisfactorily achieves this is likely to provide more reliable and representative results. Aquifer characterization should rely on a combination of both methods whenever possible.

Acknowledgments

We thank René Therrien (Laval University) for providing the HydroGeoSphere model and Philip Brunner (Flinders University) and Ian Jolly (CSIRO-Land and Water), who reviewed a draft version of the manuscript. We also thank Darren Gooddy, Murray Close, and Gary Weissmann for their constructive comments.

References

Barlebo, H.C., M.C. Hill, and D. Rosbjerg. 2004. Investigating the macrodispersion experiment (MADE) site in Columbus, Mississippi, using a three-dimensional inverse flow and transport model. *Water Resources Research* 40, W04211, doi:10.1029/2002WR001935.

Castro, M.C., and P. Gobelet. 2005. Calculation of ground water ages—a comparative analysis. *Ground Water* 43, no. 3: 368–380.

Cook, P.G., and A.L. Herczeg. 1999. *Environmental Tracers in Subsurface Hydrology*. Boston: Kluwer Academic Press, 529.

Cook, P.G., and D.K. Solomon. 1997. Recent advances in dating young groundwater: Chlorofluorocarbons, ^{3}H/^{3}He and ^{85}Kr. *Journal of Hydrology* 191, 245–265.

Cooper, H.H. Jr., and C.E. Jacob. 1946. A generalized graphical method for evaluating formation constants and summarizing well-field history. *Eos Transactions, American Geophysical Union* 27, no. 4: 526–534.

Dann, R.L., M.E. Close, L. Pang, M.J. Flintoft, and R.P. Hector. 2008. Complementary use of tracer and pumping tests to characterize a heterogeneous channelized aquifer system in New Zealand. *Hydrogeology Journal* 16, 1177–1191.

De Marsily, G. 1986. *Quantitative Hydrogeology: Groundwater Hydrology for Engineers*. Orlando, Florida: Academic Press.

Delhomme. J.P. 1979. Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach. *Water Resources Research* 15, no. 2: 269–280.

Deutsch, C.V., and A.G. Journel. 1998. *GSLIB Geostatistical Software Library and User's Guide*, 2nd ed. New York: Oxford University Press.

Drury, L.W., G.E. Calf, and J.K. Dharmasiri. 1984. Radiocarbon dating of groundwater in tertiary sediments of the Eastern Murray Basin. *Australian Journal of Soil Research* 22, no. 4: 379–387.

Ekwurzel, B., P. Schlosser, W.M. Smethie, L.N. Plummer, E. Busenberg, R.L. Michel, R. Weppernig, and M. Stute. 1994. Dating of shallow groundwater: Comparison of the transient tracers ^{3}H/^{3}He,chlorofluorocarbons, and ^{85}Kr. *Water Resources Research* 30, no. 6: 1693–1708.

Gelhar, L.W., C. Welty, and K.R. Rehfeldt. 1992. A critical review of data on field-scale dispersion in aquifers. *Water Resources Research* 28, no. 7: 1955–1974.

Gomez-Hernandez, J.J., and X.H. Wen. 1998. To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Advances in Water Resources* 21, no. 1: 47–61.

Gomez-Hernandez, J.J., and S.M. Gorelick. 1989. Effective groundwater model parameter values: Influence of spatial variability of hydraulic conductivity, leakance and recharge. *Water Resources Research* 25, no. 3: 405–419.

Hoeksema, R.J., and P.K. Kitanidis. 1985. Analysis of spatial structure of properties of selected aquifers. *Water Resources Research* 21, no. 4: 563–572.

Knudby, C., and J. Carrera. 2006. On the use of apparent hydraulic diffusivity as an indicator of connectivity. *Journal of Hydrology* 329, 377–389.

LaBolle, E.M., G.E. Fogg, J.B. Eweis, J. Gravner, and D.G. Leaist. 2008. Isotopic fractionation by diffusion in groundwater. *Water Resources Research* 44, W07405, doi:10.1029/2006WR005264.

LaBolle, E.M., G.E. Fogg, J. Gravner, and J.B. Eweis. 2006. Diffusive fractionation of ^{3}H and ^{3}He and its impact on groundwater age estimates. *Water Resources Research* 42, W07202, doi:10.1029/2005WR004756.

Lachassagne, P., E. Ledoux, and G. De Marsily. 1989. Evaluation of hydrogeological parameters in heterogeneous porous media. In *Groundwater Management: Quantity and Quality*. Wallingford, UK: *IAHS Publication*, 188: 3–18.

LeBlanc, D.R., S.P. Garabedian, K.M. Hess, L.W. Gelhar, R.D. Quadri, K.G. Stollenwrrk, and W.W. Wood. 1991. Large-scale natural gradient tracer test in sand and gravel, Cape Cod Massachusetts. 1. Experimental design and observed tracer movement. *Water Resources Research* 27, no. 5: 895–910.

Leven, C., and P. Dietrich. 2006. What information can we get from pumping tests? Comparing pumping test configurations using sensitivity coefficients. *Journal of Hydrology* 319, 199–215.

Matheron, G. 1967. *Éléments pour une théorie des milieux poreux*. Paris: Masson.

Meier, P.M., J. Carrera, and X. Sanchèz-Villa. 1998. An evaluation of Jacob's method for the interpretation of pumping tests in heterogeneous formations. *Water Resources Research* 34, no. 5: 1011–1025.

Park, J., C.M. Bethke, T. Torgensen, and T.M. Johnson. 2002. Transport modeling applied to the interpretation of ground water ^{36}Cl age. *Water Resources Research* 38, no. 5: 1043, doi:10.1029/2001WR000394.

Pearson, F.J. Jr., and D.E. White. 1967. Carbon 14 ages and flow rates of water in Carrizo Sand, Atascosa County, Texas. *Water Resources Research* 3, no. 1: 251–261.

Phillips, F.M., M.K. Tansey, and L.A. Peeters. 1989. An isotopic investigation of groundwater in the central San Juan Basin, New Mexico: Carbon 14 dating as a basis for numerical flow modeling. *Water Resources Research* 25, no. 10: 2259–2273.

Proce, C.J., R.W. Ritzi, D.F. Dominic, and Z. Dai. 2004. Modeling multiscale heterogeneity and aquifer interconnectivity. *Ground Water* 42, no. 5: 658–670.

Sanchèz-Villa, X., P. Meier, and J. Carrera. 1999. Pumping tests in heterogeneous aquifers: An analytical study of what can be obtained from their interpretation using Jacob's method. *Water Resources Research* 35, no. 4: 943–952.

Sanchèz-Villa, X., J. Carrera, and J.P. Girardi. 1996. Scale effects in transmissivity. *Journal of Hydrology* 183, 1–22.

Schad, H., and G. Teutsch. 1994. Effects of the investigation scale on pumping test results in heterogeneous porous aquifers. *Journal of Hydrology* 159, 61–77.

Theis, C.V. 1935. The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using groundwater storage. *Transactions, American Geophysical Union* 16, 519–524.

Therrien, R., R.G. McLarren, E.A. Sudicky, and S.M. Panday. 2006. *Hydrogeosphere*. Waterloo, Canada: University of Waterloo, Groundwater Simulations Group.

Varni, M., and J. Carrera. 1998. Simulation of groundwater age distributions. *Water Resources Research* 34, no. 2: 3271–3281.

Weissmann, G.C., Y. Zhang, E.M. LaBolle, and G.E. Fogg. 2002. Dispersion of groundwater age in an alluvial aquifer system. *Water Resources Research* 38, no. 10: 16-1 to 16-8; doi:10.1029/2001/WR000907.

Weissmann, G.C., and G.E. Fogg. 1999. Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework. *Journal of Hydrology* 226, 48–65.

Zuber, A., J.J. Malecki, and M. Dulinski. 2008. Groundwater ages and altitudes of recharge areas in the Polish Tatra Mts. as determined from ^{3}H, d^{18}O and d^{2}H data. *Geological Quarterly* 52, no. 1: 71–80.