

Evaluation of alternative groundwater flow models for simulating hyporheic exchange in a small mountain stream

Steven M. Wondzell^{a,*}, Justin LaNier^{b,1}, Roy Haggerty^{b,2}

^aUSDA Forest Service, Pacific Northwest Research Station, Olympia Forestry Sciences Laboratory, 3625 93rd Ave., S.W., Olympia, WA 98512, USA

^bOregon State University, Department of Geosciences, Corvallis, OR 97331, USA

ARTICLE INFO

Article history:

Received 31 January 2008

Received in revised form 3 September 2008

Accepted 20 October 2008

Keywords:

MODFLOW

MT3D

Verification

Validation

Hyporheic zone

Tracers

SUMMARY

Groundwater flow models have been used to estimate the amount of exchange flow and the residence time distribution of stream water in the hyporheic zone. However, reliability of these predictions have not been tested. We ask the questions: (1) how reliable are hyporheic groundwater models in typical applications examining hyporheic exchange flows? and (2) how does the reliability change with increased data availability and model sophistication? We developed groundwater flow models of the hyporheic zone for a mountain stream in the HJ Andrews Experimental Forest, Oregon. The models are based on surveyed topography and hydraulic conductivity (K) measurements from both slug tests and a well-to-well tracer test. We developed several models using different methods to estimate two of the most uncertain parameters – K and the depth and shape of the bedrock boundary. We first tested the goodness of fit of each model to the water levels observed in a network of wells and piezometers. Results showed that differences among models in predicted heads were quite small, whereas differences among estimated hyporheic fluxes varied by a factor of two. We then tested the model predictions of tracer arrival times to each well in the network during a stream-tracer injection. Comparison of simulated and observed travel times showed that increased model sophistication did not lead to improved model reliability, because travel time predictions from the homogeneous model were equal to, or better than, the predictions from the heterogeneous models. While general trends in solute breakthrough were correct in the models, K data from even 37 wells in a 15 m by 50 m model domain were insufficient to characterize detailed arrival times accurately. This suggests that geomorphic data may be sufficient to predict water fluxes through the subsurface and approximate travel times. However, for detailed analysis of solute transport pathways and breakthroughs, intensive sampling of the subsurface may be necessary.

Published by Elsevier B.V.

Introduction

Over the last 15 years there have been a growing number of papers exploring hyporheic exchange using numerical groundwater flow models (Harvey and Bencala, 1993; Wondzell and Swanson, 1996; Wroblicky et al., 1998; Storey et al., 2003; Kasahara and Wondzell, 2003; Cardenas and Wilson, 2007; Cardenas et al., 2004; Cardenas, 2008). This modeling approach contrasts sharply with the more frequently used stream-centric tracer injection experiments. Stream-tracer injection data are analyzed with one-dimensional advection, dispersion, and transient storage models to examine the net effect of transient storage (including hyporheic exchange) on water and solute movement within the channel

(Bencala and Walters, 1983; Runkel, 1998). The groundwater flow modeling approach offers some advantages, allowing quantitative estimates of the amounts and residence times of water exchanged between the stream and subsurface (Harvey and Bencala, 1993; Wondzell and Swanson, 1996), a spatially explicit depiction of flow paths through the hyporheic zone (Cardenas et al., 2004; Gooseff et al., 2006; Cardenas and Wilson, 2007) and examination of the factors controlling that exchange (Storey et al., 2003; Kasahara and Wondzell, 2003; Cardenas et al., 2004).

Groundwater flow models have much more intensive data requirement than do transient storage models. Numerical groundwater flow models use a 2D or 3D discretization of the area to be simulated. The models require that several parameters be specified for each grid cell and some of these parameters are difficult to measure in the field at the scale needed for the model, the most important example being hydraulic conductivity (K). Measurements are typically point values collected at observation wells. Further, the spatial distribution of aquifer properties is often highly heterogeneous, for example, point estimates of K may range over several

* Corresponding author. Tel.: +1 360 753 7691; fax: +1 360 753 7737.
E-mail addresses: swondzell@fs.fed.us (S.M. Wondzell), justin.lanier@ch2m.com (J. LaNier), haggerty@science.oregonstate.edu (R. Haggerty).

¹ Present address: 2525 Airpark Rd., Redding CA 96001, USA. Tel.: +1 530 229 3323.

² Tel.: +1 541 737 1210; fax: +1 541 737 1200.

orders magnitude within the aquifer of interest and some method must be used to interpolate point measurements to the larger spatial domain represented by the model (Anderson and Woessner, 1992). Boundary conditions must also be specified at the margins of the model domain. However, these are sometimes poorly known, resulting in simplifying assumptions that can have large effects on model solutions (Hunt et al., 1998). Finally, linking groundwater flow models to solute transport models requires additional parameters (e.g., effective porosity and dispersivity) (Wang and Anderson, 1982; Anderson and Woessner, 1992, p. 296).

Several factors limit the ability to verify the predictions of groundwater flow models. Geologic uncertainty, the number of parameters, the difficulty in measuring them, and their interpolation to the model domain all create substantial uncertainty in model predictions (Freyberg, 1988). It is commonplace to test models by comparing predicted heads to those measured from observation wells located within the model domain. Unfortunately, because groundwater flow models suffer from nonuniqueness or equifinality (*sensu* Beven, 2006), different combinations of model parameters or even different conceptual models give acceptable fits to observed heads (Freyberg, 1988; Anderson and Woessner, 1992; Poeter and Anderson, 2005; Poeter, 2007). Thus, a reasonable fit to observed heads does not validate model predictions. Tracer breakthroughs, concentrations of naturally occurring solutes or heat, or predictions under transient conditions can be used to gain confidence in a calibrated model. Alternatively, field-measured fluxes of some component of the flow system could be used to test model predictions.

Several of these alternative approaches have been used to test predictions from groundwater flow models used to examine hyporheic exchange flows. For example, a model calibrated to steady-state conditions was tested against transient conditions during a large storm event, but verification was complicated by limited information on storage parameters and difficulties in specifying transient boundary conditions (Wondzell and Swanson, 1996). Similarly, naturally occurring differences in solute concentrations from different source waters were used to test predictions of the extent of the mixing zone between short residence time stream water in the hyporheic zone and much longer residence time water sources (Lautz and Siegel, 2006) but this was insufficient to verify flux estimates. Because little data are typically available to test models, model predictions from many studies have limited verification (Wroblicky et al., 1998; Kasahara and Wondzell, 2003; Storey et al., 2003; Kasahara and Hill, 2006). Consequently, most of these studies have used sensitivity analysis to explore factors controlling hyporheic exchange and placed reduced emphasis on the absolute magnitudes of predicted fluxes. However, one of the key benefits of using numerical models is the ability to estimate water fluxes and residence times (e.g., Saenger et al., 2005; Cardenas, 2008), and use these to estimate the impact of hyporheic exchange on biogeochemical processes occurring in stream ecosystems.

The limitations of groundwater flow models, described above, are well described in the traditional groundwater flow modeling literature, and have engendered a long standing debate over model validation (Anderson and Woessner, 1992; Konikow and Bredehoeft, 1992; Bredehoeft and Konikow, 1993; Hassan, 2004). These issues have not been widely examined in the application of numerical groundwater flow models to hyporheic studies. While the general principles are similar, hyporheic studies differ from other groundwater modeling studies. First, the spatial scale of most hyporheic applications of groundwater flow models has been small, typically including stream reaches ranging in length from 10s to 100s of meters and surficial aquifers only a few meters in depth (Harvey and Bencala, 1993; Wondzell and Swanson, 1996; Wroblicky et al., 1998; Storey et al., 2003; Gooseff et al., 2006; Lautz and Siegel, 2006; Hester and Doyle, 2008). Second, stream

water surface elevations are typically used as an internal boundary which is spatially well defined, measured with a high degree of accuracy, and located close to all grid cells within the model domain. Distance from boundaries can significantly affect the accuracy of model predictions (Rubin and Dagan, 1988), and in hyporheic studies, the distance to the stream boundary is typically quite small. Therefore, we designed our study to examine the performance of a numerical groundwater flow model when used in a typical hyporheic application.

In this paper we examine a suite of several different groundwater flow models developed to predict stream–subsurface interactions at a single site. Models were developed using different conceptualizations of the spatial distribution of K and the depth and shape of the underlying bedrock boundary. We then tested the predictions of subsurface fluxes and the exchange of water between the stream and the subsurface using detailed solute transport data collected from a relatively dense network of wells and piezometers during a stream-tracer injection (Wondzell, 2006). Specifically, we examine how predictions of hyporheic exchange flows and solute transport were effected by (1) the nodal spacing used to discretize the model domain, (2) the effect of interpolation methods used to generate spatially distributed hydraulic conductivity, and (3) the effect of the shape of the aquifer–bedrock boundary.

Methods

Study site description

This study builds on the results of earlier research into hyporheic processes (Kasahara and Wondzell, 2003; Wondzell, 2006) conducted in the lower portion of Watershed 1 (WS1) in the H. J. Andrews Experimental Forest in western Oregon (44°10'N, 122°15'W). WS1 is a small, steep-mountain stream draining a 100-ha catchment. The valley floor in the study reach of WS1 averages nearly 14 m wide and the longitudinal gradient averages 13%. Annual low flows occur at the end of the summer dry season with discharge less than 1 l/s. Base flows during the wet winter season range from 10 to 20 l/s and the flood of record generated discharge of nearly 2400 l/s.

Previous field studies at this site (WS1) used well networks to make direct observations of the water table and stream-tracer experiments to evaluate hyporheic exchange fluxes within the study reach. The well network (comprised of both wells and piezometers) was installed along a 30-m reach of the WS1 stream in the summer of 1997. Wells were made from 1- to 2-m lengths of PVC pipe “screened” over the bottom 50 cm, whereas piezometers were only screened over the bottom 5 cm. Wells and piezometers were installed as deeply as possible. Wells were located in closely-spaced transects to provide high spatial resolution of subsurface flows. Each transect typically had one piezometer located in the center of the wetted stream channel and six wells which were located on stream banks, at mid-valley floor locations and at the toe slopes of adjoining hills, on both sides of the stream (Fig. 1).

A stream-tracer experiment was conducted at low baseflow discharge ($Q = 1.2$ l/s) between 4 and 8 August, 1997 (see Wondzell, 2006 for additional details). The study reach for the tracer injection was 99.7 m, roughly centered around the well network. A concentrated solution of NaCl was injected at a constant rate for five days, at which time tracer concentrations reached a constant (or plateau) concentration at the bottom of the study reach. Tracer concentrations were measured in the middle and at the bottom of the study reach using electrical conductivity (EC) because EC was highly correlated to Cl^- ($r^2 = 0.995$, $N = 21$). Water table elevations were measured from the well networks immediately before the stream-

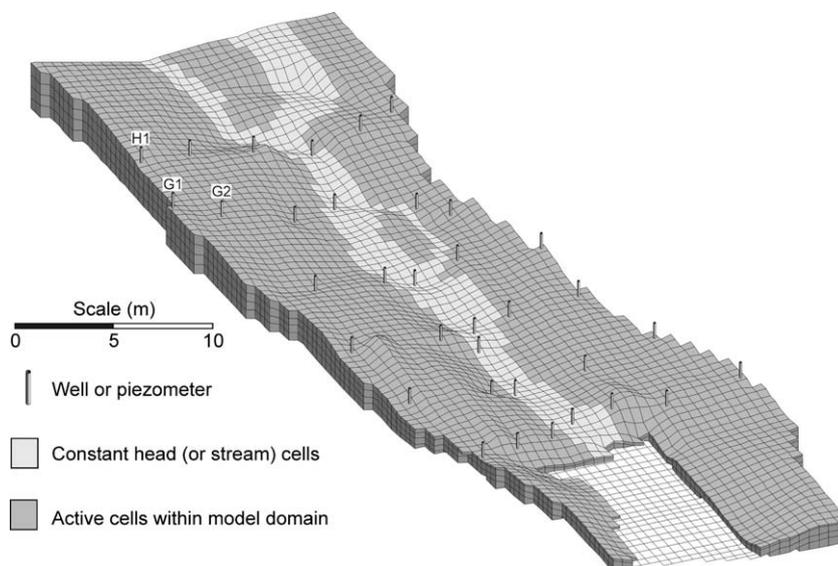


Figure 1. Discretized map of the study site, depicting the surface topography interpolated from the site surveys. The constant head cells used to simulate the stream are delineated on this topographic surface (*Note: the interpolated ground elevations do not reflect the head assigned to these cells*). The locations of both wells and piezometers are also shown as is the flat-sloping bedrock boundary which is exposed to the surface in the lower part of the study reach. The model domain shown here is the three-dimensional base model with four, variable-thickness layers, an extent of $15\text{ m} \times 50\text{ m}$, and a 0.5 m nodal spacing.

tracer experiment and the median arrival times of stream-tracers reaching each well were used to calculate median travel times of hyporheic exchange flows between the stream and each well.

The locations of all wells, primary, secondary, and back channels, and the edges of the active valley floor were surveyed and mapped to scale, thus creating a base map from which we discretized our model domain. The elevation of well heads and the ground level at each well were surveyed. Longitudinal profiles of stream channels, showing both the stream bed and stream water elevations, were also surveyed at 1-m intervals. Water table elevations were recorded from the well networks soon after surveys were completed.

Hydraulic conductivities were estimated at each well using a falling-head slug test and the Bouwer and Rice (1976) analysis method. Slug-test based estimates of K were approximately log normally distributed, with geometric mean of $7.0 \times 10^{-5}\text{ m/s}$ and $\ln K$ variance of 1.7. Slug test analyses were not completed at four wells, so these wells were assigned the mean K in all analyses. Additionally, a well-to-well tracer experiment was conducted to examine K , integrated over a larger spatial extent. A concentrated NaCl tracer solution was injected into well H1 at a constant rate for 33 h, and EC was monitored at down flow path wells (Fig. 1). Observed tracer concentrations reached plateau 26 h after the start of the injection in well G1 (located 8.6 m from the injection well) and after 29 h in well G2 (located 11.2 m from the injection well) and allowed for reliable estimates of solute transport times and K .

MODFLOW simulations

Three-dimensional groundwater flow models were built to simulate water exchange between the stream and the underlying unconfined aquifer using MODFLOW (McDonald and Harbaugh, 1988). We initially tried to reuse models developed by Kasahara and Wondzell (2003), but converting these to newer versions of GMS³ proved difficult. Therefore, we constructed new groundwater

³ Groundwater Modeling System, developed at the Environmental Modeling Research Laboratory, Brigham Young University, Provo UT, USA. Note that the use of trade or firm names in this publication is for reader information and does not imply endorsement by the US Department of Agriculture of any product or service.

flow models from the site data collected during the summer of 1997, delineating valley floor margins, location of the wetted channel, and the longitudinal profile of the channel bed and stream water surface.

All of the analyses presented here are based on a three-dimensional, four-layer base model with variable-thickness layers (Fig. 1) so that the model domain and underlying conceptual model are identical across all simulations. The land surface within the model domain is contoured from surveyed ground elevations at each well and the elevation of the streambed. We modeled stream channels using constant head cells in the uppermost model layer in the location of the wetted stream channel. Stream water elevation measured at 1-m intervals in the longitudinal survey were interpolated between points and the resulting heads were assigned to the constant head cells denoting the wetted channel. Constant head cells were also used in upstream and downstream (vertical) boundaries of the model domain. The sides of the model domain bounding adjacent hill slopes and the bedrock boundary underlying the model domain were modeled as no-flow boundaries. Model simulations were conducted only for base flow conditions when we assumed that the shape of the water table would be in equilibrium with stream water elevations. This assumption is supported by observation of the water table from the well network sites, the shape and elevation of which changed little over base flow periods. Therefore all simulations were conducted under steady-state conditions.

Grid cell size

Model domains consisted of four-layer block-center finite difference grids. A series of models were constructed with grid cells of varying size, with nodal spacings (Δx and Δy) of 2.0, 1.0, 0.5, 0.25, and 0.125 m. We also constructed one additional model with telescopic mesh refinement down the length of the model domain to provide improved spatial resolution of the wetted stream channel. Note that layer thickness (Δz) was not changed for any of these model runs, so all used a common, four-layer, variable-thickness model domain.

Effect of interpolation method

We tested four interpolation methods used to generate spatially distributed hydraulic conductivities. Because the valley floor sedi-

ments are shallow (1–2 m thick), supporting a saturated zone typically less than a meter thick, and the well screens were 50 cm in length, we could not assign discrete conductivities to separate layers within the model domain. Therefore, the aquifer was modeled as isotropic and vertically homogeneous. For the homogeneous model, K was set equal to the geometric mean of K (7.0×10^{-5} m/s).

We also constructed three heterogeneous models designed to better simulate the spatial variability in K , because estimates of K from slug tests in individual wells ranged widely (4.3×10^{-6} to 6.1×10^{-4} m/s). We examined two conceptual models using the Thiessen polygon (nearest neighbor) method to interpolate point measurements of K to the model domain, a technique that was used in a previous modeling study at this site (Kasahara and Wondzell, 2003). The first model used discrete zones of heterogeneity within Thiessen polygons drawn around each observation well (or piezometer) with the K of each zone set equal to the measured value. We also developed an “optimized” Thiessen polygon model in which K was iteratively refined through a trial and error process to minimize the error in the model fit to the heads at observation wells. During this process we maintained relatively large polygons in which K was homogeneous, rather than trying to adjust values of just a few cells surrounding individual wells. Lastly, kriging was used to interpolate point values of K to the entire model domain. A Gaussian model showed the best fit to the observed data and suggested that K had a correlation length of approximately 4 m within our study site.

The undifferentiated and unsorted sediment in the hyporheic zone of this mountain stream was deposited by an unknown number of debris flows, and the only sedimentary distinction we were able to make was between bedrock and alluvium. Consequently, none of the interpolation methods we used incorporates geologic or structural information within the alluvium because the spatial scale of the model is smaller than the scale at which we can differentiate geological processes. In many other settings – e.g., meandering or braided streams – it should be possible to incorporate a more process-based approach such as those suggested by Koltermann and Gorelick (1996) or Weissmann et al. (1999).

Sediment depth and shape of bedrock boundary

We tested three methods of characterizing the sediment depth and shape of bedrock boundary underlying valley floor sediment. In all cases, this boundary was treated as a no-flow boundary. Our simplest characterization was an inclined boundary with longitudinal gradient equal to the reach-averaged longitudinal gradient of the stream, with zero gradient across the valley, and with sediment depth equal to the average depth of penetration of all wells and piezometers ($N = 41$; average depth = 1.03 m; max = 1.73 m; min = 0.65 m). Our second characterization used a constant depth model, in which the valley floor sediment was given a constant depth equal to the average refusal depth of all wells

and piezometers, thus creating a bedrock boundary that exactly mirrored the topography of the surface of the flood plain. The third characterization used a contoured bedrock boundary, where the shape of the boundary was contoured from the refusal depth of each well and piezometer. When installing wells, anomalously shallow wells were removed and reinstalled in nearby locations to minimize the chance that well penetration was limited by buried boulders. Thus, the refusal depth was assumed to represent the underlying bedrock boundary.

Model simulations and analyses

There are an infinite number of possible combinations of nodal spacings, interpolation methods, and shapes and depths of the bedrock boundary that could be used to simulate our study site. To keep the number of models we developed manageable, we first selected a nodal spacing (Δx and Δy) to be used in all subsequent model runs so that comparisons among models would not be compounded by changes in nodal spacing. We then tested the effect of different interpolation methods used to distribute K throughout the model domain by constructing four models (Table 1; A1 through A4), all with the flat-sloping bedrock boundary. We then examined the effect of different depths and shapes of the bedrock boundary by comparing among three models (Table 1, B1 through B3), all with a homogeneous distribution of K . In all comparisons, models were tested for the goodness of fit by comparing predicted heads with those observed in the network of observation wells.

We then tested the ability of each model to predict tracer arrival times from the stream to wells located throughout the study domain. We used MT3D to simulate solute flux from the stream to each well or piezometer, and tested model fits by comparing predicted median arrival times to those observed in the network of wells during a stream solute injection. We did not want to condition the MT3D simulations by calibrating effective porosity and dispersivity to the stream-tracer injection data. Instead, we simulated a well-to-well tracer experiment with two models (A1/B1 and A2), using a range of potential parameter values for effective porosity and dispersivity. We examined a reasonable range of effective porosities for poorly sorted alluvial mixtures ($n_e = 0.10, 0.15,$ and 0.20), as given in Domenico and Schwartz (1990, p. 26). Similarly, we examined a reasonable range of possible dispersivities ($\alpha_L = 0.10, 0.20, 0.25, 0.30,$ and 0.35 m), given the length (8.6–11.2 m) scale of the well tracer injection experiment (Neuman, 1990), but focusing on the lower portion of the range of dispersivities observed in field studies as recommended by Gelhar et al. (1992). Models were run using various combinations of these parameters, and their fit to the observed arrival times in observation wells was evaluated using the RMSE of the times to 10%, 50%, and 90% of plateau concentrations. The combination of parameters with the lowest RMSE in the two observation wells ($\alpha_L = 0.25$ m; $n_e = 0.20$) was used in subsequent MT3D simulations of the stream-tracer injection. We also used data from the well-to-

Table 1

Comparison among models. Mean Abs. Error denotes the mean of the absolute values of the differences between the observed and predicted heads in the network of wells and piezometers. Q_{HEF} denotes the magnitude of the hyporheic exchange flux which is estimated as the sum of the flow from all constant head stream cells into the model domain, expressed in units of discharge.

Model ID	K interpolation	Bedrock boundary	Mean error (m)	Mean abs. error (m)	RMSE (m)	Q_{HEF} (l/s)
A1/B1	Homogeneous	Flat	0.09	0.14	0.17	0.71
A2	Thiessen	Flat	0.08	0.11	0.14	1.12
A3	Kriged	Flat	0.06	0.11	0.15	1.30
A4	Thiessen optimized	Flat	0.04	0.09	0.12	1.05
A1/B1 ^a	Homogeneous	Flat	0.09	0.14	0.17	0.71
B2	Homogeneous	Contoured	0.09	0.15	0.19	0.71
B3	Homogeneous	Constant depth	0.09	0.12	0.16	0.69

^a Note that the data for the A1/B1 model is repeated in row five to facilitate comparisons within the group of eight models examining the depth and shape of the bedrock boundary.

well tracer test to examine the ratio of transverse to longitudinal dispersivity. The two models (A1/B1 and A2) were run using a wide range of ratios (0.01, 0.10, 0.50, 0.75, 1.00) and evaluated for the RMSE of the combined travel times to 10%, 50%, and 90% of plateau concentrations. Model fits changed little over this range. Consequently, we set horizontal transverse dispersivity to 1/10 and vertical transverse dispersivity to 1/100 the longitudinal dispersivity following Zheng and Bennett (1995, p. 249).

Results

Effects of grid cell size

The size of the nodal spacing used to discretize the model domain had a substantial influence on the estimated magnitude of hyporheic exchange fluxes. Over the range of nodal spacings examined here ($2 \geq \Delta x, \Delta y \leq 0.125$ m) we observed a nearly linear increase in the estimated hyporheic exchange fluxes (Fig. 2). As no clear threshold was apparent, we did not have a clear guideline for selecting the most appropriate nodal spacing for our model investigations. However, previous studies (Wondzell, 2006) showed that turnover lengths of stream water with the hyporheic zone were less than 100 m in this stream during summer baseflow. As such, the estimated amount of hyporheic exchange flow (Q_{HEF}) from the model with 2.0 m nodal spacing would not be consistent with the observed turnover length. Large (2×2 m and 1×1 m) nodal spacings poorly represented the size and shape of the stream which was surveyed with points spaced 1 m apart with the detailed shape of the boundary of the wetted stream channel hand-drawn on a scale map produced from the survey data. Conversely, the very fine-scale discretizations (0.25×0.25 m or 0.125×0.125 m) provided a good representation of the outlines of the wetted channel as mapped. However, the fine-scale discretizations resulted in models with many cells which could substantially delay solutions, especially for the MT3D solute transport simulations. Therefore we selected a nodal spacing ($\Delta x, \Delta y$) of 0.5 m as a parsimonious compromise.

Effect of interpolation method and the depth and shape of bedrock boundary

We compared four interpolation methods using our 3D base model with a flat-sloping bedrock boundary. Differences in the model fits to the heads observed in the network of wells and piezometers were very small. For example, the mean of the absolute

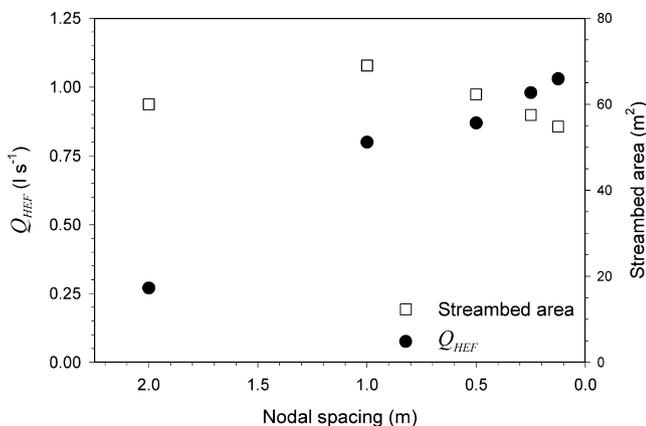


Figure 2. The effect of nodal spacing on streambed area and predicted magnitude of Q_{HEF} over the study reach. Q_{HEF} denotes the magnitude of the hyporheic exchange flux, expressed in units of discharge.

values of the errors ranged between 9 and 14 cm (Table 1). We estimated hyporheic exchange from MODFLOW's flow budget by summing the total amount of water flowing out of constant head "stream cells" and into adjacent grid cells. Despite close similarities in model fits to the observed head data, models with more complicated spatial distributions of K also had higher estimates of hyporheic exchange flows (Table 1). The Q_{HEF} was lowest in the homogeneous model (0.7 l/s) and highest in the kriged model (1.3 l/s).

We compared three methods for characterizing the boundary between that alluvial aquifer and the underlying bedrock boundary which we treated as a no-flow boundary. These comparisons were made using the 3D base model with homogeneous and isotropic sediment. Model results appeared to be insensitive to the method used to characterize the bedrock boundary because there were only minor differences in model fits between predicted and observed heads and in the estimated magnitude of hyporheic exchange (Table 1).

Overall, our simplest model (homogeneous K , isotropic, flat-sloping bedrock boundary) resulted in an acceptably good model fit between the predicted and observed heads. The homogeneous models resulted in relatively small estimates of the magnitude of hyporheic exchange flows, probably because they eliminated zones with above average hydraulic conductivity. Using more complicated spatial distributions of hydraulic conductivity within the model domain improved the fit between predicted and observed heads. However, model improvements were small relative to the increased degree of model complexity.

Solute transport simulations

We examined the ability of the six ground water flow models to simulate the flux of a conservative tracer from a 5-day stream injection to each of 35 observation wells with measurable tracer break through located in the model domain (Fig. 3). We used salt (NaCl) as our tracer, made periodic measurements of electrical conductivity (EC) in the observation wells, and analyzed the time-series data to estimate the median travel time required for the tracer to reach each observation well. All the models simulated median travel times to approximately one-third of the observation wells with reasonably small errors (Fig. 3). For example, examination of all six models shows that predicted median travel times fall

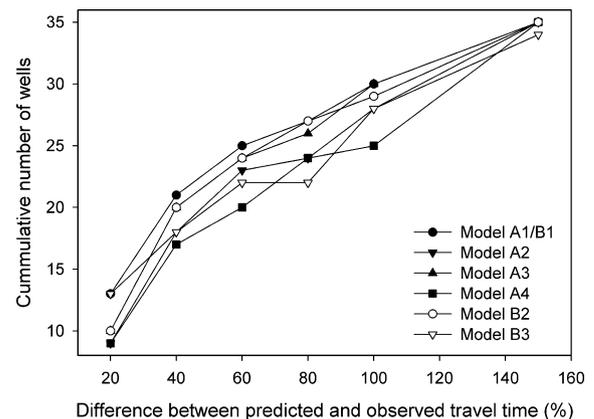


Figure 3. Comparison of predicted versus observed median travel times for each model. Data are plotted as a cumulative frequency distribution. The x-axis shows uniform-width bins in which the predicted median travel time was equal to, or smaller than, the observed median travel time, expressed as a percentage of the observed median transport time. The y-axis shows the cumulative number of wells (or piezometers) out of a total of 35. For example, predicted median travel times to 25 out of 35 wells for model A1/B1 differed by 60%, or less, from the observed travel time. Model IDs refer to the models shown in Table 1.

within $\pm 20\%$ of the observed travel times for a minimum of 9 out of 35 wells for the worst performing model and a maximum of 13 out of 35 wells for the best performing model. Differences among models were more apparent in the number of wells where simulated travel times were substantially in error. The best model (Model A1/B1, Fig. 3) simulated median travel times that were more than 100% in error at only 5 of 35 observation wells (the estimated time was less than $\frac{1}{2}$ the observed time or more than two-times the observed time). The worst model (Model A4) had simulated travel times that were more than 100% in error at 10 wells (Fig. 3).

Several results stand out. First, the simplest model, A1/B1, with homogeneous isotropic sediment and with a flat-sloping bedrock boundary had the overall best fit to the tracer-travel-time data (Fig. 3). Two other models using additional information about the study site – one using a kriged spatial distribution of K (model A3) and one using a contoured bedrock contact (model B2) – fit the observed data nearly as well as model A1/B1. But clearly, using additional data to develop a more complex model did not result in a substantial improvement in the simulated median travel times. Lastly, model A4 which we iteratively refined the spatial distribution of K via a hand-fitting procedure to minimize the error between the predicted and observed heads resulted in the worst fit to the observed travel time data (Fig. 3) despite having the best overall fit to the observed heads (Table 1).

Further analysis of model A1/B1 – the best fitting model

The simulated patterns of tracer transport through the hyporheic zone of the WS1 study site show broad general agreement with the patterns observed during the stream-tracer test (Fig. 4). Agreement between the simulated and observed movement of tracer was especially good in the upper half the model domain in a

zone of complex channel patterns caused by log jams both immediately above and in the upper part of the study reach. Differences between the simulated and observed tracer movement are especially notable around log steps in the lower half of the study reach. The observed data showed extensive penetration of tracer into the hyporheic zone above the log step 26 h after the start of the tracer injection. In contrast, the model simulations showed rapid arrival of tracer at piezometer DE4, immediately below the log step (white arrow, Fig. 4). The model also simulated more extensive movement of tracer into the lower floodplain on the left side of the stream (facing downstream) than was observed during the tracer test.

The magnitude of errors in the median travel times simulated with the A1/B1 model did not appear to be related to the error in the simulated heads or the distance from the stream to the well. Inaccuracies in simulated heads should influence local head gradients that control the direction and velocity of water and conservative tracer movement through the aquifer. However, relative magnitude of the error in the head predicted at each well did not appear related to the errors in the simulated travel times (Fig. 5). Similarly, the location of the observation well relative to the channel did not appear related to magnitude of errors in the simulated travel times (Fig. 6). We expected that the MT3D simulations would accurately predict median travel times to stream channel piezometers and stream bank wells where the distance to the stream boundary and the source of tracer would be relatively short. However, median travel times to observation wells located further from the channel were equally well predicted. The only overall trend in the simulated travel times to the observation wells is a tendency for the simulated solute transport times in the A1/B1 model to be longer than the observed times. Even so, errors at most observation wells (30 of 35 wells) fall within 100% of the observed value.

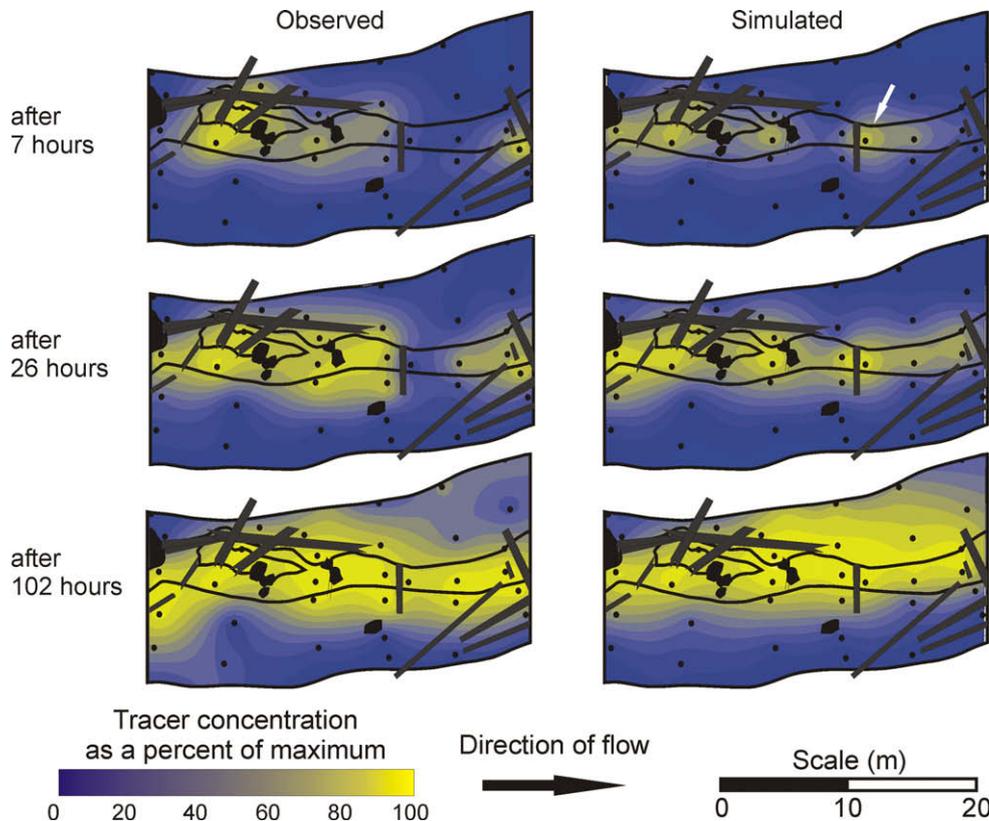


Figure 4. Contour map of E.C. used as a surrogate to measure concentration of NaCl tracer at 7, 26, and 102 h after the start of the stream-tracer injection (observed) and the movement of a solute through the model domain simulated by model A1/B1 (simulated). The white arrow shows the location of Piezometer DE4.

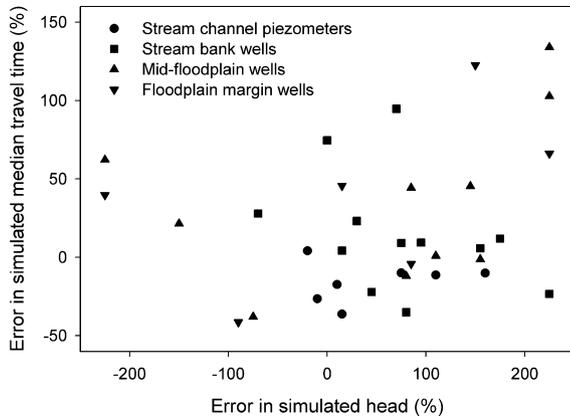


Figure 5. The relationship between the errors in the heads and median arrival times simulated in the A1/B1 model. Symbols are coded to show location of the wells.

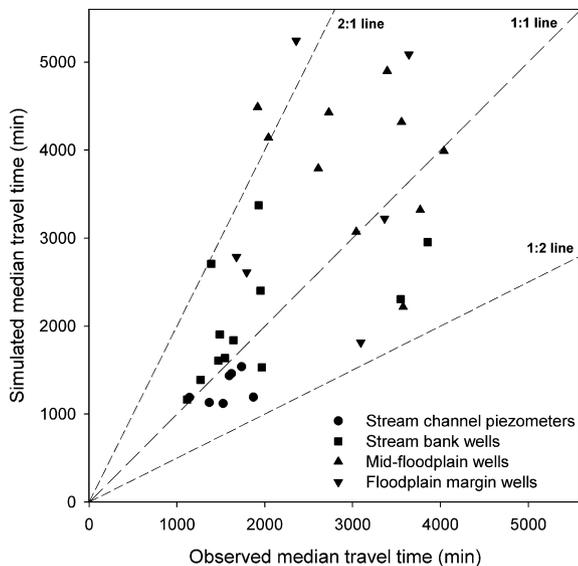


Figure 6. Simulated median travel times from the stream to individual wells ($N = 35$) from a MT3D simulation based on the A1/B1 MODFLOW model compared to the median travel times observed during a stream-tracer injection in 1997. Symbols are coded to show location of the wells and dashed lines indicate the relative error in the predicted travel times.

Comparisons between the simulated stream water fluxes into the hyporheic zone (negative Q_{HEF} , or downwelling) or return flows from the hyporheic zone to the stream (positive Q_{HEF} , or upwelling) and the vertical hydraulic gradients observed in stream channel piezometers shows reasonable agreement (Fig. 7). In general, the model simulates downwelling zones upstream of steps in the stream profile and upwelling zones just below such steps. The VHG observed in two piezometers, E4 and H4 (Fig. 7) do not agree with model simulations. A small error in locating piezometer E4 during discretization could explain the discrepancy between the observed and simulated values, as adjacent model rows show strong downwelling. Piezometer H4, however, is located in a large zone simulated with relatively neutral exchange flux, despite strongly negative VHG observed in that piezometer.

Model A1/B1 simulated a net 0.23 l/s increase in stream discharge over the stream reach contained within the model domain. Although we do not have measurements of discharge across the same reach, our stream-tracer injection suggested that stream discharge increased by 0.41 l/s over a 99.7-m long reach containing the model domain (Wondzell, 2006). Our most downstream dis-

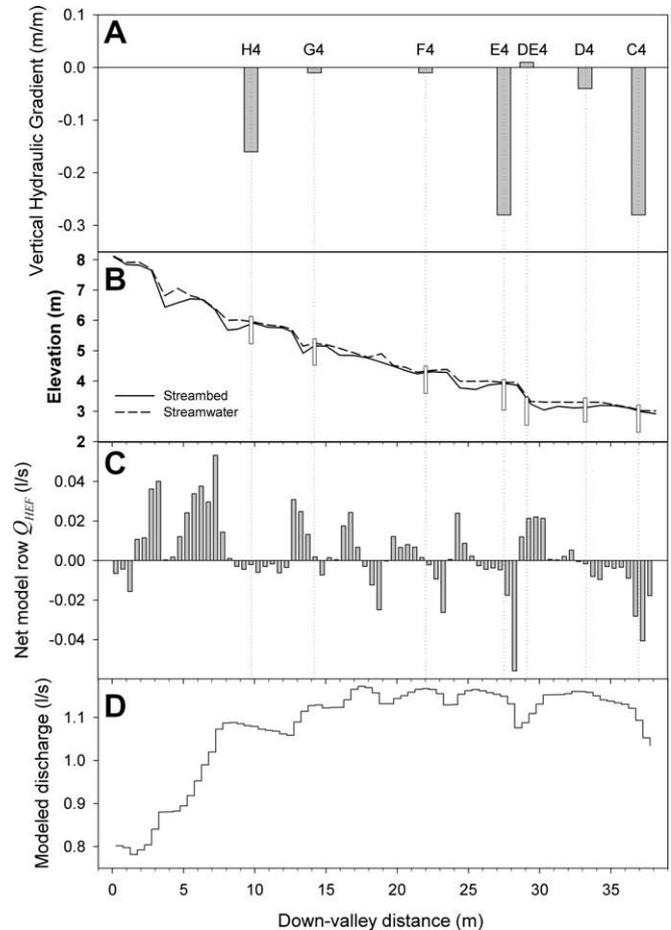


Figure 7. Comparison among observed and simulated parameters for Model A1/B1: (A) vertical hydraulic gradient observed in streambed piezometers during low baseflow, (B) longitudinal profile of the stream bed and water surface elevations, with locations and depths of piezometers shown to scale, (C) net hyporheic exchange in each model row (multiple stream cells), and (D) net change in discharge over the modeled stream reach.

charge measurement site during the stream-tracer experiment was approximately 20 m downstream of the model domain, and measurements were made in a notch where the entire valley width was scoured to bedrock, thereby collecting all stream, hyporheic, and groundwater flows moving down the valley. The simulated changes in stream discharge did not include inflows to the stream from model cells located on the downstream portion of the model domain where the valley floor is scoured to bedrock (Fig. 1). Therefore, the simulated 0.23 l/s increase in discharge over the model domain appears reasonable.

Most of the simulated increase in stream discharge occurs within the first 8 m of the model domain (Fig. 7). Although we did not recognize it at the time we established our study site and well network, the upper portion of the model domain is located just below an old, large log jam that creates a substantial step in the longitudinal profile of the stream and forces a braid in the stream channel. The stream channel often goes dry, upstream of the log jam, in the late afternoon and early evening during summer low baseflow – the conditions for which we calibrated our models. Given these observations, it seems reasonable that the upper portion of the model domain would be a major zone of hyporheic return flows to the stream. In retrospect, we also note that this was a poor choice of location in which to establish the boundary of the model domain as steep gradients in head across the boundary are likely to add substantial uncertainty to our specified boundary conditions.

Discussion

Effect of grid cell size

Numerical ground water flow models, like MODFLOW, estimate the net flux into and out of individual cells when calculating the flow budget. In hyporheic applications, individual cells may encompass zones where both upwelling and downwelling occur. Because only the net flux can be calculated, which is based on a smoothed version of true topography, some flows are not included in flux estimates. At smaller nodal spacings, improved spatial resolution captures these local flows as discrete components of the flow budget. Thus as nodal spacing decreases estimated, fluxes across boundaries tend to increase (Anderson and Woessner, 1992, p. 153). The ideal nodal spacing for hyporheic studies would be slightly smaller than the smallest scale needed to reasonably represent the channel morphologic features driving hyporheic exchange flow. Our results suggest that hyporheic exchange increases linearly with decreasing nodal spacing and increased resolution of channel morphology. Our choice of a 0.5 m nodal spacing results in estimated Q_{HEF} only 20% smaller than that from a model with 0.125 m nodal spacing (Fig. 2), a relatively small error given the differences in Q_{HEF} observed among methods of distributing K to the model domain. However, this linear relationship may not be accurate. Wörman et al. (2007) presented evidence that subsurface-surface fluxes are concentrated in the smallest topographic features and that these are fractal. If Wörman is correct, Q_{HEF} may be scale-dependent and larger Q_{HEF} would always be expected with smaller node spacing if that smaller node spacing comes with increasing topographic resolution.

We did not explore the influence of vertical nodal spacing on estimates of Q_{HEF} . While subsurface flow paths through much of the model domain are dominantly horizontal, strong vertical head gradients exist in many locations near the stream (Fig. 7). In these locations, limiting the model to only four layers may have provided insufficient resolution to simulate vertical fluxes between the stream and the subsurface. Our results showed that horizontal nodal spacing had a substantial influence on the simulated amount Q_{HEF} . Vertical nodal spacing is likely to have a substantial influence on simulated hyporheic exchange and would be an important topic for analysis in future efforts to model hyporheic exchange processes.

Evaluation of model fits

We expect that any of the six models examined in this study would have been judged as providing an acceptable fit to head data, given the small errors between the predicted and observed head distributions. For example, RMSE ranged from 0.12 m to 0.19 m and the mean absolute value of the errors ranged from 0.09 m to 0.15 m. These error ranges are quite small considering that the range of observed water table elevations was more than 4 m (Well H6, 6.523 m; Well C2, 2.502 m) over a horizontal distance of only 27 m. Similarly, stream gradients are quite steep, dropping 2.931 m in elevation over this same distance.

Our simplest model (homogeneous K , isotropic, flat-sloping bedrock boundary) resulted in an acceptably good model fit between the predicted and observed heads. Constructing models with more complicated spatial distributions of K within the model domain improved the fit between predicted and observed heads. However, model improvements were small relative to the increased model complexity.

Only model A4 was hand-fit through an iterative trial and error refinement procedure to optimize the fit between the predicted and observed heads. In this case, changes in the spatial distribution

of hydraulic conductivity were applied to individual Thiessen polygons to avoid a model solution achieved by local-scale refinement around individual wells that would not be parsimonious with the density of observed data and that would be insensitive to flows through the larger model domain (Freyberg, 1988).

Even though we developed heterogeneous models, our observation data were sparse, and did not include any information on vertical changes in sediment properties. Thus it is likely that fine-scale details of flows and tracer transport through the model were influenced by heterogeneity of the porous media that was too fine to be observed in our well network and that could not be simulated by the model, even at very small nodal spacing.

The suite of models we constructed with homogeneous K resulted in relatively small estimates of the magnitude of hyporheic exchange flows, possibly because they eliminated zones with above average hydraulic conductivity. The models we constructed with more complicated spatial distributions of K had higher estimates of hyporheic exchange flows.

Overall, the method used to characterize the bedrock boundary had very little effect on model fits and estimates of hyporheic exchange flows. Heads are less sensitive to the distance to an impermeable than they are to the distance to a constant head boundary (cf. Rubin and Dagan, 1988, 1989; also see Oliver and Christakos, 1996). This may be very good news for hyporheic flow modeling, because the lower boundary is the most difficult and expensive to characterize accurately.

Effect of boundary conditions in hyporheic investigations

Hyporheic investigations provide an example of a rare luxury in groundwater studies: highly constrained boundary conditions that are close to the modeling domain of interest. Stream water elevations provide a constant head boundary to the groundwater model, and this is measurable by survey. Furthermore, these surveyed locations are generally very close to the part of the model domain where we wish to estimate hyporheic flux.

Rubin and Dagan (1988) showed that the accuracy of model heads decrease with distance from the boundary normalized by correlation scale of $\ln(K)$. Within one correlation scale, the accuracy of the model is approximately five times that at 10 correlation scales from a boundary. The correlation scale of $\ln(K)$ within our well field was approximately 5 m. Most of our model domain is within 5 m of the stream, and all of it is within 10 m. Consequently, the accuracy of the head field is improved significantly by the proximity of the boundary conditions.

Our findings are probably typical for groundwater models used to examine hyporheic exchange flows in small streams. Groundwater flow modeling studies of hyporheic systems reported in the literature are typically focused on small systems, often in areas with relatively narrow valley floors so that distances to lateral boundaries are short. This results in tightly bounded systems in which models will generally produce good fits to the observed head data provided that the correlation scale is large relative to the size of the system.

Accuracy of modeled hyporheic discharge

The six models in the study provide a fairly well-constrained estimate of hyporheic discharge. We base this on the following arguments. First, as discussed in the previous sub-section, the models were provided with highly accurate and nearby boundary conditions. Second, spatially averaged hydraulic conductivity values are generally sufficient for prediction of groundwater flow (Wang and Anderson, 1982; Konikow and Mercer, 1988; Anderson and Woessner, 1992). Third, we believe that our estimate of mean K is accurate. Our slug test estimates of hydraulic conductivity ap-

pear quite robust as the average K estimated from tracer travel times measured in a well-to-well tracer injection (4.52×10^{-5} m/s) agreed very well with the slug-test based estimate of K averaged from the wells used in the well-to-well tracer test (6.66×10^{-5} m/s). Fourth, the model estimate of hyporheic discharge agrees with data for net change in stream discharge estimated from the stream-tracer experiment. Fifth, our results agree with other recent work (e.g., Boano et al., 2006; Cardenas and Wilson, 2007; Cardenas, 2008) showing that the morphology of the stream-hyporheic boundary is a key to flux prediction. Consequently, we believe that our models provide reasonable constraints on hyporheic discharge.

It should be pointed out that while good fits to head data are to be expected in small streams, accurate estimates of hyporheic discharge may not always be expected. Because of the highly constrained boundary conditions, the accuracy of the hyporheic discharge estimate will largely be controlled by the accuracy of the estimate of mean hydraulic conductivity. Hydraulic conductivity is difficult to measure at scales that are representative of the field, especially with slug tests in shallow, small diameter wells. Wherever possible, measurements of K should be backed up with multiple sources of data (e.g., slug tests and tracer tests).

Other thoughts on applying groundwater flow models to hyporheic investigations

Konikow and Bredehoeft (1992), in their classic paper stated that “ground water models cannot be validated”. More recently, Beven (2006) returned to this discussion, focusing on the equifinality of numerical models – that is to say that many different conceptualizations of a numerical groundwater flow model may, for a variety of reasons, fit the observed data equally well. The results presented here support these views. We developed six models, any one of which is a reasonable conceptual model for our field site and all six models had acceptable fits to the head data collected from a network of observation wells. Even when tested against independently collected solute transport data, differences among models tended to be small and at least three different conceptual models provided equally good predictions of solute transport. Unfortunately, there was no relation between the model fits to the observed heads and the model fits to the observed median solute transport times. In fact, the model with the best fit to the head data resulted in the worst prediction of travel times; the model with the best fit to the travel times had relatively mediocre fits to the head data in this group of six models.

One of the primary advantages to using groundwater flow models to explore the hyporheic zone is that these models can provide quantitative estimates of both the amounts and residence times of water exchanged between the stream channel and the subsurface. As our results show, we were unable to identify an optimal conceptual model. And while our estimates of Q_{HEF} are reasonably constrained, the range in predicted Q_{HEF} among acceptable models is relatively large. For example, the predictions of hyporheic exchange flows from the three models that best fit the solute transport data varied by nearly a factor of two, from a minimum of 0.7 l/s to a maximum of 1.3 l/s. This range in Q_{HEF} , simulated within a 50-m long stream reach, is large for predicting the effect solute transformations and other biogeochemical processes on stream water quality, considering that stream discharge was only 0.9 l/s.

Our results show that the properties of interest (Q_{HEF} , residence time distribution, the flow net) predicted by the groundwater flow models are strongly influenced by the modelers choice of conceptual model, nodal spacing, and methods used to interpolate spatially distributed parameters to the model domain. We agree with Konikow and Bredehoeft (1992) and Beven (2006) – as long as there is no way to identify a unique model solution and invali-

date all other solutions, then the uses made of groundwater flow models must be constrained by their inherent uncertainty.

Critical questions remain about basic hyporheic hydrology: What is the relative magnitude of hyporheic exchange flows in any given stream? What is the residence time distribution of stream water in the hyporheic zone? What are the physical factors driving hyporheic exchange flows? Poeter (2007) and others have said, “All models are wrong; some models are useful.” The results presented here suggest that groundwater flow models can be used to provide rough answers to some hyporheic questions, allowing better understanding of the hyporheic zone and a thorough analysis of the physical factors controlling the development of the hyporheic zone (Kasahara and Wondzell, 2003; Storey et al., 2003; Cardenas et al., 2004; Cardenas and Wilson, 2007; Cardenas, 2008). Our results, however, suggest that the models developed here would not be sufficiently accurate to predict the movement of solutes through the hyporheic zone.

Acknowledgements

We thank Ryan Ulrich, Bryan McFadin and especially Tamao Kasahara for help with field work. This work was supported by the National Science Foundation’s Hydrologic Sciences Program (EAR-9506669 and EAR-9909564), the H.J. Andrews Long-Term Ecological Research Program (DEB-9632921), and by the U.S.D.A. Forest Service, Pacific Northwest Research Station’s Aquatic and Land Interactions Program.

References

- Anderson, M.P., Woessner, W.W., 1992. Applied Groundwater Modeling: Simulation of Flow and Advective Transport. Academic Press Inc., Harcourt Brace Jovanovich Publishers, New York, NY. p. 381.
- Bencala, K.E., Walters, R.A., 1983. Simulation of solute transport in a mountain pool-and-riffle stream: a transient storage model. *Water Resources Research* 19, 718–724.
- Beven, K., 2006. A manifesto for the equifinality thesis. *Journal of Hydrology* 320, 18–36.
- Boano, F., Camporeale, C., Revelli, R., Ridolfi, L., 2006. Sinuosity-driven hyporheic exchange in meandering rivers. *Geophysical Research Letters* 33, L18406. doi:10.1029/2006GL027630.
- Bouwer, H., Rice, R.C., 1976. A slug test for determining hydraulic conductivities of unconfined aquifers with completely or partially penetrating wells. *Water Resources Research* 12, 423–428.
- Bredehoeft, J.D., Konikow, L.F., 1993. Ground-water models: validate or invalidate. *Ground Water* 31, 178–179.
- Cardenas, M.B., 2008. Surface-ground water interface geomorphology leads to scaling of residence times. *Geophysical Research Letters* 35, L08402. doi:10.1029/2008GL033753.
- Cardenas, M.B., Wilson, J.L., 2007. Dunes turbulent eddies and interfacial exchange with permeable sediments. *Water Resources Research* 43, W08412. doi:10.1029/2006WR005787.
- Cardenas, M.B., Wilson, J.L., Zlotnik, V.A., 2004. Impact of heterogeneity bed forms and stream curvature on subchannel hyporheic exchange. *Water Resources Research* 40. doi:10.1029/2004WR003008.
- Domenico, P.A., Schwartz, F.W., 1990. Physical and Chemical Hydrogeology. John Wiley and Sons, NY. p. 824.
- Freyberg, D.L., 1988. An exercise in ground-water model calibration and prediction. *Ground Water* 26, 350–360.
- Gelhar, L., Welty, C., Rehfeldt, K., 1992. A critical review of data on field-scale dispersion in aquifers. *Water Resources Research* 28, 1955–1974.
- Gooseff, M.N., Anderson, J.K., Wondzell, S.M., LaNier, J., Haggerty, R., 2006. A modeling study of hyporheic exchange pattern and sequence, size, and spacing of stream bedforms in mountain stream networks. *Hydrological Processes* 20, 2443–2457.
- Harvey, J.W., Bencala, K.E., 1993. The effect of streambed topography on surface-subsurface water exchange in mountain catchments. *Water Resources Research* 29, 89–98.
- Hassan, A.E., 2004. Validation of numerical ground water models used to guide decision making. *Ground water* 42, 277–290.
- Hester, E.T., Doyle, M.W., 2008. In-stream geomorphic structures as drivers of hyporheic exchange. *Water Resources Research* 44, W03417. doi:10.1029/2006WR005810.
- Hunt, R.J., Anderson, M.P., Kelson, V.A., 1998. Improving a complex finite-difference ground water flow model through the use of an analytic element screening model. *Ground water* 36, 1011–1017.

- Kasahara, T., Hill, A.R., 2006. Hyporheic exchange flows induced by constructed riffles and steps in lowland streams in southern Ontario, Canada. *Hydrological Processes* 20, 4287–4305.
- Kasahara, T., Wondzell, S.M., 2003. Geomorphic controls on hyporheic exchange flow in mountain streams. *Water Resources Research* 39 (1), 1005. doi:10.1029/2002WR001386.
- Koltermann, C.E., Gorelick, S.M., 1996. Heterogeneity in sedimentary deposits: a review of structure-imitating, process-imitating, and descriptive approaches. *Water Resources Research* 32 (9), 2617–2658.
- Konikow, L.F., Bredehoeft, J.D., 1992. Ground water models cannot be validated. *Advances in Water Resources* 15, 75–83.
- Konikow, L.F., Mercer, J.W., 1988. Groundwater flow and transport modeling. *Journal of Hydrology* 100, 379–409.
- Lautz, L.K., Siegel, D.I., 2006. Modeling surface and ground water mixing in the hyporheic zone using MODFLOW and MT3D. *Advances in Water Resources* 29, 1618–1633.
- McDonald, M.G., Harbaugh, A.W., 1988. A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model. *Techniques of Water Resources Investigations of the United States Geological Survey*. US Department of the Interior, US Geological Survey, Denver, Colorado (Book 6, Chapter A1).
- Neuman, S., 1990. Universal scaling of hydraulic conductivities and dispersivities in geologic media. *Water Resources Research* 26, 1749–1758.
- Oliver, L.D., Christakos, G., 1996. Boundary condition sensitivity analysis of the stochastic flow equation. *Advances in Water Resources* 19 (2), 109–120.
- Poeter, E., 2007. All models are wrong, how do we know which are useful? *Ground Water* 45, 390–391.
- Poeter, E., Anderson, D., 2005. Multimodel ranking and inference in ground water modeling. *Ground Water* 43, 597–605.
- Rubin, Y., Dagan, G., 1988. Stochastic analysis of boundaries effects on head spatial variability in heterogeneous aquifers, 1. Constant head boundary. *Water Resources Research* 24, 1689–1697.
- Rubin, Y., Dagan, G., 1989. Stochastic analysis of boundaries effects on head spatial variability in heterogeneous aquifers, 2. Impervious boundaries. *Water Resources Research* 25 (4), 707–712.
- Runkel, R.L., 1998. One-dimensional transport with inflow and storage (OTIS): a solute transport model for streams and rivers. US Geological Survey, Water-Resources Investigations Report 98-4018, Denver CO., 73 pp.
- Saenger, N., Kitanidis, P.K., Street, R.L., 2005. A numerical study of surface-subsurface exchange processes at a riffle-pool pair in the Lahn River, Germany. *Water Resources Research* 41, W12424. doi:10.1029/2004WR003875.
- Storey, R.G., Howard, K.W.F., Williams, D.D., 2003. Factors controlling riffle-scale hyporheic exchange flows and their seasonal changes in a gaining stream: a three-dimensional groundwater flow model. *Water Resources Research* 39, 1034. doi:10.1029/2002WR001367.
- Wang, H.F., Anderson, M.P., 1982. *Introduction to Groundwater Modeling: Finite Difference and Finite Element Methods*. W.H. Freeman & Co, New York. p. 237.
- Weissmann, G., Carle, S., Fogg, G., 1999. Three-dimensional hydrofacies modeling based on soil surveys and transition probability geostatistics. *Water Resources Research* 35 (6), 1761–1770.
- Wondzell, S.M., 2006. Effect of morphology and discharge on hyporheic exchange flows in two small streams in the Cascade Mountains of Oregon, USA. *Hydrological Processes* 20, 267–287.
- Wondzell, S.M., Swanson, F.J., 1996. Seasonal and storm dynamics of the hyporheic zone of a 4th-order mountain stream. I: hydrologic processes. *Journal of the North American Benthological Society* 15, 1–19.
- Wörman, A., Packman, A.I., Marklund, L., Harvey, J.W., Stone, S.H., 2007. Fractal topography and subsurface water flows from fluvial bedforms to the continental shield. *Geophysical Research Letters* 34, L07402. doi:10.1029/2007GL029426.
- Wroblicky, G.J., Campana, M.E., Valett, H.M., Dahm, C.N., 1998. Seasonal variation in surface-subsurface water exchange and lateral hyporheic area of two stream-aquifer systems. *Water Resources Research* 34, 317–328.
- Zheng, C., Bennett, G.D., 1995. *Applied Contaminant Transport Modeling*. Van Nostrand Reinhold, New York. p. 440.