Reply to comment by F. Molz et al. on “Investigating the Macrodispersion Experiment (MADE) site in Columbus, Mississippi, using a three-dimensional inverse flow and transport model”

Mary C. Hill,1 Heidi Christiansen Barlebo,2 and Dan Rosbjerg3

Received 29 September 2005; revised 27 February 2006; accepted 14 March 2006; published 14 June 2006.


[1] We are pleased to have the opportunity to discuss further what we think are important methods and results presented by Barlebo et al. [2004]. We are especially pleased to enter into a public discussion with four scientists who we regard with such high esteem concerning an important topic of mutual interest.

[2] The main goal of our paper is to present and demonstrate a novel method of using dependent-variable data (often referred to as observations) and models to investigate the accuracy of data related to input parameters, here, hydraulic conductivity. The idea of using conservation of mass equations to test different types of data against one another rigorously is not new to science in general, but to our knowledge, the only comparable attempt at testing hydraulic conductivity data rigorously using groundwater models was reported by Barth et al. [2001] using data from a controlled laboratory experiment. Using data from a complex field system such as that at the MADE site, is, of course, much more difficult. Less rigorous evaluations using groundwater models have been conducted, as discussed by Ingebritsen and Sanford [1998, pp. 14–15]. Our results suggest that while the dual-domain processes proposed by Molz et al. [2006] may be important, difficulties in the data set considered indicate that it does not demonstrate the dominance of dual-domain processes as vividly as has been presented in the literature. We went on to discuss what site-scale data would be needed to provide more conclusive evidence.

[3] Our reading of Molz et al. [2006] yields five issues that we address in the following five sections.

1. Misrepresented Results [Molz et al., 2006, paragraphs 3, 4, and 11]

[4] Barlebo et al. [2004, paragraphs 61, 85, and 92 and Table 2] indicate that their results are not as definitive as suggested by Molz et al. The statements from the abstract and conclusions cited by Molz et al. include qualifiers, and the difficulties are described explicitly in the text of the article. Considering the acknowledged difficulties, it does not seem to us that the article was “claiming a positive [result]”. Rather, we explored an alternative that is important to understanding measurements and simulation of groundwater transport. The results were not definitive but suggested some possibilities that we tried to present clearly.


[5] In this work we consider important questions about the precision and accuracy of the hydraulic conductivity data and about the extent to which imprecise or inaccurate hydraulic conductivity data might affect the support for hypotheses being considered in the literature, such as dual porosity. In our minds, the issue addressed is not whether such mechanisms occur but in what ways this data set conclusively demonstrates their prevalence and what the data set can reveal about conducting more definitive experiments. Neglecting alternative mechanisms was not “arbitrary.” It was required for the scientific inquiry conducted.

3. Use of Zones of Constant Value Given the Locally Heterogeneous Texture of the Material [Molz et al., 2006, paragraph 8]

[6] As stated by Molz et al., larger-scale hydraulic conductivity trends appeared to dominate flow and transport at the MADE site. To build a model with which the measured hydraulic conductivity data could be evaluated, the hydraulic conductivity data could not be used directly to construct the hydraulic conductivity distribution, as had been done in other studies [e.g., Feehley et al., 2000; Julian et al., 2001]). The head data clearly show abrupt changes in hydraulic gradient indicating that a hydraulic conductivity distribution represented using zones of constant value might be an approximation with some promise. The paper clearly shows that we were fully aware of the advantages and difficulties of this approximation, and we discussed in some detail what the data implied about the continuity of the hydraulic conductivity field [see Barlebo et al., 2004, paragraphs 48 and 82–84].
Figure 1. Random numbers demonstrating how noise obscures identification of large-scale trends. The random numbers are lognormally distributed with a standard deviation of half an order of magnitude. (a) The data could easily be interpreted erroneously as coming from a distribution with a linear trend. (b) The true mean values used to generate the numbers are shown.

[7] Defining zones using the hydraulic conductivity data needs to be considered in the context of the replicate data from the MADE site presented by Rehfeldt et al. [1989]. The replicate data suggest errors between 3 and 10 times the measured value. To demonstrate the consequences such errors have when trying to identify large-scale trends, consider the example shown in Figure 1. In practice, we have data sets that are far sparser and trends would be more difficult to identify.

[8] One concern we have is that investigations of alternative mechanisms such as dual porosity have not, to our knowledge, tried to explicitly consider the clearly documented replicate errors in the hydraulic conductivity data. We wonder what this means about the efficacy of the alternative mechanisms being proposed. Does it mean that they overwhelm the errors in the hydraulic conductivity data? Does it mean that they produce the right concentration distribution but they are not representative of the actual processes, for which the errors in the hydraulic conductivity data would need to be considered? We believe these questions need to be addressed to truly understand what measured hydraulic conductivity values mean and what mechanisms govern subsurface transport. Barlebo et al. [2004] present one attempt to address some of these questions. Clearly, more work is required in this area.

4. Implausibility of the Estimated Hydraulic Conductivity of Zone II [Molz et al., 2006, paragraph 8, 9, and 10]

[9] To investigate the concern about plausibility of the hydraulic conductivity of zone II presented in the paper, all values involved are presented in Table 1. In addition, we have included characteristic values for different types of deposits as needed to address the comment about the type of materials found at the site.

[10] The pump test values are reasonably similar to the geometric mean of the flowmeter measurements for zone II, and this consistency might be thought to lend credence to the flowmeter measurements. However, an important issue is what material is measured by the pump tests. The material represented by zone II is distinct enough to result in hydraulic gradients that are flatter than in neighboring areas [see Barlebo et al., 2004, Figure 4], and on average, the hydraulic conductivities in zone II are expected to be higher than in the surrounding deposits. On the basis of remarks made by Molz et al. and Boggs et al. [1992], it seems likely that the pump test values should be lower than local zone II values of hydraulic conductivity (see notes 3 and 4 of Table 1). The consistency between the pump test and flowmeter values thus supports the idea that the flowmeter measurements may be biased such that they are lower, on average, than the actual values in zone II. Further evaluation of this issue would require an analysis of the pump test data using a numerical model of the system, which is beyond the scope of this reply.

[11] Molz et al. suggest that the materials in zone II are not the clean medium gravels indicated by the model calibrated hydraulic conductivity of $82.8 \times 10^{-4}$ m/s (715 m/d). We would be very interested in whether the adjusted value of $41.4 \times 10^{-4}$ m/s (358 m/d), which we used to estimate the maximum possible bias of a factor of 5, is thought to be consistent with the field material. The estimated value was adjusted to account for model error. The adjustment is well within the range of its uncertainty as reflected by its coefficient of variation [Barlebo et al., 2004, Table 1]. If the adjusted value is not considered to be plausible, obtaining a plausible range from Molz et al. would assist us in determining a more accurate estimate of maximum possible bias.

[12] An attempt to create a simulation using the adjusted value of hydraulic conductivity for zone II would require not only changing that parameter value but also consideration of alternative models, and especially alternative ways of representing the questionable boundary conditions. Such an evaluation, though important, was considered to be beyond the scope of this reply. We felt that the model we produced was sufficiently plausible to produce some important insights, and we stopped with those.

[13] Molz et al.’s statement that the large confidence interval on the hydraulic conductivity of zone II “calls into question whether the model itself rests on solid hydrogeologic grounds” is curious. It is easy to create situations using synthetic models in which the model is known to accurately represent a system yet large confidence intervals
HILL ET AL.: COMMENTARY

Table 1. Hydraulic Conductivity Values Associated With Zone II of Barlebo et al. [2004]a

<table>
<thead>
<tr>
<th>Source of Estimate</th>
<th>Notesb</th>
<th>Hydraulic Conductivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Meters per Second</td>
</tr>
<tr>
<td>Geometric mean of flowmeter measurements</td>
<td>Cited by Barlebo et al. [2004]</td>
<td>1</td>
</tr>
<tr>
<td>Arithmetic mean of flowmeter measurements</td>
<td>Cited by Barlebo et al. [2004]</td>
<td>1</td>
</tr>
<tr>
<td>Model calibration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used to conclude that at the very most, K measurements may be biased by a factor of 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old pump test, using full thickness</td>
<td>Cited by Molz et al. [2006]</td>
<td>3</td>
</tr>
<tr>
<td>Old pump test, using limited thickness</td>
<td>Cited by Molz et al. [2006]</td>
<td>3</td>
</tr>
<tr>
<td>New pump test, using full thickness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New pump test, using limited thickness</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Text Book Values for Relevant Materials

<table>
<thead>
<tr>
<th>Source of Estimate</th>
<th>Notesb</th>
<th>Hydraulic Conductivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Meters per Second</td>
</tr>
<tr>
<td>Gravel</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Gravel</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Coarse sand</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Sand</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Medium sand</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Fine sand</td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

Notes: 1, “The geometric mean is an approximate lower bound of the effective value of horizontal hydraulic conductivity for a three-dimensional system if the material is randomly distributed, while the arithmetic mean applies to material organized in layers parallel to the flow direction. The situation at the MADE site is expected to be between these two extremes.” [Barlebo et al., 2004, paragraph 73]; 2, Barlebo et al. [2004, Table 2 and paragraph 86] assume that model error, and especially difficulties with boundary conditions, could easily result in the estimate being off by a factor of 2, so the smaller value was used to derive an estimate of the maximum possible bias; 3, from Boggs and Adams [1992, Table 3]; the test is described as a large-scale test carried out over 192 hours at a pumpage rate of 208 L/min; it seems very likely that these values are affected by nearby lower hydraulic conductivity deposits; 4, Molz et al. mention that nearby areas with lower values of hydraulic conductivity affect pump test results; 5, from Freeze and Cherry [1979, p. 29]; 6, from Domenico and Schwartz [1990, p. 65] and Zheng and Bennett [2002, p. 296].

Values are reported with units of meters per second (m/s) used by Barlebo et al. [2004] and units of meters per day (m/d) used by Molz et al. [2006].

on parameters are produced. The large confidence intervals do indicate a poorly constrained parameter value, and, as noted in the next paragraph, this made it difficult to obtain definitive conclusions.

[14] Molz et al. suggest that a solution with parameter values that more closely match measured values and are more consistent with the texture of the material would be more convincing. Of course we agree with this suggestion. Also, if the coefficients of variation on the parameters had been smaller, then the unreasonable parameter values would have clearly demonstrated the importance of alternative mechanisms, and that definitive result would have been welcome. Our results were between these two definitive possibilities. We used the model we had to learn as much as possible about the simulated flow conditions that were producing what we considered to be a reasonably good fit to the concentrations using advective-dispersive processes.

5. Model Fit to Concentrations [Molz et al., 2006, paragraphs 10 and 11]

[15] Barlebo et al. [2004] clearly discuss the ways in which their model fit and did not fit the concentration data. We state clearly that it is the simultaneous fitting of high concentrations near the source and extensive spreading of the plume that is important to consider. The extensive spreading is simulated quite well. The misfit at the source is discussed in section 5 of Barlebo et al. (paragraph 96) as follows: “In both the model and the measurements, the highest remaining concentration occurs at the source, but the maximum simulated value is 0.4% of the starting simulated concentration while the maximum measured concentration is six times that, or 2.4%. How much of the discrepancy is caused by difficulties with simulating the initial concentration distribution and numerical dispersion would determine the role that would need to be played by alternative mechanisms in maintaining high concentrations at the source.”

6. Interpretation of Mass Balance Analysis Presented in Table 4 of Barlebo et al. [Molz et al., 2006, paragraph 12]

[16] In Table 4 of Barlebo et al. [2004] we presented a mass balance to investigate the possibility that at late times the hypothesized mass deficit might be partly explained by mass having passed beyond the sampling network. Our analysis suggests that this effect may indeed be a significant contributing factor. We agree completely with Molz et al. that a good mass balance measures the accuracy of the numerical approximation, not the validity of a simulation.

7. Another Issue

[17] An interesting issue not raised by Molz et al. is that Julian et al. [2001] were able to produce a head distribution that was similar to that considered by Barlebo et al. [2004] using a hydraulic conductivity distribution that was more
closely related to the measured values of hydraulic conductivity. Julian et al. [2001], however, do not compare simulated and measured hydraulic conductivities, do not discuss the simulated flow field, and do not discuss any effort to consider what might be a trade-off between deviations from measured hydraulic conductivities and the importance of dual porosity processes. In addition, Julian et al. [2001] neglect recharge. While the reasons stated might justify that assumption for their model, the effects of neglecting recharge on dual-porosity importance were not evaluated. The results from Barlebo et al. [2004] as well as common hydrologic principles suggest that the recharge rate and distribution can be important to transport. Further analysis with the model presented by Julian et al. [2001] might help to address some of the questions posed and not definitively resolved by Barlebo et al. [2004] as well as help to understand how recharge and dual-porosity processes interact. For example, such an effort could be used to identify aspects of transport that can be equivalently produced by variations in hydraulic conductivity, recharge, or dual porosity, and aspects of transport that are unique to these system characteristics.

Finally, there is one detail that needs to be corrected. Molz et al. suggest that we only considered concentration data from 328 days in our regressions, and this is incorrect. We included selected concentration data from all four of the snapshots, as indicated in paragraphs 24 and 45 of Barlebo et al. [Molz et al., 2006, paragraph 4].

8. Conclusions

We believe that our results in some ways support the importance of dual-porosity processes, especially to maintaining high concentrations at the injection site. However, our results also suggest that dual-porosity processes may not play as dominant a role as sometimes suggested. Understanding the interplay of large-scale heterogeneity and dual-porosity processes is important to accurate predictions of groundwater transport.

At a fundamental level, our goal is that Barlebo et al. [2004] and Molz et al. [2006] focus attention on measurements and how they can be used in conjunction with conservation principles embodied in quantitative solutions such as process-based numerical models to test hypotheses about groundwater systems. Groundwater systems are perplexing and are likely to remain so for some time. Yet careful, ever vigilant inquiry is slowly providing answers.

[21] We appreciate the interest and effort of Molz et al. and look forward to fruitful discussions in the future.

References
ABSTRACT The first Macrodispersion Experiment (MADE1) at Columbus Air Force Base in northern Mississippi is utilized to perform numerical simulations of solute transport in an aquifer. The purpose is to illustrate the capability of the coupled Markov chain (CMC) model in delineating the complex geometrical configuration at the site for solute transport simulations under the lack of geological information. The CMC model is also used to study the effect of reducing geological information on the transport predictions in terms of plume configuration, first and second spatial moments and macrodispersion. The macrodispersion experiment (MADE) site in Columbus, Mississippi, is one of the most studied highly heterogeneous sites in the world (with a variance of In K = 6.6) [Dogan et al., 2011; Bohling et al., 2012]. Natural gradient tracer tests at this site have fueled an interest in delineating the complex groundwater flow and transport processes. Three large-scale tracer tests were performed at the MADE site. Here we focus on MADE 1, conducted from October 1986 to June 1988 [Boggs, 1991]. The experiment involved injecting a calcium bromide tracer solution into five wells, aligned approximately perpendicular to the natural groundwater flow direction. More than 11,000 water samples were collected from 258 monitoring wells (Figure 1) in eight synoptic snapshots. DOI: 10.1007/s11242-012-0031-z.