Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-532-RC3, 2017

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Aggregation effects on tritium-based mean transit times and young water fractions in spatially heterogeneous catchments and groundwater systems, and implications for past and future applications of tritium" *by* M. K. Stewart et al.

Anonymous Referee #3

Received and published: 12 January 2017

General comments

The results summarized in the first paragraph of the abstract and Figures 1-9 of the manuscript itself are, to the best of my knowledge, broadly correct.

However, the results reported in the second paragraph (and the related discussion and analyses in section 3.4 and 4) need to be reconsidered.

1. The statement that "well-chosen compound lumped parameter models should be

used as they will eliminate potential aggregation errors due to the application of simple lumped parameter models" directly implies that aggregation errors only arise with simple LPM's, but not with compound LPM's, or at least not with "well-chosen" ones (whatever that means).

1.a. This is not consistent with the analysis presented elsewhere in the paper. Figures 3-5 show clear aggregation errors from the use of simple lumped parameter models, but they would also show clear aggregation errors if more complex lumped parameter models were used. For example, a gamma model with alpha=0.3 closely approximates a compound LPM, but it is clearly vulnerable to aggregation errors, as shown in Figures 4 and 5.

1.b. Since the analyses in Figures 3-5 have been used to demonstrate aggregation errors in simple LPM's, exactly the same analyses must be applied to compound LPM's to demonstrate that these aggregation errors disappear. Until this is done, the claims in the abstract have not been demonstrated, and must be removed.

2. In some of the examples that are presented, the compound LPM's clearly fit the data better, but of course they should, because they have more free parameters. Whether these parameters are fitted by formal calibration or by "expert judgment" and fitting by hand makes little practical difference; in either case they make the fitted curves more flexible and thus more conformable to the data. (This comes at the cost of greater parameter uncertainty; more about that below.)

2.a. In the case of the "DDM" in Figure 10 and Table 2, for example, there are FIVE adjustable parameters: b, tau_s, tau_d, P_Ds, and P_Dd (incorrectly labeled as a second "P_Ds" in the table). So Figure 10 shows a five-parameter fit to just six data points (which are themselves not fully indepenent of one another). Is it any surprise that the curve fits well? The other models have at least two parameters, for a data set that effectively has only two or three unique values; those near the peak and those in the 2000's. Again, it is not at all surprising that these can be calibrated to fit the data.

3. Figure 10c is presented as evidence that "the mean residence times were sharply constrained close to 8 years". This is at best unproven and at worst misleading.

3.a. Consider, for example, the red curve for the DDM. In the DDM, the mean residence time (MRT) is a function of three parameters (b, tau_s, and tau_d), and the tritium curve, and thus the fit to the data (SD) is determined by these three parameters, plus two others (P_Ds and P_Dd). It is mathematically impossible for the relationship between MRT (which depends on three independent parameters) and SD (which depends on five independent parameters) to be described by a single curve. There will be multiple combinations of b, tau_s, and tau_d that give the same MRT but different values of SD, and the range of SD will be inflated further by variations in P_Ds and P_Dd.

3.b. The same problem arises, in simpler form, for the EPM and DM. The DM, for example, depends on a residence time and a dispersion parameter P_D; for any individual value of the dispersion parameter, one can draw a curve relating the residence time to the misfit parameter SD. But to describe the relationship between SD and the residence time, one needs a full family of curves, to represent the range of possible values of the dispersion parameter.

3.c. It is impossible to know for sure (since the methods are unacceptably vague on this point), but it seems likely that Figure 10c was generated by choosing fixed values for all-but-one parameter in each model, and then varying just one parameter and tracing out the resulting relationship between MRT and SD.

3.d. From a parameter estimation standpoint, this is a fundamentally flawed procedure, because (1) it ignores the extra degrees of freedom from the other parameters that are arbitrarily held constant, and (2) it therefore underestimates the uncertainty in the MRT, possibly by large factors. This is true even if the parameters were fixed by "expert judgment" rather than algorithms, as long as the experts were free to revise their "judgment" based on whether the tritium curves made sense.

3.e. Methods for multi-variable parameter estimation and uncertainty analysis are widely available. There is no valid excuse for not using them. The revised manuscript must eliminate all claims (explicit or implied) about MRT's estimated from tritium measurements using multi-parameter models, unless and until proper parameter estimation and uncertainty analysis are done.

3.f. There is likewise no valid reason for ignoring the uncertainties in the tritium measurements themselves, and their consequences for parameter uncertainties. Looking at the error bars in Figure 10a, for example, one can estimate that the pooled standard deviation (due to the measurement uncertainties themselves) is about 1-2 TU. Therefore, Figure 10c implies that the MRT is only constrained within about plus or minus two years (for a standard deviation of 1 TU) or about plus or minus four years (for a standard deviation of 2 TU), which is quite a contrast to the paper's assertions that the MRT is "sharply constrained". And this estimate does not even begin to account for the additional uncertainty introduced by the other four parameters. Again, there are standard methods for propagating these uncertainties in parameter estimation, and there is no valid excuse for not using them.

4. As was also pointed out by another reviewer, the claims that compound LPM's have less aggregation bias are not supported by clear lines of reasoning. For example:

4.a. In 3.4.1, the manuscript says that there is little aggregation bias because the simple and compound LPM's have similar mean residence times. But why does this imply an absence of aggregation bias, rather than a similar aggregation bias across all three LPM's? The manuscript also argues that we should expect little aggregation bias because the two model components have MRT's that are similar to, or shorter than, the half-life of tritium. This is only a valid argument if we have independent evidence about the ages of the system components. What evidence do we have that the deep aquifer really contributes 74% of the flow and has a MRT of 10.2 years, instead of (say) 35% of the flow with a MRT of 100 years? If such independent information exists, the reader should be made aware of it. Alternatively, the manuscript needs to demonstrate that the MRT's of the individual system components can be reliably constrained through parameter estimation (which will not be easy).

4.b. In 3.4.2 and 3.4.3, the claim seems to be that the simple LPM's are subject to aggregation bias because they disagree with each other or with the compound LPM, which fits the data better. But again, the compound LPM has at least twice as many parameters as the simple LPM's, so one would need to somehow show that the better fit does not simply arise from this rather obvious explanation. And of course the simple LPM's will disagree with each other; they have different shapes, so it is unsurprising that they may have different MRT's when fitted to data.

5. One needs to recognize that the abstract's claim that "The choice of a suitable lumped parameter model can be assisted by matching simulations to time series of tritium measurements (underlining the value of long series of tritium measurements)" is mostly a statement about the past, and is misleading as a generalization about the future.

5.a. In the (few) springs and aquifers where tritium analyses were performed decades ago, during and after the bomb peak, those analyses have turned out to be quite useful for comparison with the more recent measurements. Indeed, as Figure 12 shows, it is these early samples that allow one to distinguish between the differently shaped LPM's, and the more recent samples have almost no power to discriminate between those same LPM's.

5.b. And that is precisely the problem: going forward into the future, long time series will be much less useful, for the simple reason that the bomb pulse tritium is largely gone and we are approaching an equilibrium between tritium production and decay. Thus, going forward, long time series will not help, because tritium concentrations are becoming less and less dynamic over time. As the bomb pulse tritium vanishes, we will just be measuring the same value over and over.

5.c. I am sure the authors know this, and it is disingenuous not to make it clear to the

reader, particularly because they celebrate the one clear benefit of the fading of the bomb pulse (the end of double solutions for many tritium models).

5.d. The fading of the bomb pulse will make the parameter estimation problem outlined above even more impossible than it is already. Consider the red curves in Figure 10 as an example. As mentioned above, these are five-parameter fits to six data points. In the future, anywhere that we do not already have measurements of bomb pulse tritium, we will instead have a five-parameter fit to what is effectively just ONE data point (because in equilibrium, all future measurements are redundant).

5.e. There will still be value in sampling across a range of discharges in order to quantify how modeled tritium ages vary with different wetness conditions, as previous work from the New Zealand group has very nicely demonstrated.

Specific comments

1. As other reviewers have pointed out, the organization and clarity of the presentation must be improved. Many necessary details have also been left unmentioned.

2. Needless confusion is created by the alphabet soup of acronyms. Saying "dispersion model", "exponential model", "lumped parameter model", and so on is preferable to forcing your readers to learn a dozen acronyms just so they can get through your paper.

3. Inconsistencies abound. The double exponential piston flow model is called both DEPM and (apparently) BMM. Using inconsistent terminology like this is bad enough, but what's even worse is that readers are never told, and are left to figure this out for themselves. Most of the text uses MTT but some of the figures and captions use MRT, and again readers are never told whether these are the same things or different things. These are just a few examples of a general problem, and it should not be a reviewer's job to flag all these issues.

4. The wiggles in the black curves in Figures 3b and 3c are obvious numerical artifacts,

since the real theoretical curves should be smooth. It is troubling that such visually obvious numerical errors have not been noticed and corrected. One naturally wonders whether there are other technical issues that are less visually obvious, and also have not been caught.

5. at line 8 on page 3, Bethke and Johnson (2008) should be cited; otherwise it looks like the authors are taking credit for this observation.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-532, 2016.