We thank Julien Farlin and Jim Kirchner for interesting Comments on our paper (Stewart et al.). These raise important questions, some of which are beyond the scope of our paper, but we give some observations below in response to their comments.

Comment SC1 by J. Farlin

We agree with Farlin's deduction from Fig. 3 in Stewart et al. that "The relationships between mean transit time and tritium activity shown on Figure 3 clearly display approximately linear segments over which the mixture of water coming in equal volume from two different subcatchments would lead to a negligible underestimation of the true MTT (for instance, on figure 3d for MTTs between 0 and 20 years)."

Farlin asks: "What degree of heterogeneity, and hence how large a difference in subcatchments' MTTs, can usually be expected in real-world catchments?". Stewart et al. looked for evidence of real-world aggregation error by examining some tritium dating case studies from the literature. These were chosen to cover the range of tritium dating (0-200 years) and were cases where we knew there were "few and distinct" heterogeneities in the catchments (in the words of Luther and Haitjema, 1998). These cases showed substantial aggregation error for MTTs greater than about 10-20 years. Other case studies in which there were apparently no "few and distinct" heterogeneities could have been chosen; we expect that these would have shown considerably less aggregation error.

Farlin questions conclusions drawn from the use of Kirchner's (2016) and Stewart et al's virtual experimental model (water from two subsystems with different MTTs mixing to give the system outflow) as being too conservative for real-world systems (i.e. as indicating too much aggregation error). The model produces aggregation errors that range from zero when the MTTs of the subsystems are the same to very large when the MTTs are very different. Luther and Haitjema (1998) is cited to assert that most real-world groundwater systems are effectively relatively homogeneous (describable by a simple exponential LPM at baseflow). Many catchments, on the other hand, have outflows composed of quickflow and baseflow, which could have very different MTTs leading to large aggregation errors. We note Farlin's support for our suggestion that using compound LPMs could reduce aggregation errors greatly by conceptually catering for young and old MTT waters within catchments or groundwater systems.

We also thank him for the comment about tritium input and will explain that tritium data for Trier before 1978 was calculated by regression from Vienna data.

Comment SC3 by J. Kirchner

Kirchner says that an important question is: "Given how little we know about the patterns of heterogeneity in catchments' characteristics and/or their transit time distributions, how sanguine should we be about the risk of aggregation errors?". He answers: “We know that important catchment properties (hydraulic conductivity, depth to bedrock, soil characteristic curves, etc.) typically vary by large factors, in spatially correlated fashion, across all the scales at which they can be measured. Given this pervasive multiscale heterogeneity, the burden of proof should be on those who claim that it doesn’t matter, or who want to use techniques that are prone to aggregation errors (such as estimating MTT from seasonal tracer cycles). Alternatively, we should develop – and use – methods that are much less vulnerable to aggregation errors (such as the young water fraction concept presented by Kirchner, 2016).”

Farlin had cited Luther and Haitjema (1998) to support effective homogeneity in groundwater aquifers. However, Kirchner suggests that the range of variation in hydraulic conductivity used by Luther and Haitjema was limited in the simulated cases with analytic element model GFLOW, and that the fine-scale heterogeneity assumed in the finite-difference MODFLOW model lacked spatial correlation. Hence, the finding of effective homogeneity, although reasonable for the tested conditions, would not be applicable to many real-world situations where there would be “significant and distinct” heterogeneity.

We believe that the use of compound LPMs could strongly reduce aggregation errors in hydrological systems with “significant and distinct” heterogeneity. For example, we consider a simple case of a catchment split into two parts by two very different rock types that produce waters with very different MTTs; i.e. the most extreme “significant and distinct” heterogeneity one can imagine. A binary LPM is ideally suited for this type of system, and when optimised with suitable data would very effectively separate the young MTT from the old MTT waters in the catchment outflow, and therefore minimise aggregation errors in MTT. If we now consider a catchment split into four parts with two areas of each rock type, the binary LPM when optimised is still very effective for separating the two types of water, while the potential for aggregation errors is smaller. In systems which are split into eight, sixteen, etc. parts the binary LPM retains its effectiveness and the potential for aggregation errors becomes very much smaller because the system starts to look homogeneous at larger scales. There is, of course, a wide range of types of hydrological systems, but the binary LPM is likely to remain effective in cases of “significant and distinct” heterogeneity, which are the ones of most concern for aggregation error.

References

Kirchner JW, Aggregation in environmental systems – Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments, Hydrology and Earth System Sciences, 20, 279-297, doi:10.5194/hess-20-279-2016, 2016.

Luther KH and Haitjema HM, Numerical experiments on the residence time distributions of heterogeneous groundwatersheds, Journal of Hydrology 207, 1-17, 1998).