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.

J. Kirchner

jwkirchner@gmail.com

In his interactive comment on the manuscript by Mike Stewart and colleagues, Julien Farlin substantially misrepresents a recent paper of mine (Kirchner, 2016), as well as earlier work by Luther and Haitjema (1998).

Farlin writes that "Kirchner only mentions in passing a factor of 2 as characteristic for "true heterogeneity" without further elaboration." This misrepresents a passage in

Kirchner (2016) that says nothing about real-world patterns of heterogeneity, but instead describes Monte Carlo trials of hypothetical pairs of subcatchments, which were combined to determine whether their runoff yielded reliable estimates of mean transit time (MTT). Monte Carlo trials can, by chance, lead to cases where both subcatchments are nearly identical, and therefore the resulting combined catchment is not meaningfully heterogeneous. Therefore, as Kirchner (2016) says, "Pairs with MTTs that differed by a factor of two were excluded, so that the entire sample consisted of truly heterogeneous catchments" (that is, so that the sample excluded catchments that by chance were nearly homogeneous). This statement very clearly refers to hypothetical Monte Carlo trials (and about how variable these hypothetical catchments need to be, in order to be considered as heterogeneous rather than homogeneous in that context). Instead it has been misrepresented by Farlin as a making a claim about how much heterogeneity actually exists in the real world.

Farlin characterizes my analysis as a "toy model" and asks "whether the toy model adopted is appropriate at all to study the effects of heterogeneities on the transit time distribution and hence on the estimates of the mean transit time." This misrepresents the point of the analysis, which was not "to study the effects of heterogeneities ON the transit time distribution", but rather to study how spatial variations IN transit time distributions (whatever their cause) will affect the reliability of MTT estimates. Farlin, in short, complains that my analysis is not appropriate, but for a task that was never its goal in the first place: a classic straw-man argument.

These confusions could have been cleared up in minutes via phone or e-mail, so it is regrettable that Farlin did not contact me directly before launching his public attack.

Nonetheless, Farlin's comment indirectly raises two potentially important questions. First, how is heterogeneity in catchments' characteristics related to heterogeneity in their transit time distributions? Neither Kirchner (2016) nor the present manuscript by Stewart et al. address this question, but it is an interesting one. However, even if this relationship were known, it would simply substitute one factor that is often poorly

quantified (the spatial heterogeneity in catchments' characteristics) for another that is also poorly quantified (the spatial heterogeneity is their transit time distributions). The second question raised by Farlin's comment 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?

My answer is the following. 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's answer appears to be different: "One could be tempted to answer that since we do not know how to quantify the degree of heterogeneity, it could be anything, and consequently assuming a large difference is conservative. I am concerned however that too much conservativeness leads to confusing or over cautious results, but additionally, there IS (at least) one study that addressed this question mechanistically for a number of case in an heterogeneous aquifer, namely that of Luther and Haitjema in Journal of Hydrology (1998). The authors show that in many cases ("stratified, unstratified, confined or unconfined [aquifers]"), the simple exponential distribution (i.e. a special case of the gamme [sic] function with the alpha term being equal to 1) is a good approximation of the transit time distribution (TTD) of heterogeneous catchments as long as heterogeneity is "not significant and distinct" [...] For groundwater systems, the results of Luther and Haitjema show that the "homogeneous assumption" holds in many real-world situations. Thus, Kirchner's conclusion that "MTT's estimated from seasonal tracer cycles are fundamentally unreliable" is too broad and must be corrected urgently."

C3

The crucial misrepresentation here is that Luther and Haitjema never say that their results are relevant to "many real-world situations", perhaps because, in fact, they aren't. For example, in Luther and Haitjema's horizontally stratified simulations, the simulated transit time distributions deviated significantly from the exponential distribution (what Farlin calls the "homogeneous assumption"), even though the maximum conductivity difference was always less than a factor of 10. In the one horizontally stratified case shown in the paper (Case H, Fig. 8), the largest conductivity contrast between the layers was less than a factor of 3. A factor of 3, or even 10, is vastly less than the 10,000-fold variation in conductivity that one finds just among different types of sand, or the 1,000,000-fold variation in conductivity among glacial tills, or the 5 to 8 orders-of-magnitude variability in hydraulic conductivity that one finds even within individual lithologic groups, or the roughly 10-14 orders of magnitude that separate igneous rocks from gravel (Gleeson et al., 2011).

As another example, when Luther and Haitjema simulated random variations in aquifer properties, they varied conductivity, recharge, and porosity by small factors (whereas in the real world, conductivity alone can vary by orders of magnitude), and they then assigned these random aquifer properties individually to the roughly 73,000 cells within the model watershed, with zero spatial correlation. Thus it is unsurprising that the resulting simulations conformed to the "homogeneous assumption", for the simple reason that at every scale larger than a few grid cells, the model watershed was extremely homogeneous indeed. As a representation of real-world heterogeneity, the variation in aquifer properties was too small to be realistic; furthermore, the lack of spatial correlation is inconsistent with every set of field data that I have ever seen.

In characterizing Luther and Haitjema's analysis as "serious" and "physically based", Farlin has apparently overlooked its obviously nonphysical assumptions. To note just one example, Luther and Haitjema allow their "confined" aquifers to receive spatially uniform recharge. Thus the "confining" layer must somehow allow recharge to pass through vertically, while simultaneously confining the aquifer by preventing vertical flow.

C4

How, exactly, is this supposed to work?

Farlin has also apparently overlooked the abundant evidence (e.g., Kirchner et al., 2000, Godsey et al., 2010, Kirchner and Neal, 2013, Aubert et al., 2014) showing that tracer fluctuations in a wide variety of real-world catchments have spectral signatures that are inconsistent with exponential transit time distributions. This empirical evidence would seem to refute the notion that Luther and Haitjema's results hold "in many real-world situations".

Thus even if Luther and Haitjema's results are correct for idealized (and/or nonphysical) theoretical cases, with heterogeneity that is "not significant or distinct", they are inapplicable to much of the real world, in which heterogeneity is both significant and distinct (that is, correlated up to the hillslope or catchment scale).

In summary, Farlin's strident claim that "... Kirchner's conclusion that "MTT's estimated from seasonal tracer cycles are fundamentally unreliable" is too broad and must be corrected urgently" is not supported by any scientific evidence, either from his letter or from the work that he cites.

Aubert A, Kirchner JW, Gascuel-Odoux C, Faucheux M, Gruau G, and Mérot M, Fractal water quality fluctuations spanning the periodic table in an intensively farmed watershed, Environmental Science and Technology, 48, 930-937, doi: 10.1021/es403723r, 2014.

Gleeson T, Smith L, Moosdorf N, Hartmann J, Dürr HH, Manning AH, van Beek LPH, and Jellinek AM, Mapping permeability over the surface of the Earth, Geophysical Research Letters, 38, L02401, doi:10.1029/2010GL045565, 2011.

Godsey SE, Aas W, Clair T, Dennis I, de Wit H, Fernandez I, Kahl S, Malcolm I, Neal C, Neal M, Nelson S, Norton S, Palucis S, Tetzlaff D, Skjelkvaale BL, Soulsby C, and Kirchner JW, Generality of fractal 1/f scaling in catchment tracer time series, and its implications for catchment travel time distributions, Hydrological Processes, 24, 1660–

1671, 2010.

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.

Kirchner JW and Neal C, Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection, Proceedings of the National Academy of Sciences, 110 (30), 12213-12218, doi: 10.1073/pnas.1304328110, 2013.

Kirchner JW, Feng XH, and Neal C, Fractal stream chemistry and its implications for contaminant transport in catchments, Nature, 403, 524-527, 2000.

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