M. Weiler (Referee)

I really enjoyed reading this paper, however, I have to admit that it took me a while to find enough time to read through over 100 pages of the two papers combined. The paper nicely and very elegantly addresses the question how TTD in heterogeneous catchments will change the MTT, a question I have also thought a lot in the past, but I was unable to come up with such a great way to address this question. The paper is very well written, however, too long and is certainly of high relevance to the readers of HESS. I have a couple of concerns and ideas and hope that JK can resolve these so the paper can be published in HESS.

I thank my colleague Markus Weiler for his thoughtful comments and suggestions. These will help in formulating the revisions to the manuscript.

There are two reasons that the papers appear somewhat long. First, I am trying to introduce a substantial analysis based on a new concept, so I have to tell the whole story. But secondly, the "over 100 pages" are an artifact of Copernicus Publications' policy of publishing discussion papers in what is effectively a half-page format, thus more than doubling the page count (and, perhaps not coincidentally, more than doubling the page charges that authors pay to Copernicus.)

For example, Seeger and Weiler (2014) was 51 pages (or actually half-pages) in HESSD, but the final paper was only 21 pages in HESS. This one example suggests that the page count in HESSD is inflated, relative to the page count in HESS, by a factor of 2.5. The first of my two papers is 45 pages in HESSD, so one may expect that it would be about 18 pages in HESS. The second paper is 63 pages in HESSD, and should run about 26 pages in HESS.

General comments:

1) James Kirchner (JK) uses a simple convolution version to compute concentrations in the stream (eq 1) without considering inflow (precipitation) variation and/or evapotranspiration (e.g. Steward and McDonnell, 1991; Weiler et al., 2003). Particularly in catchments with a strong seasonality, this will markedly change the resulting tracer signal – a very strong change can be observed in snow dominated catchments (e.g. Seeger and Weiler, 2014). Under those conditions, the simple sine wave approach JK selects for his analysis may be flawed, since the observed sine wave in precipitation is very different to the input concentration into the catchments. Most success with the sine wave approach was in humid catchment without a strong seasonality (Scotland, Wales, East Coast US). It would be helpful to frame the results of this paper either by additional analysis in the context of these kind of catchments or at least discuss this in more detail with the related assumptions and consequences.

The points that Markus Weiler (MW) makes here are valid, but they are not specific to the simple sine wave approach; instead, they apply to convolution approaches in general. The manuscript already says that these approaches assume steady state and ignore evapotranspiration (see p. 3066), which are the main points that MW mentions in his comment. I had considered adding the point that when precipitation volumes vary over time, the concentrations in precipitation may not reflect the volume-weighted

inputs to the catchment. One approach to handle this is to volume-weight the sine wave fitting.

These points could be added to the introductory text (at the cost of making the paper longer). One could also go into much greater detail about all the assumptions behind convolution methods, but this has already been done elsewhere (e.g. by McGuire and McDonnell, 2006). The point of this paper is not to catalogue all of the possible factors that can complicate tracer-based transit time estimates, but instead to look in detail at the particular problem of aggregation across heterogeneous catchments.

2) It was very interesting to see, that age of the young water fraction of 0.2 years JK derived from his analysis is very close to the duration Seeger and Weiler (2014) derived for the time all catchments and models produce a very similar “discharge fractions after certain elapsed times”, which is equal to the young water fraction of this paper. In S&W we came up with a so called young water fraction of 2-3 months based on observations and applications of different convolutions models. WE also argued that this young water fraction should be used instead of the MTT. So I believe this supports greatly the results of JK and he may be able to strengthen his paper including these additional information.

I appreciate the suggestion and will see what I can do here.

The closest statement I can find to "*we came up with a so called young water fraction of 2-3 months based on observations and applications of different convolution models*" in Seeger and Weiler (2014) is "We observed a high agreement between the cumulated TTD fractions of the first 3 months (hereafter CF3M) for GM and TPLR (see Fig. 9). On the other hand, the TTDs tailings and MTTs varied notably between different models and proved to be less identifiable." (p. 4762, where GM and TPLR stand for Gamma Model and Two Parallel Linear Reservoirs, respectively, two different TTD models that were fitted to isotope data)

The closest statement I can find to "*we also argued that this young water fraction should be used instead of the MTT*" in Seeger and Weiler (2014) is "Therefore we decided to include CF3M as an apparently more consistent transit time metric than MTT into this analysis." (p. 4762)

Seeger and Weiler actually concluded, on page 4767, that what should replace the MTT is not the young water fraction, but instead the so-called transit time proxy *TTP*, defined as the ratio of the standard deviations of the tracer concentrations in precipitation and discharge. Interestingly, if the tracer concentration time series are seasonal cycles, *TTP* is numerically equal to the amplitude ratio, and in this paper I show that the amplitude ratio can be used to estimate the young water fraction. However, there is nothing in Seeger and Weiler (2014) that indicates that these quantities are connected.

Specific comments:

Title: not sure if aggregation really captures the main idea to other people and reflects the main message of the papers – see also paper 2. In addition, I would remove the -but not mean transit time-

Both papers are centrally concerned with how transit time estimates are affected by aggregation of tracer signals in heterogeneous catchments. Hence the link to aggregation is important, although I understand the point that our community may not be used to thinking in these terms.

Roughly half of the paper is devoted to demonstrating that seasonal tracer cycles yield strongly biased MTT estimates, so I think that this point about mean transit time really needs to be in the title.

I will think about whether the title can be streamlined, although I've already given this quite a lot of thought. The titles of the two papers also need to be linked by common phrases, which constrains the feasible possibilities.

Equations 3a-3d are not necessary since they are not used again in the paper.

Equation 3a is used in equation (10), and all four equations are important as support for the interpretations that are stated in the paragraph immediately below them.

The implications are quite long and it may help to provide subheading to better structure them.

This is a good point, which can be straightforwardly handled in the revision.

The figure captions are very long and often too detailed – I agree that a figure should be understood only with the figure caption, but JK sometimes includes interpretation of the figure and could shorten the captions in general.

The figure captions are written this way as part of a deliberate communication strategy. Minimalist figure captions often lead to unnecessary workload and confusion for the reader, who must jump back and forth between the figure and the text (perhaps several pages away) in order to understand what the figure says. Furthermore, readers often scan papers by looking at the figures without reading the text, meaning that the figures should be able to stand on their own.

Putting interpretations in figure captions can be a great help to readers, who can thereby get a sense of what the figures *mean* rather than just what they *are*. Experience has shown that authors often think that their figures will be self-evident (which of course they are *for the authors*, who already know what they are trying to say), and fail to comprehend how divergent a reader's understanding may be. Thus it is a smart communication strategy to lean in the direction of over-explaining rather than under-explaining.

Summary and Conclusion: Since the paper is already very long, I would highly recommend to shorten the S&C. I think it is not necessary to repeat the main ideas and steps and relate them to the figures – which is a very uncommon format anyway.

I disagree with MW's assertion that the paper is "very long". It is, for comparison, six pages shorter than the HESSD version of Seeger and Weiler, 2014 (in the Copernicus half-page format). I also disagree that the summary and conclusions section is overly

long. Again, for comparison, the conclusions section in the HESSD version of Seeger and Weiler is 28 lines long and mine is 39. Is 11 lines such an important difference?

I do agree that it is unconventional to refer to individual figures in the conclusions, but again this is a deliberate strategy. Often when they encounter a particular statement in the conclusions at the end of a complex paper, readers often wonder, "Wait, did the authors really show that? *Where* did they show that?" Providing this information gives readers a thumbnail index showing where the main points of the paper are covered. This can save them from searching through pages of dense text. It is also a great help to many readers, who follow the "first-last-middle" strategy of reading the abstract first and the conclusions second, then scanning the figures, and then perhaps reading the text.