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 we should deal with hydrologic nonstationarity to estimate MTT and TTD. The paper is very well written, however, much too long (see comments below) and is certainly of high relevance to the readers of HESS. I have a couple of concerns, of which one could be a major factor changing the main outcome of the paper – if JK can resolve this I think the paper can be published in HESS.

I thank my colleague Markus Weiler (hereafter MW) 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.)

I disagree that the paper is "much too long", particularly in relation to the substantial ground that it covers. For comparison, Seeger and Weiler (2014) was 51 pages (or actually half-pages) in HESSD; the present manuscript is 63 pages, or only 20% longer.

General comments: 1) I agree with the general approach to use any kind of model to test his approach if the sine wave fitting and calculated change in amplitude and phase shift using the MTT and Fyw will support the one or other approach. However, I miss a very relevant component in the model – evapotranspiration – which will either concentrate certain compounds in the catchments (Cl) or “remove” certain compounds (stable water isotopes). Several studies (e.g. Hrachowitz et al., 2014, Sprenger et al., 2015, HESS) have shown how relevant this part of the hydrological cycle can be including a very strong change in the sine wave of the precipitation signal. I hardly doubt, that a model including ET will come to the same results regarding the strong relation between sine wave dampening and young water fraction.

As stated on p. 3109, the analysis here ignores evapotranspiration in the interests of simplicity. Most analyses based on convolution methods do likewise. Including ET effects in a realistic way is a nontrivial exercise, and the details will be specific to individual tracers; to take just one example, evaporation and transpiration fractionate isotopes differently. Why does MW believe that including ET is important if, as he says, he doesn't think it will substantially affect the results? The point of this paper is to look at how heterogeneity and nonstationarity affect travel time estimates, not to look in general at all the things that could or should be included in a complete process model for tracers in catchments.

I am currently working on a separate paper on evapotranspiration effects on seasonal cycles of stable isotopes, and it will probably be nearly as long as the present manuscript. Including this material here would thus nearly double the length of the paper, which MW already says is too long.

2) The paper is very detailed with a lot of information and great thoughts of JK and many additional information that alone make the papers certainly worthwhile to read. Unfortunately, I fear, that a lot of these comments and thoughts will not be read, since they are hidden in this very long paper. I will make a couple of suggestion to shorten it, but JK may consider to move some parts of the paper to a new paper or an appendix with a clear message, where the reader is able to find his thoughts and ideas much better.

These are options that I already considered when I wrote the original manuscript, and for various reasons they create problems of their own. In any case, the current paper is not really so unusually long; it is only 20% longer than the HESSD version of Seeger and Weiler (2014).

Specific comments:

The introduction builds very much on paper 1 – however, it misses a more thorough literature review to this topic of non-stationarity and the need to develop other approaches and find solutions to infer catchments TTDs.

It is hard to know how to respond without a more specific idea of what MW thinks is missing here. Given that there is a lot of material to cover, I was trying to get down to the task without a long didactic introduction. I will consider whether there is an efficient way to provide more context here.

Section2.1 – Again, very informative and a lot of relevant and interesting ideas – however, it drifts a bit away from the main goal of the paper and makes the paper much longer to read and may result that less mathematically informed readers may stop here. Could JK not move much of this chapter in an Appendix – so modellers can have a more detailed look, but other readers can more easily grasp the main message.

I appreciate the point. Moving this material to the appendix will not save any length (since the appendix will also be part of the paper), and there are a number of logical links that might get broken. A smarter strategy is probably to provide a very clear textual signpost at the beginning of section 2.1, telling readers that if they are not interested in the details they can skip to section 2.3.

Section2.3 – I would propose to name the 3 input time series according to the Köppen climate zones they are in or any other short information. The catchment information is only relevant in this chapter, but not for the whole paper, so JK could avoid to always refer to the catchment names and the climate characteristic of the time series.

I recognize that the place names (Smith River, Broad River, Plynlimon) may not be directly relevant to the readers. It is relevant, however, that the three sites have varying degrees of seasonality, which is why I keep referring to the climate characteristics of the sites, wherever this information is important for the interpretation. Perhaps the primary designation should be the climate type rather than the place name (although this would imply that these particular places and time series were representative of those climatic zones, rather than just being examples of them). In any case, unless readers are well versed in the Köppen classification system, referring to the time series by only their Köppen codes (specifically, Csb in place of mentioning Smith River's

Mediterranean climate, or Cfb in place of Plynlimon's maritime temperate climate, or Cfa in place of Broad River's humid temperate climate) could be confusing.

Section 3.2 – there has been quite a bit of work done on this field – JK cites some of it, but it would help the chapter to include more of this work. In my opinion, the chapter is too close to the model and its assumptions and the parameters sets derived – if he could provide a more general conclusion, the chapter would certainly be helpful, but at the moment, it is mostly a very nice and interesting side way – I would move it to a Appendix chapter.

Again, it would help greatly to have more specific information about what MW feels is missing from the cited literature here. Section 3.2 is important for several reasons. First, it sheds important light on what controls long-term memory (vs. fast hydraulic response – the old water paradox) in the context of this simple model. Second, it provides a clear demonstration of how simple scaling arguments can yield useful insight into system behavior, including characteristic time scales. In the revision, I will try to make these implications more generally accessible to the reader.

Section 3.3 – Interesting, in particular for catchment modellers – however, it includes a new idea, which distracts from the main idea of the paper. If JK would have included the potential of tracer data or young water fraction to constrain the parameters, then the chapter may be helpful to support his ideas, but at the moment it is not. Maybe I also miss some things, as it is not completely clear to me how he derived Figure 8.

The results shown here are, MW and I agree, not essential for the analysis that follows. However, as we both also agree, they make an interesting point. In theory one could write a separate paper devoted to just that point, repeating the description of the model and so forth. As someone who tries to avoid such "salami tactics", I would prefer to keep this material here instead.

I believe MW's question about Figure 8 may be asking how I came up with the particular combinations of parameters. This was essentially trial and error, but informed by the scaling analysis of Section 3.2. I will try to make this clearer in the revision.

Section 3.6 – I would propose to shorten this section, in particular in relation to the figures related to the section. They do not show more detailed information than Figure 10-13. I think the main message of this approach can be summarized in a table or in the text and the figures could be moved to an appendix. I also believe that the illustration in Figure 14 is not necessary.

The point here is not to show "more detailed information than Figures 10-13", but rather to show <u>*different*</u> information than Figures 10-13. Figures 15-16 look superficially similar to 10-13, but they represent a fundamentally different situation (in which there is not just one nonstationary two-box model, but a whole collection of them, with different parameters. It is not obvious *a priori* that such a heterogeneous collection of nonstationary subcatchments would behave like the earlier much simpler system (in the sense that Fyw can still be reliably estimated, and MTT cannot). That is what these figures show. I also think Figure 14 is useful for explaining what has been done.

Remember that moving these figures to an appendix will not save any length; it will only mean that readers will have to flip back and forth between the appendix and the main text.

Section 3.7 – Interesting, but sometimes not clear how JK derived the data for Fyw of the different flow percentiles. This should be better explained in a method chapter, so the reader is able to follow the ways he calculates the data from the three catchments, which are not explained in the beginning in detail.

In principle a more tutorial explanation of the methods is possible, but at the cost of more text and possibly also another explanatory figure. In any case, I think any such explanation should be done here, not in a methods chapter that would not be understood without the necessary context (which readers only will understand once they have seen the results), and which probably would not be remembered by the time it is finally needed.

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 would expect from JK the highlights and his vision for the future following his ideas.

I disagree with MW's assertion that the paper is "very long". It is, for comparison, only about 20% longer than the HESSD version of Seeger and Weiler, 2014.

I do agree that it is unconventional to refer to individual figures in the conclusions, but 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.

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 includes already interpretation of the figure. In addition, shortening the names of the precipitation time series would help as well.

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.