

## ***Interactive comment on “Quantifying new water fractions and transit time distributions using ensemble hydrograph separation: theory and benchmark tests” by J. Kirchner***

**J.W. Kirchner**

kirchner@ethz.ch

Received and published: 8 October 2018

Let me begin by thanking Anonymous Referee #1 (hereafter AR1) for taking on the job of reading, understanding, and evaluating this manuscript. The paper is long, technical, and at times complex, and I appreciate the effort involved in reviewing it.

AR1 presents a nice overview of some salient features of the ensemble hydrograph separation approach, and points out that it could find application in re-analyzing many tracer data sets that have been collected in experimental catchments. I would, however, qualify AR1's statement that "*The method can take advantage of all these existing data sets to possibly refine our view of what MTTs are for low-order catchments in*

C1

*different regions, and what the form of the TTDs actually look like (at least the early-time portions of TTDs; the long tails remain difficult to constrain).*" I share AR1's view with respect to TTD's, but would like to comment further on the implications of the difficulty in constraining the long tails.

The difficulty in constraining long tails (from conservative tracer data) is not unique to ensemble hydrograph separation, but instead is shared by every estimation method. This is not a feature of the estimation methods, but instead is intrinsic in the conservative tracers themselves, because catchment tracer inputs lack sufficient long-timescale variability to leave a measurable long-timescale signal in the output [DeWalle et al., 1997; Seeger and Weiler, 2014; Kirchner, 2016]. Because the long tails of the TTD exert significant leverage over the MTT, the estimation of MTT's from typical conservative tracer data is simply infeasible by any method that does not invoke strong and unverifiable assumptions (such as an assumed shape of a transit time distribution). Thus I would not try to "refine our view of what MTT's are", beyond pointing out that we should seek more reliably quantifiable transit time metrics instead (such as new water fractions introduced here, or the young water fraction  $F_{yw}$  that I introduced in 2016).

AR1 raises an interesting challenge: "*I would ask the author to consider, and possibly elaborate in the discussion, about what might be the advantage of this method over the contemporary methods of lumped-parameter-transport modeling based on time-variable TTDs and SAS distributions/functions? For those researchers organizing new field studies and networks of sites, and that are reasonably competent and diligent with field-data collection (i.e. avoiding lots of data gaps), is there any reason to consider this method? One may argue that it is not any simpler to implement, and cannot achieve similar temporal resolution when estimating time-variable TTDs and SAS distributions. As discussed by the author, application to short subsets of data (for comparison of TTDs at different times) reduces sample size and enhances uncertainty in the regression approach.*"

Advantages of this approach, many of which are already mentioned in the text, include:

C2

a) First and foremost, this approach has been rigorously benchmark tested, and most other approaches have not. Thus we know how well it should work, and what its capabilities, biases, and uncertainties are likely to be, when applied to real-world data. This information simply does not exist for most other approaches.

b) This approach is not based on integration and therefore errors in the inputs do not accumulate over time, as they do in essentially every other method for estimating transit time distributions (except the spectral method of Kirchner et al. (2000, 2001). This is a significant advantage, because it is notoriously difficult to accurately measure catchment-scale inputs (either in quantity or concentration), even with "reasonably competent and diligent field-data collection".

c) What AR1 dismisses as a minor matter of being "reasonably competent and diligent with field-data collection" and "avoiding lots of data gaps" is actually a serious issue that is rarely taken seriously. Most existing methods for analyzing catchment data are based on convolutions (either explicitly, or embedded in a model). Convolutions require absolutely continuous input data - not just "avoiding lots of data gaps". There is a presumption in our field that small numbers of data gaps don't matter, as long as one fills them using some kind of ad hoc procedure. Whether this presumption is justified or not is anybody's guess, because it hasn't been tested. But simple mathematics would suggest that this presumption may not, in fact, be justified, because in convolutions, errors accumulate.

d) The existing methods for estimating time-variable TTDs generally do not estimate time-variable TTDs of real-world catchments at all; instead they measure time-variable TTDs of models that are fitted to catchment data, often with rather large prediction errors that the authors somehow manage to just dismiss with a wave of the hand ("large" errors mean those that are much larger than the putative model uncertainties). Whether these fitted TTDs are informative or misleading depends on whether the underlying mechanisms are actually correct, which is unknown and unknowable. (It's not good enough that the mechanisms "seem reasonable", because that says only that

C3

they agree with our preconceptions, not that they accurately describe the real world.)

e) AR1 says that, "One may argue that it is not any simpler to implement." I would argue the opposite. New water fractions can be estimated using bog-standard linear regression; really, how much simpler could that be? And TTD's in this approach are estimated with multiple linear regression, albeit with a few minor twists. I am now writing and testing an R script that handles all the calculations so that users don't have to figure out the details themselves (this script will be published separately, after it goes through several rounds of beta testing in our group). By contrast, the implementations of many other methods are anything but straightforward, in part because they often involve a series of ad hoc choices (among model structures, simulation algorithms, parameter calibration procedures, data assimilation methods, gap-filling protocols, etc. etc.) whose implications are unknown, because they haven't been tested.

f) AR1 also says that, "application to short subsets of data (for comparison of TTDs at different times) reduces sample size and enhances uncertainty in the regression approach," as if this issue were somehow unique to my approach. But of course it isn't unique; instead, it's inherent in any data-driven analysis. The crucial difference is that my analysis recognizes this problem and quantifies its implications, because I've bothered to look. The fact that others may not have bothered to look doesn't mean that their approaches are magically freed from the problem of data scarcity. Of course, one can disguise the data scarcity problem by assuming some rule (such as a SAS function) that specifies how the TTD varies over time, and then fitting this globally to the whole data set. But that puts the answer in from the beginning, because the relationships among the TTD's are determined by the researcher's assumptions about them (e.g. that they obey a SAS function), rather than letting the data determine what the answer is.

Specific Comments:

(1) *Page 5; line 24: Another complication would seem to be the fact that CQ(j)*

C4

*will be strongly temporally correlated with  $CQ(j-1)$ . Most perennial streams will be hydraulically connected to a riparian aquifer, and receiving discharge from that aquifer so long as the stream is gaining. The water volume in that aquifer may be 1 or 2 orders of magnitude greater than the volume of water associated with typical precipitation events.  $CQ(j-1)$  (i.e., old water) and  $CQ(j)$  should then be strongly temporally correlated. In other cases  $C_{new}(j)$  might also be strongly temporally correlated with  $CQ(j)$ , for example, in catchments with widespread impervious surfaces that induce infiltration-excess-overland flow. Whether or not the latter type of temporal correlation would exist would also depend on the temporal resolution of sampling. Is this potential for temporal correlation within the explanatory and dependent variables problematic for the linear regression approach here?*

No, for the simple reason that in ensemble hydrograph separation,  $CQ(j)$  and  $C_{new}(j)$  are both normalized by subtracting  $CQ(j-1)$ . Of course AR1 is correct that streams are mostly fed by aquifers with residence times that are much longer than individual precipitation events, and therefore tracer concentrations in streamflow are serially correlated. This has been known for decades, and many of us have written about it. The whole point of subtracting  $CQ(j-1)$  is to removing any effects that are common to both  $CQ(j)$  and  $CQ(j-1)$ , including legacies of past precipitation inputs, measurement biases, isotopic fractionation effects, and so forth. Subtracting  $CQ(j-1)$  from  $C_{new}(j)$  is usually not consequential in practice, because  $C_{new}$  will typically be much more variable than  $CQ$ , but it is done for the sake of consistency, and also to make the derivation conceptually analogous to conventional hydrograph separation.

In the paper's benchmark tests (e.g., Fig. 1), the temporal correlations described by AR1 are present (and often quite strong). Nonetheless, as the benchmark tests show, ensemble hydrograph separation yields quantitatively accurate estimates of the new water fractions and transit time distributions. That is the power of benchmark tests: one does not need to wonder, based on intuition, whether issues like this are, as AR1 puts it, "problematic". The tests show whether they are, or not.

C5

(2) *Page 6, line 3-5: What would qualify as small?*

As one can see from Eq. A7, "small" in this case means small in comparison to  $F_{new}$  itself. I will add this to the text.

*The condition in line 5 would not necessarily be true would it? In the case of stable-isotopes as tracers, if 1) rainfall is distributed uniformly throughout the year, 2) there is seasonal variation in the mean  $C_{new}(j)$ , and 3) streamflow is generated preferentially by precipitation that falls in the cold season, then this difference would not be expected to be zero, would it? These conditions are fairly common.*

AR1 is correct that one can imagine scenarios in which the mean of  $C_{new}(j)-CQ(j-1)$  may not be nearly zero (and I will change "should be nearly zero" to "should usually be small" to account for such cases). But the condition in line 5 is just one of two conditions, which are multiplied together. The second condition is that the slope of the relationship between  $F_{new}(j)$  and  $C_{new}(j)-CQ(j-1)$  should be small. This will normally be the case, and it makes the net effect on  $F_{new}$  small, regardless of what happens to the first term.

Again this points to the importance of rigorous benchmark tests - which, unfortunately, have not been conducted for other methods that have been used to estimate transit times from tracer data. One doesn't need to speculate about whether these issues are significant or not; one can simply do the test. Unlike other benchmark tests, the ones that I have conducted do not start by assuming that the assumptions underlying the method actually hold. Instead, they start with a wide range of reasonable assumptions that are not tailored to the method being tested, and test whether the method still works, even when its underlying assumptions do not hold.

(3) *Page 13, lines 8-10: Could you comment on the rationale for randomly assigning  $iA^2 d^2 H$  values to precipitation in this way, in light of the fact that actual  $iA^2 d^2 H$  values are often correlated with cumulative rainfall amounts/unit time? I ask only because the existence of this correlation would seemingly have bearing on condition 2 on page*

C6

6. If cumulative rainfall amount during a storm is not correlated with  $\delta^{18}O$ , then in principle either very negative  $\delta^{18}O$  values, or values near zero (a plausible upper boundary of the variable's range), could be observed for large storms that generate greater  $F_{new}$ . In that case condition 2 on page 6 might be questionable. Seemingly the random assignment here in some sense violates that condition, yet the results in Figures 1 and 2 are quite good?

One often sees extreme isotopic compositions at the beginning or end of a storm event, when precipitation rates are small. The water that makes up the bulk of the event usually has less extreme isotope values. But in any case, ensemble hydrograph separation will typically be applied to daily and weekly tracer time series (since these are by far the most widely available isotope records), and within-storm isotopic dynamics will be largely masked on those time scales.

(4) Page 26, lines 28-30 and Page 27, lines 1-3: Worst-case scenario based on what? Estimates of actual evaporative enrichment measured in surficial soil water? Or plausible ranges from models [e.g. Allison et al., 1983; Barnes and Allison, 1983]? Certainly in semi-arid environments, or where vegetation is interspersed (e.g. tree plantations or agricultural settings) there could be more than 20% enrichment of soil water before it percolates below the root zone. The fractionation effect could vary strongly at daily and weekly time scales depending on storm frequency and intensity, and potential evaporation (from plant surfaces and surficial soil). The fractionation effect could in fact be strongly correlated with storm size— $\delta^{18}O$  not a constant offset or random fluctuation that introduce no bias. And of course this simulation makes no consideration of spatial variability of the fractionation effect, which might occur due to substantial land-cover variability (e.g. catchments with mixed bare-soil and vegetated cover; wetland-meadow-forest transitions; partially snow-covered with spatially- and temporally-variable melt dynamics, etc.). Finally, it's worth noting that the isotope composition of infiltrating water can be different than precipitation (greater or lesser) due to evaporation from, or storage within, plant canopies [Allen et al., 2016]. These

C7

effects can be of variable sign and magnitude during individual storms. I would argue it is these storm-specific effects that matter not an average offset over the long term because these storm-specific effects influence the individual values of  $F_{new}$  - Cold in the regression model.

I think the concluding statement about this analysis (page 27, lines 1-3) should be more suggestive than definitive. People will read this and make the convenient assumption that a single precipitation collector is all that is needed in the field, rather than the more labor-intensive alternative of having multiple precipitation/throughfall/snowfall collectors (to capture some essence of spatial variability, and thus leading to more samples to analyze in the lab). This is an aspect of the method that should at least be examined in a few field studies in catchments that vary climatically and ecologically, rather than assumed unimportant based on this useful, but inconclusive, simulation. The same problem exists among most historical studies of mean transit times in forested catchments: the input data sets in the convolution integral are all biased to some unknown degree due to the spatial heterogeneity of storage, throughfall, and evaporation.

I agree that the concluding statement (page 27, lines 1-3) should be suggestive rather than definitive; that's why the statement said, "Figure 10 thus suggests that ensemble hydrograph separation should yield realistic estimates of new water fractions. . .". I will nonetheless remove the term "worst-case scenario", and instead say, "even with substantial confounding by evaporative fractionation."

Nonetheless, the scenario I used was indeed designed to be a worst-case scenario, in the specific sense that it is designed to lead to biased estimates of new water fractions. It intentionally creates an artifactual correlation between the degree of (seasonal) isotopic fractionation and the (seasonal) variation in the true isotopic ratio of the measured precipitation - and such an artifactual correlation could potentially skew the regressions that ensemble hydrograph separation relies on. By contrast, the effects of random variations in isotopic fractionation will tend to average out, and will not lead to biased new water fractions. Likewise, persistent biases from isotopic fractionation will also not bias

C8

the new water fractions, because regressions ignore constant offsets. Isotopic fractionation that is correlated with the true isotopic ratio of the measured precipitation has the greatest potential to mess up the results, which is why I've used that kind of fractionation here.

Of course one can imagine an wide variety of isotopic fractionation scenarios, as well as lots of other complicating factors like spatial heterogeneity. But let's keep this all in perspective. These complications will cause problems for *every* method for inferring transit times from tracer data. The standard of practice in our field seems to be to dream up new techniques and rush immediately to applications with field data, without bothering to test whether they really measure what we think they do, or without even quantifying any of the uncertainties involved. As far as I know (and I would be happy to be corrected on this), the kind of mathematical analyses that are performed in Appendix A - as well as, for that matter, anything nearly as rigorous as the benchmark tests in the rest of the paper - have not been performed for any of the other methods for inferring transit times from tracer data.

(5) *Page 29, line 10: Could you clarify here if lag time  $m$  corresponds to the most distant past time when a measurement of CP is available, or can the value of  $m$  be chosen even among those past times when CP was measured? You seem to be implying the latter case, but I don't understand why you would potentially lump multiple measured values of CP into the Colder term, rather than using the values individually in the regression analysis. Perhaps this is clarified later on in the text.*

The lag time  $m$  is simply the longest lag time for which one wants to estimate the TTD. This should not be the most distant time for which one has C\_P measurements, because any estimates on that time scale will be massively uncertain (due to a loss of statistical degrees of freedom). In any case, the method does not "lump multiple measured values of C\_P into the C\_older term" at all. Instead, as shown later on that page and the next page, the method uses C\_Q( $j-m-1$ ) in place of C\_older, in order to factor out long-term patterns in stream concentrations due to legacy effects of prior

C9

inputs (potentially even from before the beginning of the data series). In other words, we use the catchment to integrate over those past inputs - we don't do it ourselves.

(6) *Page 40, lines 15-20: This seems related to the question I posed in specific comment 1. If that is correct, perhaps you could foreshadow for the reader at that location in the text that this issue is discussed later.*

A better approach is to simply mention the function of C\_Q( $j-1$ ) as a "filter" for legacy effects of past inputs, directly at the point where the reader needs to know it (between equations 8 and 9). I'll do that.

(7) *Page 45, lines 15-25: Could you clarify in this case how the parameter  $m$  is determined? Seemingly if you want to isolate some subsets of the time series of Q for comparison of the average-backward TTDs, for example, Q during September versus April at Smith River, the CQj and distribution of water ages at any time during either interval could be strongly influenced by CP that occurred over many historical time intervals prior to September or April.*

This is the whole point of using a reference concentration: to correct for the legacy effects of more distant tracer inputs. The parameter  $m$  is not determined by analysis, but instead is simply chosen by the user. If you want to look at lags between 0 and 10 time steps, then  $m$  is 10. If you want to look at lags between 0 and 20 time steps,  $m$  is 20. Whatever the value of  $m$ , the streamwater concentration at lag  $m+1$  is used as a reference concentration (for both CQj and CP), to normalize for the inherited effects of precipitation tracers that fell prior to lag  $m$ . The smaller the value of  $m$ , the more important it is to subtract these legacy effects (and the more effectively the reference concentration does this).

(8) *Page 47, figures 16a,b: Regarding the orange line, it's interesting and counter-intuitive to me that these average TTDs would show that volume fractions of Q with ages of only a few days would be (slightly) more probable than volume fractions of Q with ages of several days to weeks. It rains so infrequently during the summer, most*

C10

*often there would be little to no water at all within the catchment that had residence time of only a few days. I would think there should be a slight increasing trend in these probabilities from left to right on the transit-time axis, up to a point where the trend then turns back downward. Is this possibly because those few storms that do occur during the summer (and deliver some water to the stream with age of only a few days) represent a disproportionately large fraction of total Q over the summer months?*

Intuition is a tricky thing. Remember that infrequent rainfall means that discharge will contain little zero-day-old water (because it probably didn't rain today), and also little one-day-old water (because it probably didn't rain yesterday), and also little two-day-old water (because it probably didn't rain the day before yesterday), and so forth, all the way out to the beginning of the dry season. The TTD's average these probabilities over the few days that had rain and the many days that didn't. Thus it's not possible, on average, to have (for example) more rain two weeks ago than rain one day ago, since each rain event is "two weeks ago" for a day two weeks into the future, and "one day ago" tomorrow.

(9) *Page 49, line 4: Certainly there are several good examples of where storage dependence has been examined, even one case in the same climate as the Smith River. These are worth acknowledging [Benettin et al., 2013; Harman, 2015; Heidbuchel et al., 2013; Heidbuchel et al., 2012; Rodriguez et al., 2018; van der Velde et al., 2010; van der Velde et al., 2012].*

The statement in question is, "Antecedent wetness has been recognized as a controlling factor in catchment storm response (e.g., Detty and McGuire, 2010; Merz et al., 2006; Penna et al., 2011), but its effects on solute transport at the catchment scale have not been widely explored." I meant to refer to solute transport effects estimated from catchment data (of which there have indeed been few or none), as distinct from the behavior of calibrated models (of which there have been several, as pointed out by AR1). All of the references mentioned by AR1 concern inferences drawn from the behavior of calibrated models, where it is generally unclear whether the storage-dependence is

C11

strongly constrained by the data, or whether it would come and go within the plausible ranges of the model parameters, or even whether it results from arbitrary model choices (such as the frequent use of linear reservoirs). None of the mentioned references provide anything remotely resembling Figs. 18a and c, which show the functional dependence of "new water" on antecedent moisture, directly from catchment data (and as far as I can tell after re-reading these papers, some of them do not address the general topic directly at all).

In any case, a more precise statement would be, "Antecedent wetness has been recognized as a controlling factor in catchment storm response (e.g., Detty and McGuire, 2010; Merz et al., 2006; Penna et al., 2011), but its effects on solute transport at the catchment scale have rarely been quantified, outside of the context of calibrated simulation models (e.g., Heidbüchel et al., 2012; van der Velde et al., 2012; Harman, 2015; Rodriguez et al., 2018)." This will be adopted in the revision.

(10) *Page 51, line 19: I would argue that point 5 here should perhaps be omitted. While suggestive, I don't think the simulation that leads to this conclusion is a very realistic representation of isotope fractionation effects over time or space. See specific comment 4 above. Also, while I'm no proponent of traditional hydrograph separation, I don't think the first sentence in point 5 is really true. The method is not necessarily vulnerable to biases in tracer measurements resulting from fractionation. If fractionation has occurred in the end member, the effect should be apparent in the measured delta values. Fractionation doesn't inhibit accurate quantification of the tracer concentration in the end member.*

The statement in question refers to biases (i.e., persistent offsets) in tracer measurements, and as such, it is mathematically correct as stated.

Concerning conventional hydrograph separation, the problem isn't that fractionation "inhibits accurate quantification" of the tracer concentration, it's that the accurately quantified tracer concentration does not represent what actually goes into the catchment.

C12

One measures the tracer concentration in a rainfall collector, for example, not the (fractionated, or to put it more precisely, differently fractionated) throughfall or soil moisture that eventually becomes streamflow. The difference between what is measured and what is the actual end-member can potentially distort the results of traditional hydrograph separation, whether that difference fluctuates randomly (and thus adds random noise to the results) or is persistent over time (and thus adds persistent bias to the results). Neither random fractionation nor persistent fractionation bias poses a significant problem in ensemble hydrograph separation (the first will tend to average out, and the second will not affect the regression slopes on which the method depends). There are some particular patterns of fractionation effects that could pose a problem for ensemble hydrograph separation (and the paper specifies what they are), but they would also pose at least as big a problem for conventional hydrograph separation.

(11) *Page 52, lines 7-9: Perhaps put "time-invariant" in italics here for emphasis, since the focus of both of those papers was of course to demonstrate that, while potentially more temporally stable than TTDs, even SAS distributions should be considered time-variant in most cases.*

Emphasizing "time-invariant" in this way could lead readers to believe that time-variant SAS functions somehow avoid the estimation problems mentioned in this paragraph, which is not correct. The better solution is to simply remove the phrase "time-invariant" as it pertains to SAS functions.

(12) *Page 52, lines 11-14: And note some important works preceding those you have cited [e.g. Ali et al., 2014; Fiori and Russo, 2008; Fiori et al., 2009; Rinaldo et al., 2011; Russo and Fiori, 2009].*

These are interesting papers but they are not relevant to the statement in the text, which is, "Yet another approach that is coming into more frequent use is to calibrate a conceptual or physically based model to reproduce, as closely as possible, the observed hydrograph and streamflow tracer time series, and then infer the catchment transit time

C13

distribution from particle tracking within the model." The papers that AR1 mentions all concern simulations of transit time distributions under various physical or conceptual assumptions, but none of them infer real-world transit time distributions by calibrating those models to observational data. Thus none of them are relevant citations for the specific statement made in the text. (If the topic were simulations of transit time distributions more generally, then I would also cite Kirchner et al. 2001, which precedes any of the papers mentioned by AR1 by almost a decade.)

(13) *Page 52, lines 15-19: This paper [Pangle et al., 2017] provides a clear illustration of your point, where multiple hydrologic variables, and a tracer breakthrough curve, can be simulated quite accurately with traditional flow and transport models, while still not accurately reproducing the known age distributions of water in the flow out of the system.*

If Pangle et al.'s (2017) results show this, it's unfortunate that Pangle et al. don't come right out and say so. In particular, although their section 3.3 says that different parameterizations reproduce the observed Q, S, and H (discharge, storage, and hydraulic head), it does not say that these parameterizations also reproduce the tracer breakthrough curve. Thus one needs to "read between the lines" to draw the conclusion that AR1 draws from Pangle et al. (a conclusion which the authors themselves do not draw, at least in any way that I can find).

Technical Edits:

(1) *Page 2, equation 2: Should you have subscript j along with subscripts "new" and "old" attached to C terms in equations 2 and 3? In practice they are not uniquely measured at every time step, but must at least be assumed at each time step.*

No. Equation 2 describes conventional hydrograph separation, in which (see line 15) "one assumes that streamflow is a mixture of two end-members of fixed composition" (emphasis added). If the composition is fixed then there is no subscript.

C14

(2) *Page 61, equation A5: Is it redundant to use overbars and angled brackets on the same term?*

In general it's not. Angled brackets indicate that whatever is contained between them is averaged. Overbars indicate individual terms that are averaged. Thus, for example,  $\langle \bar{x}y \rangle \neq \langle xy \rangle$ . In one case in equation A5, the angled brackets are indeed redundant, because  $\langle \bar{\beta x} \rangle = \bar{\beta x}$ . But in this case the brackets are retained for clarity, since all the angle-bracketed terms in (A5) result from the expansion of the angle-bracketed term in (A4).

(3) *Figures 4, 6, and 8: Some kind of discontinuity in the vertical axes of the some of the graphs. Maybe due to pdf rendering? Not important, just bringing it to attention in case it can be easily fixed.*

This is an annoying pdf rendering issue. The original eps files do not have these defects, so they will not appear in the final version.

(4) *Page 28, line 24: Should the subscripts be j here rather than i? Since times associated with sampling P have been denoted with i whereas those for Q with j?*

Correct! You have clearly read the manuscript very carefully. This is a silly typo that will be fixed in the final version.

(5) *Page 30, line 17: Looks like a typo on the subscript of second term on the right-hand side of the equal sign.*

Good eyes! This is a rendering error in MS Word's equation editor. It will be fixed.

–

–

DeWalle, D. R., Edwards, P. J., Swistock, B. R., Aravena, R., and Drimmie, R. J.: Seasonal isotope hydrology of three Appalachian forest catchments, *Hydrol. Process.*, 11, 1895-1906, 1997.

C15

Seeger, S. and Weiler, M.: Reevaluation of transit time distributions, mean transit times and their relation to catchment topography, *Hydrol. Earth Syst. Sci.*, 18, 4751-4771, doi:10.5194/hess-18-4751-2014, 2014.

Kirchner, J. W.: Aggregation in environmental systems - Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments, *Hydrol. Earth Syst. Sci.*, 20, 279-297, doi:10.5194/hess-20-279-2016, 2016.

Kirchner, J. W., Feng, X., and Neal, C.: Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations, *J. Hydrol.*, 254, 81-100, 2001.

C16