Author's response (HESS-2015-62)

My responses to the reviews were submitted as part of the open review process; they are reproduced below for reference.

Summary of changes in the revised manuscript:

As requested, I...:

1) analyzed the sensitivity of young water fractions to the assumed shape of the transit time distribution (section 4.2 and Fig. 12 in the revised manuscript),

2) added more explicit step-by-step instructions for how to estimate young water fractions from amplitude and phase information (in section 4.4 of the revised manuscript),

3) added subheadings to section 5, and

4) outlined several ways that the young water fraction could be put to use in catchment hydrology (section 5.5 in the revised manuscript).

Author's response to Anonymous Referee #1

I appreciate Anonymous Referee #1's comments and suggestions. Where possible, these will be used to improve the manuscript on revision. Specific responses to individual comments are detailed below.

GENERAL COMMENTS

This is an interesting paper that explores which information can be gained, in terms of catchment transit times, from the analysis of seasonal tracer cycles. The paper is written in a clear form that makes it easy to read. The contents can be divided into two parts: 1) it is shown, through rigorous benchmark tests based on a virtual experiments, that the stationary travel time distributions estimated from seasonal tracer cycles are typically unreliable and biased towards younger mean transit times. 2) a new metric (the “young water fraction”) is introduced that can be more accurately derived from tracer cycle information. The results suggest that, for a range of plausible TTDs (apparently, every TTD that can be derived by combining gamma distributions with shape parameter alpha in [0.2, 2]), the amplitude ratio derived from sine wave fitting is representative of the fraction of water younger than about 1.5 – 3 months.

Thank you for your supportive comments and concise summary. To be precise, seasonal tracer cycles have not (and cannot) be used to estimate "stationary travel time distributions" ***per se*****; instead, they have been used to estimate** ***parameters*** **for travel time distributions whose shape must be assumed** ***a priori*****.**

Both the parts are of good scientific significance, but while part 1 is also straightforward and easy to understand, part 2 is at times unclear to me. Considering that part 2 is the basis for Paper 2, and that potentially the method will be widely used in the future by the scientific community, it would be advisable to revise part 2, so as to permit a better understanding of the contents. Below, I included some comments that may help making the manuscript clearer:

i) There is some ambiguity between the general idea of “young water fraction” used in common speaking and the specific definition “young water fraction” developed by the author (the fraction of particles younger than 2-3 months). It may be desirable using a different name for the new variable defined by the author, to avoid this ambiguity.

I'm not sure what "general idea... in common speaking" is being referred to, so it is difficult to comment in detail. Clearly it is important to avoid, where possible, semantic confusions in science. I adopted the term "young water" precisely because, when it has been previously used in hydrology, its meaning has been consistent with the sense in which I use the term here. Obviously the ***threshold*** **that separates "young" and "old" water will vary depending on the dating technique that is used (^{3}H, ^{14}C, CFC's, and so forth), but the general concept is the same.**

Regarding the specific term "young water fraction", a Web of Science search for this phrase gives no hits at all, suggesting that the risk of confusion is low.

ii) The Fyw is an interesting and promising concept, but its definition in real catchments is not easy to digest because it is affected by the imprecision in determining the threshold age (on the other hand MTT has a very intuitive definition, but it is an uncertain metric). The paper would benefit from a deeper analysis of how the threshold age varies in the virtual experiment when the tributaries are aggregated (see Detailed Comments on Section 4.1).

Thanks for this comment. The MTT does seem more intuitive, and it is certainly more familiar because it has been used in catchment hydrology for a long time. However, in principle it is no more precisely defined than the young water fraction is, given that both depend on the shape of the assumed travel time distribution (TTD). Indeed the MTT is arguably much *less* precisely defined than Fyw is, because plausible variations in the shape of the TTD lead to order-of-magnitude uncertainties in MTT (but much smaller uncertainties in Fyw) for any given rainfall and runoff tracer time series.

In its most basic sense, the young water concept marks a shift from trying to estimate the *statistical moments* of the TTD (which are sensitive to the shape of the distribution across its entire range), to determining individual *fractions* of the distribution. (The precise statistical term is *fractiles*, but I don't mention this in the paper because it is too easily confused with *fractals*). In principle, these fractions can be more reliably determined than MTT because they depend only on the total mass of the distribution that lies above or below the threshold age, and not on how *far* above or below that threshold it lies.

Nonetheless there are key distinctions to be drawn between (a) the general concept of a young water fraction, namely, the fraction younger than a threshold age, (b) a particular young water fraction, namely, the fraction younger than some *specific* threshold age, and (c) the result of a particular procedure designed to *estimate* this fraction.

iii) The author often mentions the catchment "spatial heterogeneity" and the related "aggregation error". However it is not clear what the author's definition of "heterogeneous" and "homogeneous" is. This has implications, because the essence of the problem with the traditional derivation of MTT from sine wave fitting methods is the use of a wrong assumption on the TTD shape. I would call this an error caused by the wrong assumption of using a simple TTD for a complex system, and I don't see why the author calls this an "aggregation" error.

"Aggregation error" is a technical term referring to the idea that a method of analysis that works correctly at one level of aggregation may fail at a higher level of aggregation (see O'neill and Rust , Aggregation error in ecological models, Ecological Modelling, 7, 91-105, 1979; Gardner et al., Robust analysis of aggregation error, Ecology, 63, 1771-1779, 1982; Rastetter et al., Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems, Ecological Applications, 2, 55-70, 1992; Kaminski, et al., On aggregation errors in atmospheric transport inversions, Journal of Geophysical Research-Atmospheres, 106, 4703-4715, 2001).

I use the term "aggregation error" in this case because traditional estimation methods that correctly determine the MTT for subcatchments with a specified TTD shape (say, for example, exponential) will fail when applied to larger catchments that aggregate those subcatchments (even if each subcatchment has exactly the same shape, but just a

different MTT). Such a catchment is "heterogeneous" in the specific sense that the subcatchments have different MTT's. The traditional estimation methods would work correctly across a range of scales, but only if the catchment were homogeneous (in the specific sense that the subcatchments had the same shapes and MTT's).

When subcatchments with the same TTD shape, but different MTT's, are aggregated together, one gets a different TTD shape at the higher level of aggregation, as I point out in the paper (see Figure 6). Thus one could say that this is just a problem of not assuming the correct TTD. But then how are we going to assume the correct shape, and how will we know when we have done so? We almost never have good catchment-specific constraints on the shape of the TTD, either from physically based theory or from data, and the results of any MTT calculation will be quite sensitive to whatever one assumes about the TTD shape (see, e.g., Kirchner et al., Comparing chloride and water isotopes as hydrological tracers in two Scottish catchments, Hydrological Processes, 24, 1631-1645, 2010). particularly about the long tail, which cannot be constrained by conservative tracer data).

iv) The paper presents several interesting inferences on the relationship between the amplitude ratio and the Fyw. As these are not causal relationships, one would expect to see a paragraph with a summary of the fundamental working hypotheses (e.g., the shape parameters alpha in [0.2, 2]), that can guide the reader towards the limits of applicability of the outlined method.

Defining limits of applicability in a rigorous mathematical sense is tricky, because there are all sorts of exotic TTD shapes that are theoretically possible, even if they are unrealized in practice. Strictly speaking the paper only demonstrates the stated results for gamma distributions with shape factors of 0.2-2. This is rather explicitly stated in section 2.1, and shown in Figure 2 (with shape factor ranges specified on page 3068, lines 18-20). But the general principles should hold for many different TTD's that have similar overall shapes to those specified. To avoid any confusion, I will remove the alpha=8 curves from Figure 2 and state in the caption that this figure shows the range of TTD's considered in this analysis.

v) Sections 4.1 and 4.3 include details that are not always clear to me, and should be better explained (see Detailed Comments). In particular, I could not find in the manuscript a description of how to incorporate the phase shift information in the determination of Fyw.

You are right, this is not as explicit as it should be. In the revised paper, I will include a "cookbook" procedure for determining Fyw using phase shift information.

DETAILED COMMENTS

3063, l. 3: It would be important to better define the working framework at the beginning of the paper. The author may mention here that the flowpaths and the catchment connectivity change in time, potentially by large factors. The catchment has no stationary behavior and stationarity is a legitimate assumption, but it must be stated that it is an explicit assumption, which allows the use of one TTD instead of several TTDs. The author may also move up here lines 3-14 of page 3066.

I would prefer to briefly mention it here, and keep the longer treatment on 3066. One needs to be careful about breaking other logical connections when one moves things around.

l. 7: (connected to comment on line 3) “have simply assumed that the TTD is stationary and has a given shape”

Something like this can be done.

l. 8: it is not so “obvious” to me that MTT is the ratio between storage and fluxes. While it surely is for a well-mixed system (which produces an exponential TTD), I am not so confident that the same holds for other storage mixing hypotheses.

This is a general mathematical result. It does not require that the system is well mixed; it just requires that the system is stationary and that no component of the system is completely immobile (and thus has infinite residence time).

l. 16-19: it may be appropriate referring to the recent commentary on WRR by McDonnell and Beven (2014) on this topic.

The point goes back at least five decades, to Hewlett and Hibbert (1967) and Horton and Hawkins (1965). I can add some references.

3070 l. 5: Eq. (7) is not enough to derive Eq. (8). Maybe start the sentence with “from Eq. (1) and Eq. (7), using the Fourier Transform properties, one can. . . ”

Sure.

3075 l. 6-14: This is a very important result, and should be better explained. As the author says, it is not intuitively obvious (and it is actually quite surprising) that the tracer cycle amplitude in the mixture is almost exactly equal to the average of the tracer amplitudes in the two tributaries. This looks like an interesting property of the gamma filtering for the shape parameters investigated by the author, where the damping of the tracer cycle prevails over the shifting. Other filters would not behave the same (e.g. gamma distributions with shape parameter alpha>2?), suggesting what the limits of applicability of the method are. Indeed, one may expect the same behavior from the advection-dispersion model TTDs derived by Kirchner et al., 2001, and not from the lognormal TTDs reported by Selle et al., 2015.

I'm not sure why the reviewer says that this behavior would not be expected for the log-normal TTD's reported by Selle et al.; one would need to do the analysis. I have tried these mixing experiments and have found that the reported result is widely observed. The exceptions are cases where the distribution is narrow, with a large offset from zero (such as a gamma distribution with a shape factor alpha much larger than 2). However, these do not seem to be plausible shapes for catchment transit time distributions. Indeed, the distributions reported by Selle et al. were obtained only with highly un-natural experimental conditions, in which the tracer was applied only to a small part of the catchment, and no tracer was applied close to the channel or close to the catchment outlet. Thus the data shown by Selle et al. resemble point-source breakthrough curves more than they resemble whole-catchment transit time distributions.

Section 4.1: 3076 l. 15: at this point in the paper it seems like it is the opposite: you look for the threshold age for which the Fyw closely approximates As/Ap across a wide range of scale factors. I would suggest stressing that the existence itself of one single threshold age, which verifies almost exactly the equality Fyw=As/Ap for very different scale parameters, is already an interesting result.

I don't understand the first statement (the opposite of what?). I agree that it is interesting that, for a given shape factor, one can define a threshold age for which Fyw is approximately As/Ap. But if this were all that could be shown, we would have the same problem that we have with MTT estimation: we could estimate Fyw for any TTD shape, but how do we know what that shape is?

Therefore the really important result is that, across a wide range of TTD shapes (as shown in Figure 2), this threshold age varies so little. The implication of this result is that we don't need to know the TTD shape, as long as it looks something like any of the curves in Figure 2 (except the one with alpha=8, which I will remove in the revision). No matter what the shape is, as long as it looks something like these, we can quantify Fyw and know that we are talking about the fraction of water younger than something like 2-3 months.

And an even more important result, which unfortunately can't be introduced until section 4.2, is that this same principle holds for TTD's that are created by mixing together widely varying gamma distributions, with different shape factors and scale factors. These aggregated distributions are not gamma distributions! This implies that the same general principles hold for a very broad class of distributions, well beyond the gammas for which the results in 4.1 were derived.

3076 l. 20 to 3077 l. 4: in this paragraph there is a fundamental perspective shift that needs to be explicitly clarified. Before this point, the young water fraction was defined to be equal to the amplitude ratio. After this point, due to the results shown in Figure 9, the perspective changes and the amplitude ratio will be always assumed to be a good predictor of the relative amount of water younger than 2-3 months. If this is not stated clearly, the sentence will sound circular (the amplitude ratio is a good predictor of a new variable that has been explicitly defined to be equal to the amplitude ratio!).

Yes, I get the point and can revise the text accordingly. The point is not that Fyw is *<u>defined as</u>* the amplitude ratio, but that the amplitude ratio is a good *<u>estimator for</u>* Fyw. Furthermore, for many different TTD shape factors, the "young water" threshold falls in the narrow range of about 2-3 months. Thus the amplitude ratio is a good estimator of the fraction of water younger than 2-3 months.

3077 l. 11: “leads to the important result”. Is it not a hypothesis that is going to be demonstrated, rather than a result?

It is a theoretical result, which is then numerically tested. Perhaps "implication" is a more precise word than "result".

3077 l. 15-20: from the same procedure used to determine Fyw for the gamma distribution, it would be possible to determine the “real” young water fraction (as well as the “real” threshold age) in the mixed runoff. So why did the author not perform this test? It would make the statement “the amplitude ratio predicts the young water fraction also in the combined runoff from heterogeneous landscapes” much stronger.

If I correctly understand what the reviewer is saying here, this is in fact what I did. The key issue again seems to be the distinction between what Fyw *is* (namely, the fraction of water younger than some threshold age), and ways that the *value* of Fyw can be *estimated* (for example, from As/Ap for certain threshold ages).

Moreover, it would be interesting to see the effect of the aggregation on the threshold ages (particularly from tributaries with different shape parameters). Does a single threshold age still verify the equality As/Ap=Fyw, for different parameters alpha? Do the threshold ages fall in the same 2-3 months range in the mixed runoff? Do they average linearly?

Perhaps it's now my turn to be confused! Threshold ages are not properties of the system, unless one specifies a criterion for setting the threshold. If we specify the criterion as, "the threshold age is the one for which Fyw, the fraction younger than this age, is closely approximated by the amplitude ratio As/Ap", then this threshold age is about 2-3 months. This also holds for the mixed runoff (otherwise the aggregation of Fyw wouldn't work correctly).

I don't understand what is meant by "Does a single threshold age still verify the equality As/Ap=Fyw, for different parameters alpha?" There is a range of threshold ages, not a single threshold age, for different alpha values.

Section 4.2: same comment as 3077 l. 15-20: the “real” young water fractions and threshold ages could be determined from equation 16. So is the amplitude ratio a good predictor of the “real” young water fraction?

Yes, that's exactly the point.

This would really make the young water fractions independent from the gamma distributions they were initially defined from.

Yes, that's exactly the point.

Also, is there any hint on what causes the larger departures from the 1:1 line in Figure 11 and Figure 12? Could it suggest anything for the limits of applicability of the method?

I have not analyzed this comprehensively, but in general, larger contributions from subcatchments with higher alphas produce more scatter, and if one extended the range of alpha values to (say) 4, there would be visibly much greater scatter. Let's be clear: the "larger departures" in Fyw estimates in Figures 11 and 12 are on the order of single-digit percents, whereas the uncertainties in MTT are hundreds of percent.

Section 4.3: it is really unclear how the phase shift can affect the determination of the young water fraction, as it does not appear anywhere in its definition. So I am not able to interpret Figure 13 a-c.

As noted above, in the revised manuscript I will provide step-by-step instructions on how can include phase information in estimating Fyw.

3083 l. 19: “the most useful metric” seems like an overstatement.

Sorry, that came out sounding rather immodest, didn't it? I only meant to say that Fyw was a more useful metric than MTT.

Section 5: The uncertainty induced by sine-wave fitting is not mentioned (while it is, briefly, in Paper 2). In my opinion, the manuscript would benefit from a simple analysis on how the uncertainty in sine wave fitting translates into uncertainty in the estimation of the young water fractions.

If there were "a simple analysis" I would have included it, but to treat this rigorously probably requires another ~6 pages, ~8 equations, and several figures to illustrate the results... and the paper is rather long already. In any case, probably the most important sources of uncertainty are not in the data-fitting itself, but in the assumptions underlying the interpretation of the data (as outlined on p. 3084, l. 17-23).

Besides showing that Fyw is a reliable metric while MTT is not, the paper does not suggest what the young water fractions can be used for. This is partially addressed in Paper 2 (section 3.7), but some hints also in paper 1 would make the impact of the manuscript stronger.

Thanks for this suggestion. I will see if I can helpfully foreshadow the applications that are outlined in Paper 2. Most obviously, Fyw directly quantifies the fraction of water flowing by relatively fast flowpaths (where "relatively fast" means faster than a few months).

TECHNICAL CORRECTIONS

3078 l. 3: minimal

Figure 11 caption: horizontal axes

Literature cited:

McDonnell, J. J., & Beven, K. (2014). Debates - The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. Water Resources Research. http://doi.org/10.1002/2013WR015141

Kirchner, J. W., Feng, X., & Neal, C. (2001). Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations. Journal of Hydrology, 254(1-4), 82–101. http://doi.org/10.1016/S0022-1694(01)00487-5

Selle, B., Lange, H., Lischeid, G., & Hauhs, M. (2015). Transit times of water under steady stormflow conditions in the Gårdsjön G1 catchment. Hydrological Processes, (in press)

Author's response to Anonymous Referee #2:

I appreciate Anonymous Referee #2's comments and suggestions. Where possible, these will be used to improve the manuscript on revision. Specific responses to individual comments are detailed below.

This paper explores how mean transit times (MTT) derived from seasonal tracer cycles aggregate when scaling up by adding small catchments to represent a larger catchment. Kirchner finds that the MTT does not scale well at all by performing thorough benchmark tests and he proposes a new metric: the young water fraction that by its definition scales as good as possible with spatial heterogeneity. Thus 2 main messages in this paper:1) From heron, do never use MTT again, 2) use Fyw instead. This new metric is interesting, but at the same time challenging to use as its definition contains uncertainty (i.e. the type of transit time distribution, which is always unknown).

Furthermore I fully agree with the call for thorough benchmarking of simple hydrological models in the face of spatial and temporal heterogeneity. The paper is well written and is an important contribution to hydrology.

Many thanks for your kind remarks about the paper. Regarding the point that the transit time distribution is always unknown: yes, but this is also true (and much more consequential) for mean transit time determinations, where it leads to order-of-magnitude uncertainties in MTT.

However, several questions remained after reading the paper:

How can I use the Fyw (fraction young water) with my data? Is the approach something like: First guess a range of alphas that are likely to represent my system, Let's say 0.3- 1.5. Then derive the Thresholds Times with Eq 14,–> 0.12-0.22 years. Next derive from data the As/Ap. For example 0.3. This then means that around 30% of my stream water is younger than 0.12 to 0.22 years? Next we can refine this approach by including the phase shift? Between catchments we can now compare this Fyw. I think this could be explained more clear in this paper, for example with the data of figure 1.

The suggestion of a "worked example" is a good one. I will see whether I can fit it in, without making the manuscript too much longer.

What is the advantage of comparing Fyw over As/Ap between catchments?

The advantage is that Fyw tells you something about transit times, and As/Ap doesn't (at least not directly). In the second paper, for example, I show how these methods can be used to (for example) quantify how the young water fraction (and thus the fraction of water flowing by relatively fast flowpaths) varies between high flow and low flow. Furthermore the relative fractions of young and old water (and their variations with discharge regime) can be compared to stream chemistry, to define the chemical fingerprints of "young" and "old" end-members. Following Referee #1's suggestion, I

will try to give the reader a taste of these potential applications already in the first paper, although they will not be spelled out in detail until the second paper.

What about evapotranspiration? Are the proposed methods valid when half of the water balance goes to evapotranspiration? You convincingly proofed that the MTT does not scale up well, but does FYW still scale well with evapotranspiration? Page 3069, line 9, seems to suggest it does not, but with amplitudes I can imagine it does work. Does this need further benchmarking?

Do you mean line 20 on page 3066 instead? In practice, convolution-based approaches (including those used to estimate MTT) ignore evapotranspiration (ET). Estimating how ET would affect Fyw determinations is not straightforward, because this will depend on how ET alters the concentrations of the conservative tracer. Thus this effect will differ, for example, between stable isotopes and chemical tracers. I am currently working on a manuscript that looks at this question for stable isotopes, but this is a rather complex topic that is well beyond the scope of the current paper.

Following your own reasoning on page 3070, line 20, a catchment consists of almost infinite number of flow routes, each with own travel times. All these flow routes are grouped to yield the catchment TTD. You showed that Fyw scales well for 8 subcatchments, but does it still scale well for 1.000.000 sub-flow routes?

By extension it should, but numerically demonstrating this would be computationally tedious, and well beyond the scope of the current paper.

Is there any chance that due to the central limit theorem an infinite number of weakly-related gamma distributions for each flow route (log-transforming them, adding them yielding a normal distribution, and back-transforming them to yield a log-normal distribution) yields a log-normal TTD distribution at the catchment outlet of which the MTT does scale well as long as we assume that the central limit theorem holds at all the subcatchments as well at the catchment? I dont think so, but Im also not entirely sure that the Fyw does much better.

I don't understand the reasoning here. The point is not whether MTT scales well (by definition, the mean will always aggregate linearly), but whether *a procedure for estimating* MTT will work correctly, when the only inputs are observable behaviors (like tracer concentrations) that come from heterogeneous aggregates of subcatchments.

Minor comments: Title: As the authors refers in both papers to “paper 1” and “paper 2”, it would be good to include this number somewhere in the title of the paper as the papers are likely to end up in reverse order on a website (like now on HESSD).

The original manuscripts had such numbers in the title, but these had to be removed because of problems that would be created for any future papers in this series, which may appear separately. In the revision, I will try to clarify how each paper refers to the other one.

Page 3066, line 16: one can relax... flow-equivalent time. I dont think it is possible to express time as flow-equivalent time when sine wave fitting. Thus this statement is confusing to me in the context of this paper.

Obviously, a mathematical sine wave will no longer be a pure sine wave if the time base is locally stretched and shrunk in a non-uniform way. But in practice sine waves are fit to rather noisy tracer data, so it's not clear how much this will affect the fitted sine wave. In any case, the statement is a general one about convolution methods, and is not specific to sine-wave fitting (which is introduced three pages later).

Page 3066, line 13: “However in practical applications”: this statement renders all the above references impractical, while the objective of using time variant TTD actually is to be a bit more practical. To me considering a catchment as a stationary flux field is totally theoretical and only suited for catchment intercomparison studies. These stationary studies hardly have any practical relevance in helping to understand how to lower or mitigate solute fluxes.

What I meant was, "in applications using real-world data". I did not intend to label time-variant TTD approaches as impractical, since that is not an issue one needs to get into here. Your comment does, however, point to an important issue. There is a rich theoretical literature on time-varying TTD's, but it is only now starting to come to grips with the important problem of how we can determine what these TTD's actually are, in the real world, based on real-world data. This is not at all a simple problem. Since the first requirement of any approach to practical problems is that we must be able to use it reliably with real-world data, this represents a substantial challenge for time-variant TTD methods. Although the present paper deals with stationary (but heterogeneous) systems, the second paper shows that these methods can also be applied to data from nonstationary (and heterogeneous) systems.

Page 3078 line 25. Following your reasoning on page 3071, line20, each subcatchment consists of an almost infinite number of independent flow paths that contribute to stream discharge. Do you think you still get the results of figure 12 for an infinite number of subcatchments? Is this what you are saying on page 3079, line 3?

Yes, that is what I am saying.

Page 3080, line 25 MTT values derived from seasonal tracer cycles

Correct. But note that (as explained on pp. 3081-3082), there is little reason for optimism that other methods of estimating MTT from tracer data will be any more reliable. There are several reasons for this. First of all, MTT depends strongly on how long the long-time tail is, but conservative tracers are insensitive to such long-term behavior (either because the tracers themselves don't exhibit much decadal-scale variation, or because we don't have measurements that run that long). Secondly, to the extent that the seasonal cycle is the dominant feature of many tracer time series, that cycle will largely control the results obtained from those time series, no matter what methods are used to fit or interpret the data. Sine-wave fitting just happens to be the simplest and most analytically tractable of those methods, which is why I have studied it here.

Page 3082, line 10. Im not entirely sure what you mean with the time series convolution approach. If it refers to methods that solely use the waterbalance (water storage and water fluxes time series) to calculate the MTT, this aggregation bias is likely to be absent. At a larger or smaller scales this approach leads to a new water balance with new storage and water fluxes, which lead to a new MTT independent of the aggregation (close to [average Storage] / [average precip].

I mean time-domain convolution of tracer time series. Note that water balance methods lead to highly biased estimates of MTT, since at best they only measure dynamic storage and not passive storage.

However, I fully agree that MTT is an awful and often meaningless metric to use. Median traveltimes or indeed Fyw are much more meaningful.

Page 8030, line 19. You mean to say that Fyw is more useful than MTT?

I think you mean page 3083, and yes, that's what I meant to say.

Author's response to Referee #3 (Markus Weiler)

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.

Aggregation in environmental systems: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments

J. W. Kirchner[1,2]

[1]{ETH Zürich, Zürich, Switzerland}

[2]{Swiss Federal Research Institute WSL, Birmensdorf, Switzerland}

Correspondence to: J. W. Kirchner (kirchner@ethz.ch)

Abstract

Environmental heterogeneity is ubiquitous, but environmental systems are often analyzed as if they were homogeneous instead, resulting in aggregation errors that are rarely explored and almost never quantified. Here I use simple benchmark tests to explore this general problem in one specific context: the use of seasonal cycles in chemical or isotopic tracers (such as Cl^-, $\delta^{18}O$, or $\delta^{2}H$) to estimate timescales of storage in catchments. Timescales of catchment storage are typically quantified by the mean transit time, meaning the average time that elapses between parcels of water entering as precipitation and leaving again as streamflow. Longer mean transit times imply greater damping of seasonal tracer cycles. Thus, the amplitudes of tracer cycles in precipitation and streamflow are commonly used to calculate catchment mean transit times. Here I show that these calculations will typically be wrong by several hundred percent, when applied to catchments with realistic degrees of spatial heterogeneity. This aggregation bias arises from the strong nonlinearity in the relationship between tracer cycle amplitude and mean travel time. I propose an alternative storage metric, the young water fraction in streamflow, defined as the fraction of runoff with transit times of less than roughly 0.2 years. I show that this young water fraction (not to be confused with event-based "new water" in hydrograph separations) is accurately predicted by seasonal tracer cycles within a precision of a few percent, across the entire range of mean transit times from

Style Definition: Normal: Left
Style Definition: Heading 1: Left
Style Definition: Heading 2: Left
Style Definition: Heading 3: Left
Style Definition: Heading 4: Left
Style Definition: Heading 5: Left
Style Definition: Heading 6: Left
Style Definition: Heading 7: Left
Style Definition: Heading 8: Left
Style Definition: Heading 9: Left
Style Definition: Footer: Left
Style Definition: Title: Left
Style Definition: Subtitle: Left
Style Definition: Comment Text: Left
Style Definition: Comment Subject: Left
Style Definition: Balloon Text: Left
Style Definition: equation: Left
Style Definition: Header: Left
Deleted: 1
Deleted: $\delta^{18}H$

almost zero to almost infinity. Importantly, this relationship is also virtually free from aggregation error. That is, seasonal tracer cycles also accurately predict the young water fraction in runoff from highly heterogeneous mixtures of subcatchments with strongly contrasting transit time distributions. Thus, although tracer cycle amplitudes yield biased and unreliable estimates of catchment mean travel times in heterogeneous catchments, they can be used reliably to estimate the fraction of young water in runoff.

Keywords: transit time, travel time, residence time, isotope tracers, residence time, convolution, catchment hydrology, aggregation error, aggregation bias

1 Introduction

Environmental systems are characteristically complex and heterogeneous. Their processes and properties are often difficult to quantify at small scales, and difficult to extrapolate to larger scales. Thus translating process inferences across scales, and aggregating across heterogeneity, are fundamental challenges for environmental scientists. These ubiquitous aggregation problems have been a focus of research in some environmental fields, such as ecological modelling (e.g., Rastetter et al., 1992), but have received surprisingly little attention elsewhere. In the catchment hydrology literature, for example, spatial heterogeneity has been widely recognized as a fundamental problem, but has rarely been the subject of rigorous analysis.

Instead, it is often tacitly assumed (although *hoped* might be a better word) that any problems introduced by spatial heterogeneity will be solved or masked by model parameter calibration. This is an intuitively appealing notion. After all, we are often not particularly interested in understanding or predicting point-scale processes within the system, but rather in predicting the resulting ensemble behavior at the whole-catchment scale, such as stream flow, stream chemistry, evapotranspiration losses, ecosystem carbon uptake, and so forth. Furthermore, we rarely have point-scale information from the system under study, and when we do, we have no clear way to translate it to larger scales. Instead, often our most reliable and readily available measurements are at the whole-catchment scale: stream flow, stream chemistry, weather variables, etc. Wouldn't it be nice if these whole-catchment measurements could be used to estimate spatially aggregated model parameters that somehow subsume the spatial heterogeneity of the system, at least well enough to generate reliable predictions of whole-catchment behavior?

This is a testable proposition, and the answer will depend partly on the nature of the underlying model. All models obscure a system's spatial heterogeneity to some degree, and many conceptual models obscure it completely, by treating spatially heterogeneous catchments as if they were spatially homogeneous instead. Doing so is not automatically disqualifying, but neither is it obviously valid. Rather, this spatial aggregation is a modelling choice, whose consequences should be explicitly analyzed and quantified. What do I mean by "explicitly analyzed and quantified?" As an example, consider Kirchner et al.'s (1993) analysis of how spatial heterogeneity affected a particular geochemical model for estimating catchment buffering of acid deposition. The authors began by noting that spatial heterogeneities will not "average out" in nonlinear model equations, and by showing that the resulting aggregation bias will be proportional to the nonlinearity in the model equations (which can be directly estimated), and proportional to the variance in the heterogeneous real-world parameter values (which is typically unknown, but may at least be given a plausible upper bound). They then showed that **their geochemical model's** governing equations were sufficiently linear that the effects of spatial heterogeneity were likely to be small. They then confirmed this theoretical result by mixing measured runoff chemistry time series from random pairs of geochemically diverse catchments **(**which do not flow together in the real world**)**. They showed that the geochemical model correctly predicted the buffering behavior of these spatially heterogeneous pseudo-catchments, without knowing that those catchments were heterogeneous, and without knowing anything about the nature of their heterogeneities.

Deleted: the

Deleted: of their particular geochemical model

Deleted: ,

Here I use similar thought experiments to explore the consequences of spatial heterogeneity for catchment mean transit time estimates derived from seasonal tracer cycles in precipitation and streamflow. Catchment *transit time*, or, equivalently, *travel time* – the time that it takes for rainfall to travel through a catchment and emerge as streamflow – is a fundamental hydraulic parameter that controls the retention and release of contaminants and thus the downstream consequences of pollution episodes (Kirchner et al., 2000; McDonnell et al., 2010). In many geological settings, catchment transit times also control chemical weathering rates, geochemical solute production and the long-term carbon cycle (Burns et al., 2003; Godsey et al., 2009; Maher, 2010; Maher and Chamberlain, 2014).

A catchment is characterized by its travel time distribution (TTD), which reflects the diversity of flowpaths (and their velocities) connecting each point on the landscape with the stream. **Because these flowpaths and velocities change with hydrologic forcing, the TTD is non-stationary (Kirchner et al., 2001; Tetzlaff et al., 2007; Botter et al., 2010; Hrachowitz et al., 2010a; Van der Velde et al., 2010; Birkel et al., 2012; Heidbüchel et al., 2012; Peters et al., 2014), but time-varying TTD's are difficult to estimate in practice, so most catchment studies have focused on estimating time-averaged TTD's instead.** Both the shape of the TTD and its corresponding mean travel time (MTT) reflect storage and mixing processes in the catchment (Kirchner et al., 2000, 2001; Godsey et al., 2010; Hrachowitz et al., 2010a). However, due to the difficulty in reliably estimating the shape of the TTD, and the volumes of data required to do so, many catchment studies have simply assumed that the TTD has a given shape, and **have** estimated only **its** MTT. As a result, and **also** because of its obvious physical interpretation as the ratio between the storage volume and the average water flux (in steady state), the MTT is by far the most universally reported parameter in catchment travel-time studies. Estimates of MTT's have been correlated with a wide range of catchment characteristics, including drainage density, aspect, hillslope gradient, depth to groundwater, hydraulic conductivity, and the prevalence of hydrologically responsive soils (e.g., McGuire et al., 2005; Soulsby et al., 2006; Tetzlaff et al., 2009; Broxton et al., 2009; Hrachowitz et al., 2009; Hrachowitz et al., 2010b; Asano and Uchida, 2012; Heidbüchel et al., 2013),

Deleted:

Formatted: English (U.S.)

Deleted: the

Travel time distributions and mean travel times cannot be measured directly, and they differ – often by orders of magnitude – from the hydrologic response timescale, because the former is determined by the velocity of water flow, and the latter is determined by the celerity of hydraulic potentials **(Horton and Hawkins, 1965; Hewlett and Hibbert, 1967; Beven, 1982; Kirchner et al., 2000; McDonnell and Beven, 2014**). Nor can travel time characteristics be reliably determined *a priori* from theory. Instead, they must be determined from chemical or isotopic tracers, such as Cl^-, ^{18}O, and ^2H, in precipitation and streamflow. These passive tracers "follow the water"; thus their temporal fluctuations reflect the transport, storage, and mixing of rainfall as it is transformed into runoff. (Groundwaters can also be dated using dissolved gases such as CFC's and $^3H/^3He$, but these tracers are not conserved in surface waters or in the vadose zone, so they are not well suited to estimating whole-catchment travel times.)

Deleted: (

As reviewed by McGuire and McDonnell (2006), three methods are commonly used to infer catchment travel times from conservative tracer time series: 1) time-domain convolution of the input time series to simulate the output time series, with parameters of the convolution **kernel (the** travel-time distribution**)** fitted by iterative search techniques, 2) Fourier transform spectral analysis of the input and output time series, and 3) sine-wave fitting to the seasonal tracer variation in the input and output. In all three methods, the greater the damping of the input signal in the output, the longer the inferred mean travel time. Sine-wave fitting can be viewed as the simplest possible version of both spectral analysis (examining the Fourier transform at just the annual frequency) and time-domain convolution (approximating the input and output as sinusoids, for which the convolution relationship is particularly easy to calculate). Whereas time-domain convolution methods require continuous, unbroken precipitation isotopic records spanning at least several times the MTT (McGuire and McDonnell, 2006; Hrachowitz et al., 2011), and spectral methods require time series spanning a wide range of time scales (Feng et al., 2004), sine-wave fitting can be performed on sparse, irregularly sampled data sets. Because sine-wave fitting is mathematically straightforward, and because its data requirements are modest compared to the other two methods, it is arguably the best candidate for comparison studies based on large multi-site datasets of isotopic measurements in precipitation and river flow. For that reason – and because it presents an interesting test case of the general aggregation issues alluded to above, in which some key results can be derived analytically – the sinusoidal fitting method will be the focus of my analysis.

The isotopic composition of precipitation varies seasonally as shifts in meridional circulation alter atmospheric vapor transport pathways (Feng et al., 2009**),** and as shifts in temperature and storm intensity alter the degree of rainout-driven fractionation that air masses undergo (Bowen, 2008). The resulting seasonal cycles in precipitation (e.g., Fig. 1a) are damped and phase-shifted as they are transmitted through catchments (e.g., Fig. 1b), by amounts that depend on – and thus can be used to infer **properties of** – the travel-time distribution. Figure 1 shows an example of sinusoidal fits to seasonal $\delta^{18}O$ cycles in precipitation and baseflow at one particular field site. The visually obvious damping of the isotopic cycle in baseflow relative to precipitation implies, in this case, an estimated MTT of 1.4 years (DeWalle et al., 1997**) under the assumption that the TTD is exponential.**

Deleted: ,

Deleted:)

Deleted:).

That particular estimate of mean transit time, like practically all such estimates in the literature, was made by methods that assume that the catchment is homogeneous, and therefore that the shape of its TTD can be straightforwardly characterized. Typical catchments violate this assumption, but the consequences for estimating MTT's have not been systematically investigated, either for sine-wave fitting or for any other methods **that infer** travel times from tracer data. Are any of these estimation methods reliable under realistic degrees of spatial heterogeneity? Are they biased, and by how much? We simply do not know, because they have not been tested. Instead, we have been directly applying theoretical results**, derived for** idealized hypothetical cases, to complex real-world situations that do not share those idealized characteristics. Methods for estimating catchment travel times urgently need benchmark testing. The work presented below is intended as one small step toward filling that gap.

Deleted: ,

Deleted: for inferring

Deleted: from

2 Mathematical preliminaries: tracer cycles in homogeneous catchments

Any method for inferring transit-time distributions (or their parameters, such as mean transit time) must make simplifying assumptions about the system under study. Most such methods assume that conservative tracers in streamflow can be modeled as the convolution of the catchment's transit time distribution with the tracer time series in precipitation (Maloszewski et al., 1983; Maloszewski and Zuber, 1993; Barnes and Bonell, 1996; Kirchner et al., 2000),

$$c_S(t) = \int_0^\infty h(\tau)\ c_P(t-\tau)\ \mathrm{d}\tau\,, \qquad (1)$$

where $c_S(t)$ is the concentration in the stream at time t, $c_P(t\text{-}\tau)$ is the concentration in precipitation at any previous time $t\text{-}\tau$, and $h(\tau)$ is the distribution of transit times τ separating the arrival of tracer molecules in precipitation and their delivery in streamflow. The concentrations $c_S(t)$ and $c_P(t\text{-}\tau)$ can also represent ratios of stable isotopes in the familiar δ notation (e.g., δ^{18}O or δ^{2}H); the mathematics are the same in either case.

The transit-time distribution $h(\tau)$ expresses the fractional contribution of past inputs to present runoff. Equation (1) implicitly assumes that the catchment is a linear time-invariant system, and thus that the convolution kernel $h(\tau)$ is stationary (i.e., constant through time). This is never strictly true, most obviously because if no precipitation falls on a particular day, it cannot contribute any tracer to the stream τ days later, and because higher precipitation rates will increase the rate that water and tracers are flushed through the catchment. Thus real-world TTD's vary through time, depending on the history of prior precipitation (**Kirchner et al., 2001; Tetzlaff et al., 2007;** Botter et al., 2010; Hrachowitz et al., 2010a; Van der Velde et al., 2010; Birkel et al., 2012; Heidbüchel et al., 2012; Peters et al., 2014). However, in applications **using real-world data**, $h(\tau)$ is conventionally interpreted as a time-invariant ensemble average, taken over an ensemble of precipitation histories, which obviously will differ from one another in detail. Mathematically, the ensemble averaging embodied in Eq. (1) is equivalent to the simplifying assumption that water fluxes in precipitation and streamflow are constant over time. (One can relax this assumption somewhat by integrating over the cumulative water flux rather than time, as proposed by Niemi (1977). If the rates of transport and mixing vary proportionally to the flow rate through the catchment, this yields a stationary distribution in flow-equivalent time.) A further simplification inherent in Eq. (1) is that evapotranspiration and its effects on tracer signatures are ignored.

Deleted: practical

2.1 A class of transit-time distributions

In much of the analysis that follows, I will assume that the transit-time distribution $h(\tau)$ belongs to the family of gamma distributions,

$$h(\tau) = \frac{\tau^{\alpha-1}}{\beta^{\alpha}\,\Gamma(\alpha)}\,e^{-\tau/\beta} = \frac{\tau^{\alpha-1}}{(\bar{\tau}/\alpha)^{\alpha}\,\Gamma(\alpha)}\,e^{-\alpha\,\tau/\bar{\tau}}, \qquad (2)$$

where α and β are a shape factor and scale factor, respectively, τ is the transit time, and $\bar{\tau} = \alpha\,\beta$ is the mean transit time. I make this assumption mostly so that some key results can be calculated exactly, but as I show below, the key results extend beyond this (already broad) class of distributions.

Figure 2 shows gamma distributions spanning a range of shape factors α. For the special case of α=1, the gamma distribution becomes the exponential distribution. Exponential distributions describe the behavior of continuously mixed reservoirs of constant volume, and they have been widely used to model catchment storage and mixing. The gamma distribution expresses the TTD of a Nash cascade (Nash, 1957) of α identical linear reservoirs connected in series, and the analogy to a Nash cascade holds even for non-integer α, through the use of fractional integration. For α>1, the gamma distribution rises to a peak and then falls off, similarly to a typical storm hydrograph, which is why Nash cascades have often been used to model rainfall-runoff relationships. For α<1, however, the gamma distribution has a completely different shape, having maximum weight at lags near zero, and also having a relatively long tail. These characteristics represent problematic contaminant behavior: an intense spike of contamination in short time and persistent contamination in long time. Tracer time series from many catchments have been shown to exhibit fractal $1/f$ scaling, which is consistent with gamma TTD's with $\alpha \approx 0.5$ (Kirchner et al., 2000, 2001; Godsey et al., 2010; Kirchner and Neal, 2013; Aubert et al., 2014).

Deleted:

For present purposes, it is sufficient to note that the family of gamma distributions encompasses a wide range of shapes which approximate many plausible TTD's (Fig. 2). The moments of the gamma distribution vary systematically with the shape factor α (Walck, 2007):

Deleted: ,

$$\text{mean}\,(\tau) = \beta\alpha = \bar{\tau}\,, \qquad \text{(3a)}$$

$$\text{SD}\,(\tau) = \beta\sqrt{\alpha} = \bar{\tau}/\sqrt{\alpha}\,, \qquad \text{(3b)}$$

$$\text{skewness}\,(\tau) = 2/\sqrt{\alpha}\,, \qquad \text{(3c)}$$

$$\text{and kurtosis}\,(\tau) = 6/\alpha\,. \qquad \text{(3d)}$$

As α increases above 1, the standard deviation **(SD)** declines in relation to the mean, and the shape of the distribution becomes more normal. But as α decreases below 1, the **SD** grows in relation to the mean, implying greater variability in transit times for the same average (in other words: more short transit times, more long transit times, and fewer close to the mean). Likewise the skewness and kurtosis grow with decreasing α, reflecting greater dominance by the tails of the distribution.

Deleted: standard deviation

Studies that have used tracers to constrain the shape of catchment TTD's have generally found shape factors α ranging from 0.3 to 0.7, **corresponding** to spectral slopes of the transfer function between roughly 0.6 and 1.4 (Kirchner et al., 2000, 2001; Godsey et al., 2010; Hrachowitz et al., 2010a; Kirchner and Neal, 2013; Aubert et al., 2014). Other studies – including those that have used annual tracer cycles to estimate mean transit times – have assumed that the catchment is a well-mixed reservoir and thus that α=1. Here I will assume that α falls in the range of 0.5 to 1 for typical catchment transit-time distributions, but I will

Deleted: which corresponds

also show some key results for the somewhat wider range of α=0.2-2, for illustrative purposes. The results reported here will not necessarily apply to TTD's that rise to a peak after a long delay, such as the gamma distribution with α>>2. However, one would not expect such a distribution to characterize whole-catchment TTD's in the first place, because except in very unusual catchments a substantial amount of precipitation can fall close to the stream and enter it relatively quickly, thus producing a strong peak at a short lag (Kirchner et al., 2001).

Deleted: =8 shown in Fig.

2.2 Estimating mean transit time from tracer cycles

Because convolutions (Eq. 1) are linear operators, they transform any sinusoidal cycle in the precipitation time series $c_P(t)$ into a sinusoidal cycle of the same frequency, but a different amplitude and/or phase, in the streamflow time series $c_S(t)$. Real-world transit-time distributions $h(\tau)$ are causal (i.e., $h(\tau)$=0 for t<0) and mass-conserving (i.e., $\int h(\tau)=1$), implying that $c_S(t)$ will be damped and phase-shifted relative to $c_P(t)$, and also implying that one can use the relative amplitudes and phases of cycles in $c_S(t)$ and $c_P(t)$ to infer **characteristics** of $h(\tau)$. This mathematical property forms the basis for sine-wave fitting, and also for the spectral methods of Kirchner et al. (2000, 2001), which can be viewed as sine-wave fitting across many different time scales.

Deleted: properties

The amplitudes A and phases ϕ of seasonal cycles in precipitation and streamflow can be estimated by nonlinear fitting,

$$\begin{aligned} c_P(t) &= A_P \sin(2\pi f t - \phi_P) + k_P \\ c_S(t) &= A_S \sin(2\pi f t - \phi_S) + k_S \end{aligned}, \qquad (4)$$

or by determining the cosine and sine coefficients a and b via multiple linear regression,

$$\begin{aligned} c_P(t) &= a_P \cos(2\pi f t) + b_P \sin(2\pi f t) + k_P \\ c_S(t) &= a_S \cos(2\pi f t) + b_S \sin(2\pi f t) + k_S \end{aligned}, \qquad (5)$$

and then calculating the amplitudes and phases using the conventional identities

$$A_P = \sqrt{a_P^2 + b_P^2}\,,\; A_S = \sqrt{a_S^2 + b_S^2}\,,\; \phi_P = \arctan(b_P / a_P) \text{ and } \phi_S = \arctan(b_S / a_S). \qquad (6)$$

Formatted: English (U.S.)

Formatted: English (U.S.)

Formatted: English (U.S.)

Deleted: /

In Eqs. (4)-(6) above, t is time, f is the frequency of the cycle (f=1 yr^{-1} for a seasonal cycle), and the subscripts P and S refer to precipitation and streamflow. **In fitting sinusoidal cycles**

to real-world data, robust estimation techniques such as iteratively reweighted least squares (IRLS) regression can help in limiting the influence of outliers. Also, because precipitation and streamflow rates vary through time, it may be useful to weight each tracer sample by its associated volume, for example to reduce the influence of small rainfall events (for more on the implications of volume-weighting, see Kirchner, 2015). An R script for performing volume-weighted IRLS is available from the author.

The key to calculating the amplitude damping and phase shift that will result from convolving a sinusoidal input with a gamma-distributed $h(\tau)$ is the gamma distribution's Fourier transform, also called, in this context, its "characteristic function" (Walck, 2007):

$$H(f) = \left(1 - i2\pi f \beta\right)^{-\alpha} = \left(1 - i2\pi f \bar{\tau}/\alpha\right)^{-\alpha} . \qquad (7)$$

From Eq. (7), one can derive how the shape factor α and the mean transit time $\bar{\tau}$ affect the amplitude ratio A_S/A_P between the streamflow and precipitation cycles,

$$\frac{A_S}{A_P} = \left(1 + \left(2\pi f \beta\right)^2\right)^{-\alpha/2} , \qquad (8)$$

and also the phase shift between them,

$$\phi_S - \phi_P = \alpha \arctan(2\pi f \beta) \quad , \qquad (9)$$

where $\beta = \bar{\tau}/\alpha$. Figures 3a and 3b show the expected amplitude ratios and phase shifts for a range of shape factors and mean transit times.

If the shape factor α is known (or can be assumed), the mean transit time can be calculated directly from the amplitude ratio A_S/A_P by inverting Eq. (8):

$$\bar{\tau} = \alpha\,\beta \quad , \quad \beta = \frac{1}{2\pi f}\sqrt{\left(A_S/A_P\right)^{-2/\alpha} - 1} \quad . \qquad (10)$$

Equation (10), with α=1, is the standard tool for estimating MTT's from seasonal tracer cycles in precipitation and streamflow. Alternatively, as Fig. 3c shows, both the shape factor α and the mean transit time $\bar{\tau}$ can be jointly determined from the phase shift ϕ_S-ϕ_P and the amplitude ratio A_S/A_P, if these can both be quantified with sufficient accuracy. Mathematically, this joint solution can be achieved by substituting Eq. (10) in Eq. (9), yielding the following implicit expression for α,

$$\phi_S - \phi_P = \alpha \arctan\left(\sqrt{\left(\frac{A_S}{A_P}\right)^{-2/\alpha} - 1}\right), \tag{11}$$

which can be solved using nonlinear search techniques such as Newton's method. Once α has been determined, the mean transit time $\bar{\tau}$ can be calculated straightforwardly using Eq. (10). However, when precipitation is episodic, the phase shift ϕ_S-ϕ_P may be difficult to estimate accurately, which can result in large errors in α and thus $\bar{\tau}$, particularly if the phase shift is near zero. Perhaps for this reason, or perhaps because (to the best of my knowledge) the relevant math has not previously been presented, tracer cycle phase information has not typically been used in estimating α and MTT.

3 Transit times and tracer cycles in heterogeneous catchments: a thought experiment

The methods outlined above can be applied straightforwardly in a homogeneous catchment characterized by a single transit time distribution. Real-world catchments, however, are generally heterogeneous; they combine different landscapes with different characteristics, and thus different TTD's. The implications of this heterogeneity can be demonstrated with a simple thought experiment. What if, instead of a single homogeneous catchment, we have two subcatchments with different MTT's, and therefore different tracer cycles, which then flow together, as shown in Fig. 4? If we observed only the tracer cycle in the combined runoff (the solid blue line in Fig. 4), and not the tracer cycles in the individual subcatchments (the red and orange lines in Fig. 4), would we correctly infer the whole-catchment MTT? Note that although I refer to the different runoff sources as "subcatchments", they could equally well represent alternate slopes draining to the same stream channel, or even independent flow paths down the same hillslope; nothing in this thought experiment specifies the scale of the analysis. And, of course, real-world catchments are much more complex than the simple thought experiment diagrammed in Fig. 4, but this two-component model is sufficient to illustrate the key issues at hand.

From assumed MTT's $\bar{\tau}$ and shape factors α for each of the subcatchments, one can calculate the amplitude ratios A_S/A_P and phase shifts ϕ_S-ϕ_P of their tracer cycles using Eqs. (8)-(9) above, and then average these cycles together using the conventional trigonometric identities.

(Equivalently, one can estimate the cosine and sine coefficients of the individual subcatchments' tracer cycles from the real and imaginary parts of Eq. (7) and algebraically average them together.) The shares of the two subcatchments in the average will depend on their relative drainage areas and/or water yields. For simplicity, I combine the runoff from the two subcatchments in a 1:1 ratio; this also guarantees that the combined runoff will be as different as possible from each of the two sources. I then ask the question: from the tracer behavior in the combined runoff (the solid blue line in Fig. 4), would I correctly estimate the mean transit time for the whole catchment? That is, would I infer a MTT that is close to the average of the MTT's of the two subcatchments?

One can immediately see that this situation is highly prone to aggregation bias, following Kirchner et al.'s (1993) rule of thumb that the degree of aggregation bias is proportional to the nonlinearity in the governing equations and the variance in the heterogeneous parameters. The amplitude ratios A_S/A_P and phase shifts ϕ_S-ϕ_P of seasonal tracer cycles are strongly nonlinear functions of the MTT (see Eqs. 8 and 10), as illustrated in Figs. 3a-b. And, importantly, the likely range of variation in subcatchment MTT's (from, say, fractions of a year to perhaps several years) straddles the nonlinearity in the governing equations. Thus we should expect to see significant aggregation bias in estimates of MTT.

Figure 5 illustrates the crux of the problem. The plotted curve shows the relationship between A_S/A_P and MTT for exponential transit time distributions (α=1); other realistic transit time distributions will give somewhat different relationships, but they will all be curved. Seasonal cycles from the two subcatchments (the red and orange squares) will mix along the dashed gray line (which is nearly straight but not exactly so, owing to phase differences between the two cycles). A 50:50 mixture of tracer cycles from the two subcatchments will plot as the solid blue square, with an amplitude ratio A_S/A_P of 0.43 and a MTT of just over 2 years in this particular example. But the crux of the problem is that if we use this amplitude ratio to infer the corresponding MTT, we will do so where the amplitude ratio intersects with the black curve (Eq. 10), yielding an inferred MTT of only 0.33 yr (the open square), which underestimates the true MTT of the mixed runoff by more than a factor of six. Bethke and Johnson (2008) pointed out that nonlinear averaging can lead to bias in groundwater dating by radioactive tracers; Fig. 5 illustrates how a similar bias can also arise in age determinations based on fluctuation damping in passive tracers.

Combining flows from two subcatchments with different mean transit times will result in a combined TTD that differs in shape, not just in scale, from the TTD's of either of the subcatchments. For example, combining two exponential distributions with different mean transit times does not result in another exponential distribution, but rather a hyperexponential distribution, as shown in Fig. 6. The characteristic function of the hyperexponential distribution (Walck, 2007) yields the following expression for the amplitude ratio of tracer cycles in precipitation and streamflow,

$$\frac{A_S}{A_P} = \left(\left(\frac{p}{1+(2\pi f \bar{\tau}_1)^2} + \frac{q}{1+(2\pi f \bar{\tau}_2)^2} \right)^2 + \left(\frac{p\, 2\pi f \bar{\tau}_1}{1+(2\pi f \bar{\tau}_1)^2} + \frac{q\, 2\pi f \bar{\tau}_2}{1+(2\pi f \bar{\tau}_2)^2} \right)^2 \right)^{1/2} , \qquad (12)$$

where $\bar{\tau}_1$ and $\bar{\tau}_2$ are the mean transit times of the two exponential distributions, and p and q=1-p are their proportions in the mixed runoff. Equation (12) describes the dashed grey line in Fig. 5, and one can see by inspection that in a 1:1 mixture (p=q), the amplitude ratio A_S/A_P will be determined primarily by the shorter of the two mean transit times. As Fig. 5 shows, the amplitude ratio implied by Eq. (12) is greater – often much greater – than Eq. (8) would predict for an exponential distribution with an equivalent mean transit time $\bar{\tau} = p\bar{\tau}_1 + q\bar{\tau}_2$. In other words, when amplitude ratios are interpreted as if they were generated by individual uniform catchments (i.e., Eq. 8) rather than a heterogeneous collection of subcatchments (i.e., Eq. 12), the inferred mean transit time will be underestimated, potentially by large factors.

To test the generality of this result, I repeated the thought experiment outlined above for 1000 hypothetical pairs of subcatchments, each with individual MTT's randomly chosen from a uniform distribution of logarithms spanning the interval between 0.1 and 20 years (Fig. 7). Pairs with MTT's that differed by less than a factor of two were excluded, so that the entire sample consisted of truly heterogeneous catchments. I then applied Eq. (10) to calculate the apparent MTT from the inferred runoff. As Fig. 7 shows, apparent MTT's calculated from the combined runoff of the two subcatchments can underestimate true whole-catchment MTT's by an order of magnitude or more, and this strong underestimation bias persists across a wide range of shape factors α. MTT's are reliably estimated (with values close to the 1:1 line in Fig. 7) only when both subcatchments have MTT's of much less than 1 year.

In most real-world cases, unlike these hypothetical thought experiments, one will only have measurements or samples from the whole catchment's runoff. **The** properties of the

Deleted: , and the

individual subcatchments, and thus the degree of heterogeneity in the system, will generally be unknown. And even if data were available for the subcatchments, those subcatchments would be composed of sub-sub-catchments, which would themselves be heterogeneous to some unknown degree, and so on. Thus it will generally be difficult or impossible to characterize the system's heterogeneity, but that is no justification for pretending that this heterogeneity does not exist. Nonetheless, in such situations it will be tempting to treat the whole system as if it were homogeneous, perhaps using terms like "apparent age" or "model age" to preserve a sense of rigor. But whatever the semantics, as Fig. 7 shows, assuming homogeneity in heterogeneous catchments will result in strongly biased estimates of whole-catchment mean transit times.

4 Quantifying the young water component of streamflow

The analysis above demonstrates what can be termed an "aggregation error": in heterogeneous systems, mean transit times estimated from seasonal tracer cycles yield inconsistent results at different levels of aggregation. The aggregation bias demonstrated in Figs. 5 and 7 implies that seasonal cycles of conservative tracers are unreliable estimators of catchment mean transit times. This observation raises the obvious question: is there anything *else* that can be estimated from seasonal tracer cycles, and that is relatively free **from** the aggregation bias that afflicts estimates of mean transit times?

Deleted: of

One hint is provided by the observation that when two tributaries are mixed, the tracer cycle amplitude in the mixture will almost exactly equal the average of the tracer cycle amplitudes in the two tributaries (Fig. 8). This is not intuitively obvious, because the tributary cycles will generally be somewhat out of phase with each other, so their amplitudes will not average **exactly** linearly. But when the tributary cycles are far out of phase (because the subcatchments have markedly different mean transit times or shape factors), the two amplitudes will also generally be very different, and thus the phase angle between the tributary cycles will have little effect on the amplitude of the mixed cycle.

Deleted: be

Deleted: to

Because tracer cycle amplitudes will average almost linearly when two streams merge, and thus are virtually free from aggregation bias (Fig. 8), anything that is proportional to tracer cycle amplitude will also be virtually free from aggregation bias. So, what is proportional to tracer cycle amplitude? One hint is provided by the observation that in Fig. 5, for example, the tracer cycle amplitude in the mixture is highly sensitive to transit times that are much shorter than the period of the tracer cycle (for a seasonal cycle, this period is T=1 yr), **but** highly insensitive to transit times that are much longer than the period of the tracer cycle. **As a thought experiment**, one can imagine a catchment in which some fraction of precipitation bypasses storage entirely (and thus transmits the precipitation tracer cycle directly to the stream), while the remainder is stored and mixed over very long time scales (and thus its tracer cycles are completely obliterated by mixing). In this idealized catchment, the amplitude ratio A_S/A_P between the tracer cycles in the stream and precipitation will be proportional to (indeed it will be exactly *equal to*) the fraction of precipitation that bypasses storage (and thus has a near-zero transit time).

Deleted: and

Deleted: In the limiting case

4.1 Young water

These lines of reasoning lead to the conjecture that for many realistic transit-time distributions, the amplitude ratio A_S/A_P may be **a good estimator of the** fraction of **streamflow** that is younger than some threshold age. This "young water" threshold should be expected to vary somewhat with the shape of the TTD. It should also be proportional to the tracer cycle period T because, as dimensional scaling arguments require, and as Eq. (8) shows for the specific case of gamma distributions, convolving the tracer cycle with the TTD will yield amplitude ratios A_S/A_P that are functions of $f\,\bar{\tau} = \bar{\tau}/T$.

Deleted: nearly equal to the

Deleted: the transit-time distribution

Numerical experiments verify these conjectures for gamma distributions spanning a wide range of shape factors (see Fig. 9). I define the "young water" fraction F_{yw} as the proportion of the transit-time distribution younger than a threshold age τ_{yw}, and calculate this proportion via the regularized lower incomplete gamma function,

$$F_{yw} = P(\tau < \tau_{yw}) = \Gamma(\tau_{yw},\, \alpha,\, \beta) = \int_{\tau=0}^{\tau_{yw}} \frac{\tau^{\alpha-1}}{\beta^{\alpha}\,\Gamma(\alpha)}\, e^{-\tau/\beta}\, d\tau \ , \qquad (13)$$

where, as before, $\beta = \bar{\tau}/\alpha$. I then numerically search for the threshold age for which (for a given shape factor α) the amplitude ratio A_S/A_P closely approximates F_{yw} across a wide range of scale factors β (or equivalently, a wide range of mean transit times $\bar{\tau}$). As Fig. 9 shows, this young water fraction nearly equals the amplitude ratio A_S/A_P, with the threshold for "young" water varying from 1.7 to 2.7 months as the shape factor α ranges from 0.5 to 1.5.

The amplitude ratio A_S/A_P and the young water fraction F_{yw} are both dimensionless and they both range from 0 to 1, so they can be directly compared without further calibration, beyond the determination of the threshold age τ_{yw}. As Fig. 10 shows, the best-fit threshold age varies modestly as a function of the shape factor α,

$$\tau_{yw}/T \approx 0.0949 + 0.1065\,\alpha - 0.0126\,\alpha^2 \quad . \qquad (14)$$

Across the entire range of α=0.2 to α=2 shown in Fig. 10, and across the entire range of amplitude ratios from 0 to 1 (and thus mean transit times from zero to near-infinity), the amplitude ratio A_S/A_P **estimates** the "young water" fraction with a root mean square error of less than 0.023, or 2.3 percent.

The young water fraction F_{yw}, as defined here, has the inevitable drawback that, because the shape factors of individual tributaries will usually be unknown, the threshold age τ_{yw} will necessarily be somewhat imprecise. However, F_{yw} has the considerable advantage that it is virtually immune to aggregation bias in heterogeneous catchments, because it is nearly equal to the amplitude ratio A_S/A_P (Fig. 9), which itself aggregates with very little bias, and also with very little random error (Fig. 8). This observation leads to the important **implication** that ***A_S/A_P* should reliably estimate** F_{yw}, not only in individual subcatchments, but also in the combined runoff from heterogeneous landscapes. To test this proposition, I calculated the young water fractions F_{yw} for 1000 heterogeneous pairs of synthetic subcatchments (with the same MTT's and shape factors shown in Fig. 7) using Eqs. (13) and (14), and compared each pair's average F_{yw} to the amplitude ratio A_S/A_P in the merged runoff. Figure 11 shows that, as hypothesized, A_S/A_P **estimates** the young water fraction in the merged runoff with very little scatter or bias. The root-mean-square error in Fig. 11 is roughly two percent or less, in marked contrast to errors of several hundred percent shown in Fig. 7 for estimates of mean transit time from the same synthetic catchments.

4.2 Sensitivity to assumed TTD shape and threshold age

The analysis presented in Sect. 4.1 shows that the amplitude ratio A_S/A_P accurately estimates the fraction of streamflow younger than a threshold age. But this threshold age depends on the shape factor α of the subcatchment TTDs, which will generally be uncertain. Consider, for example, a hypothetical case where we measure an amplitude ratio of A_S/A_P=0.2 in the seasonal tracer cycles in a particular catchment, but we don't

Deleted: span a

Deleted: predicts

Deleted: ¶

Deleted: result

Deleted: A_S/A_P will be approximately equal to

Deleted: predicts

know whether its subcatchments are characterized by α=1, α=0.5, or a mixture of distributions between these shape factors. How much does this uncertainty in α, and thus in the threshold age, affect the inferences we can draw from A_S/A_P? We can approach this question from two different perspectives.

We can interpret the uncertainty in α as creating ambiguity in either the threshold age τ_{yw} (which defines "young" in "young water fraction"), or in the proportion of water younger than any fixed threshold age (the "fraction" in "young water fraction").

First, from Fig. 10 we can estimate how uncertainty in α affects the threshold age τ_{yw} that defines what counts as 'young' streamflow. One can see that across the plausible range of shape factors, the young water threshold (that is, the threshold defining whatever young water fraction will aggregate correctly) varies from about τ_{yw}=1.75 months for α=0.5 to τ_{yw}=2.27 months for α=1. Thus the ambiguity in α translates into an ambiguity of 0.52 months (or about two weeks) in the threshold that defines "young" water. If some subcatchments are characterized by α=0.5 and others by α=1, and still others by values in between, then the effective threshold age for the ensemble will lie somewhere between 1.75 and 2.27 months. If the range of uncertainty in α is wider, then the range of uncertainty in τ_{yw} will be wider as well, spanning over a factor of two (1.37 to 3.10 months) for values of α spanning the full order-of-magnitude range shown in Fig. 2 (α=0.2 to 2).

Alternatively, we can treat the uncertainty in α as creating, for any fixed threshold age, an ambiguity in the fraction of streamflow that is younger than that age. Consider the hypothetical case outlined above, in which A_S/A_P=0.2. If we assume that the subcatchments are characterized by α=1 (and thus τ_{yw}=2.27 months), then we would infer that roughly 20% of streamflow is younger than 2.27 months (the exact young water fraction, using Eqs. (10) and (13), is 0.215). But if the subcatchments are characterized by α=0.5 instead, then according to Eqs. (10) and (13) the fraction younger than 2.27 months will be 0.242 instead of 0.215. Thus the uncertainty in α corresponds to an uncertainty in the young water fraction of 3% (of the range of a priori uncertainty in F_{yw}, which is between 0 and 1), or 13% (of the original estimate for α=1).

For comparison, we can contrast this uncertainty with the corresponding uncertainty in the mean transit time $\bar{\tau}$ calculated from Eq. (10). A seasonal tracer cycle amplitude

ratio A_S/A_P=0.2 implies a mean transit time of $\bar{\tau} = 0.80$ years if α=1, but $\bar{\tau} = 1.99$ years if α=0.5. Thus the uncertainty in the mean transit time is a factor of 2.5, compared to a few percent for the young water fraction.

We can extend these sample calculations over a range of shape factors α and amplitude ratios A_S/A_P (see Fig. 12). As Fig. 12 shows, when the shape factor is uncertain in the range of 0.5<α<1, the corresponding uncertainty in the young water fraction F_{yw} is typically several percent, but the corresponding uncertainty in the MTT is typically a factor of two or more. For a factor-of-ten uncertainty in the shape factor (0.2<α<2), the uncertainty in the young water fraction is consistently less than a factor of two, whereas the uncertainty in the MTT can exceed a factor of 100.

Similar sensitivity of mean transit time to model assumptions was also observed by Kirchner et al. (2010) in two Scottish streams, and by Seeger and Weiler (2014), in their study calibrating three different transit time models to monthly δ^{18}O time series from 24 mesoscale Swiss catchments. Seeger and Weiler's three transit time models yielded MTT estimates that were often inconsistent by orders of magnitude, but yielded much more consistent estimates of the fraction of water younger than 3 months, foreshadowing the sensitivity analysis presented here.

Formatted: Bullets and Numbering

4.3 Young water estimation with non-gamma distributions

Because both the young water fraction F_{yw} and the tracer cycle amplitude ratio A_S/A_P aggregate nearly linearly, the results shown in Fig. 11 will also approximately hold at higher levels of aggregation. That is, we can merge each catchment in Fig. 11, which has two tributaries, with another two-tributary catchment to form a four-tributary catchment, which we can merge with another four-tributary catchment to form an eight-tributary catchment, and so on. Figure **13** shows the outcome of this thought experiment. One can see that just like in the two-tributary case, the tracer cycle amplitude ratio A_S/A_P in the merged runoff predicts the average young water fraction F_{yw} with relatively little scatter. There is a slight underestimation bias, which is more visible in Fig. **13** than for the two-tributary case in Fig. 11. In contrast to the **minimal** estimation bias in F_{yw}, MTT is underestimated by large factors in both the two-tributary case and the 8-tributary case.

Deleted: 12

Deleted: 12

Deleted: (

Deleted: minimial

Deleted:)

It is important to recognize that the two-tributary catchments that were merged in Fig. 13 are not characterized by gamma transit time distributions (although their tributaries are), because mixing two gamma distributions does not create another gamma distribution. Thus Fig. 13 demonstrates the important result that although the analysis presented here was based on gamma distributions for mathematical convenience, the general principles developed here – namely, that the amplitude ratio A_S/A_P estimates the young water fraction F_{yw}, and that estimates of F_{yw} are relatively immune to aggregation bias in heterogeneous catchments – are not limited to distributions within the gamma family.

Deleted: 12

Deleted: a mixture of

Deleted: is

Deleted: 12

For example, as Fig. 6 showed, mixing two exponential distributions will not create another exponential distribution, nor any other member of the gamma family, but rather a hyperexponential distribution. Thus Fig. 13b implies that A_S/A_P also estimates F_{yw} accurately for mixtures of exponentials, that is, for any distribution of the form,

Deleted: the mixture of

Deleted: is

Deleted: 12b

$$h(\tau) = \frac{1}{\sum_{i=1}^{n} k_i} \sum_{i=1}^{n} \frac{k_i}{\bar{\tau}_i} e^{-\tau/\bar{\tau}_i} \qquad (15)$$

where the weights k_i and mean transit times $\bar{\tau}_i$ can take on any positive real values. Likewise Fig. 13c implies that A_S/A_P estimates F_{yw} reasonably accurately for mixtures of gamma distributions, that is, for any distribution of the form,

Deleted: 12c

$$h(\tau) = \frac{1}{\sum_{i=1}^{n} k_i} \sum_{i=1}^{n} \frac{k_i \, \tau^{\alpha_i - 1}}{(\bar{\tau}_i/\alpha_i)^{\alpha_i} \, \Gamma(\alpha_i)} e^{-\alpha_i \tau/\bar{\tau}_i} \qquad (16)$$

where, as above, the weights k_i and mean transit times $\bar{\tau}_i$ can take on any positive real values, and the shape factors α_i can take on any values between 0.2 and 2. In the continuum limit, n could potentially be infinite in Eq. (15) or (16), whereupon the summations become integrals. Equations (15) and (16) describe very broad classes of distributions, suggesting that the results reported here also apply to a very wide range of catchment transit time distributions, well beyond the (already broad) family of gamma distributions with shape factors α<2.

4.4 Incorporating phase information in estimating young water fractions and mean transit times

One interpretation of the strong aggregation bias in mean transit time estimates, as documented in Figs. 7 and 13, is that when the transit time distributions of the individual tributaries are averaged together, the result has a different shape (i.e., averages of exponentials are not exponentials, and averages of gamma distributions are not gamma-distributed). Thus it is unsurprising that a formula for estimating mean travel times based on exponential distributions (for example) will be inaccurate when applied to non-exponential distributions. The practical issue in the real world, of course, is that the shape of the transit time distribution will usually be unknown, so the problem of fitting the "wrong" distribution will be difficult to solve.

In the specific case of fitting seasonal sinusoidal patterns, the only information one has for estimating the transit time distribution is the amplitude ratio and the phase shift of streamflow relative to precipitation. The phase shift has heretofore been ignored as a source of additional information. Could it be helpful?

As described in Sect. 2.2 above, **one can use** the amplitude ratio and phase shift to jointly **estimate** the shape factor α **by iteratively solving Eq. 11, and then estimate** the scale factor β **via Eq. 10. The mean transit time can then be estimated as $\alpha\beta$ (Eq. 3a). From the fitted value of α, one can also use Eq. 14 to estimate the threshold age τ_{yw} for young water fractions that should aggregate nearly linearly,** and **then finally estimate the young water fraction as $F_{yw}=\Gamma(\tau_{yw}, \alpha, \beta)$ (Eq. 13). The lower incomplete gamma function $\Gamma(\tau_{yw}, \alpha, \beta)$ is readily available in many software packages (for example, the Igamma function in R or the GAMMA.DIST function in Microsoft Excel).**

This approach assumes that the catchment's transit times are gamma-distributed. To test whether it can nonetheless improve estimates of the mean transit time or the young water fraction, even in catchments whose transit times are not gamma-distributed, I applied this method to the 8-tributary synthetic catchments shown in Fig. **13. As pointed out in Sect. 4.3 above, the TTD's of these catchments (and even their two-subcatchment tributaries) will be sums of gammas and thus not gamma-distributed themselves.** Figure **14** shows **the** new estimates **based on amplitude ratios and phase shifts** (in dark blue), superimposed on the previous estimates from Fig. **13, based on amplitude ratios alone,** as

Formatted: Bullets and Numbering

Deleted: 12

Deleted: In section

Deleted: I outlined how

Deleted: can be used

Deleted: determine

Deleted: ,

Deleted: , and the mean transit time $\overline{\tau}$ (Eqs. 10

Deleted: 11).

Deleted: 12, to assess whether it substantially improved the resulting estimates of either the mean transit time or the young water fraction.

Deleted: 13

Deleted: these

Deleted: 12

reference (in light blue). Mean transit time estimates based on both phase and amplitude information are somewhat more accurate than those based on amplitude ratios alone (Fig. 14d-14f), but they still exhibit very large aggregation bias. Incorporating phase information in estimates of F_{yw} (Fig. 14a-14c) eliminates much of the (already small) bias in F_{yw} estimates obtained from amplitude ratios alone. (The logarithmic axes of Figs. 14a-14c make this bias more visible than it is on the linear axes of Figs. 13a-13c). The top and bottom rows of Fig. 14 are plotted on consistent axes (both are logarithmic scales spanning a factor of 50), so they provide a direct visual comparison of the reliability of estimates of F_{yw} and MTT.

Deleted: 13d-13f

Deleted: 13a-13c

Deleted: 13a-13c

Deleted: 12a-12c).

Deleted: 13

5 Implications

Two main results emerge from the analysis presented above. First, mean transit times (MTT's) estimated from seasonal tracer cycles exhibit severe aggregation bias in heterogeneous catchments, underestimating the true MTT by large factors. Second, seasonal tracer cycle amplitudes accurately reflect the fraction of "young" water in streamflow and exhibit very little aggregation bias. Both of these results have important implications for catchment hydrology.

5.1 Biases in mean transit times

Figures 7, 13, and 14 indicate that in spatially heterogeneous catchments (which is to say, all real-world catchments), MTT's estimated from seasonal tracer cycles are fundamentally unreliable. The relationship between true and inferred MTT shown in these figures is not only strongly biased, but also wildly scattered – so much so, that it can only be visualized on logarithmic axes. The huge scatter in the relationship means that there is little point in trying to correct the bias with a calibration curve, because most of the resulting estimates would still be wrong by large factors. This scatter also implies that one should be careful about drawing inferences from site-to-site comparisons of MTT values **derived from seasonal cycles**, since a large part of their variability may be aggregation noise.

Deleted: 12

Deleted: 13

The underestimation bias in MTT estimates arises because, as Figs. 3a and 5 show, travel times significantly shorter than one year have a much bigger effect on seasonal tracer cycles than travel times of roughly one year and longer. DeWalle et al. (1997) calculated that an exponential TTD with a MTT of 5 years would result in such a small isotopic cycle in streamflow that it would approach the analytical detection limit of isotope measurements. But while this may be the hypothetical upper limit to MTT's determined from seasonal isotope cycles, my results show that even MTT's far below that limit cannot be reliably estimated in heterogeneous landscapes. Indeed, Figure 7 shows that MTT's can only be reliably estimated (that is, they will fall close to the 1:1 line) in heterogeneous systems where the MTT is roughly 0.2 years or so – in other words, only when most of the streamflow is "young" water.

It is becoming widely recognized that stable isotopes are effectively blind to the long tails of travel time distributions (Stewart et al., 2010; Stewart et al., 2012; Seeger and Weiler, 2014). The results presented here reinforce this point, showing how in heterogeneous catchments, any stable isotope cycles from long-MTT subcatchments (or flowpaths) will be overwhelmed by much larger cycles from short-MTT subcatchments (or flowpaths). Furthermore, the nonlinearities in the governing equations (Figs. 3 and 5) imply that the shorter-MTT components will dominate MTT estimates, which will thus be biased low. This underestimation bias may help to explain the discrepancy between MTT estimates derived from stable isotopes and those derived from other tracers, such as tritium (Stewart et al., 2010; Stewart et al., 2012). However, one should note that, like any radioactive tracer, tritium ages should themselves be vulnerable to underestimation bias in heterogeneous systems (Bethke and Johnson, 2008). Until tritium ages are subjected to benchmark tests like those I have presented here for stable isotopes, one cannot estimate how much they, too, are distorted by aggregation bias.

5.2 Other methods for estimating MTT's from tracers

Sine-wave fitting to seasonal tracer cycles is just one of several methods for estimating MTT's from tracer data. I have focused on this method because the relevant calculations are easily posed, and several key results can be obtained analytically. My results show that MTT estimates from sine-wave fitting are subject to severe aggregation bias, but they do not show whether other methods are better or worse in this regard. This is unknown at present, and needs to be tested. But until this is done, there is little basis for optimism that other methods will be immune to the biases identified here. One would expect that the results presented here should translate straightforwardly to spectral methods for estimating MTT's, as these methods essentially perform sine-wave fitting across a range of time scales. Thus one should expect aggregation bias at each time scale. The upper limit of reliable MTT estimates should be expected to be a fraction of the longest observable cycles in the data (as it is for the annual cycles measured here). Thus this upper limit will depend on the lengths of the tracer time series, and also on whether they contain significant input and output variability on long wavelengths (longer records will not help, unless the tracer concentrations are actually variable on those longer time scales). The same principles are likely to apply to convolution modeling **of tracer time series**, due to the formal equivalence of the time and frequency domains under Fourier's theorem. **Furthermore, to the extent that seasonal cycles are the dominant features of many natural tracer time series, convolution modeling of tracer time series may effectively be an elaborate form of sine-wave fitting, with all the attendant biases outlined here.** Until these conjectures are tested, however, they will remain speculative. Given the severe aggregation bias identified here, there is an urgent need for benchmark testing of the other common methods for MTT estimation.

Deleted: time-series

It should also be noted that methods for estimating MTT's assume not only homogeneity but also stationarity, and real-world catchments violate both of these assumptions. The results presented here suggest that nonstationarity (which is, very loosely speaking, heterogeneity in time) is likely to create its own aggregation bias, in addition to the spatial aggregation bias identified here. This aggregation bias can also be characterized using benchmark tests, as I show in a companion paper (Kirchner, 2015).

5.3 Implications for mechanistic interpretations of MTT's

The analysis presented here implies that many literature values of MTT are likely to be underestimated by large factors, or, in other words, that typical catchment travel times are probably several times longer than we previously thought they were. This result sharpens the "rapid mobilization of old water" paradox: how do catchments store water for weeks or months, and then release it within minutes or hours in response to precipitation events (Kirchner, 2003)? This result also sharpens an even more basic puzzle: where can catchments store so much water, that it can be so old, on average?

Many studies have sought to link **MTT's** to catchment characteristics, often with inconsistent results. For example, McGuire et al. (2005) reported that MTT was positively correlated with the ratio of flow path distance to average hillslope gradient at experimental catchments in Oregon, but Tetzlaff et al. (2009) reported that MTT was *negatively* correlated with the same ratio, and positively correlated with the extent of hydrologically responsive soils, at several Scottish catchments. Hrachowitz et al. (2009) reported that MTT was related to precipitation

Deleted: that have estimated catchment MTT's

Deleted: them

intensity, soil characteristics, drainage density, and topographic wetness index across a larger network of Scottish catchments, whereas Asano and Uchida (2012) reported that subsurface flow path depth was the main control on baseflow MTT at their Japanese field sites. Heidbüchel et al. (2013) reported that MTT was correlated with soil depth, hydraulic conductivity, or planform curvature, with different characteristics becoming more important under different rainfall regimes. And most recently, Seeger and Weiler (2014) reported that most of the observed correlations between MTT and terrain characteristics across 24 Swiss catchments became non-significant when the variation in mean annual discharge was taken into account. My analysis casts much of this literature in a different light. Given that a large component of MTT estimates in the literature may be aggregation noise (Figs. 7, **13** and **14**), one should not be surprised if MTT estimates exhibit weak and inconsistent correlations with catchment characteristics, even if those characteristics are important controls on real-world MTT's.

Deleted: 12

Deleted: 13

5.4 The young water fraction F_{yw} as an alternative travel time metric

More generally, though, my analysis implies that the **young water fraction F_{yw} is a more** useful metric of catchment travel time than MTT **is**, for the simple reason that F_{yw} can be reliably determined in heterogeneous catchments **but** MTT cannot. Of course, if we know the young water fraction in runoff, we obviously also know the fraction of "old" water as well (meaning water older than the "young water" threshold). But we do not know – and my analysis implies that we generally *cannot* know – how old this "old" water is, at least from analyses of seasonal tracer cycles.

Deleted: most

Deleted: is the young water fraction F_{yw} rather

Deleted: , whereas

Of course, because F_{yw} is nearly equal to the amplitude ratio, and MTT can also be expressed as a function of the amplitude ratio for travel time distributions (TTD's) of any known shape, one might conclude that MTT and F_{yw} are just transforms of one another. But that conclusion presumes that the shape of the TTD is known, and my analysis shows that in heterogeneous catchments, the shape of the TTD will be unpredictable. Because the MTT is sensitive to the shape of the TTD – and in particular to the long-time tail, which is particularly poorly constrained – it cannot be reliably estimated. By contrast, my analysis shows that despite the uncertainty in the shape of the TTD in heterogeneous catchments, the F_{yw} can be reliably estimated from the amplitude ratio of seasonal tracer cycles in precipitation and runoff. The fact that this is possible is neither a miracle nor a fortuitous accident; instead F_{yw} has been

Deleted: directly

defined with exactly this result in mind. The F_{yw} entails an unavoidable ambiguity in what, exactly, the threshold age of young water is (because this depends on the shape of the TTD, which is usually unknown), but this uncertainty is small (Fig. 10) compared to the very large uncertainty in the MTT.

It should be kept in mind that in real-world data, unlike the thought experiments analyzed here, the tracer measurements themselves will be somewhat uncertain, and this uncertainty will also flow through to estimates of either MTT or F_{yw}. In particular, although my analysis has focused on the effects of spatial heterogeneity in catchment properties (as reflected in the TTD's of the individual tributary subcatchments), it has ignored any spatial heterogeneity in the atmospheric inputs themselves. Furthermore, estimates of MTT or F_{yw} typically assume that any patterns in stream tracer concentrations arise only from the convolution of varying input concentrations, and not, for example, from seasonal evapoconcentration effects (for chemical tracers) or evaporative fractionation (for isotopes). If this assumption is violated, the resulting structural errors are potentially much more consequential than random errors in tracer measurements.

5.5 Potential applications for young water fractions

Since young water fractions are estimated from amplitude ratios and phase shifts of seasonal tracer cycles, one could ask whether they add any new information, or whether we could characterize catchments equally well by their amplitude ratios and phase shifts instead. One obvious answer is that amplitude ratios and phase shifts, by themselves, are purely phenomenological descriptions of input-output behavior. Young water fractions, by contrast, offer a mechanistic explanation for how that behavior arises, showing how it is linked to the fraction of precipitation that reaches the stream in much less than one year. Not only is this potentially useful for understanding the transport of contaminants and nutrients, it also directly quantifies the importance of relatively fast flowpaths in the catchment. These fast flowpaths are likely to be shallow [since permeability typically decreases rapidly with depth; \Brooks, 2004 #2208; Bishop, 2011 #2207], and to originate relatively close to flowing channels. One would expect F_{yw} to increase under wetter conditions, as the water table rises into more permeable near-surface zones, and as the flowing channel network extends to more finely dissect the landscape (Godsey and Kirchner, 2014), thus shortening the path length of subsurface

flows, as well as multiplying the wetted catchment area in riparian zones. In a companion paper (Kirchner, 2015), I show that young water fractions can be estimated separately for individual flow regimes, allowing one to infer how shifts in hydraulic forcing alter the fraction of streamflow that is generated via fast flowpaths. I further demonstrate how one can estimate the chemistry of "young water" and "old water" end members, based on comparisons of F_{yw} and solute concentrations across different flow regimes.

Because one can estimate F_{yw} from irregularly and sparsely sampled tracer time series, it can be used to facilitate intercomparisons among many catchments that lack more detailed tracer data. For example, Jasechko et al. (in review) have recently used the approach outlined here to calculate young water fractions for hundreds of catchments around the globe, ranging from small research watersheds to continental-scale river basins, and to examine how they respond to variations in catchment characteristics.

One final note: it has not escaped my notice that because the "young water" threshold is defined as a fraction of the period of the fitted sinusoid (here, an annual cycle), and because spectral analysis is equivalent to fitting sinusoids across a range of time scales, the input and output spectra of conservative tracers can be re-expressed as a series of young water fractions for a series of young water thresholds. In principle, then, this cascade of young water fractions (and their associated threshold ages) should directly express the catchment's cumulative distribution of travel times, thus solving the longstanding problem of measuring the shape of the transit time distribution. A proof-of-concept study of this direct approach to deconvolution is currently underway.

6 Summary and conclusions

I used benchmark tests with data from simple synthetic catchments (Fig. 4) to test how catchment heterogeneity affects estimates of mean transit times (MTT's) derived from seasonal tracer cycles in precipitation and streamflow (e.g., Fig. 1). The relationship between tracer cycle amplitude and MTT is strongly nonlinear (Fig. 3), with the result that tracer cycles from heterogeneous catchments will underestimate their average MTT's (Fig. 5). In heterogeneous catchments, furthermore, the shape of the transit time distribution (TTD) in the mixed runoff will differ from that of the tributaries; e.g., mixtures of exponential distributions are not exponentials (Fig. 6), and mixtures of gamma distributions are not gamma-distributed. These two effects combine to make seasonal tracer cycles highly unreliable as estimators of MTT's, with large scatter and strong underestimation bias in heterogeneous catchments (Figs. 7 and **13**). These results imply that many literature values of MTT are likely to be underestimated by large factors, and thus that typical catchment travel times are much longer than previously thought.

Deleted: 12

However, seasonal tracer cycles can be used to reliably estimate the *young water fraction* (F_{yw}) in runoff, defined as the fraction younger than approximately 0.15-0.25 years (i.e., ~2-3 months), depending on the shape of the underlying travel-time distribution (Figs. 9-10). The amplitude ratio of seasonal tracer cycles in precipitation and runoff predicts F_{yw} with an accuracy of roughly 2 percent or better, across the entire range of plausible TTD shape factors from α=0.2 to α=2, and across the entire range of mean transit times from nearly zero to near-infinity (Fig. 9). Most importantly, this relationship is virtually immune to aggregation bias, so the amplitude ratio reliably predicts the young water fraction in the combined runoff from heterogeneous landscapes, with little bias or scatter (Figs. 11 and **13**). Incorporating phase as well as amplitude information virtually eliminates the (already small) bias in F_{yw} estimates obtained from amplitude information alone (Fig. **14**). Thus my analysis not only reveals large aggregation errors in MTT, which has been **widely** used to characterize catchment transit time; it also proposes an alternative metric, F_{yw}, which should be reliable in heterogeneous catchments.

Deleted: "

Deleted: "

Formatted: Font: Italic

Formatted: Font: Italic

Deleted: 12

Deleted: 13

Deleted: universally

More generally, these results vividly illustrate how the pervasive heterogeneity of environmental systems can confound the simple conceptual models that are often used to analyze them. **But** not all properties of environmental systems are equally susceptible to aggregation error. **Although environmental** heterogeneity makes some measures (like MTT) highly unreliable, **it** has little effect on others (like F_{yw}). Benchmark tests are essential for determining which measures are highly susceptible to aggregation error, and which are relatively immune. Thus these results highlight the broader need for benchmark testing to diagnose aggregation errors in environmental measurements and models, beyond the specific illustrative case analyzed here.

Deleted: However, my results also demonstrate that

Deleted: Environmental

Deleted: but

Deleted: and quantify

Acknowledgements

This analysis was motivated by intensive discussions with Scott Jasechko and Jeff McDonnell; I thank them for their encouragement, and for many insightful comments. **I also appreciate the comments by Markus Weiler and two anonymous reviewers, which spurred improvements in the final version of the manuscript.**

References

Asano, Y., and Uchida, T.: Flow path depth is the main controller of mean base flow transit times in a mountainous catchment, Water Resour. Res., 48, W03512, doi: 10.1029/2011wr010906, 2012.

Aubert, A. H., Kirchner, J. W., Gascuel-Odoux, C., Facheux, M., Gruau, G., and Merot, P.: Fractal water quality fluctuations spanning the periodic table in an intensively farmed watershed, Environmental Science & Technology, 48, 930-937, doi: 10.1021/es403723r, 2014.

Barnes, C. J., and Bonell, M.: Application of unit hydrograph techniques to solute transport in catchments, Hydrological Processes, 10, 793-802, 1996.

Bethke, C. M., and Johnson, T. M.: Groundwater age and groundwater age dating, Annual Review of Earth and Planetary Sciences, 36, 121-152, doi: 10.1146/annurev.earth.36.031207.124210, 2008.

Beven, K.: On subsurface stormflow: predictions with simple kinematic theory for saturated and unsaturated flows, Water Resour. Res., 18, 1627-1633, 1982.

Birkel, C., Soulsby, C., Tetzlaff, D., Dunn, S., and Spezia, L.: High-frequency storm event isotope sampling reveals time-variant transit time distributions and influence of diurnal cycles, Hydrological Processes, 26, 308-316, doi: 10.1002/hyp.8210, 2012.

Botter, G., Bertuzzo, E., and Rinaldo, A.: Transport in the hydrological response: Travel time distributions, soil moisture dynamics, and the old water paradox, Water Resour. Res., 46, W03514, doi: 10.1029/2009WR008371, 2010.

Bowen, G. J.: Spatial analysis of the intra-annual variation of precipitation isotope ratios and its climatological corollaries, Journal of Geophysical Research-Atmospheres, 113, D05113, doi: 10.1029/2007jd009295, 2008.

Broxton, P. D., Troch, P. A., and Lyon, S. W.: On the role of aspect to quantify water transit times in small mountainous catchments, Water Resour. Res., 45, W08427, doi: 10.1029/2008wr007438, 2009.

Burns, D. A., Plummer, L. N., McDonnell, J. J., Busenberg, E., Casile, G. C., Kendall, C., Hooper, R. P., Freer, J. E., Peters, N. E., Beven, K. J., and Schlosser, P.: The geochemical evolution of riparian ground water in a forested piedmont catchment, Ground Water, 41, 913-925, 2003.

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

Feng, X. H., Kirchner, J. W., and Neal, C.: Spectral analysis of chemical time series from long-term catchment monitoring studies: Hydrochemical insights and data requirements, Water, Air, and Soil Pollution: Focus, 4, 221-235, 2004.

Feng, X. H., Faiia, A. M., and Posmentier, E. S.: Seasonality of isotopes in precipitation: A global perspective, Journal of Geophysical Research-Atmospheres, 114, D08116, doi: 10.1029/2008jd011279, 2009.

Godsey, S. E., Kirchner, J. W., and Clow, D. W.: Concentration-discharge relationships reflect chemostatic characteristics of US catchments, Hydrological Processes, 23, 1844-1864, doi: 10.1002/hyp.7315, 2009.

Godsey, S. E., Aas, W., Clair, T. A., de Wit, H. A., Fernandez, I. J., Kahl, J. S., Malcolm, I. A., Neal, C., Neal, M., Nelson, S. J., Norton, S. A., Palucis, M. C., Skjelkvåle, B. L., Soulsby, C., Tetzlaff, D., and Kirchner, J. W.: Generality of fractal 1/f scaling in catchment tracer time series, and its implications for catchment travel time distributions, Hydrological Processes, 24, 1660-1671, doi: 10.1002/hyp.7677, 2010.

Godsey, S. E., and Kirchner, J. W.: Dynamic, discontinuous stream networks: hydrologically driven variations in active drainage density, flowing channels, and stream order, Hydrological Processes, 28, 5791-5803, doi: 10.1002/hyp.10310, 2014.

Heidbüchel, I., Troch, P. A., Lyon, S. W., and Weiler, M.: The master transit time distribution of variable flow systems, Water Resour. Res., 48, W06520, doi: 10.1029/2011WR011293, 2012.

Heidbüchel, I., Troch, P. A., and Lyon, S. W.: Separating physical and meteorological controls of variable transit times in zero-order catchments, Water Resour. Res., 49, 7644-7657, doi: 10.1002/2012wr013149, 2013.

Hewlett, J. D., and Hibbert, A. R.: Factors affecting the response of small watersheds to precipitation in humid regions, in: Forest Hydrology, edited by: Sopper, W. E., and Lull, H. W., Pergamon Press, Oxford, 275-290, 1967.

Horton, J. H., and Hawkins, R. H.: the path of rain from the soil surface to the water table, Soil Science, 100, 377-383, 1965.

Hrachowitz, M., Soulsby, C., Tetzlaff, D., Dawson, J. J. C., and Malcolm, I. A.: Regionalization of transit time estimates in montane catchments by integrating landscape controls, Water Resour. Res., 45, W05421, doi: 10.1029/2008wr007496, 2009.

Hrachowitz, M., Soulsby, C., Tetzlaff, D., Malcolm, I. A., and Schoups, G.: Gamma distribution models for transit time estimation in catchments: Physical interpretation of parameters and implications for time-variant transit time assessment, Water Resour. Res., 46, W10536, doi: 10.1029/2010wr009148, 2010a.

Hrachowitz, M., Soulsby, C., Tetzlaff, D., and Speed, M.: Catchment transit times and landscape controls-does scale matter?, Hydrological Processes, 24, 117-125, doi: 10.1002/hyp.7510, 2010b.

Hrachowitz, M., Soulsby, C., Tetzlaff, D., and Malcolm, I. A.: Sensitivity of mean transit time estimates to model conditioning and data availability, Hydrological Processes, 25, 980-990, doi: 10.1002/hyp.7922, 2011.

Jasechko, S., Kirchner, J. W., Welker, J. M., and McDonnell, J. J.: Substantial young streamflow in global rivers, Nature Geoscience, in review.

Kirchner, J. W., Dillon, P. J., and LaZerte, B. D.: Predictability of geochemical buffering and runoff acidification in spatially heterogeneous catchments, Water Resour. Res., 29, 3891-3901, 1993.

Kirchner, J. W., Feng, X., and Neal, C.: Fractal stream chemistry and its implications for contaminant transport in catchments, Nature, 403, 524-527, 2000.

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.

Kirchner, J. W.: A double paradox in catchment hydrology and geochemistry, Hydrological Processes, 17, 871-874, 2003.

Kirchner, J. W., Tetzlaff, D., and Soulsby, C.: Comparing chloride and water isotopes as hydrological tracers in two Scottish catchments, Hydrological Processes, 24, 1631-1645, doi: 10.1002/hyp.7676, 2010.

Kirchner, J. W., 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 of the United States of America, 110, 12213-12218, doi: 10.1073/pnas.1304328110, 2013.

Kirchner, J. W.: Aggregation in environmental systems**: Catchment mean transit times and** young water **fractions** under hydrologic nonstationarity, Hydrol. Earth Syst. Sci., submitted manuscript, 2015.

Maher, K.: The dependence of chemical weathering rates on fluid residence time, Earth. Planet. Sci. Lett., 294, 101-110, 2010.

Maher, K., and Chamberlain, C. P.: Hydrologic regulation of chemical weathering and the geologic carbon cycle, Science, 343, 1502-1504, 2014.

Maloszewski, P., Rauert, W., Stichler, W., and Herrmann, A.: Application of flow models in an alpine catchment area using tritium and deuterium data, J. Hydrol., 66, 319-330, 1983.

Maloszewski, P., and Zuber, A.: Principles and practice of calibration and validation of mathematical models for the interpretation of environmental tracer data in aquifers, Advances in Water Resources, 16, 173-190, 1993.

McDonnell, J. J., McGuire, K., Aggarwal, P., Beven, K. J., Biondi, D., Destouni, G., Dunn, S., James, A., Kirchner, J., Kraft, P., Lyon, S., Maloszewski, P., Newman, B., Pfister, L., Rinaldo, A., Rodhe, A., Sayama, T., Seibert, J., Solomon, K., Soulsby, C., Stewart, M., Tetzlaff, D., Tobin, C., Troch, P., Weiler, M., Western, A., Worman, A., and Wrede, S.: How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis, Hydrological Processes, 24, 1745-1754, doi: 10.1002/hyp.7796, 2010.

McDonnell, J. J., and Beven, K.: Debates-The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph, Water Resour. Res., 50, 5342-5350, doi: 10.1002/2013wr015141, 2014.

Deleted: 2: tracer cycles,

Deleted: , and catchment mean travel time

McGuire, K. J., McDonnell, J. J., Weiler, M., Kendall, C., McGlynn, B. L., Welker, J. M., and Seibert, J.: The role of topography on catchment-scale water residence time, Water Resour. Res., 41, W05002, 2005.

McGuire, K. J., and McDonnell, J. J.: A review and evaluation of catchment transit time modeling, J. Hydrol., 330, 543-563, 2006.

Nash, J. E.: The form of the instantaneous unit hydrograph, Comptes Rendus et Rapports, IASH General Assembly Toronto 1957, Int. Assoc. Sci. Hydrol. (Gentbrugge), Publ. No. 45, 3, 114-121, 1957.

Niemi, A. J.: Residence time distributions of variable flow processes, International Journal of Applied Radiation and Isotopes, 28, 855-860, 1977.

Peters, N. E., Burns, D. A., and Aulenbach, B. T.: Evaluation of high-frequency mean streamwater transit-time estimates using groundwater age and dissolved silica concentrations in a small forested watershed, Aquatic Geochemistry, 20, 183-202, 2014.

Rastetter, E. B., King, A. W., Cosby, B. J., Hornberger, G. M., O'Neill, R. V., and Hobbie, J. E.: Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems, Ecological Applications, 2, 55-70, 1992.

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.

Soulsby, C., Tetzlaff, D., Rodgers, P., Dunn, S., and Waldron, A.: Runoff processes, stream water residence times and controlling landscape characteristics in a mesoscale catchment: an initial evaluation, J. Hydrol., 325, 197-221, 2006.

Stewart, M. K., Morgenstern, U., and McDonnell, J. J.: Truncation of stream residence time: how the use of stable isotopes has skewed our concept of streamwater age and origin, Hydrological Processes, 24, 1646-1659, doi: 10.1002/hyp.7576, 2010.

Stewart, M. K., Morgenstern, U., McDonnell, J. J., and Pfister, L.: The 'hidden streamflow' challenge in catchment hydrology: a call to action for stream water transit time analysis, Hydrological Processes, 26, 2061-2066, doi: 10.1002/hyp.9262, 2012.

Tetzlaff, D., Malcolm, I. A., and Soulsby, C.: Influence of forestry, environmental change and climatic variability on the hydrology, hydrochemistry and residence times of upland catchments, J. Hydrol., 346, 93-111, 2007.

Tetzlaff, D., Seibert, J., and Soulsby, C.: Inter-catchment comparison to assess the influence of topography and soils on catchment transit times in a geomorphic province; the Cairngorm mountains, Scotland, Hydrological Processes, 23, 1874-1886, doi: 10.1002/hyp.7318, 2009.

Van der Velde, Y., De Rooij, G. H., Rozemeijer, J. C., van Geer, F. C., and Broers, H. P.: The nitrate response of a lowland catchment: on the relation between stream concentration and travel time distribution dynamics, Water Resour. Res., 46, W11534, doi: doi:10.1029/2010WR009105, 2010.

Walck, C.: Handbook on statistical distributions for experimentalists, Particle Physics Group, University of Stockholm, Stockholm, 202 pp., 2007.

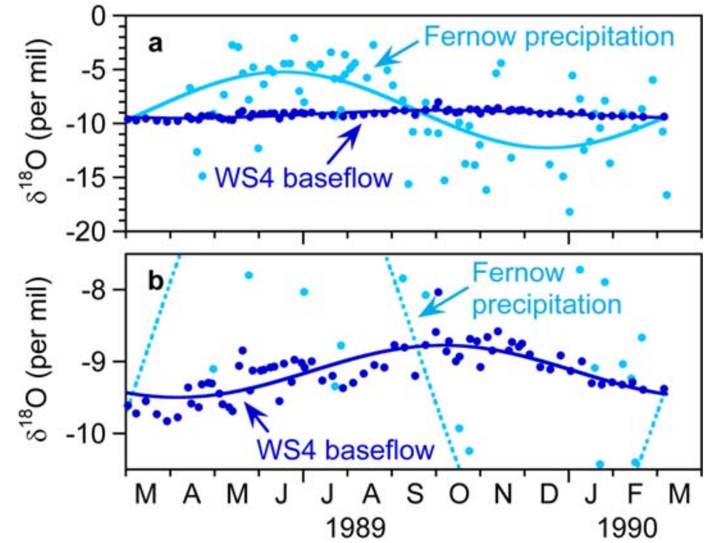

Figure 1. Seasonal cycles in $\delta^{18}O$ in precipitation and baseflow at catchment WS4, Fernow Experimental Forest, West Virginia, USA (DeWalle et al., 1997). Both panels show the same data; the axes of panel (b) are expanded to more clearly show the seasonal cycle in baseflow. Sinusoidal cycles are fitted by iteratively reweighted least squares regression (IRLS), a robust fitting technique that limits the influence of outliers.

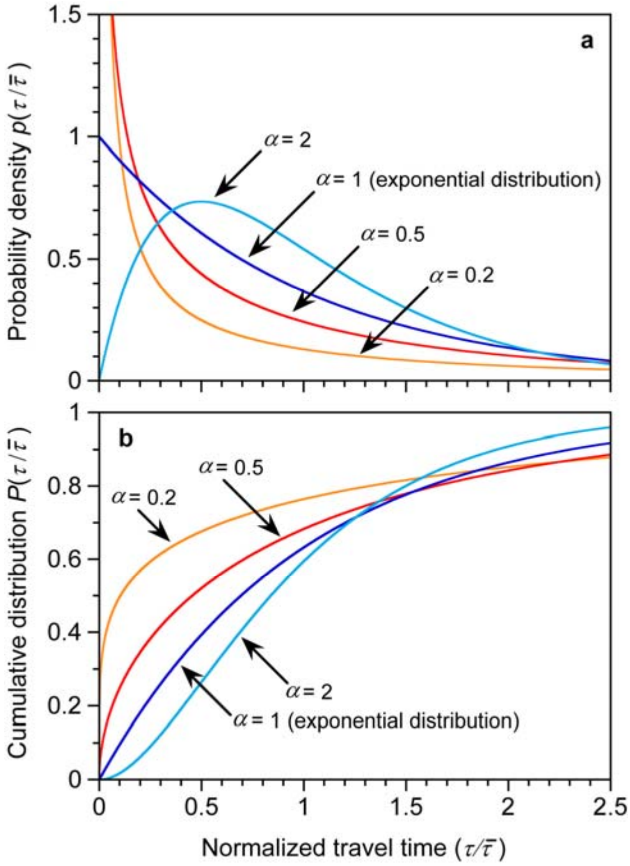

Figure 2. Gamma distributions for **the** range of shape factors **α=0.2-2 considered in this analysis.** Horizontal axes are normalized by the mean transit time $\bar{\tau}$, and thus are dimensionless.

Deleted: a

Deleted: α.

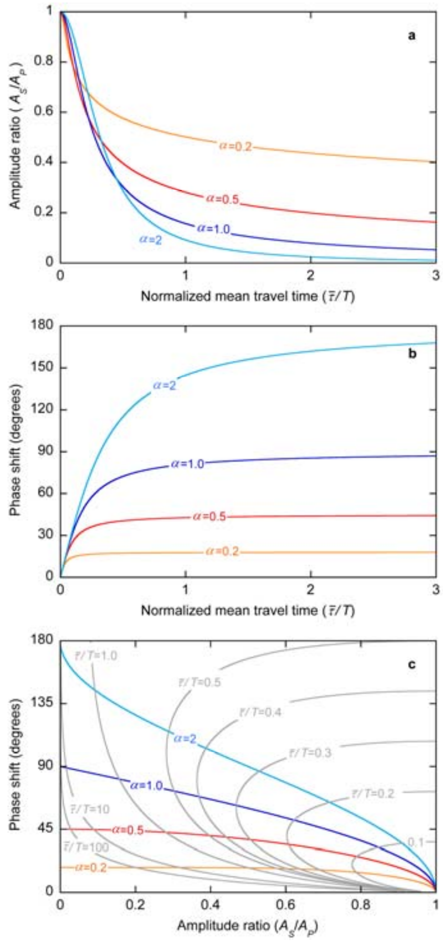

Figure 3. Amplitude ratio and phase shift between seasonal cycles in precipitation and streamflow, for gamma-distributed catchment transit time distributions with **a range of** shape factors α (colored lines). Panel (a): ratio of seasonal cycle amplitudes in streamflow and precipitation (A_S/A_P) as a function of mean transit time ($\bar{\tau}$) normalized by the period (T=1/f) of the tracer cycle. Panel (b): phase lag between streamflow and precipitation cycles, as a function of mean transit time normalized by the tracer cycle period ($\bar{\tau}/T$). Panel (c): relationship between phase lag and amplitude ratio, with contours of shape factor (α) ranging from 0.2 to 8 (colored lines), and contours of mean transit time normalized by tracer cycle period $\bar{\tau}/T$ (gray lines). For seasonal tracer cycles, T=1/f=1 yr and normalized transit time equals time in years.

Deleted: ranging from 0.2 to 8

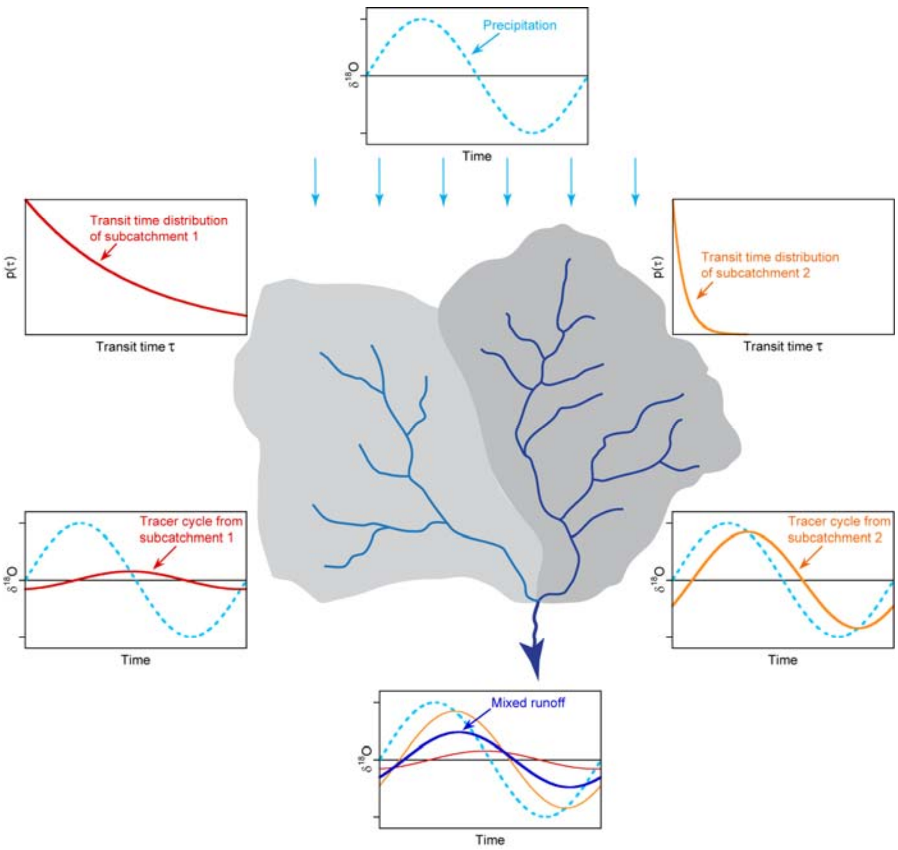

Figure 4. Conceptual diagram illustrating mixture of seasonal tracer cycles in runoff from a heterogeneous catchment, comprising two subcatchments with strongly contrasting mean transit times (MTT's), and which thus damp the tracer cycle in precipitation (light blue dashed line) by different amounts. The tracer cycle in the combined runoff from the two subcatchments (dark blue solid line) will average together the highly damped cycle from subcatchment 1, with long MTT (solid red line), and the less damped cycle from subcatchment 2, with short MTT (solid orange line).

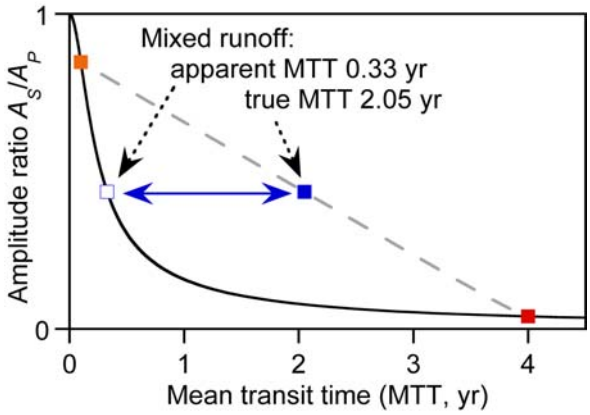

Figure 5. Illustration of the aggregation error that arises when mean transit time is inferred from seasonal tracer cycles in mixed runoff from two landscapes with contrasting transit time distributions (e.g., Fig. 4). The relationship between mean transit time (MTT) and the amplitude ratio (A_S/A_P) of annual cycles in streamflow and precipitation is strongly nonlinear (black curve). Seasonal cycles from subcatchments with MTT of 0.1 yr (A_S/A_P=0.85, orange square) and 4 yr (A_S/A_P=0.04, red square) will mix along the dashed gray line. A 50:50 mixture of the two sources will have a MTT of (4+0.1)/2=2.05 years and an amplitude ratio A_S/A_P of 0.43 (blue square). But if this amplitude ratio is interpreted as coming from a single catchment (Eq. 10), it implies a MTT of only 0.33 yr (open square), 6 times shorter than the true MTT of the mixed runoff.

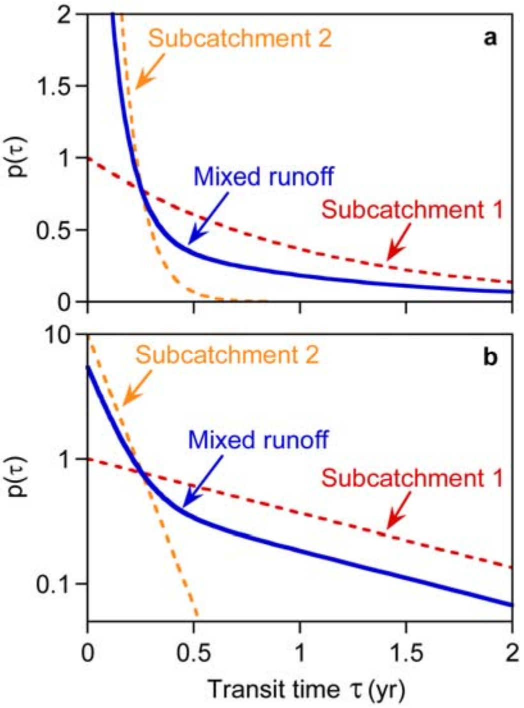

Figure 6. Exponential transit-time distributions for subcatchments 1 and 2 in Fig. 4 (with mean transit times of 1 and 0.1 yr, shown by the orange and red dashed lines, respectively), and the hyperexponential distribution formed by merging them in equal proportions (solid blue line). **Panels (a)** and **(b)** show linear and logarithmic axes.

Deleted: Left

Deleted: right panels

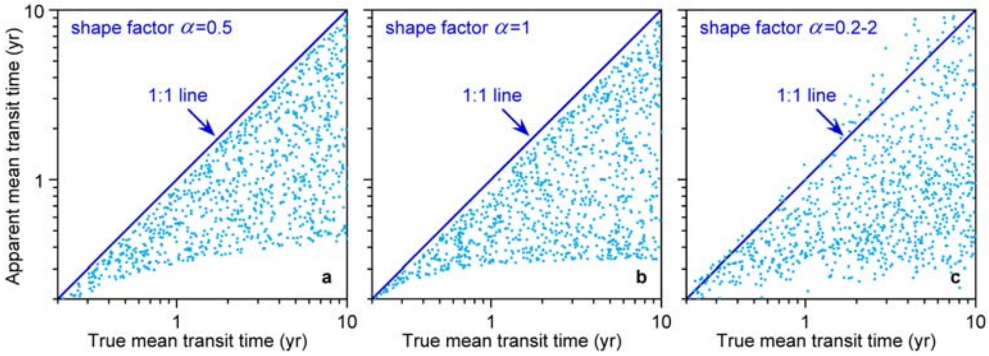

Figure 7. Apparent mean transit time (MTT) inferred from seasonal tracer cycles, showing order-of-magnitude deviations from true MTT for 1000 synthetic catchments. Each synthetic catchment comprises two subcatchments with individual MTT's randomly chosen from a uniform distribution of logarithms spanning the interval between 0.1 and 20 years, with each pair differing by at least a factor of 2. In panels (a) and (b), both subcatchments have shape factors α of 0.5 and 1, respectively; in panel (c), the subcatchments' shape factors are independently chosen from the range of 0.2 to 2. Apparent MTT's were inferred from the amplitude ratio A_S/A_P of the combined runoff using Eq. (10), with an assumed value of α=0.5 for panel (a), α=1 for panel (b), and also α=1 for panel (c), both because α=1 is close to the average of the randomized α values, and because α=1 is typically assumed whenever Eq. (10) is applied to real catchment data.

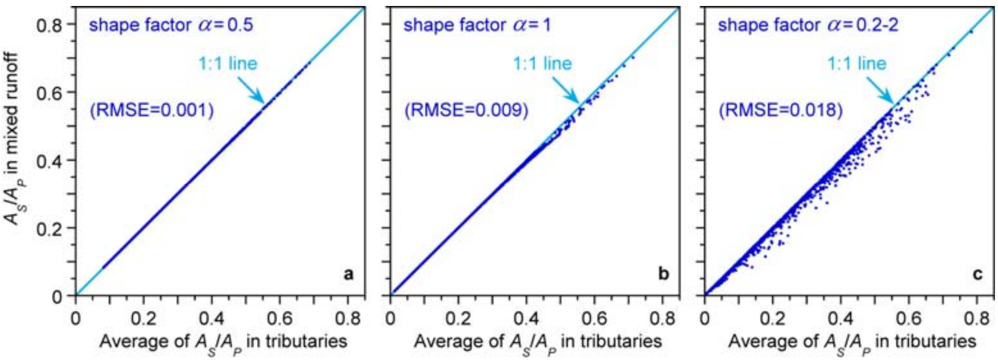

Figure 8. Amplitude ratio (A_S/A_P) of tracer cycles in precipitation and mixed runoff from the same 1000 synthetic catchments shown in Fig. 7 (vertical axes), compared to the average of the tracer cycle amplitude ratios in the two tributaries (horizontal axes). As in Fig. 7, each synthetic catchment comprises two subcatchments with individual MTT's randomly chosen from a uniform distribution of logarithms spanning the interval between 0.1 and 20 years, and with each pair of MTT's differing by at least a factor of 2. In panels (a) and (b), all subcatchments have the same shape factor α. In panel (c), shape factors for each subcatchment are randomly chosen from a uniform distribution between α=0.2 and α=2. The close fits to the 1:1 lines, and the small root-mean-square error (RMSE) values, show that the tracer cycle amplitudes from the tributaries are averaged almost exactly in the mixed runoff.

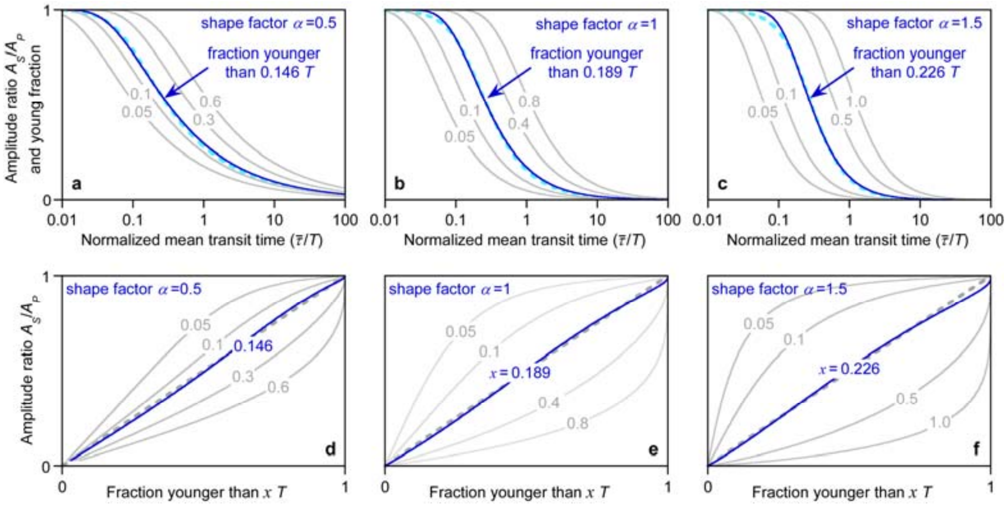

Figure 9. Panels (a)-(c) **show the** amplitude **ratios** A_S/A_P in precipitation and streamflow tracer cycles (light blue dashed line) as function of mean transit time $\bar{\tau}$, compared to the fraction of water younger than several threshold ages (gray lines), and the best-fit age threshold (dark blue line). Panels (d)-(f) **show the** relationship between amplitude ratio and the fraction of water younger than several age thresholds (gray lines) and the best-fit age threshold (dark blue line), with the 1:1 line (dashed gray) for comparison. Panels show results for three different gamma distributions, with shape factors α=0.5, α=1, and α=1.5. Root-mean-squared errors (RMSE's) for amplitude ratios A_S/A_P as predictors of the best-fit "young water" fractions are 0.012, 0.011, and 0.015 for panels (d), (e), and (f), respectively. In all panels, threshold age and mean transit time are normalized by T, the period of the tracer cycle. For seasonal tracer cycles, T=1 yr and thus threshold age and mean transit time are in years.

Deleted: -
Deleted: :
Deleted: ratio
Deleted: a
Deleted: -
Deleted: :

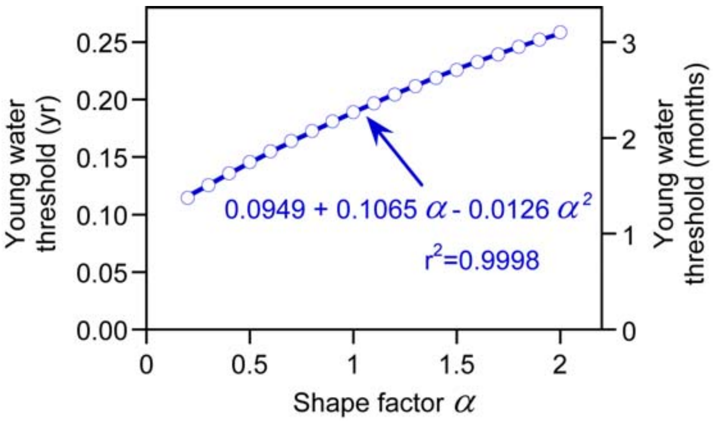

Figure 10. Best-fit "young water" thresholds for gamma transit time distributions, as a function of shape factors α ranging from 0.2 to 2.0. The young water threshold τ_{yw} is defined such that the fraction of the distribution with ages less than τ_{yw} approximately equals the amplitude ratio (A_S/A_P) of annual cycles in streamflow and precipitation (see Fig. 9).

Deleted: x

Deleted: x

Deleted: ¶

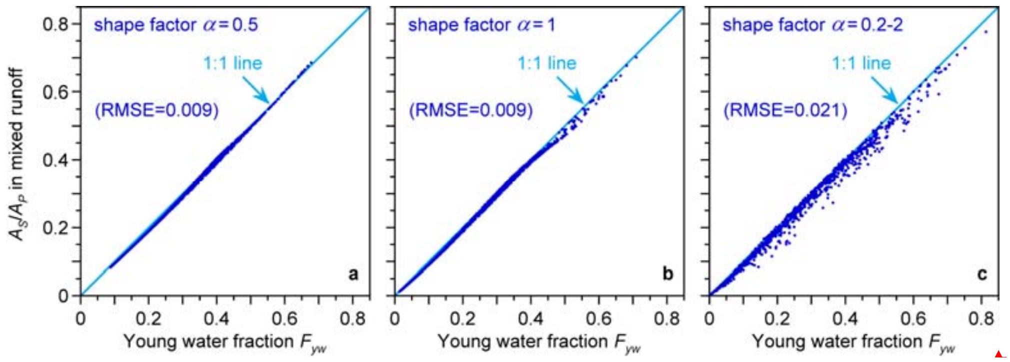

Figure 11. True and apparent "young water" fractions for the same 1000 synthetic catchments shown in Fig. 7. The tracer cycle amplitude ratio in the combined runoff of the two subcatchments (vertical axes) corresponds closely to the average young water fraction in the combined runoff (horizontal axes). As in Fig. 7, each synthetic catchment comprises two subcatchments with individual MTT's randomly chosen from a uniform distribution of logarithms spanning the interval between 0.1 and 20 years, and with each pair of MTT's differing by at least a factor of 2. In panels (a) and (b), all subcatchments have the same shape factor α. In panel (c), shape factors for each subcatchment are randomly chosen from a uniform distribution between α=0.2 and α=2.

Formatted: English (U.S.)

Deleted: horzontal

Formatted: Right: 0 cm

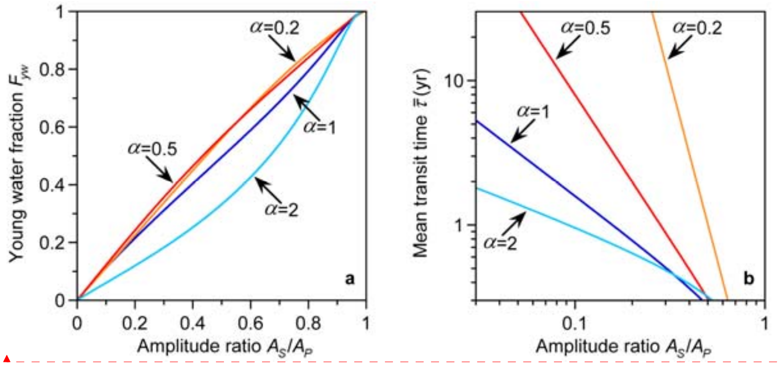

Figure 12. **Sensitivity analysis showing how variations in shape factor α affect young water fractions F_{yw} (panel a) and mean transit times $\bar{\tau}$ (panel b) inferred from the amplitude ratio A_S/A_P of seasonal tracer cycles in precipitation and streamflow. Curves are shown for the four shape factors shown in Figs. 2 and 3. For a plausible range of uncertainty in the shape factor (0.5<α<1; see Sect. 2.1), estimated young water fractions vary by a few percent (panel a), whereas estimated mean transit times vary by large multiples (note the logarithmic axes in panel b). Panel (a) shows the fractions of water younger than τ_{yw}=2.27 months, which are closely approximated by A_S/A_P if α=1 (the dark blue curve). In panel (b), the axis scales are chosen to span transit times ranging from several months to several years, as is commonly observed in transit time studies (McGuire and McDonnell, 2006).**

Formatted: English (U.K.)

Deleted: ¶

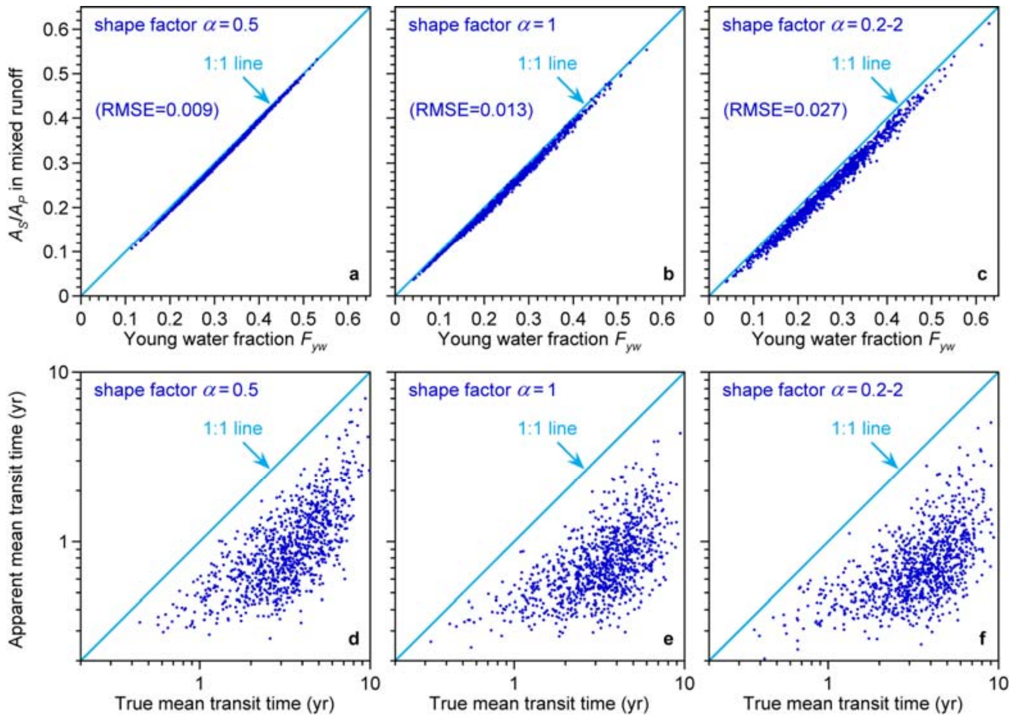

Figure 13. True and apparent "young water" fractions F_{yw} for 1000 synthetic catchments, each consisting of 8 subcatchments with randomly chosen mean transit times between 0.1 and 20 years (top panels), and true and apparent mean transit times for the same catchments (bottom panels). The tracer cycle amplitude ratio in the combined runoff predicts the true young water fraction with a slight underestimation bias (top panels). Mean transit times inferred from tracer cycle amplitude ratios show severe underestimation bias (bottom panels). In panels (a-b) and (d-e), all subcatchments have the same shape factor α. In panels (c) and (f), shape factors for each subcatchment are randomly chosen from a uniform distribution between α=0.2 and α=2.

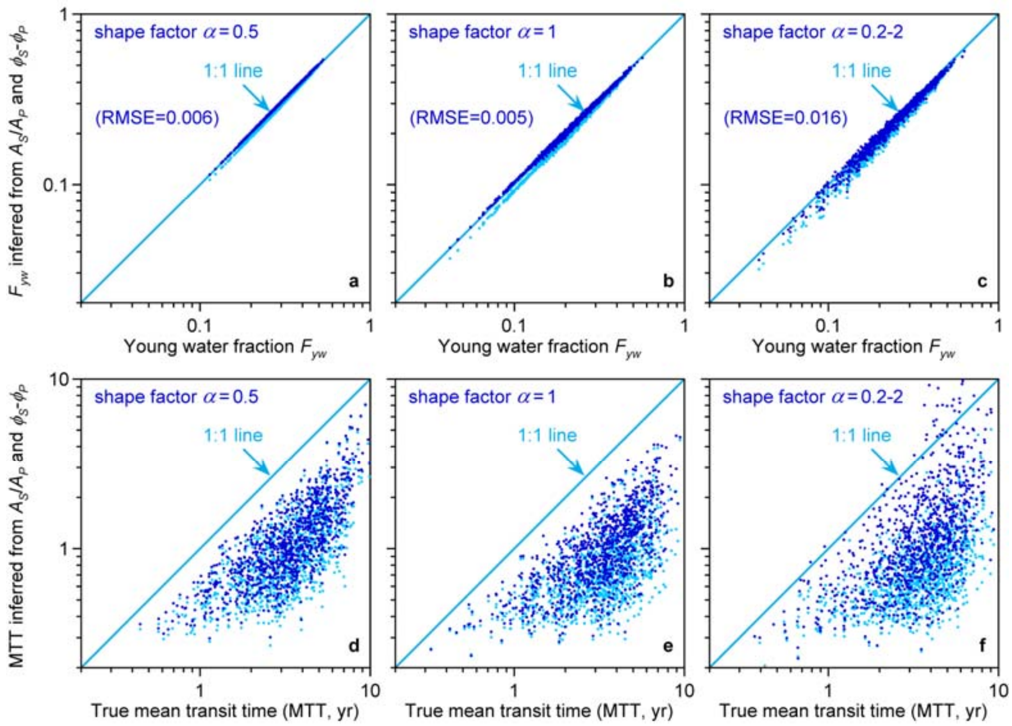

Figure 14. Effect of including phase information in estimates of young water fraction (F_{yw}) and mean transit time (MTT). Light symbols show F_{yw} and MTT estimates derived from tracer cycle amplitude ratios (A_S/A_P) alone; dark symbols show the same estimates derived from amplitude ratios and phase shifts (ϕ_S-ϕ_P). Data points come from the same 1000 synthetic catchments shown in Fig. 13, each consisting of 8 subcatchments with randomly chosen mean transit times between 0.1 and 20 years. Adding phase shift information eliminates much of the (already small) bias in F_{yw} estimates, particularly when F_{yw} is small. Adding phase information reduces the bias in MTT estimates as well, but a severe underestimation bias remains.

Deleted: 13

Deleted: 12