**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)