

1 **Aggregation in environmental systems: Seasonal tracer** 2 **cycles quantify young water fractions, but not mean transit** 3 **times, in spatially heterogeneous catchments**

4
5 **J. W. Kirchner^{1,2}**

6 [1]{ETH Zürich, Zürich, Switzerland}

7 [2]{Swiss Federal Research Institute WSL, Birmensdorf, Switzerland}

8
9 Correspondence to: J. W. Kirchner (kirchner@ethz.ch)

10 11 **Abstract**

12 Environmental heterogeneity is ubiquitous, but environmental systems are often analyzed as if
13 they were homogeneous instead, resulting in aggregation errors that are rarely explored and
14 almost never quantified. Here I use simple benchmark tests to explore this general problem in
15 one specific context: the use of seasonal cycles in chemical or isotopic tracers (such as Cl^- ,
16 $\delta^{18}\text{O}$, or $\delta^2\text{H}$) to estimate timescales of storage in catchments. Timescales of catchment
17 storage are typically quantified by the mean transit time, meaning the average time that
18 elapses between parcels of water entering as precipitation and leaving again as streamflow.
19 Longer mean transit times imply greater damping of seasonal tracer cycles. Thus, the
20 amplitudes of tracer cycles in precipitation and streamflow are commonly used to calculate
21 catchment mean transit times. Here I show that these calculations will typically be wrong by
22 several hundred percent, when applied to catchments with realistic degrees of spatial
23 heterogeneity. This aggregation bias arises from the strong nonlinearity in the relationship
24 between tracer cycle amplitude and mean travel time. I propose an alternative storage metric,
25 the young water fraction in streamflow, defined as the fraction of runoff with transit times of
26 less than roughly 0.2 years. I show that this young water fraction (not to be confused with
27 event-based "new water" in hydrograph separations) is accurately predicted by seasonal tracer
28 cycles within a precision of a few percent, across the entire range of mean transit times from
29 almost zero to almost infinity. Importantly, this relationship is also virtually free from

1 aggregation error. That is, seasonal tracer cycles also accurately predict the young water
2 fraction in runoff from highly heterogeneous mixtures of subcatchments with strongly
3 contrasting transit time distributions. Thus, although tracer cycle amplitudes yield biased and
4 unreliable estimates of catchment mean travel times in heterogeneous catchments, they can be
5 used reliably to estimate the fraction of young water in runoff.

6

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

9

10 **1 Introduction**

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

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

1 heterogeneity of the system, at least well enough to generate reliable predictions of whole-
2 catchment behavior?

3 This is a testable proposition, and the answer will depend partly on the nature of the
4 underlying model. All models obscure a system's spatial heterogeneity to some degree, and
5 many conceptual models obscure it completely, by treating spatially heterogeneous
6 catchments as if they were spatially homogeneous instead. Doing so is not automatically
7 disqualifying, but neither is it obviously valid. Rather, this spatial aggregation is a modelling
8 choice, whose consequences should be explicitly analyzed and quantified. What do I mean by
9 "explicitly analyzed and quantified?" As an example, consider Kirchner et al.'s (1993)
10 analysis of how spatial heterogeneity affected a particular geochemical model for estimating
11 catchment buffering of acid deposition. The authors began by noting that spatial
12 heterogeneities will not "average out" in nonlinear model equations, and by showing that the
13 resulting aggregation bias will be proportional to the nonlinearity in the model equations
14 (which can be directly estimated), and proportional to the variance in the heterogeneous real-
15 world parameter values (which is typically unknown, but may at least be given a plausible
16 upper bound). They then showed that their geochemical model's governing equations were
17 sufficiently linear that the effects of spatial heterogeneity were likely to be small. They then
18 confirmed this theoretical result by mixing measured runoff chemistry time series from
19 random pairs of geochemically diverse catchments (which do not flow together in the real
20 world). They showed that the geochemical model correctly predicted the buffering behavior
21 of these spatially heterogeneous pseudo-catchments, without knowing that those catchments
22 were heterogeneous, and without knowing anything about the nature of their heterogeneities.
23 Here I use similar thought experiments to explore the consequences of spatial heterogeneity
24 for catchment mean transit time estimates derived from seasonal tracer cycles in precipitation
25 and streamflow. Catchment *transit time*, or, equivalently, *travel time* – the time that it takes
26 for rainfall to travel through a catchment and emerge as streamflow – is a fundamental
27 hydraulic parameter that controls the retention and release of contaminants and thus the
28 downstream consequences of pollution episodes (Kirchner et al., 2000; McDonnell et al.,
29 2010). In many geological settings, catchment transit times also control chemical weathering
30 rates, geochemical solute production and the long-term carbon cycle (Burns et al., 2003;
31 Godsey et al., 2009; Maher, 2010; Maher and Chamberlain, 2014).

1 A catchment is characterized by its travel time distribution (TTD), which reflects the diversity
2 of flowpaths (and their velocities) connecting each point on the landscape with the stream.
3 Because these flowpaths and velocities change with hydrologic forcing, the TTD is non-
4 stationary (Kirchner et al., 2001; Tetzlaff et al., 2007; Botter et al., 2010; Hrachowitz et al.,
5 2010a; Van der Velde et al., 2010; Birkel et al., 2012; Heidbüchel et al., 2012; Peters et al.,
6 2014), but time-varying TTD's are difficult to estimate in practice, so most catchment studies
7 have focused on estimating time-averaged TTD's instead. Both the shape of the TTD and its
8 corresponding mean travel time (MTT) reflect storage and mixing processes in the catchment
9 (Kirchner et al., 2000, 2001; Godsey et al., 2010; Hrachowitz et al., 2010a). However, due to
10 the difficulty in reliably estimating the shape of the TTD, and the volumes of data required to
11 do so, many catchment studies have simply assumed that the TTD has a given shape, and
12 have estimated only its MTT. As a result, and also because of its obvious physical
13 interpretation as the ratio between the storage volume and the average water flux (in steady
14 state), the MTT is by far the most universally reported parameter in catchment travel-time
15 studies. Estimates of MTT's have been correlated with a wide range of catchment
16 characteristics, including drainage density, aspect, hillslope gradient, depth to groundwater,
17 hydraulic conductivity, and the prevalence of hydrologically responsive soils (e.g., McGuire
18 et al., 2005; Soulsby et al., 2006; Tetzlaff et al., 2009; Broxton et al., 2009; Hrachowitz et al.,
19 2009; Hrachowitz et al., 2010b; Asano and Uchida, 2012; Heidbüchel et al., 2013),
20 Travel time distributions and mean travel times cannot be measured directly, and they differ –
21 often by orders of magnitude – from the hydrologic response timescale, because the former is
22 determined by the velocity of water flow, and the latter is determined by the celerity of
23 hydraulic potentials (Horton and Hawkins, 1965; Hewlett and Hibbert, 1967; Beven, 1982;
24 Kirchner et al., 2000; McDonnell and Beven, 2014). Nor can travel time characteristics be
25 reliably determined *a priori* from theory. Instead, they must be determined from chemical or
26 isotopic tracers, such as Cl⁻, ¹⁸O, and ²H, in precipitation and streamflow. These passive
27 tracers "follow the water"; thus their temporal fluctuations reflect the transport, storage, and
28 mixing of rainfall as it is transformed into runoff. (Groundwaters can also be dated using
29 dissolved gases such as CFC's and ³H/³He, but these tracers are not conserved in surface
30 waters or in the vadose zone, so they are not well suited to estimating whole-catchment travel
31 times.)

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

22 The isotopic composition of precipitation varies seasonally as shifts in meridional circulation
23 alter atmospheric vapor transport pathways (Feng et al., 2009), and as shifts in temperature
24 and storm intensity alter the degree of rainout-driven fractionation that air masses undergo
25 (Bowen, 2008). The resulting seasonal cycles in precipitation (e.g., Fig. 1a) are damped and
26 phase-shifted as they are transmitted through catchments (e.g., Fig. 1b), by amounts that
27 depend on – and thus can be used to infer properties of – the travel-time distribution. Figure 1
28 shows an example of sinusoidal fits to seasonal $\delta^{18}\text{O}$ cycles in precipitation and baseflow at
29 one particular field site. The visually obvious damping of the isotopic cycle in baseflow
30 relative to precipitation implies, in this case, an estimated MTT of 1.4 years (DeWalle et al.,
31 1997) under the assumption that the TTD is exponential.

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

13

14 **2 Mathematical preliminaries: tracer cycles in homogeneous catchments**

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

$$20 \quad c_S(t) = \int_0^{\infty} h(\tau) c_P(t-\tau) d\tau, \quad (1)$$

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

26 The transit-time distribution $h(\tau)$ expresses the fractional contribution of past inputs to present
27 runoff. Equation (1) implicitly assumes that the catchment is a linear time-invariant system,
28 and thus that the convolution kernel $h(\tau)$ is stationary (i.e., constant through time). This is
29 never strictly true, most obviously because if no precipitation falls on a particular day, it
30 cannot contribute any tracer to the stream τ days later, and because higher precipitation rates

1 will increase the rate that water and tracers are flushed through the catchment. Thus real-
 2 world TTD's vary through time, depending on the history of prior precipitation (Kirchner et
 3 al., 2001; Tetzlaff et al., 2007; Botter et al., 2010; Hrachowitz et al., 2010a; Van der Velde et
 4 al., 2010; Birkel et al., 2012; Heidbüchel et al., 2012; Peters et al., 2014). However, in
 5 applications using real-world data, $h(\tau)$ is conventionally interpreted as a time-invariant
 6 ensemble average, taken over an ensemble of precipitation histories, which obviously will
 7 differ from one another in detail. Mathematically, the ensemble averaging embodied in Eq.
 8 (1) is equivalent to the simplifying assumption that water fluxes in precipitation and
 9 streamflow are constant over time. (One can relax this assumption somewhat by integrating
 10 over the cumulative water flux rather than time, as proposed by Niemi (1977). If the rates of
 11 transport and mixing vary proportionally to the flow rate through the catchment, this yields a
 12 stationary distribution in flow-equivalent time.) A further simplification inherent in Eq. (1) is
 13 that evapotranspiration and its effects on tracer signatures are ignored.

14 **2.1 A class of transit-time distributions**

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

$$17 \quad h(\tau) = \frac{\tau^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\tau/\beta} = \frac{\tau^{\alpha-1}}{(\bar{\tau}/\alpha)^\alpha \Gamma(\alpha)} e^{-\alpha \tau / \bar{\tau}}, \quad (2)$$

18 where α and β are a shape factor and scale factor, respectively, τ is the transit time, and
 19 $\bar{\tau} = \alpha \beta$ is the mean transit time. I make this assumption mostly so that some key results can
 20 be calculated exactly, but as I show below, the key results extend beyond this (already broad)
 21 class of distributions.

22 Figure 2 shows gamma distributions spanning a range of shape factors α . For the special case
 23 of $\alpha=1$, the gamma distribution becomes the exponential distribution. Exponential
 24 distributions describe the behavior of continuously mixed reservoirs of constant volume, and
 25 they have been widely used to model catchment storage and mixing. The gamma distribution
 26 expresses the TTD of a Nash cascade (Nash, 1957) of α identical linear reservoirs connected
 27 in series, and the analogy to a Nash cascade holds even for non-integer α , through the use of
 28 fractional integration. For $\alpha>1$, the gamma distribution rises to a peak and then falls off,

1 similarly to a typical storm hydrograph, which is why Nash cascades have often been used to
 2 model rainfall-runoff relationships. For $\alpha < 1$, however, the gamma distribution has a
 3 completely different shape, having maximum weight at lags near zero, and also having a
 4 relatively long tail. These characteristics represent problematic contaminant behavior: an
 5 intense spike of contamination in short time and persistent contamination in long time. Tracer
 6 time series from many catchments have been shown to exhibit fractal $1/f$ scaling, which is
 7 consistent with gamma TTD's with $\alpha \approx 0.5$ (Kirchner et al., 2000, 2001; Godsey et al., 2010;
 8 Kirchner and Neal, 2013; Aubert et al., 2014).

9 For present purposes, it is sufficient to note that the family of gamma distributions
 10 encompasses a wide range of shapes which approximate many plausible TTD's (Fig. 2). The
 11 moments of the gamma distribution vary systematically with the shape factor α (Walck,
 12 2007):

$$13 \quad \text{mean}(\tau) = \beta \alpha = \bar{\tau}, \quad (3a)$$

$$14 \quad \text{SD}(\tau) = \beta \sqrt{\alpha} = \bar{\tau} / \sqrt{\alpha}, \quad (3b)$$

$$15 \quad \text{skewness}(\tau) = 2 / \sqrt{\alpha}, \quad (3c)$$

$$16 \quad \text{and kurtosis}(\tau) = 6 / \alpha. \quad (3d)$$

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

23 Studies that have used tracers to constrain the shape of catchment TTD's have generally found
 24 shape factors α ranging from 0.3 to 0.7, corresponding to spectral slopes of the transfer
 25 function between roughly 0.6 and 1.4 (Kirchner et al., 2000, 2001; Godsey et al., 2010;
 26 Hrachowitz et al., 2010a; Kirchner and Neal, 2013; Aubert et al., 2014). Other studies –
 27 including those that have used annual tracer cycles to estimate mean transit times – have
 28 assumed that the catchment is a well-mixed reservoir and thus that $\alpha=1$. Here I will assume
 29 that α falls in the range of 0.5 to 1 for typical catchment transit-time distributions, but I will
 30 also show some key results for the somewhat wider range of $\alpha=0.2-2$, for illustrative
 31 purposes. The results reported here will not necessarily apply to TTD's that rise to a peak

1 after a long delay, such as the gamma distribution with $\alpha \gg 2$. However, one would not
 2 expect such a distribution to characterize whole-catchment TTD's in the first place, because
 3 except in very unusual catchments a substantial amount of precipitation can fall close to the
 4 stream and enter it relatively quickly, thus producing a strong peak at a short lag (Kirchner et
 5 al., 2001).

6 **2.2 Estimating mean transit time from tracer cycles**

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

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

$$18 \begin{aligned} c_P(t) &= A_P \sin(2\pi f t - \phi_P) + k_P \\ c_S(t) &= A_S \sin(2\pi f t - \phi_S) + k_S \end{aligned} \quad (4)$$

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

$$20 \begin{aligned} c_P(t) &= a_P \cos(2\pi f t) + b_P \sin(2\pi f t) + k_P \\ c_S(t) &= a_S \cos(2\pi f t) + b_S \sin(2\pi f t) + k_S \end{aligned} \quad (5)$$

21 and then calculating the amplitudes and phases using the conventional identities

$$22 \quad A_P = \sqrt{a_P^2 + b_P^2}, \quad A_S = \sqrt{a_S^2 + b_S^2}, \quad \phi_P = \arctan(b_P / a_P) \text{ and } \phi_S = \arctan(b_S / a_S). \quad (6)$$

23 In Eqs. (4)-(6) above, t is time, f is the frequency of the cycle ($f=1 \text{ yr}^{-1}$ for a seasonal cycle),
 24 and the subscripts P and S refer to precipitation and streamflow. In fitting sinusoidal cycles to
 25 real-world data, robust estimation techniques such as iteratively reweighted least squares
 26 (IRLS) regression can help in limiting the influence of outliers. Also, because precipitation
 27 and streamflow rates vary through time, it may be useful to weight each tracer sample by its

1 associated volume, for example to reduce the influence of small rainfall events (for more on
 2 the implications of volume-weighting, see Kirchner, 2015). An R script for performing
 3 volume-weighted IRLS is available from the author.

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

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

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

$$10 \quad \frac{A_S}{A_P} = \left(1 + (2\pi f \beta)^2\right)^{-\alpha/2} , \quad (8)$$

11 and also the phase shift between them,

$$12 \quad \phi_S - \phi_P = \alpha \arctan(2\pi f \beta) \quad , \quad (9)$$

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

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

$$17 \quad \bar{\tau} = \alpha \beta \quad , \quad \beta = \frac{1}{2\pi f} \sqrt{\left(\frac{A_S}{A_P}\right)^{-2/\alpha} - 1} \quad . \quad (10)$$

18 Equation (10), with $\alpha=1$, is the standard tool for estimating MTT's from seasonal tracer cycles
 19 in precipitation and streamflow. Alternatively, as Fig. 3c shows, both the shape factor α and
 20 the mean transit time $\bar{\tau}$ can be jointly determined from the phase shift $\phi_S - \phi_P$ and the
 21 amplitude ratio A_S/A_P , if these can both be quantified with sufficient accuracy.

22 Mathematically, this joint solution can be achieved by substituting Eq. (10) in Eq. (9),
 23 yielding the following implicit expression for α ,

$$24 \quad \phi_S - \phi_P = \alpha \arctan\left(\sqrt{\left(\frac{A_S}{A_P}\right)^{-2/\alpha} - 1}\right) , \quad (11)$$

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

8

9 **3 Transit times and tracer cycles in heterogeneous catchments: a thought** 10 **experiment**

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

26 From assumed MTT's $\bar{\tau}$ and shape factors α for each of the subcatchments, one can calculate
27 the amplitude ratios A_S/A_P and phase shifts $\phi_S - \phi_P$ of their tracer cycles using Eqs. (8)-(9)
28 above, and then average these cycles together using the conventional trigonometric identities.
29 (Equivalently, one can estimate the cosine and sine coefficients of the individual
30 subcatchments' tracer cycles from the real and imaginary parts of Eq. (7) and algebraically
31 average them together.) The shares of the two subcatchments in the average will depend on
32 their relative drainage areas and/or water yields. For simplicity, I combine the runoff from

1 the two subcatchments in a 1:1 ratio; this also guarantees that the combined runoff will be as
2 different as possible from each of the two sources. I then ask the question: from the tracer
3 behavior in the combined runoff (the solid blue line in Fig. 4), would I correctly estimate the
4 mean transit time for the whole catchment? That is, would I infer a MTT that is close to the
5 average of the MTT's of the two subcatchments?

6 One can immediately see that this situation is highly prone to aggregation bias, following
7 Kirchner et al.'s (1993) rule of thumb that the degree of aggregation bias is proportional to the
8 nonlinearity in the governing equations and the variance in the heterogeneous parameters.

9 The amplitude ratios A_S/A_P and phase shifts $\phi_S - \phi_P$ of seasonal tracer cycles are strongly
10 nonlinear functions of the MTT (see Eqs. 8 and 10), as illustrated in Figs. 3a-b. And,
11 importantly, the likely range of variation in subcatchment MTT's (from, say, fractions of a
12 year to perhaps several years) straddles the nonlinearity in the governing equations. Thus we
13 should expect to see significant aggregation bias in estimates of MTT.

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

28 Combining flows from two subcatchments with different mean transit times will result in a
29 combined TTD that differs in shape, not just in scale, from the TTD's of either of the
30 subcatchments. For example, combining two exponential distributions with different mean
31 transit times does not result in another exponential distribution, but rather a hyperexponential
32 distribution, as shown in Fig. 6. The characteristic function of the hyperexponential

1 distribution (Walck, 2007) yields the following expression for the amplitude ratio of tracer
 2 cycles in precipitation and streamflow,

$$3 \frac{A_S}{A_P} = \left(\left(\frac{p}{1 + (2\pi f \bar{\tau}_1)^2} + \frac{q}{1 + (2\pi f \bar{\tau}_2)^2} \right)^2 + \left(\frac{p 2\pi f \bar{\tau}_1}{1 + (2\pi f \bar{\tau}_1)^2} + \frac{q 2\pi f \bar{\tau}_2}{1 + (2\pi f \bar{\tau}_2)^2} \right)^2 \right)^{1/2}, \quad (12)$$

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

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

23 In most real-world cases, unlike these hypothetical thought experiments, one will only have
 24 measurements or samples from the whole catchment's runoff. The properties of the individual
 25 subcatchments, and thus the degree of heterogeneity in the system, will generally be
 26 unknown. And even if data were available for the subcatchments, those subcatchments would
 27 be composed of sub-sub-catchments, which would themselves be heterogeneous to some
 28 unknown degree, and so on. Thus it will generally be difficult or impossible to characterize
 29 the system's heterogeneity, but that is no justification for pretending that this heterogeneity
 30 does not exist. Nonetheless, in such situations it will be tempting to treat the whole system as

1 if it were homogeneous, perhaps using terms like "apparent age" or "model age" to preserve a
2 sense of rigor. But whatever the semantics, as Fig. 7 shows, assuming homogeneity in
3 heterogeneous catchments will result in strongly biased estimates of whole-catchment mean
4 transit times.

5

6 **4 Quantifying the young water component of streamflow**

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

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

22 Because tracer cycle amplitudes will average almost linearly when two streams merge, and
23 thus are virtually free from aggregation bias (Fig. 8), anything that is proportional to tracer
24 cycle amplitude will also be virtually free from aggregation bias. So, what is proportional to
25 tracer cycle amplitude? One hint is provided by the observation that in Fig. 5, for example,
26 the tracer cycle amplitude in the mixture is highly sensitive to transit times that are much
27 shorter than the period of the tracer cycle (for a seasonal cycle, this period is $T=1$ yr), but
28 highly insensitive to transit times that are much longer than the period of the tracer cycle. As
29 a thought experiment, one can imagine a catchment in which some fraction of precipitation
30 bypasses storage entirely (and thus transmits the precipitation tracer cycle directly to the
31 stream), while the remainder is stored and mixed over very long time scales (and thus its
32 tracer cycles are completely obliterated by mixing). In this idealized catchment, the

1 amplitude ratio A_S/A_P between the tracer cycles in the stream and precipitation will be
 2 proportional to (indeed it will be exactly *equal to*) the fraction of precipitation that bypasses
 3 storage (and thus has a near-zero transit time).

4 **4.1 Young water**

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

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

$$16 \quad F_{yw} = P(\tau < \tau_{yw}) = \Gamma(\tau_{yw}, \alpha, \beta) = \int_{\tau=0}^{\tau_{yw}} \frac{\tau^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\tau/\beta} d\tau, \quad (13)$$

17 where, as before, $\beta = \bar{\tau} / \alpha$. I then numerically search for the threshold age for which (for a
 18 given shape factor α) the amplitude ratio A_S/A_P closely approximates F_{yw} across a wide range
 19 of scale factors β (or equivalently, a wide range of mean transit times $\bar{\tau}$). As Fig. 9 shows,
 20 this young water fraction nearly equals the amplitude ratio A_S/A_P , with the threshold for
 21 "young" water varying from 1.7 to 2.7 months as the shape factor α ranges from 0.5 to 1.5.
 22 The amplitude ratio A_S/A_P and the young water fraction F_{yw} are both dimensionless and they
 23 both range from 0 to 1, so they can be directly compared without further calibration, beyond
 24 the determination of the threshold age τ_{yw} . As Fig. 10 shows, the best-fit threshold age varies
 25 modestly as a function of the shape factor α ,

$$26 \quad \tau_{yw} / T \approx 0.0949 + 0.1065 \alpha - 0.0126 \alpha^2. \quad (14)$$

27 Across the entire range of $\alpha=0.2$ to $\alpha=2$ shown in Fig. 10, and across the entire range of
 28 amplitude ratios from 0 to 1 (and thus mean transit times from zero to near-infinity), the

1 amplitude ratio A_S/A_P estimates the "young water" fraction with a root mean square error of
2 less than 0.023, or 2.3 percent.

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

18 **4.2 Sensitivity to assumed TTD shape and threshold age**

19 The analysis presented in Sect. 4.1 shows that the amplitude ratio A_S/A_P accurately estimates
20 the fraction of streamflow younger than a threshold age. But this threshold age depends on
21 the shape factor α of the subcatchment TTDs, which will generally be uncertain. Consider,
22 for example, a hypothetical case where we measure an amplitude ratio of $A_S/A_P=0.2$ in the
23 seasonal tracer cycles in a particular catchment, but we don't know whether its subcatchments
24 are characterized by $\alpha=1$, $\alpha=0.5$, or a mixture of distributions between these shape factors.
25 How much does this uncertainty in α , and thus in the threshold age, affect the inferences we
26 can draw from A_S/A_P ? We can approach this question from two different perspectives.

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

30 First, from Fig. 10 we can estimate how uncertainty in α affects the threshold age τ_{yw} that
31 defines what counts as 'young' streamflow. One can see that across the plausible range of

1 shape factors, the young water threshold (that is, the threshold defining whatever young water
2 fraction will aggregate correctly) varies from about $\tau_{yw}=1.75$ months for $\alpha=0.5$ to $\tau_{yw}=2.27$
3 months for $\alpha=1$. Thus the ambiguity in α translates into an ambiguity of 0.52 months (or
4 about two weeks) in the threshold that defines "young" water. If some subcatchments are
5 characterized by $\alpha=0.5$ and others by $\alpha=1$, and still others by values in between, then the
6 effective threshold age for the ensemble will lie somewhere between 1.75 and 2.27 months. If
7 the range of uncertainty in α is wider, then the range of uncertainty in τ_{yw} will be wider as
8 well, spanning over a factor of two (1.37 to 3.10 months) for values of α spanning the full
9 order-of-magnitude range shown in Fig. 2 ($\alpha=0.2$ to 2).

10 Alternatively, we can treat the uncertainty in α as creating, for any fixed threshold age, an
11 ambiguity in the fraction of streamflow that is younger than that age. Consider the
12 hypothetical case outlined above, in which $A_S/A_P=0.2$. If we assume that the subcatchments
13 are characterized by $\alpha=1$ (and thus $\tau_{yw}=2.27$ months), then we would infer that roughly 20%
14 of streamflow is younger than 2.27 months (the exact young water fraction, using Eqs. (10)
15 and (13), is 0.215). But if the subcatchments are characterized by $\alpha=0.5$ instead, then
16 according to Eqs. (10) and (13) the fraction younger than 2.27 months will be 0.242 instead of
17 0.215. Thus the uncertainty in α corresponds to an uncertainty in the young water fraction of
18 3% (of the range of a priori uncertainty in F_{yw} , which is between 0 and 1), or 13% (of the
19 original estimate for $\alpha=1$).

20 For comparison, we can contrast this uncertainty with the corresponding uncertainty in the
21 mean transit time $\bar{\tau}$ calculated from Eq. (10). A seasonal tracer cycle amplitude ratio
22 $A_S/A_P=0.2$ implies a mean transit time of $\bar{\tau} = 0.80$ years if $\alpha=1$, but $\bar{\tau} = 1.99$ years if $\alpha=0.5$.
23 Thus the uncertainty in the mean transit time is a factor of 2.5, compared to a few percent for
24 the young water fraction.

25 We can extend these sample calculations over a range of shape factors α and amplitude ratios
26 A_S/A_P (see Fig. 12). As Fig. 12 shows, when the shape factor is uncertain in the range of
27 $0.5 < \alpha < 1$, the corresponding uncertainty in the young water fraction F_{yw} is typically several
28 percent, but the corresponding uncertainty in the MTT is typically a factor of two or more.
29 For a factor-of-ten uncertainty in the shape factor ($0.2 < \alpha < 2$), the uncertainty in the young
30 water fraction is consistently less than a factor of two, whereas the uncertainty in the MTT
31 can exceed a factor of 100.

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

8 **4.3 Young water estimation with non-gamma distributions**

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

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

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

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

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

$$5 \quad h(\tau) = \frac{1}{\sum_{i=1}^n k_i} \sum_{i=1}^n \frac{k_i \tau^{\alpha_i-1}}{(\bar{\tau}_i/\alpha_i)^{\alpha_i} \Gamma(\alpha_i)} e^{-\alpha_i \tau/\bar{\tau}_i} \quad (16)$$

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

12 **4.4 Incorporating phase information in estimating young water fractions and** 13 **mean transit times**

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

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

1 As described in Sect. 2.2 above, one can use the amplitude ratio and phase shift to jointly
2 estimate the shape factor α by iteratively solving Eq. 11, and then estimate the scale factor β
3 via Eq. 10. The mean transit time can then be estimated as $\alpha\beta$ (Eq. 3a). From the fitted value
4 of α , one can also use Eq. 14 to estimate the threshold age τ_{yw} for young water fractions that
5 should aggregate nearly linearly, and then finally estimate the young water fraction as
6 $F_{yw}=\Gamma(\tau_{yw}, \alpha, \beta)$ (Eq. 13). The lower incomplete gamma function $\Gamma(\tau_{yw}, \alpha, \beta)$ is readily
7 available in many software packages (for example, the Igamma function in R or the
8 GAMMA.DIST function in Microsoft Excel).

9 This approach assumes that the catchment's transit times are gamma-distributed. To test
10 whether it can nonetheless improve estimates of the mean transit time or the young water
11 fraction, even in catchments whose transit times are not gamma-distributed, I applied this
12 method to the 8-tributary synthetic catchments shown in Fig. 13. As pointed out in Sect. 4.3
13 above, the TTD's of these catchments (and even their two-subcatchment tributaries) will be
14 sums of gammas and thus not gamma-distributed themselves. Figure 14 shows the new
15 estimates based on amplitude ratios and phase shifts (in dark blue), superimposed on the
16 previous estimates from Fig. 13, based on amplitude ratios alone, as reference (in light blue).
17 Mean transit time estimates based on both phase and amplitude information are somewhat
18 more accurate than those based on amplitude ratios alone (Fig. 14d-14f), but they still exhibit
19 very large aggregation bias. Incorporating phase information in estimates of F_{yw} (Fig. 14a-
20 14c) eliminates much of the (already small) bias in F_{yw} estimates obtained from amplitude
21 ratios alone. (The logarithmic axes of Figs. 14a-14c make this bias more visible than it is on
22 the linear axes of Figs. 13a-13c). The top and bottom rows of Fig. 14 are plotted on
23 consistent axes (both are logarithmic scales spanning a factor of 50), so they provide a direct
24 visual comparison of the reliability of estimates of F_{yw} and MTT.

25 **5 Implications**

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

1 **5.1 Biases in mean transit times**

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

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

21 It is becoming widely recognized that stable isotopes are effectively blind to the long tails of
22 travel time distributions (Stewart et al., 2010; Stewart et al., 2012; Seeger and Weiler, 2014).
23 The results presented here reinforce this point, showing how in heterogeneous catchments,
24 any stable isotope cycles from long-MTT subcatchments (or flowpaths) will be overwhelmed
25 by much larger cycles from short-MTT subcatchments (or flowpaths). Furthermore, the
26 nonlinearities in the governing equations (Figs. 3 and 5) imply that the shorter-MTT
27 components will dominate MTT estimates, which will thus be biased low. This
28 underestimation bias may help to explain the discrepancy between MTT estimates derived
29 from stable isotopes and those derived from other tracers, such as tritium (Stewart et al., 2010;
30 Stewart et al., 2012). However, one should note that, like any radioactive tracer, tritium ages
31 should themselves be vulnerable to underestimation bias in heterogeneous systems (Bethke
32 and Johnson, 2008). Until tritium ages are subjected to benchmark tests like those I have

1 presented here for stable isotopes, one cannot estimate how much they, too, are distorted by
2 aggregation bias.

3 **5.2 Other methods for estimating MTT's from tracers**

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

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

1 **5.3 Implications for mechanistic interpretations of MTT's**

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

9 Many studies have sought to link MTT's to catchment characteristics, often with inconsistent
10 results. For example, McGuire et al. (2005) reported that MTT was positively correlated with
11 the ratio of flow path distance to average hillslope gradient at experimental catchments in
12 Oregon, but Tetzlaff et al. (2009) reported that MTT was *negatively* correlated with the same
13 ratio, and positively correlated with the extent of hydrologically responsive soils, at several
14 Scottish catchments. Hrachowitz et al. (2009) reported that MTT was related to precipitation
15 intensity, soil characteristics, drainage density, and topographic wetness index across a larger
16 network of Scottish catchments, whereas Asano and Uchida (2012) reported that subsurface
17 flow path depth was the main control on baseflow MTT at their Japanese field sites.
18 Heidbüchel et al. (2013) reported that MTT was correlated with soil depth, hydraulic
19 conductivity, or planform curvature, with different characteristics becoming more important
20 under different rainfall regimes. And most recently, Seeger and Weiler (2014) reported that
21 most of the observed correlations between MTT and terrain characteristics across 24 Swiss
22 catchments became non-significant when the variation in mean annual discharge was taken
23 into account. My analysis casts much of this literature in a different light. Given that a large
24 component of MTT estimates in the literature may be aggregation noise (Figs. 7, 13 and 14),
25 one should not be surprised if MTT estimates exhibit weak and inconsistent correlations with
26 catchment characteristics, even if those characteristics are important controls on real-world
27 MTT's.

28 **5.4 The young water fraction F_{yw} as an alternative travel time metric**

29 More generally, though, my analysis implies that the young water fraction F_{yw} is a more
30 useful metric of catchment travel time than MTT is, for the simple reason that F_{yw} can be
31 reliably determined in heterogeneous catchments but MTT cannot. Of course, if we know the

1 young water fraction in runoff, we obviously also know the fraction of "old" water as well
2 (meaning water older than the "young water" threshold). But we do not know – and my
3 analysis implies that we generally cannot know – how old this "old" water is, at least from
4 analyses of seasonal tracer cycles.

5 Of course, because F_{yw} is nearly equal to the amplitude ratio, and MTT can also be expressed
6 as a function of the amplitude ratio for travel time distributions (TTD's) of any known shape,
7 one might conclude that MTT and F_{yw} are just transforms of one another. But that conclusion
8 presumes that the shape of the TTD is known, and my analysis shows that in heterogeneous
9 catchments, the shape of the TTD will be unpredictable. Because the MTT is sensitive to the
10 shape of the TTD – and in particular to the long-time tail, which is particularly poorly
11 constrained – it cannot be reliably estimated. By contrast, my analysis shows that despite the
12 uncertainty in the shape of the TTD in heterogeneous catchments, the F_{yw} can be reliably
13 estimated from the amplitude ratio of seasonal tracer cycles in precipitation and runoff. The
14 fact that this is possible is neither a miracle nor a fortuitous accident; instead F_{yw} has been
15 defined with exactly this result in mind. The F_{yw} entails an unavoidable ambiguity in what,
16 exactly, the threshold age of young water is (because this depends on the shape of the TTD,
17 which is usually unknown), but this uncertainty is small (Fig. 10) compared to the very large
18 uncertainty in the MTT.

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

1 **5.5 Potential applications for young water fractions**

2 Since young water fractions are estimated from amplitude ratios and phase shifts of seasonal
3 tracer cycles, one could ask whether they add any new information, or whether we could
4 characterize catchments equally well by their amplitude ratios and phase shifts instead. One
5 obvious answer is that amplitude ratios and phase shifts, by themselves, are purely
6 phenomenological descriptions of input-output behavior. Young water fractions, by contrast,
7 offer a mechanistic explanation for how that behavior arises, showing how it is linked to the
8 fraction of precipitation that reaches the stream in much less than one year. Not only is this
9 potentially useful for understanding the transport of contaminants and nutrients, it also
10 directly quantifies the importance of relatively fast flowpaths in the catchment. These fast
11 flowpaths are likely to be shallow [since permeability typically decreases rapidly with depth;
12 Brooks, 2004 #2208; Bishop, 2011 #2207], and to originate relatively close to flowing
13 channels. One would expect F_{yw} to increase under wetter conditions, as the water table rises
14 into more permeable near-surface zones, and as the flowing channel network extends to more
15 finely dissect the landscape (Godsey and Kirchner, 2014), thus shortening the path length of
16 subsurface flows, as well as multiplying the wetted catchment area in riparian zones. In a
17 companion paper (Kirchner, 2015), I show that young water fractions can be estimated
18 separately for individual flow regimes, allowing one to infer how shifts in hydraulic forcing
19 alter the fraction of streamflow that is generated via fast flowpaths. I further demonstrate how
20 one can estimate the chemistry of "young water" and "old water" end members, based on
21 comparisons of F_{yw} and solute concentrations across different flow regimes.

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

28 One final note: it has not escaped my notice that because the "young water" threshold is
29 defined as a fraction of the period of the fitted sinusoid (here, an annual cycle), and because
30 spectral analysis is equivalent to fitting sinusoids across a range of time scales, the input and
31 output spectra of conservative tracers can be re-expressed as a series of young water fractions
32 for a series of young water thresholds. In principle, then, this cascade of young water

1 fractions (and their associated threshold ages) should directly express the catchment's
2 cumulative distribution of travel times, thus solving the longstanding problem of measuring
3 the shape of the transit time distribution. A proof-of-concept study of this direct approach to
4 deconvolution is currently underway.

5

6 **6 Summary and conclusions**

7 I used benchmark tests with data from simple synthetic catchments (Fig. 4) to test how
8 catchment heterogeneity affects estimates of mean transit times (MTT's) derived from
9 seasonal tracer cycles in precipitation and streamflow (e.g., Fig. 1). The relationship between
10 tracer cycle amplitude and MTT is strongly nonlinear (Fig. 3), with the result that tracer
11 cycles from heterogeneous catchments will underestimate their average MTT's (Fig. 5). In
12 heterogeneous catchments, furthermore, the shape of the transit time distribution (TTD) in the
13 mixed runoff will differ from that of the tributaries; e.g., mixtures of exponential distributions
14 are not exponentials (Fig. 6), and mixtures of gamma distributions are not gamma-distributed.
15 These two effects combine to make seasonal tracer cycles highly unreliable as estimators of
16 MTT's, with large scatter and strong underestimation bias in heterogeneous catchments (Figs.
17 7 and 13). These results imply that many literature values of MTT are likely to be
18 underestimated by large factors, and thus that typical catchment travel times are much longer
19 than previously thought.

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

1 time; it also proposes an alternative metric, F_{yw} , which should be reliable in heterogeneous
2 catchments.

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

12

13 **Acknowledgements**

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

1 **References**

2

3 Asano, Y., and Uchida, T.: Flow path depth is the main controller of mean base flow transit
4 times in a mountainous catchment, *Water Resour. Res.*, 48, W03512, doi:
5 10.1029/2011wr010906, 2012.

6 Aubert, A. H., Kirchner, J. W., Gascuel-Oudou, C., Facheux, M., Gruau, G., and Merot, P.:
7 Fractal water quality fluctuations spanning the periodic table in an intensively farmed
8 watershed, *Environmental Science & Technology*, 48, 930-937, doi: 10.1021/es403723r,
9 2014.

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

12 Bethke, C. M., and Johnson, T. M.: Groundwater age and groundwater age dating, *Annual*
13 *Review of Earth and Planetary Sciences*, 36, 121-152, doi:
14 10.1146/annurev.earth.36.031207.124210, 2008.

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

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

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

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

26 Broxton, P. D., Troch, P. A., and Lyon, S. W.: On the role of aspect to quantify water transit
27 times in small mountainous catchments, *Water Resour. Res.*, 45, W08427, doi:
28 10.1029/2008wr007438, 2009.

1 Burns, D. A., Plummer, L. N., McDonnell, J. J., Busenberg, E., Casile, G. C., Kendall, C.,
2 Hooper, R. P., Freer, J. E., Peters, N. E., Beven, K. J., and Schlosser, P.: The geochemical
3 evolution of riparian ground water in a forested piedmont catchment, *Ground Water*, 41, 913-
4 925, 2003.

5 DeWalle, D. R., Edwards, P. J., Swistock, B. R., Aravena, R., and Drimmie, R. J.: Seasonal
6 isotope hydrology of three Appalachian forest catchments, *Hydrological Processes*, 11, 1895-
7 1906, 1997.

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

11 Feng, X. H., Faiia, A. M., and Posmentier, E. S.: Seasonality of isotopes in precipitation: A
12 global perspective, *Journal of Geophysical Research-Atmospheres*, 114, D08116, doi:
13 10.1029/2008jd011279, 2009.

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

17 Godsey, S. E., Aas, W., Clair, T. A., de Wit, H. A., Fernandez, I. J., Kahl, J. S., Malcolm, I.
18 A., Neal, C., Neal, M., Nelson, S. J., Norton, S. A., Palucis, M. C., Skjelkvåle, B. L., Soulsby,
19 C., Tetzlaff, D., and Kirchner, J. W.: Generality of fractal 1/f scaling in catchment tracer time
20 series, and its implications for catchment travel time distributions, *Hydrological Processes*,
21 24, 1660-1671, doi: 10.1002/hyp.7677, 2010.

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

25 Heidebüchel, I., Troch, P. A., Lyon, S. W., and Weiler, M.: The master transit time distribution
26 of variable flow systems, *Water Resour. Res.*, 48, W06520, doi: 10.1029/2011WR011293,
27 2012.

28 Heidebüchel, I., Troch, P. A., and Lyon, S. W.: Separating physical and meteorological
29 controls of variable transit times in zero-order catchments, *Water Resour. Res.*, 49, 7644-
30 7657, doi: 10.1002/2012wr013149, 2013.

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

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

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

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

13 Hrachowitz, M., Soulsby, C., Tetzlaff, D., and Speed, M.: Catchment transit times and
14 landscape controls-does scale matter?, *Hydrological Processes*, 24, 117-125, doi:
15 10.1002/hyp.7510, 2010b.

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

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

21 Kirchner, J. W., Dillon, P. J., and LaZerte, B. D.: Predictability of geochemical buffering and
22 runoff acidification in spatially heterogeneous catchments, *Water Resour. Res.*, 29, 3891-
23 3901, 1993.

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

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

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

1 Kirchner, J. W., Tetzlaff, D., and Soulsby, C.: Comparing chloride and water isotopes as
2 hydrological tracers in two Scottish catchments, *Hydrological Processes*, 24, 1631-1645, doi:
3 10.1002/hyp.7676, 2010.

4 Kirchner, J. W., and Neal, C.: Universal fractal scaling in stream chemistry and its
5 implications for solute transport and water quality trend detection, *Proceedings of the*
6 *National Academy of Sciences of the United States of America*, 110, 12213-12218, doi:
7 10.1073/pnas.1304328110, 2013.

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

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

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

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

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

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

26 McDonnell, J. J., and Beven, K.: Debates-The future of hydrological sciences: A (common)
27 path forward? A call to action aimed at understanding velocities, celerities and residence time
28 distributions of the headwater hydrograph, *Water Resour. Res.*, 50, 5342-5350, doi:
29 10.1002/2013wr015141, 2014.

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

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

6 Nash, J. E.: The form of the instantaneous unit hydrograph, *Comptes Rendus et Rapports*,
7 IASH General Assembly Toronto 1957, Int. Assoc. Sci. Hydrol. (Gentbrugge), Publ. No. 45,
8 3, 114-121, 1957.

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

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

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

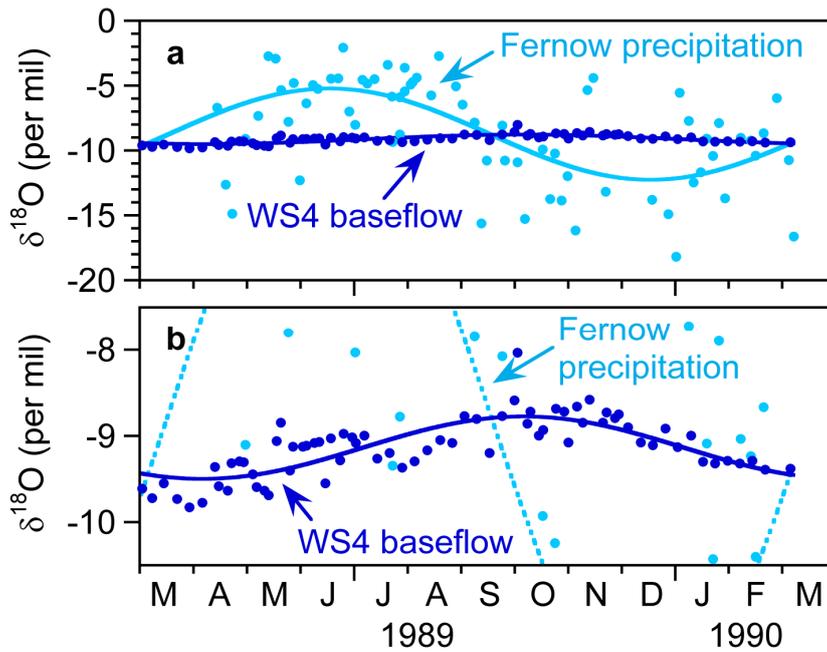
17 Seeger, S., and Weiler, M.: Reevaluation of transit time distributions, mean transit times and
18 their relation to catchment topography, *Hydrol. Earth Syst. Sci.*, 18, 4751-4771, doi:
19 10.5194/hess-18-4751-2014, 2014.

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

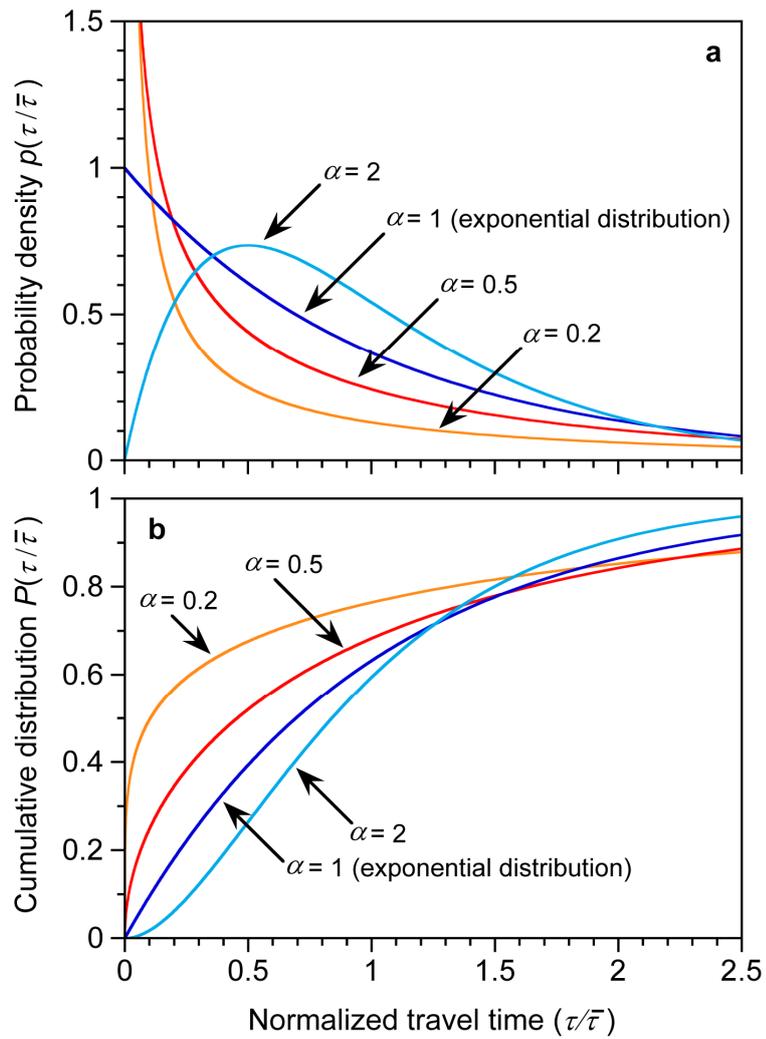
23 Stewart, M. K., Morgenstern, U., and McDonnell, J. J.: Truncation of stream residence time:
24 how the use of stable isotopes has skewed our concept of streamwater age and origin,
25 *Hydrological Processes*, 24, 1646-1659, doi: 10.1002/hyp.7576, 2010.

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

- 1 Tetzlaff, D., Malcolm, I. A., and Soulsby, C.: Influence of forestry, environmental change and
2 climatic variability on the hydrology, hydrochemistry and residence times of upland
3 catchments, *J. Hydrol.*, 346, 93-111, 2007.
- 4 Tetzlaff, D., Seibert, J., and Soulsby, C.: Inter-catchment comparison to assess the influence
5 of topography and soils on catchment transit times in a geomorphic province; the Cairngorm
6 mountains, Scotland, *Hydrological Processes*, 23, 1874-1886, doi: 10.1002/hyp.7318, 2009.
- 7 Van der Velde, Y., De Rooij, G. H., Rozemeijer, J. C., van Geer, F. C., and Broers, H. P.: The
8 nitrate response of a lowland catchment: on the relation between stream concentration and
9 travel time distribution dynamics, *Water Resour. Res.*, 46, W11534, doi:
10 doi:10.1029/2010WR009105, 2010.
- 11 Walck, C.: Handbook on statistical distributions for experimentalists, Particle Physics Group,
12 University of Stockholm, Stockholm, 202 pp., 2007.
- 13

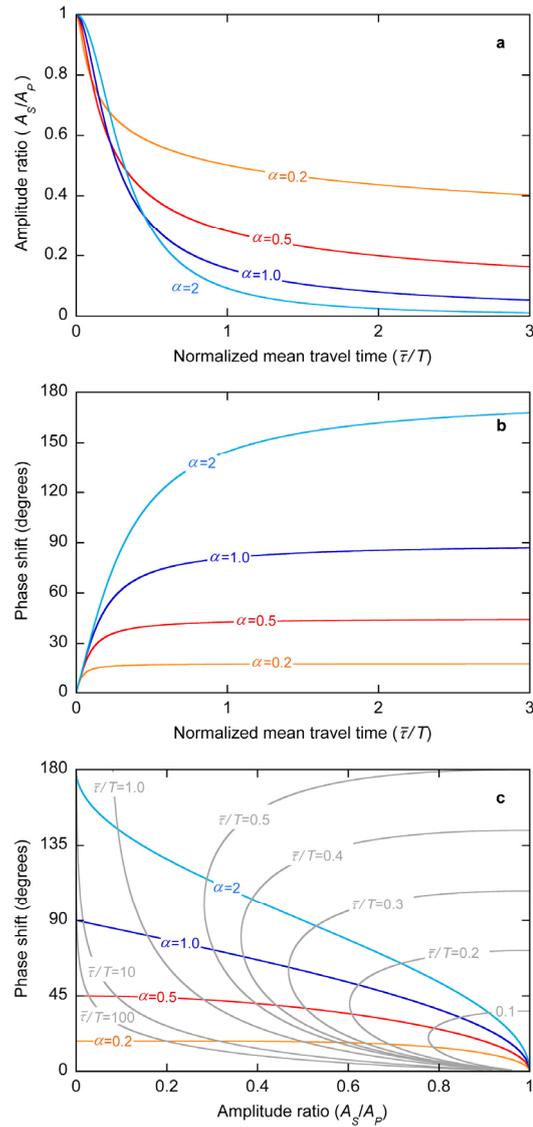


1
 2 Figure 1. Seasonal cycles in $\delta^{18}\text{O}$ in precipitation and baseflow at catchment WS4, Fernow
 3 Experimental Forest, West Virginia, USA (DeWalle et al., 1997). Both panels show the same
 4 data; the axes of panel (b) are expanded to more clearly show the seasonal cycle in baseflow.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



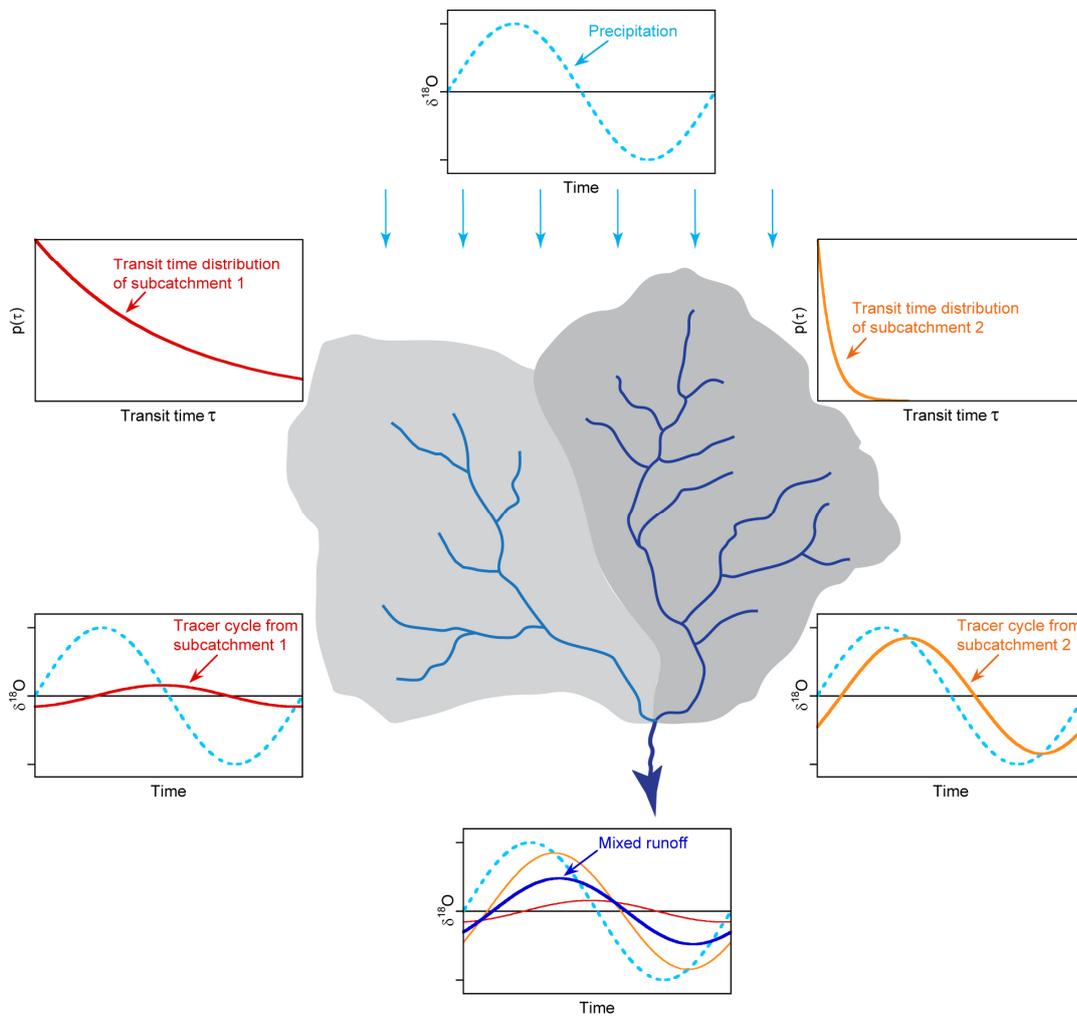
1

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

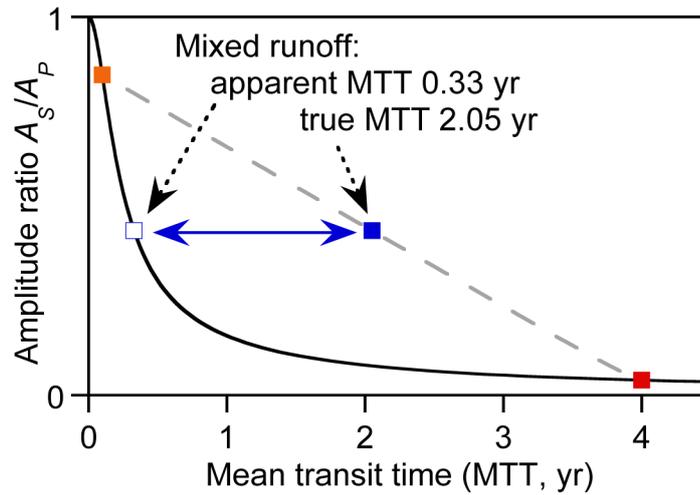


1

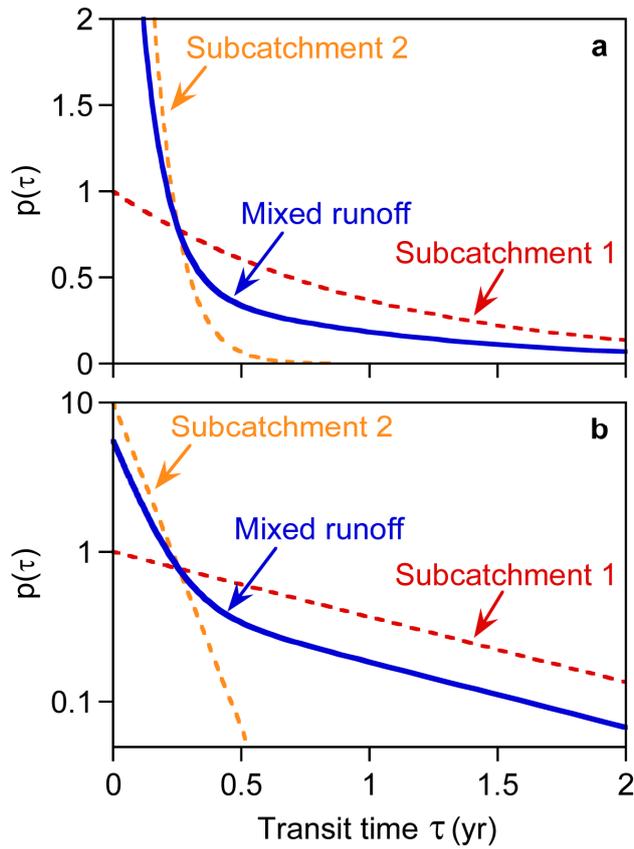
2 Figure 3. Amplitude ratio and phase shift between seasonal cycles in precipitation and
 3 streamflow, for gamma-distributed catchment transit time distributions with a range of shape
 4 factors α (colored lines). Panel (a): ratio of seasonal cycle amplitudes in streamflow and
 5 precipitation (A_S/A_P)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



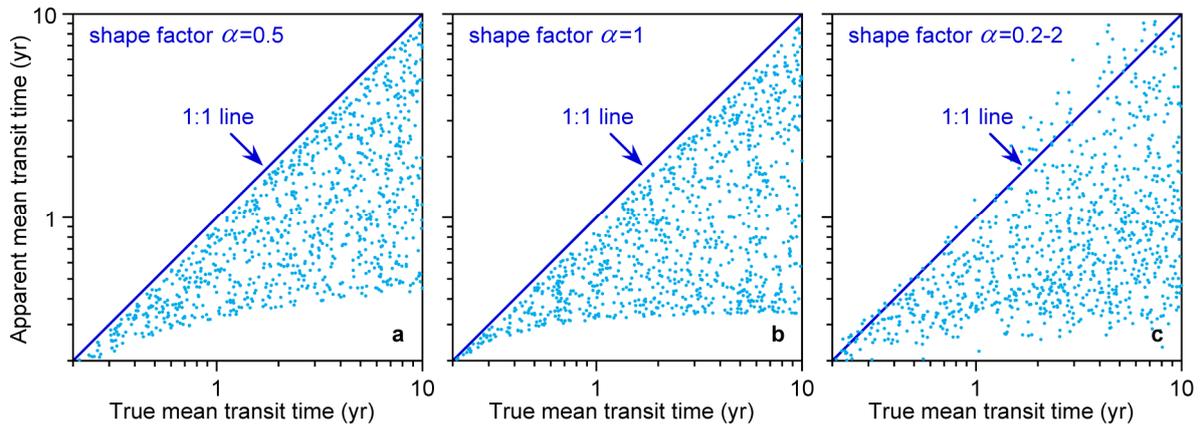
1
 2 Figure 4. Conceptual diagram illustrating mixture of seasonal tracer cycles in runoff from a
 3 heterogeneous catchment, comprising two subcatchments with strongly contrasting mean
 4 transit times (MTT's), and which thus damp the tracer cycle in precipitation (light blue dashed
 5 line) by different amounts. The tracer cycle in the combined runoff from the two
 6 subcatchments (dark blue solid line) will average together the highly damped cycle from
 7 subcatchment 1, with long MTT (solid red line), and the less damped cycle from
 8 subcatchment 2, with short MTT (solid orange line).
 9



1
2 Figure 5. Illustration of the aggregation error that arises when mean transit time is inferred
3 from seasonal tracer cycles in mixed runoff from two landscapes with contrasting transit time
4 distributions (e.g., Fig. 4). The relationship between mean transit time (MTT) and the
5 amplitude ratio (A_S/A_P) of annual cycles in streamflow and precipitation is strongly nonlinear
6 (black curve). Seasonal cycles from subcatchments with MTT of 0.1 yr ($A_S/A_P=0.85$, orange
7 square) and 4 yr ($A_S/A_P=0.04$, red square) will mix along the dashed gray line. A 50:50
8 mixture of the two sources will have a MTT of $(4+0.1)/2=2.05$ years and an amplitude ratio
9 A_S/A_P of 0.43 (blue square). But if this amplitude ratio is interpreted as coming from a single
10 catchment (Eq. 10), it implies a MTT of only 0.33 yr (open square), 6 times shorter than the
11 true MTT of the mixed runoff.

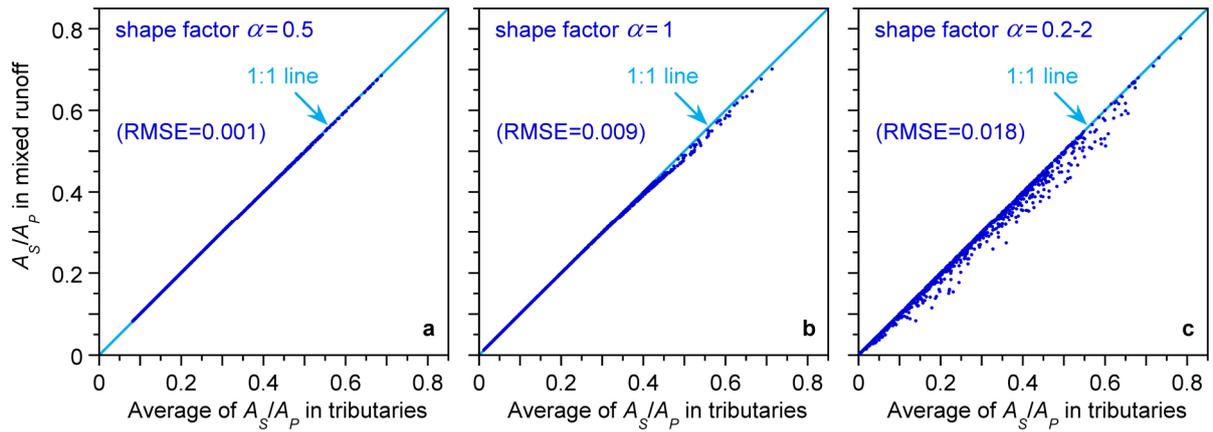


1
 2 Figure 6. Exponential transit-time distributions for subcatchments 1 and 2 in Fig. 4 (with
 3 mean transit times of 1 and 0.1 yr, shown by the orange and red dashed lines, respectively),
 4 and the hyperexponential distribution formed by merging them in equal proportions (solid
 5 blue line). Panels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

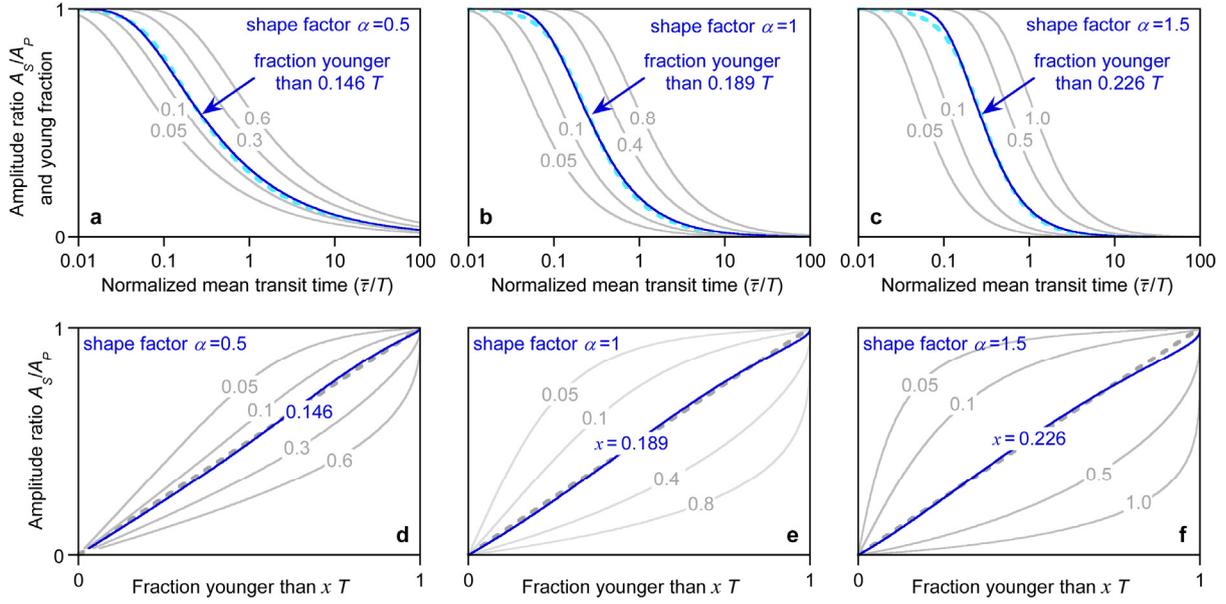


1
 2 Figure 7. Apparent mean transit time (MTT) inferred from seasonal tracer cycles, showing
 3 order-of-magnitude deviations from true MTT for 1000 synthetic catchments. Each synthetic
 4 catchment comprises two subcatchments with individual MTT's randomly chosen from a
 5 uniform distribution of logarithms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

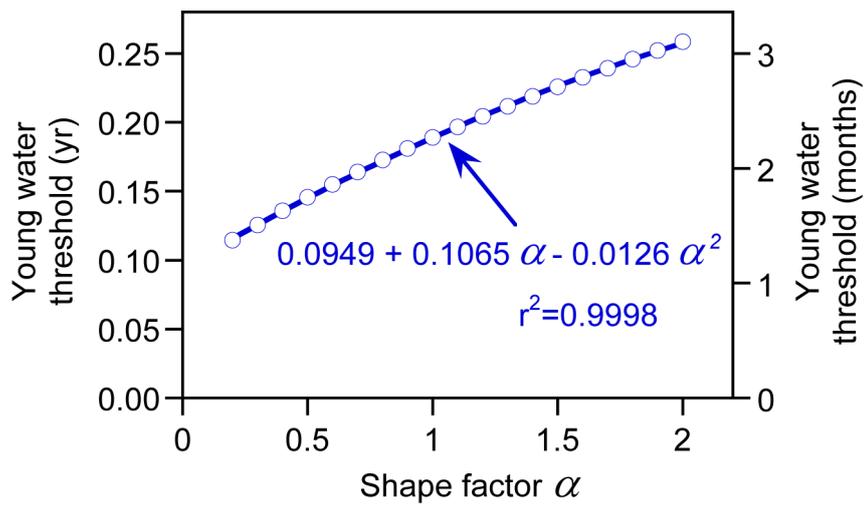
13



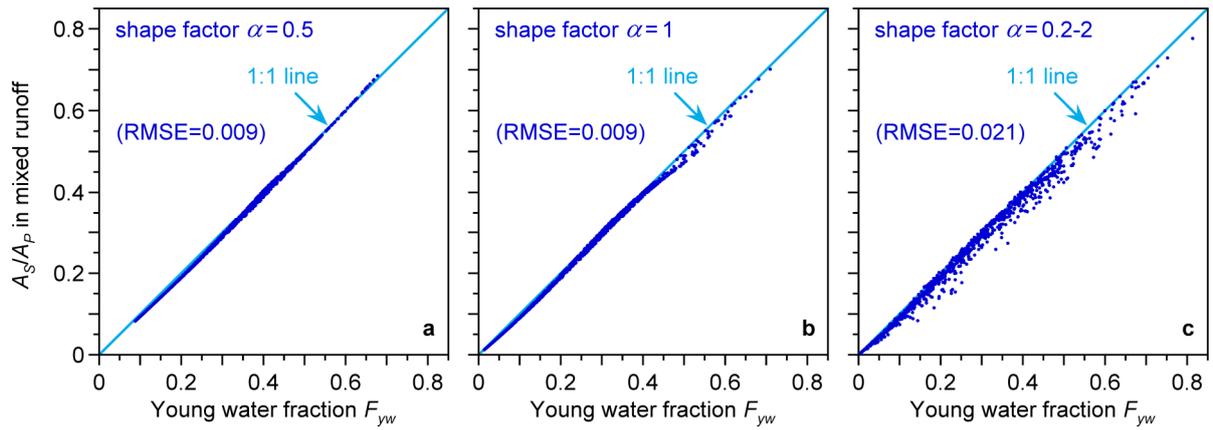
1
 2 Figure 8. Amplitude ratio (A_S/A_P) of tracer cycles in precipitation and mixed runoff from the
 3 same 1000 synthetic catchments shown in Fig. 7 (vertical axes), compared to the average of
 4 the tracer cycle amplitude ratios in the two tributaries (horizontal axes). As in Fig. 7, each
 5 synthetic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



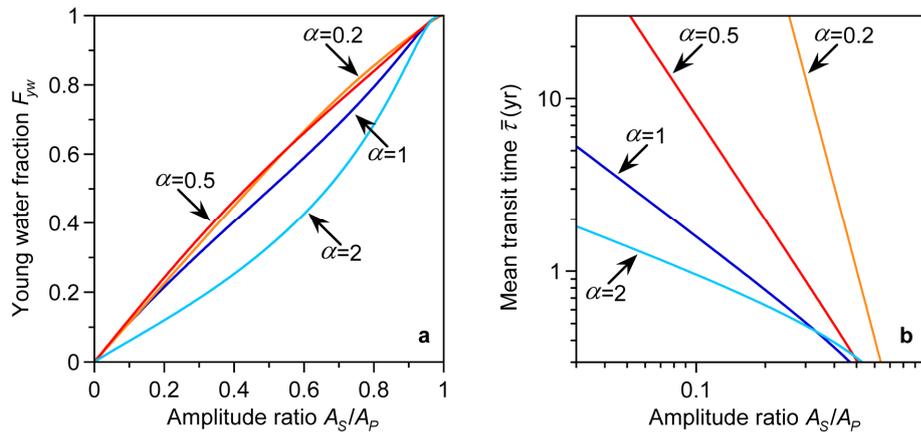
1
2 Figure 9. Panels (a)-(c) show the amplitude ratios A_s/A_p in precipitation and streamflow tracer
3 cycles (light blue dashed line) as function of mean transit time $\bar{\tau}$, compared to the fraction of
4 water younger than several threshold ages (gray lines), and the best-fit age threshold (dark
5 blue line). Panels (d)-(f) show the relationship between amplitude ratio and the fraction of
6 water younger than several age thresholds (gray lines) and the best-fit age threshold (dark
7 blue line), with the 1:1 line (dashed gray) for comparison. Panels show results for three
8 different gamma distributions, with shape factors $\alpha=0.5$, $\alpha=1$, and $\alpha=1.5$. Root-mean-
9 squared errors (RMSE's) for amplitude ratios A_s/A_p as predictors of the best-fit "young water"
10 fractions are 0.012, 0.011, and 0.015 for panels (d), (e), and (f), respectively. In all panels,
11 threshold age and mean transit time are normalized by T , the period of the tracer cycle. For
12 seasonal tracer cycles, $T=1$ yr and thus threshold age and mean transit time are in years.
13



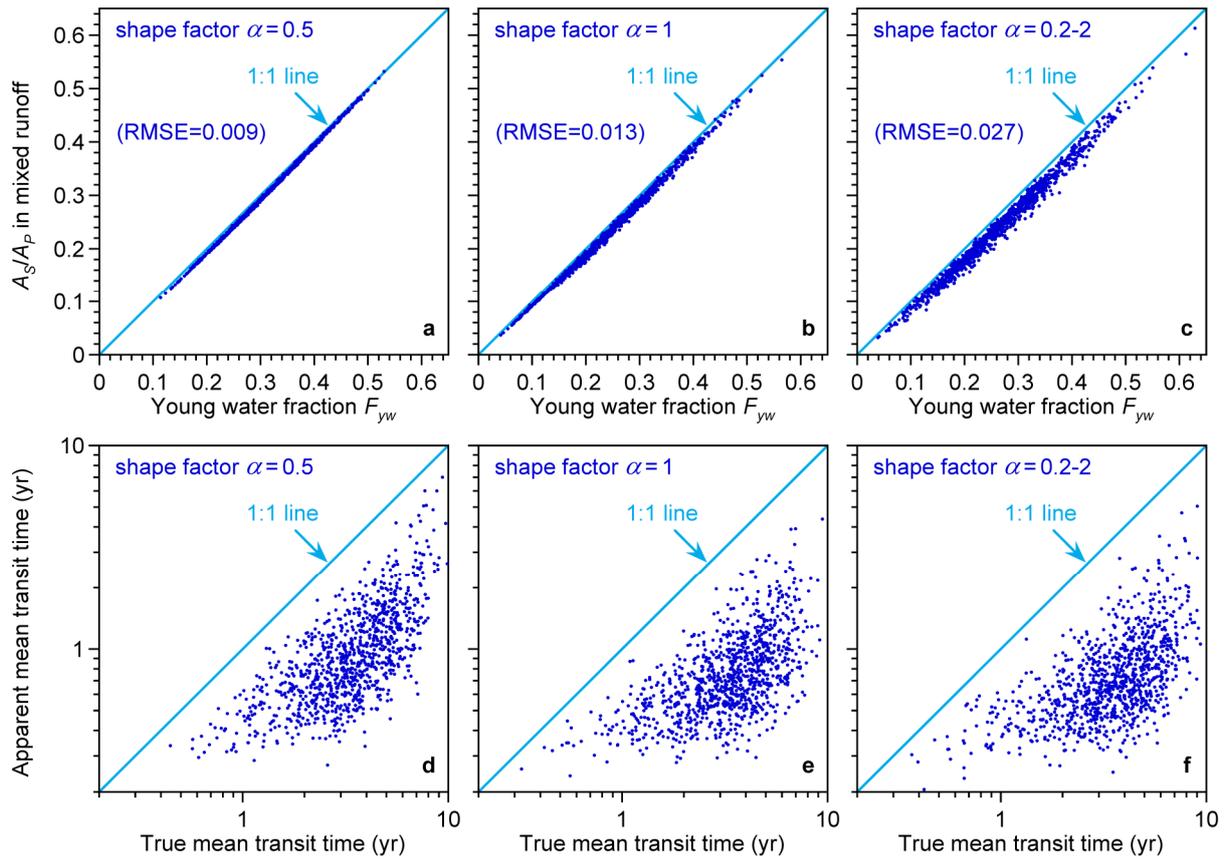
1
2 Figure 10. Best-fit "young water" thresholds for gamma transit time distributions, as a
3 function of shape factors α ranging from 0.2 to 2.0. The young water threshold τ_{yw} is
4 defined such that the fraction of the distribution with ages less than τ_{yw} approximately equals
5 the amplitude ratio (A_S/A_P) of annual cycles in streamflow and precipitation (see Fig. 9).



1
 2 Figure 11. True and apparent "young water" fractions for the same 1000 synthetic catchments
 3 shown in Fig. 7. The tracer cycle amplitude ratio in the combined runoff of the two
 4 subcatchments (vertical axes) corresponds closely to the average young water fraction in the
 5 combined runoff (horizontal axes). As in Fig. 7, each synthetic catchment comprises two
 6 subcatchments with individual MTT's randomly chosen from a uniform distribution of
 7 logarithms spanning the interval between 0.1 and 20 years, and with each pair of MTT's
 8 differing by at least a factor of 2. In panels (a) and (b), all subcatchments have the same
 9 shape factor α . In panel (c), shape factors for each subcatchment are randomly chosen from a
 10 uniform distribution between $\alpha=0.2$ and $\alpha=2$.

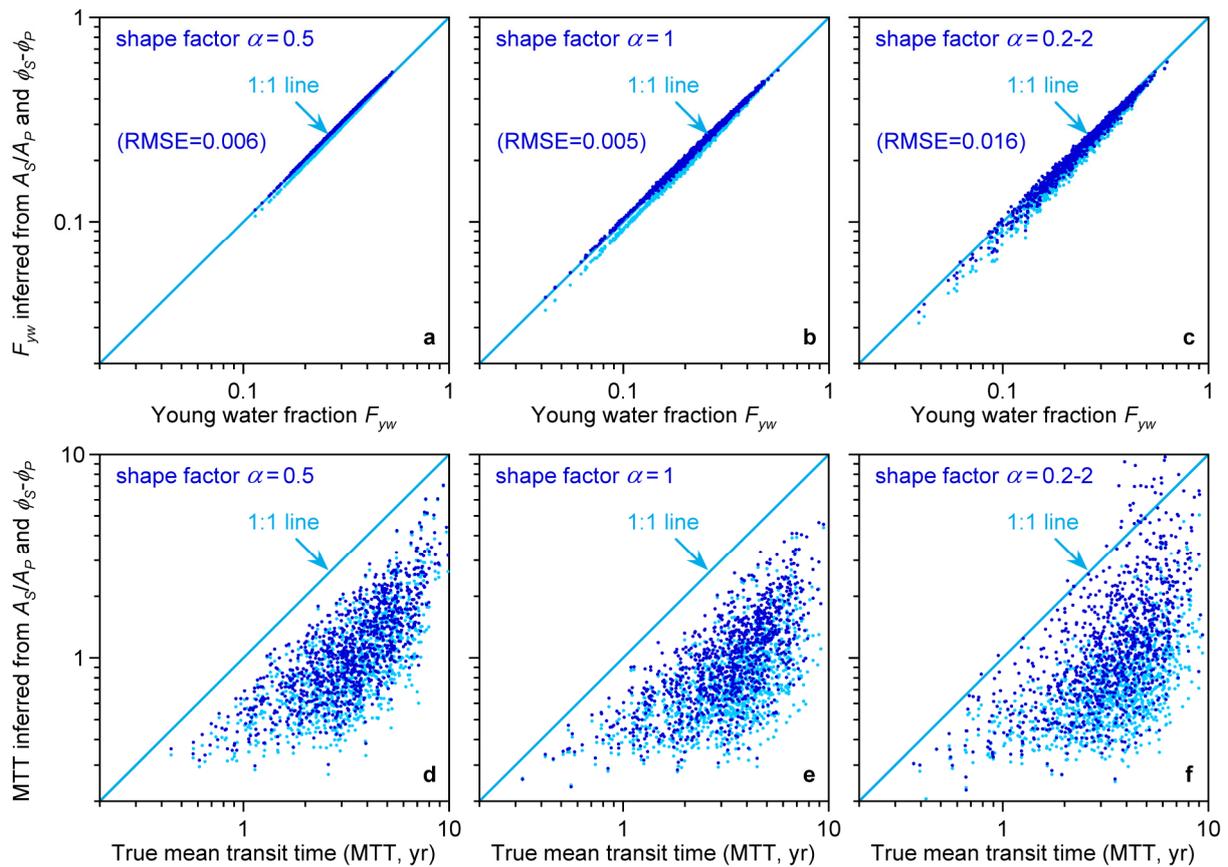


1
2 Figure 12. Sensitivity analysis showing how variations in shape factor α affect young water
3 fractions F_{yw} (panel a) and mean transit times $\bar{\tau}$ (panel b) inferred from the amplitude ratio
4 A_S/A_P of seasonal tracer cycles in precipitation and streamflow. Curves are shown for the four
5 shape factors shown in Figs. 2 and 3. For a plausible range of uncertainty in the shape factor
6 ($0.5 < \alpha < 1$; see Sect. 2.1), estimated young water fractions vary by a few percent (panel a),
7 whereas estimated mean transit times vary by large multiples (note the logarithmic axes in
8 panel b). Panel (a) shows the fractions of water younger than $\tau_{yw}=2.27$ months, which are
9 closely approximated by A_S/A_P if $\alpha=1$ (the dark blue curve). In panel (b), the axis scales are
10 chosen to span transit times ranging from several months to several years, as is commonly
11 observed in transit time studies (McGuire and McDonnell, 2006).



1
2

3 Figure 13. True and apparent "young water" fractions F_{yw} for 1000 synthetic catchments,
 4 each consisting of 8 subcatchments with randomly chosen mean transit times between 0.1 and
 5 20 years (top panels), and true and apparent mean transit times for the same catchments
 6 (bottom panels). The tracer cycle amplitude ratio in the combined runoff predicts the true
 7 young water fraction with a slight underestimation bias (top panels). Mean transit times
 8 inferred from tracer cycle amplitude ratios show severe underestimation bias (bottom panels).
 9 In panels (a-b) and (d-e), all subcatchments have the same shape factor α . In panels (c) and
 10 (f), shape factors for each subcatchment are randomly chosen from a uniform distribution
 11 between $\alpha=0.2$ and $\alpha=2$.



1
2
3
4
5
6
7
8
9
10
11
12

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 ($\phi_S - \phi_P$). 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.