Hydrology and
Earth System
Sciences

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

J. W. Kirchner[1,2]

[1]ETH Zürich, Zurich, Switzerland
[2]Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

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

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

1 Introduction

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

Instead, it is often tacitly assumed (although *hoped* might be a better word) that any problems introduced by spatial heterogeneity will be solved or masked by model parameter calibration. This is an intuitively appealing notion. After all, we are often not particularly interested in understanding or predicting point-scale processes within the system, but rather in predicting the resulting ensemble behavior at the whole-catchment scale, such as streamflow, stream chemistry, evap-

otranspiration losses, ecosystem carbon uptake, and so forth. Furthermore, we rarely have point-scale information from the system under study, and when we do, we have no clear way to translate it to larger scales. Instead, often our most reliable and readily available measurements are at the whole-catchment scale: streamflow, stream chemistry, weather variables, etc. Would it not be nice if these whole-catchment measurements could be used to estimate spatially aggregated model parameters that somehow subsume the spatial heterogeneity of the system, at least well enough to generate reliable predictions of whole-catchment behavior?

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

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

cle (Burns et al., 2003; Godsey et al., 2009; Maher, 2010; Maher and Chamberlain, 2014).

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

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

As reviewed by McGuire and McDonnell (2006), three methods are commonly used to infer catchment travel times from conservative tracer time series: (1) time-domain convolution of the input time series to simulate the output time series, with parameters of the convolution kernel (the travel-time distribution) fitted by iterative search techniques; (2) Fourier transform spectral analysis of the input and output

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

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

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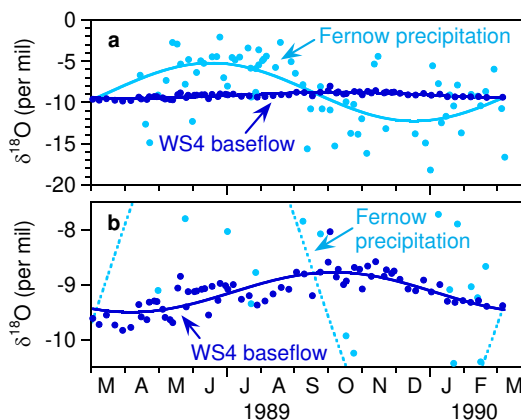

Figure 1. Seasonal cycles in δ^{18}O in precipitation and baseflow at catchment WS4, Fernow Experimental Forest, West Virginia, USA (DeWalle et al., 1997). Both panels show the same data; the axes of **(b)** are expanded to more clearly show the seasonal cycle in baseflow. Sinusoidal cycles are fitted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

times urgently need benchmark testing. The work presented below is intended as one small step toward filling that gap.

2 Mathematical preliminaries: tracer cycles in homogeneous catchments

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

$$c_{\mathrm{S}}(t) = \int_0^\infty h(\tau) c_{\mathrm{P}}(t - \tau)\mathrm{d}\tau, \tag{1}$$

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

The transit-time distribution $h(\tau)$ expresses the fractional contribution of past inputs to present runoff. Equation (1) implicitly assumes that the catchment is a linear time-invariant system and, thus, that the convolution kernel $h(\tau)$ is stationary (i.e., constant through time). This is never strictly true, most obviously because if no precipitation falls on a partic-

ular day, it cannot contribute any tracer to the stream τ days later, and because higher precipitation rates will increase the rate at which water and tracers are flushed through the catchment. Thus, real-world TTDs vary through time, depending on the history of prior precipitation (Kirchner et al., 2001; Tetzlaff et al., 2007; Botter et al., 2010; Hrachowitz et al., 2010a; Van der Velde et al., 2010; Birkel et al., 2012; Heidbüchel et al., 2012; Peters et al., 2014). However, in applications using real-world data, $h(\tau)$ is conventionally interpreted as a time-invariant ensemble average, taken over an ensemble of precipitation histories, which obviously will differ from one another in detail. Mathematically, the ensemble averaging embodied in Eq. (1) is equivalent to the simplifying assumption that water fluxes in precipitation and streamflow are constant over time. (One can relax this assumption somewhat by integrating over the cumulative water flux rather than time, as proposed by Niemi (1977). If the rates of transport and mixing vary proportionally to the flow rate through the catchment, this yields a stationary distribution in flow-equivalent time.) A further simplification inherent in Eq. (1) is that evapotranspiration and its effects on tracer signatures are ignored.

2.1 A class of transit-time distributions

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

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

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

Figure 2 shows gamma distributions spanning a range of shape factors α. For the special case of $\alpha = 1$, the gamma distribution becomes the exponential distribution. Exponential distributions describe the behavior of continuously mixed reservoirs of constant volume, and they have been widely used to model catchment storage and mixing. The gamma distribution expresses the TTD of a Nash cascade (Nash, 1957) of α identical linear reservoirs connected in series, and the analogy to a Nash cascade holds even for noninteger α, through the use of fractional integration. For $\alpha > 1$, the gamma distribution rises to a peak and then falls off, similarly to a typical storm hydrograph, which is why Nash cascades have often been used to model rainfall–runoff relationships. For $\alpha < 1$, however, the gamma distribution has a completely different shape, having maximum weight at lags near zero and a relatively long tail. These characteristics represent problematic contaminant behavior, with rapid release of an intense contaminant spike followed by persistent lower-level contamination far into the future. Tracer time series from
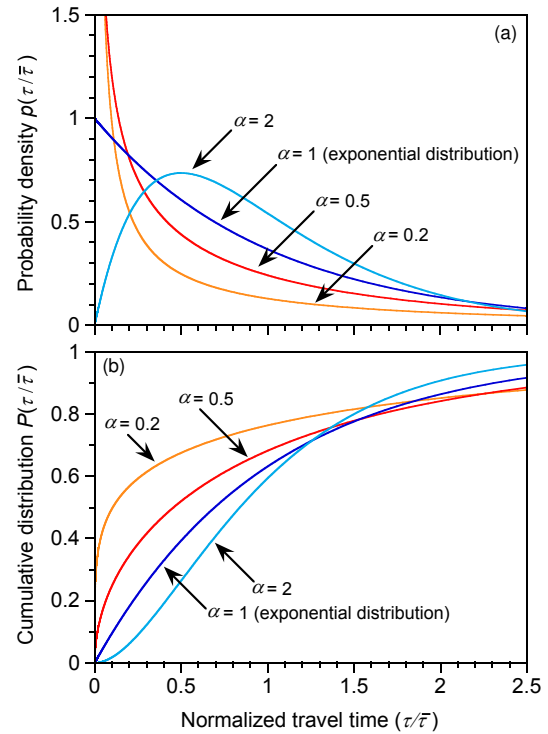

Figure 2. Gamma distributions for the range of shape factors $\alpha = 0.2$–2 considered in this analysis. Horizontal axes are normalized by the mean transit time $\overline{\tau}$ and thus are dimensionless.

many catchments have been shown to exhibit fractal $1/f$ scaling, which is consistent with gamma TTDs with $\alpha \approx 0.5$ (Kirchner et al., 2000, 2001; Godsey et al., 2010; Kirchner and Neal, 2013; Aubert et al., 2014).

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

$$\text{mean}(\tau) = \beta\alpha = \overline{\tau}, \tag{3a}$$

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

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

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

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

Studies that have used tracers to constrain the shape of catchment TTDs have generally found shape factors α rang-

ing from 0.3 to 0.7, corresponding to spectral slopes of the transfer function between roughly 0.6 and 1.4 (Kirchner et al., 2000, 2001; Godsey et al., 2010; Hrachowitz et al., 2010a; Kirchner and Neal, 2013; Aubert et al., 2014). Other studies – including those that have used annual tracer cycles to estimate mean transit times – have assumed that the catchment is a well-mixed reservoir and thus that $\alpha = 1$. Here I will assume that α falls in the range of 0.5–1 for typical catchment transit-time distributions, but I will also show some key results for the somewhat wider range of $\alpha = 0.2$–2, for illustrative purposes. The results reported here will not necessarily apply to TTDs that rise to a peak after a long delay, such as the gamma distribution with $\alpha \gg 2$. However, one would not expect such a distribution to characterize whole-catchment TTDs in the first place because, except in very unusual catchments, a substantial amount of precipitation can fall close to the stream and enter it relatively quickly, thus producing a strong peak at a short lag (Kirchner et al., 2001).

2.2 Estimating mean transit time from tracer cycles

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

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

$$c_P(t) = A_P \sin(2\pi f t - \varphi_P) + k_P,$$
$$c_S(t) = A_S \sin(2\pi f t - \varphi_S) + k_S, \tag{4}$$

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

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

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

$$A_P = \sqrt{a_P^2 + b_P^2}, \ A_S = \sqrt{a_S^2 + b_S^2}, \ \varphi_P = \arctan(b_P/a_P)$$
$$\text{and } \varphi_S = \arctan(b_S/a_S). \tag{6}$$

In Eqs. (4)–(6) above, t is time, f is the frequency of the cycle ($f = 1 \text{ year}^{-1}$ for a seasonal cycle), and the subscripts P

and S refer to precipitation and streamflow. In fitting sinusoidal cycles to real-world data, robust estimation techniques such as iteratively reweighted least squares (IRLS) regression can help in limiting the influence of outliers. Also, because precipitation and streamflow rates vary through time, it may be useful to weight each tracer sample by its associated volume, for example to reduce the influence of small rainfall events (for more on the implications of volume-weighting, see Kirchner, 2016). An R script for performing volume-weighted IRLS is available from the author.

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

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

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

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

and also the phase shift between them,

$$\varphi_S - \varphi_P = \alpha \arctan(2\pi f \beta), \tag{9}$$

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

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

$$\overline{\tau} = \alpha\beta, \ \beta = \frac{1}{2\pi f} \sqrt{(A_S/A_P)^{-2/\alpha} - 1}. \tag{10}$$

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

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

which can be solved using nonlinear search techniques such as Newton's method. Once α has been determined, the mean transit time $\overline{\tau}$ can be calculated straightforwardly using Eq. (10). However, when precipitation is episodic, the phase shift $\varphi_S - \varphi_P$ may be difficult to estimate accurately, which
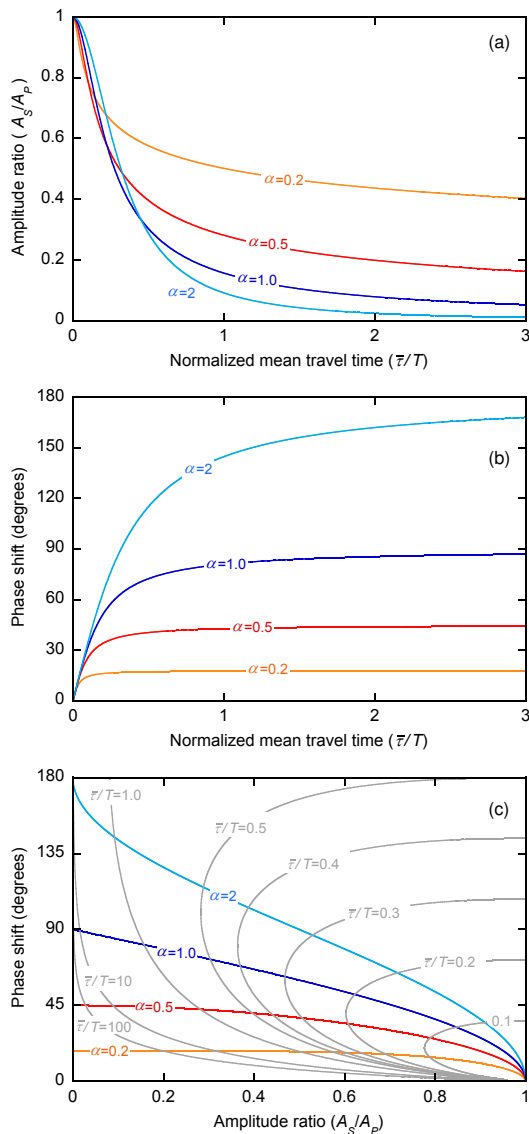
Figure 3. Amplitude ratio and phase shift between seasonal cycles in precipitation and streamflow, for gamma-distributed catchment transit-time distributions with a range of shape factors α (colored lines). **(a)** Ratio of seasonal cycle amplitudes in streamflow and precipitation (A_S/A_P) as a function of mean transit time ($\overline{\tau}$) normalized by the period ($T = 1/f$) of the tracer cycle. **(b)** Phase lag between streamflow and precipitation cycles, as a function of mean transit time normalized by the tracer cycle period ($\overline{\tau}/T$). **(c)** Relationship between phase lag and amplitude ratio, with contours of shape factor (α) ranging from 0.2 to 8 (colored lines), and contours of mean transit time normalized by tracer cycle period $\overline{\tau}/T$ (gray lines). For seasonal tracer cycles, $T = 1/f = 1$ year and normalized transit time equals time in years.

can result in large errors in α and thus $\overline{\tau}$, particularly if the phase shift is near zero. Perhaps for this reason or because (to the best of my knowledge) the relevant math has not previ-

ously been presented, tracer cycle phase information has not typically been used in estimating α and MTT.

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

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

From assumed MTTs $\overline{\tau}$ and shape factors α for each of the subcatchments, one can calculate the amplitude ratios A_S/A_P and phase shifts $\varphi_S - \varphi_P$ of their tracer cycles using Eqs. (8) and (9), and then average these cycles together using the conventional trigonometric identities. (Equivalently, one can estimate the cosine and sine coefficients of the individual subcatchments' tracer cycles from the real and imaginary parts of Eq. (7) and algebraically average them together.) The shares of the two subcatchments in the average will depend on their relative drainage areas and/or water yields. For simplicity, I combine the runoff from the two subcatchments in a $1:1$ ratio; this also guarantees that the combined runoff will be as different as possible from each of the two sources. I then ask the question: from the tracer behavior in the combined runoff (the solid blue line in Fig. 4), would I correctly estimate the mean transit time for the whole catchment? That is, would I infer a MTT that is close to the average of the MTTs of the two subcatchments?

One can immediately see that this situation is highly prone to aggregation bias, following the Kirchner et al. (1993) rule of thumb that the degree of aggregation bias is proportional to the nonlinearity in the governing equations and the variance in the heterogeneous parameters. The amplitude ratios A_S/A_P and phase shifts $\varphi_S - \varphi_P$ of seasonal tracer cycles are strongly nonlinear functions of the MTT (see Eqs. 8 and 10),
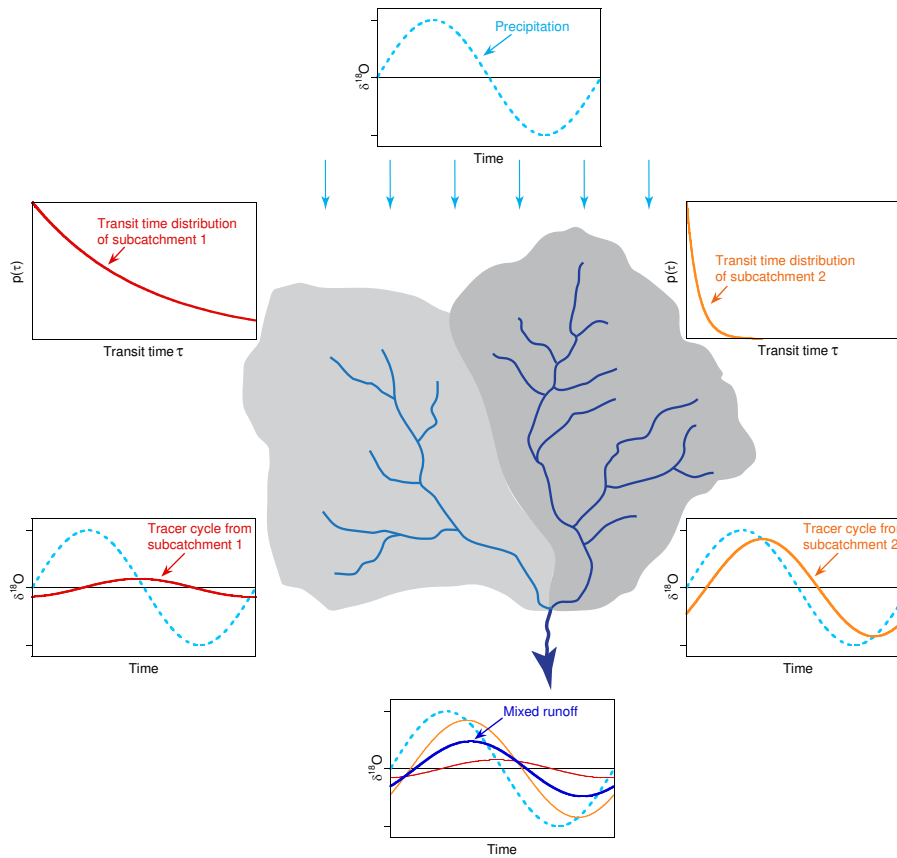
Figure 4. Conceptual diagram illustrating the mixture of seasonal tracer cycles in runoff from a heterogeneous catchment, comprising two subcatchments with strongly contrasting MTTs, and which thus damp the tracer cycle in precipitation (light blue dashed line) by different amounts. The tracer cycle in the combined runoff from the two subcatchments (dark blue solid line) will average together the highly damped cycle from subcatchment 1, with long MTT (solid red line), and the less damped cycle from subcatchment 2, with short MTT (solid orange line).

as illustrated in Fig. 3a and b. And, importantly, the likely range of variation in subcatchment MTTs (from, say, fractions of a year to perhaps several years) straddles the nonlinearity in the governing equations. Thus, we should expect to see significant aggregation bias in estimates of MTT.

Figure 5 illustrates the crux of the problem. The plotted curve shows the relationship between A_S/A_P and MTT for exponential transit-time distributions ($\alpha = 1$); other realistic transit-time distributions will give somewhat different relationships, but they will all be curved. Seasonal cycles from the two subcatchments (the red and orange squares) will mix along the dashed gray line (which is nearly straight but not exactly so, owing to phase differences between the two cycles). A 50 : 50 mixture of tracer cycles from the two subcatchments will plot as the solid blue square, with an amplitude ratio A_S/A_P of 0.43 and a MTT of just over 2 years in this particular example. But the crux of the problem is that if we use this amplitude ratio to infer the corresponding MTT, we will do so where the amplitude ratio intersects with the black curve (Eq. 10), yielding an inferred MTT of

only 0.33 years (the open square), which underestimates the true MTT of the mixed runoff by more than a factor of six. Bethke and Johnson (2008) pointed out that nonlinear averaging can lead to bias in groundwater dating by radioactive tracers; Fig. 5 illustrates how a similar bias can also arise in age determinations based on fluctuation damping in passive tracers.

Combining flows from two subcatchments with different mean transit times will result in a combined TTD that differs in shape, not just in scale, from the TTDs of either of the subcatchments. For example, combining two exponential distributions with different mean transit times does not result in another exponential distribution but rather a hyperexponential distribution, as shown in Fig. 6. The characteristic function of the hyperexponential distribution (Walck, 2007) yields the following expression for the amplitude ratio of tracer cycles in precipitation and streamflow,
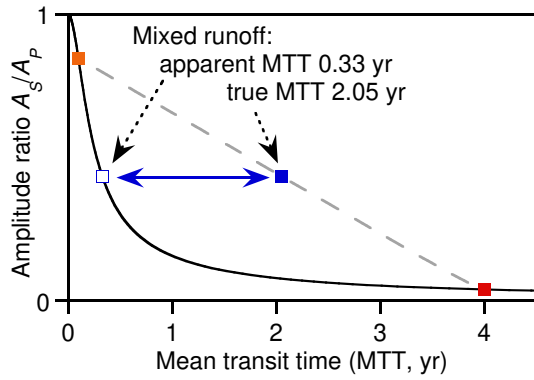
Figure 5. Illustration of the aggregation error that arises when mean transit time is inferred from seasonal tracer cycles in mixed runoff from two landscapes with contrasting transit-time distributions (e.g., Fig. 4). The relationship between MTT and the amplitude ratio (A_S/A_P) of annual cycles in streamflow and precipitation is strongly nonlinear (black curve). Seasonal cycles from sub-catchments with MTT of 0.1 years ($A_S/A_P = 0.85$, orange square) and 4 years ($A_S/A_P = 0.04$, red square) will mix along the dashed gray line. A 50 : 50 mixture of the two sources will have a MTT of $(4 + 0.1)/2 = 2.05$ years and an amplitude ratio A_S/A_P of 0.43 (blue square). But if this amplitude ratio is interpreted as coming from a single catchment (Eq. 10), it implies a MTT of only 0.33 years (open square), 6 times shorter than the true MTT of the mixed runoff.

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

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

To test the generality of this result, I repeated the thought experiment outlined above for 1000 hypothetical pairs of subcatchments, each with individual MTTs randomly chosen from a uniform distribution of logarithms spanning the interval between 0.1 and 20 years (Fig. 7). Pairs with MTTs that differed by less than a factor of 2 were excluded, so that the entire sample consisted of truly heterogeneous catchments. I
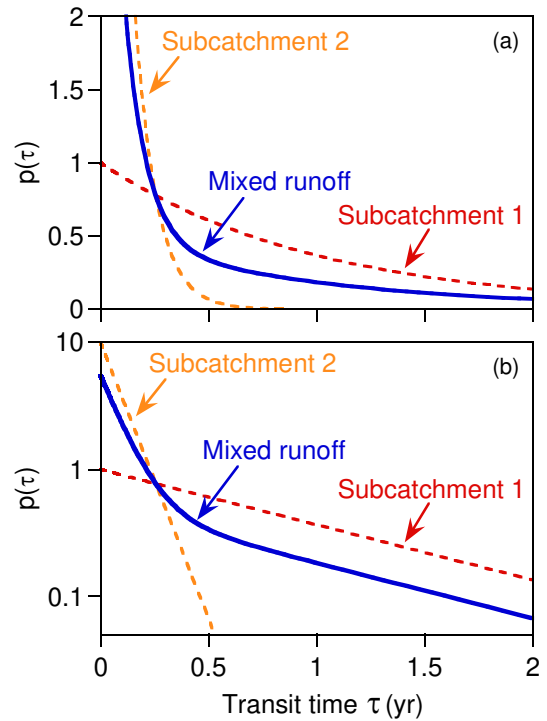

Figure 6. Exponential transit-time distributions for subcatchments 1 and 2 in Fig. 4 (with mean transit times of 1 and 0.1 years, shown by the orange and red dashed lines, respectively), and the hyperexponential distribution formed by merging them in equal proportions (solid blue line). **(a)** and **(b)**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

then applied Eq. (10) to calculate the apparent MTT from the inferred runoff. As Fig. 7 shows, apparent MTTs calculated from the combined runoff of the two subcatchments can underestimate true whole-catchment MTTs by an order of magnitude or more, and this strong underestimation bias persists across a wide range of shape factors α. MTTs are reliably estimated (with values close to the 1 : 1 line in Fig. 7) only when both subcatchments have MTTs of much less than 1 year.

In most real-world cases, unlike these hypothetical thought experiments, one will only have measurements or samples from the whole catchment's runoff. The properties of the individual subcatchments and thus the degree of heterogeneity in the system will generally be unknown. And even if data were available for the subcatchments, those subcatchments would be composed of sub-subcatchments, which would themselves be heterogeneous to some unknown degree, and so on. Thus, it will generally be difficult or impossible to characterize the system's heterogeneity, but that is no justification for pretending that this heterogeneity does not exist. Nonetheless, in such situations it will be tempting to treat the whole system as if it were homogeneous, perhaps using terms like "apparent age" or "model age" to preserve a sense of rigor. But whatever the semantics, as Fig. 7 shows, assum-
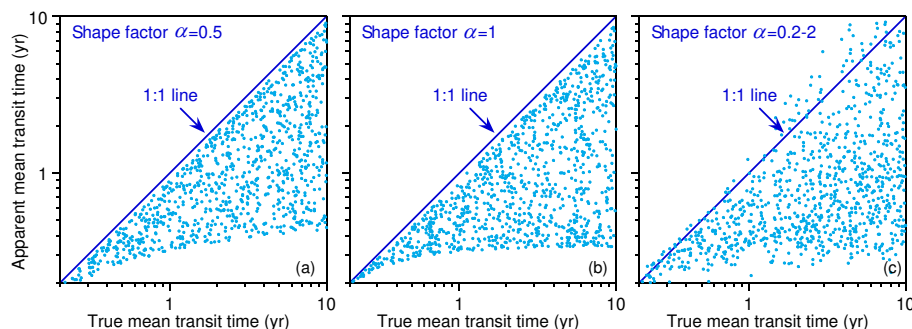
Figure 7. Apparent MTT inferred from seasonal tracer cycles, showing order-of-magnitude deviations from true MTT for 1000 synthetic catchments. Each synthetic catchment comprises two subcatchments with individual MTTs randomly chosen from a uniform distribution of logarithms spanning the interval between 0.1 and 20 years, with each pair differing by at least a factor of 2. In **(a)** and **(b)**, both subcatchments have shape factors α of 0.5 and 1, respectively; in **(c)**, the subcatchments' shape factors are independently chosen from the range of 0.2–2. Apparent MTTs were inferred from the amplitude ratio A_S/A_P of the combined runoff using Eq. (10), with an assumed value of $\alpha = 0.5$ for **(a)**, $\alpha = 1$ for **(b)**, and also $\alpha = 1$ for **(c)**, both because $\alpha = 1$ is close to the average of the randomized α values and because $\alpha = 1$ is typically assumed whenever Eq. (10) is applied to real catchment data.
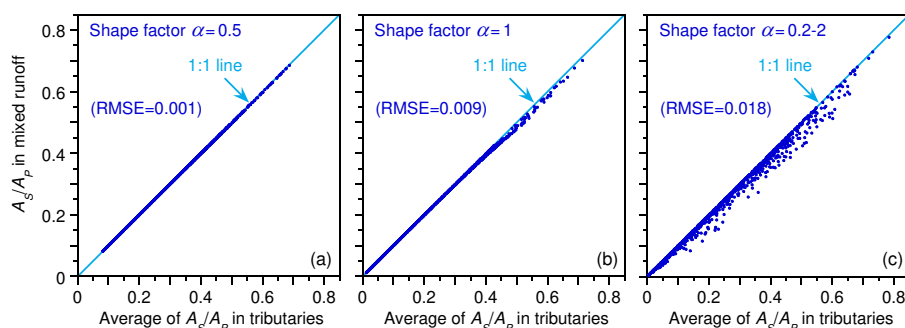

Figure 8. Amplitude ratio (A_S/A_P) of tracer cycles in precipitation and mixed runoff from the same 1000 synthetic catchments shown in Fig. 7 (vertical axes), compared to the average of the tracer cycle amplitude ratios in the two tributaries (horizontal axes). As in Fig. 7, each synthetic catchment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

ing homogeneity in heterogeneous catchments will result in strongly biased estimates of whole-catchment mean transit times.

4 Quantifying the young water component of streamflow

The analysis in Sect. 3 demonstrates what can be termed an "aggregation error": in heterogeneous systems, mean transit times estimated from seasonal tracer cycles yield inconsistent results at different levels of aggregation. The aggregation bias demonstrated in Figs. 5 and 7 implies that seasonal cycles of conservative tracers are unreliable estimators of catchment mean transit times. This observation raises the obvious question: is there anything *else* that can be estimated from

seasonal tracer cycles and that is relatively free from the aggregation bias that afflicts estimates of mean transit times?

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

Because tracer cycle amplitudes will average almost linearly when two streams merge and thus are virtually free
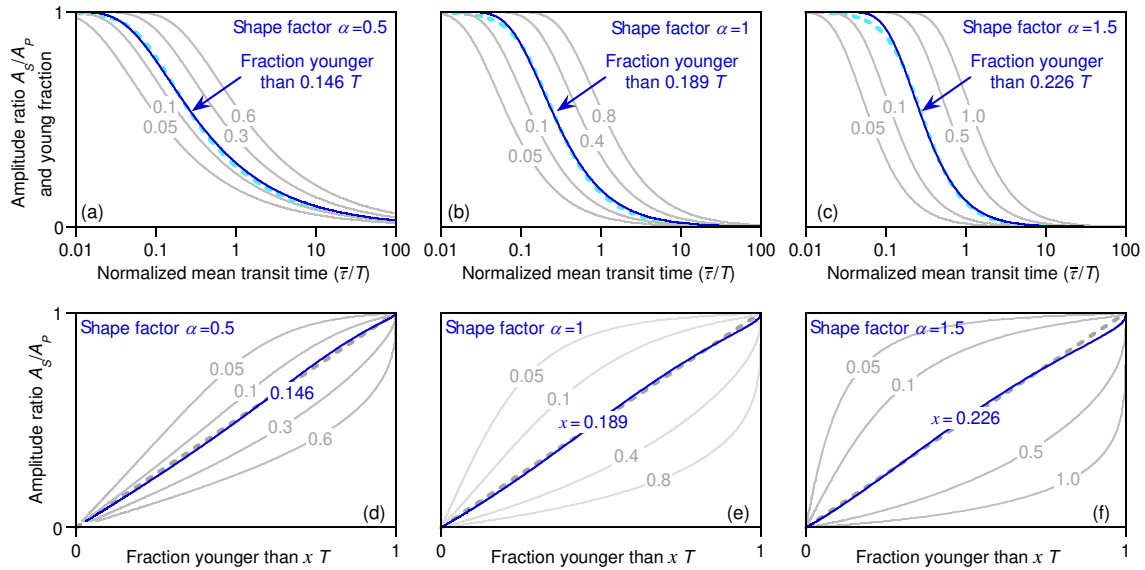
Figure 9. (a)–(c) show the amplitude ratios A_S/A_P in precipitation and streamflow tracer cycles (light blue dashed line) as function of mean transit time $\overline{\tau}$, compared to the fraction of water younger than several threshold ages (gray lines), and the best-fit age threshold (dark blue line). **(d)–(f)** show the relationship between amplitude ratio and the fraction of water younger than several age thresholds (gray lines) and the best-fit age threshold (dark blue line), with the 1 : 1 line (dashed gray) for comparison. Panels show results for three different gamma distributions, with shape factors $\alpha = 0.5$, $\alpha = 1$, and $\alpha = 1.5$. Root-mean-squared errors (RMSEs) for amplitude ratios A_S/A_P as predictors of the best-fit young water fractions are 0.012, 0.011, and 0.015 for **(d)–(f)**, respectively. In all panels, threshold age and mean transit time are normalized by T, the period of the tracer cycle. For seasonal tracer cycles, $T = 1$ year and thus threshold age and mean transit time are in years.

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

4.1   Young water

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

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

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

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

the best-fit threshold age varies modestly as a function of the shape factor α:

$$\tau_{yw}/T \approx 0.0949 + 0.1065\alpha - 0.0126\alpha^2. \tag{14}$$

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

The young water fraction F_{yw}, as defined here, has the inevitable drawback that, because the shape factors of individual tributaries will usually be unknown, the threshold age τ_{yw} will necessarily be somewhat imprecise. However, F_{yw} has the considerable advantage that it is virtually immune to aggregation bias in heterogeneous catchments because it is nearly equal to the amplitude ratio A_S/A_P (Fig. 9), which itself aggregates with very little bias and also with very little random error (Fig. 8). This observation leads to the important implication that A_S/A_P should reliably estimate F_{yw}, not only in individual subcatchments but also in the combined runoff from heterogeneous landscapes. To test this proposition, I calculated the young water fractions F_{yw} for 1000 heterogeneous pairs of synthetic subcatchments (with the same MTTs and shape factors shown in Fig. 7) using Eqs. (13) and (14), and compared each pair's average F_{yw} to the amplitude ratio A_S/A_P in the merged runoff. Figure 11 shows that, as hypothesized, A_S/A_P estimates the young water fraction in the merged runoff with very little scatter or bias. The root-mean-square error in Fig. 11 is roughly 2 % or less, in marked contrast to errors of several hundred percent shown in Fig. 7 for estimates of mean transit time from the same synthetic catchments.

4.2 Sensitivity to assumed TTD shape and threshold age

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

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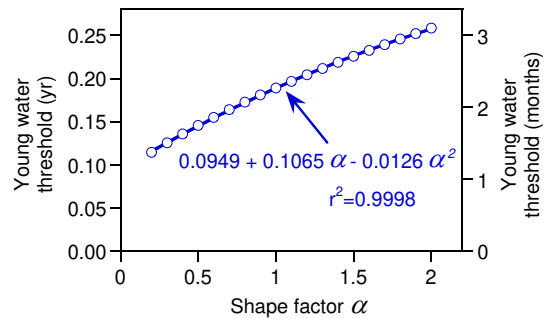

Figure 10. Best-fit young water thresholds for gamma transit-time distributions, as a function of shape factors α ranging from 0.2 to 2.0. The young water threshold τ_{yw} is defined such that the fraction of the distribution with ages less than τ_{yw} approximately equals the amplitude ratio (A_S/A_P)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

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

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

For comparison, we can contrast this uncertainty with the corresponding uncertainty in the mean transit time $\overline{\tau}$ calculated from Eq. (10). A seasonal tracer cycle amplitude ratio of $A_S/A_P = 0.2$ implies a mean transit time of $\overline{\tau} = 0.80$ years
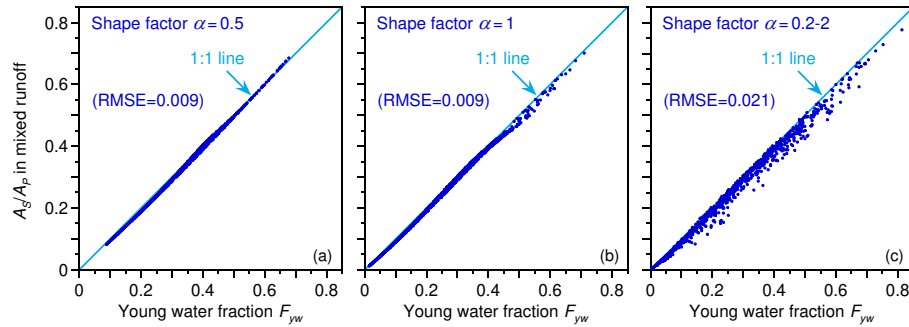
Figure 11. True and apparent young water fractions for the same 1000 synthetic catchments shown in Fig. 7. The tracer cycle amplitude ratio in the combined runoff of the two subcatchments (vertical axes) corresponds closely to the average young water fraction in the combined runoff (horizontal axes). As in Fig. 7, each synthetic catchment comprises two subcatchments with individual MTTs randomly chosen from a uniform distribution of logarithms spanning the interval between 0.1 and 20 years, and with each pair of MTTs differing by at least a factor of 2. In **(a)** and **(b)**, all subcatchments have the same shape factor α. In **(c)**, shape factors for each subcatchment are randomly chosen from a uniform distribution between $\alpha = 0.2$ and $\alpha = 2$.

if $\alpha = 1$, but $\overline{\tau} = 1.99$ years if $\alpha = 0.5$. Thus, the uncertainty in the mean transit time is a factor of 2.5, compared to a few percent for the young water fraction.

We can extend these sample calculations over a range of shape factors α and amplitude ratios A_S/A_P (see Fig. 12). As Fig. 12 shows, when the shape factor is uncertain in the range of $0.5 < \alpha < 1$, the corresponding uncertainty in the young water fraction F_{yw} is typically several percent, but the corresponding uncertainty in the MTT is typically a factor of 2 or more. For a factor of 10 uncertainty in the shape factor ($0.2 < \alpha < 2$), the uncertainty in the young water fraction is consistently less than a factor of 2, whereas the uncertainty in the MTT can exceed a factor of 100.

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

4.3 Young water estimation with nongamma distributions

Because both the young water fraction F_{yw} and the tracer cycle amplitude ratio A_S/A_P aggregate nearly linearly, the results shown in Fig. 11 will also approximately hold at higher levels of aggregation. That is, we can merge each catchment in Fig. 11, which has two tributaries, with another two-tributary catchment to form a four-tributary catchment, which we can merge with another four-tributary catchment to form an eight-tributary catchment, and so on. Figure 13 shows the outcome of this thought experiment. One can see that just like in the two-tributary case, the tracer cycle am-
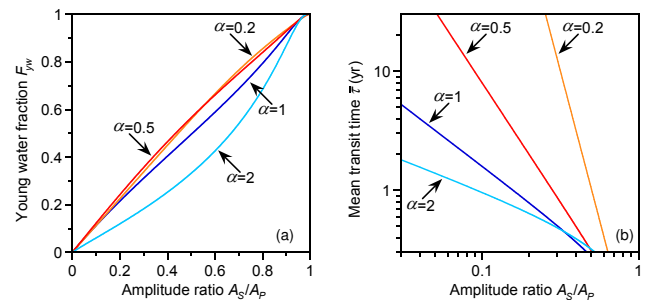

Figure 12. Sensitivity analysis showing how variations in shape factor α affect young water fractions F_{yw} **(a)** and mean transit times $\overline{\tau}$ **(b)** inferred from the amplitude ratio A_S/A_P of seasonal tracer cycles in precipitation and streamflow. Curves are shown for the four shape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

plitude ratio A_S/A_P in the merged runoff predicts the average young water fraction F_{yw} with relatively little scatter. There is a slight underestimation bias, which is more visible in Fig. 13 than for the two-tributary case in Fig. 11. In contrast to the minimal estimation bias in F_{yw}, MTT is underestimated by large factors in both the two-tributary case and the eight-tributary case.

It is important to recognize that the two-tributary catchments that were merged in Fig. 13 are not characterized by gamma transit-time distributions (although their tributaries are), because mixing two gamma distributions does not cre-
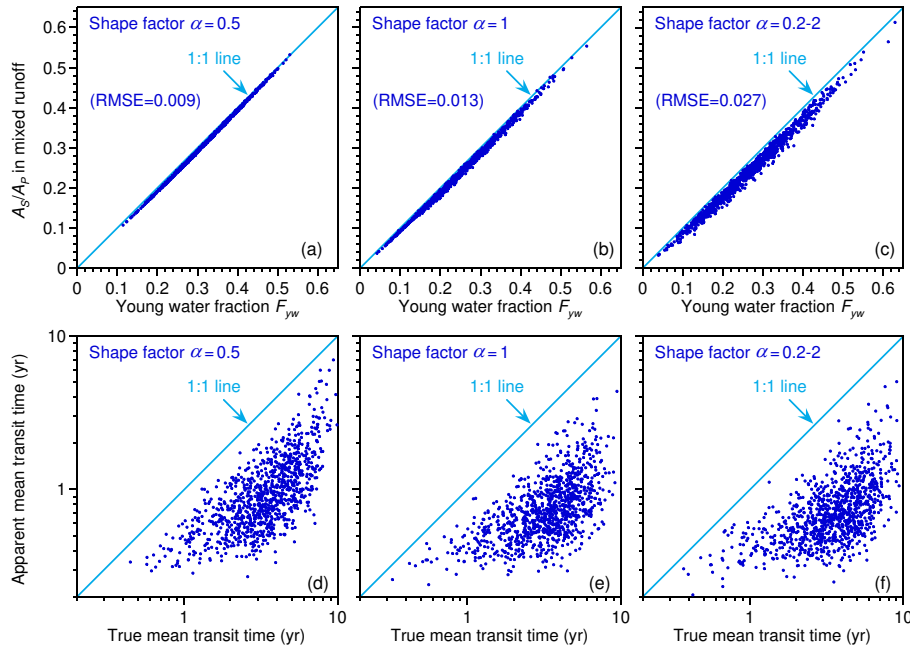
Figure 13. True and apparent young water fractions F_{yw} for 1000 synthetic catchments, each consisting of eight subcatchments with randomly chosen mean transit times between 0.1 and 20 years (top panels) and true and apparent mean transit times for the same catchments (bottom panels). The tracer cycle amplitude ratio in the combined runoff predicts the true young water fraction with a slight underestimation bias (top panels). Mean transit times inferred from tracer cycle amplitude ratios show severe underestimation bias (bottom panels). In **(a)**, **(b)**, **(d)**, and **(e)**, all subcatchments have the same shape factor α. In **(c)** and **(f)**, shape factors for each subcatchment are randomly chosen from a uniform distribution between $\alpha = 0.2$ and $\alpha = 2$.

ate another gamma distribution. Thus, Fig. 13 demonstrates the important result that although the analysis presented here was based on gamma distributions for mathematical convenience, the general principles developed here – namely, that the amplitude ratio $A_{\mathrm{S}}/A_{\mathrm{P}}$ estimates the young water fraction F_{yw}, and that estimates of F_{yw} are relatively immune to aggregation bias in heterogeneous catchments – are not limited to distributions within the gamma family.

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

$$h(\tau) = \frac{1}{\sum\limits_{i=1}^{n} k_i} \sum_{i=1}^{n} \frac{k_i}{\overline{\tau}_i} e^{-\tau/\overline{\tau}_i}, \tag{15}$$

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

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

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

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

One interpretation of the strong aggregation bias in mean transit-time estimates, as documented in Figs. 7 and 13, is that when the transit-time distributions of the individual tributaries are averaged together, the result has a different shape (i.e., averages of exponentials are not exponentials and averages of gamma distributions are not gamma-distributed). Thus, it is unsurprising that a formula for estimating mean travel times based on exponential distributions (for example) will be inaccurate when applied to nonexponential distributions. The practical issue in the real world, of course, is that the shape of the transit-time distribution will usually be un-
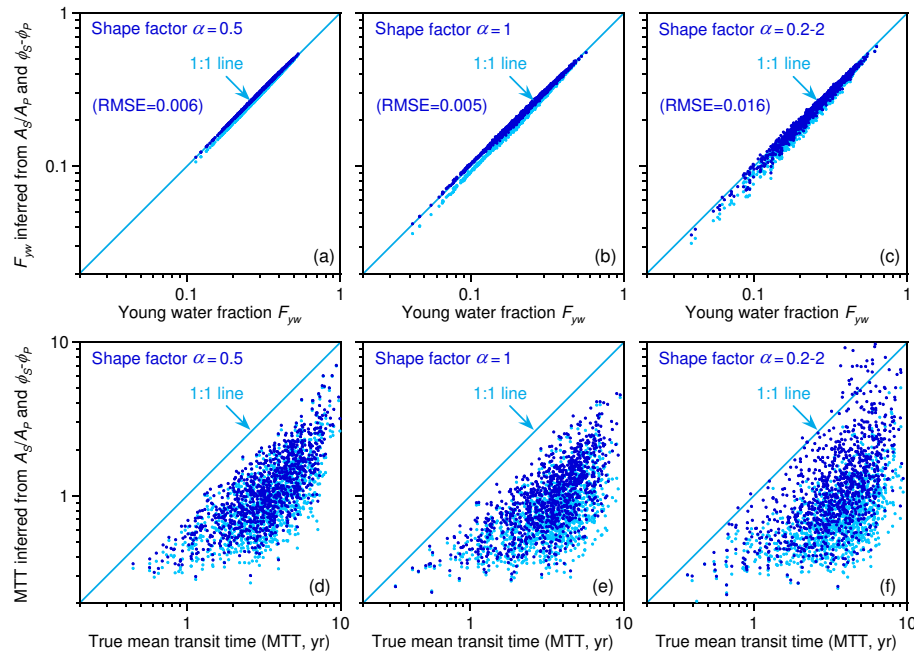
Figure 14. Effect of including phase information in estimates of young water fraction (F_{yw}) and MTT. Light symbols show F_{yw} and MTT estimates derived from tracer cycle amplitude ratios (A_S/A_P) alone; dark symbols show the same estimates derived from amplitude ratios and phase shifts ($\varphi_S - \varphi_P$). Data points come from the same 1000 synthetic catchments shown in Fig. 13, each consisting of eight 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.

known, so the problem of fitting the "wrong" distribution will be difficult to solve.

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

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

This approach assumes that the catchment's transit times are gamma-distributed. To test whether it can nonetheless improve estimates of the mean transit time or the young water fraction, even in catchments whose transit times are not gamma-distributed, I applied this method to the eight-tributary synthetic catchments shown in Fig. 13. As pointed

out in Sect. 4.3, the TTDs of these catchments (and even their two-subcatchment tributaries) will be sums of gammas and thus not gamma-distributed themselves. Figure 14 shows the new estimates based on amplitude ratios and phase shifts (in dark blue), superimposed on the previous estimates from Fig. 13, based on amplitude ratios alone, as reference (in light blue). Mean transit-time estimates based on both phase and amplitude information are somewhat more accurate than those based on amplitude ratios alone (Fig. 14d–f), but they still exhibit very large aggregation bias. Incorporating phase information in estimates of F_{yw} (Fig. 14a–c) eliminates much of the (already small) bias in F_{yw} estimates obtained from amplitude ratios alone. (The logarithmic axes of Fig. 14a–c make this bias more visible than it is on the linear axes of Fig. 13a–c). The top and bottom rows of Fig. 14 are plotted on consistent axes (both are logarithmic scales spanning a factor of 50), so they provide a direct visual comparison of the reliability of estimates of F_{yw} and MTT.

5   Implications

Two main results emerge from the analysis presented above. First, MTTs estimated from seasonal tracer cycles exhibit severe aggregation bias in heterogeneous catchments, underestimating the true MTT by large factors. Second, sea-

sonal tracer cycle amplitudes accurately reflect the fraction of young water in streamflow and exhibit very little aggregation bias. Both of these results have important implications for catchment hydrology.

5.1 Biases in mean transit times

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

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

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

ble isotopes, one cannot estimate how much they, too, are distorted by aggregation bias.

5.2 Other methods for estimating MTTs from tracers

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

It should also be noted that methods for estimating MTTs assume not only homogeneity but also stationarity, and real-world catchments violate both of these assumptions. The results presented here suggest that nonstationarity (which is, very loosely speaking, heterogeneity in time) is likely to create its own aggregation bias, in addition to the spatial aggregation bias identified here. This aggregation bias can also be characterized using benchmark tests, as I show in a companion paper (Kirchner, 2016).

5.3 Implications for mechanistic interpretations of MTTs

The analysis presented here implies that many literature values of MTT are likely to be underestimated by large factors or, in other words, that typical catchment travel times are probably several times longer than we previously thought

they were. This result sharpens the "rapid mobilization of old water" paradox: how do catchments store water for weeks or months, and then release it within minutes or hours in response to precipitation events (Kirchner, 2003)? This result also sharpens an even more basic puzzle: where can catchments store so much water, that it can be so old, on average?

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

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

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

Of course, because F_{yw} is nearly equal to the amplitude ratio and because MTT can also be expressed as a function of the amplitude ratio for TTDs of any known shape, one might conclude that MTT and F_{yw} are just transforms of one another. But that conclusion presumes that the shape of the TTD is known, and my analysis shows that in heterogeneous catchments the shape of the TTD will be unpredictable. Be-

cause the MTT is sensitive to the shape of the TTD – and in particular to the long-time tail, which is particularly poorly constrained – it cannot be reliably estimated. By contrast, my analysis shows that despite the uncertainty in the shape of the TTD in heterogeneous catchments, the F_{yw} can be reliably estimated from the amplitude ratio of seasonal tracer cycles in precipitation and runoff. The fact that this is possible is neither a miracle nor a fortuitous accident; instead, F_{yw} has been defined with exactly this result in mind. The F_{yw} entails an unavoidable ambiguity in what, exactly, the threshold age of young water is (because this depends on the shape of the TTD, which is usually unknown), but this uncertainty is small (Fig. 10) compared to the very large uncertainty in the MTT.

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

5.5 Potential applications for young water fractions

Since young water fractions are estimated from amplitude ratios and phase shifts of seasonal tracer cycles, one could ask whether they add any new information or whether we could characterize catchments equally well by their amplitude ratios and phase shifts instead. One obvious answer is that amplitude ratios and phase shifts, by themselves, are purely phenomenological descriptions of input–output behavior. Young water fractions, by contrast, offer a mechanistic explanation for how that behavior arises, showing how it is linked to the fraction of precipitation that reaches the stream in much less than 1 year. Not only is this potentially useful for understanding the transport of contaminants and nutrients, it also directly quantifies the importance of relatively fast flowpaths in the catchment. These fast flowpaths are likely to be shallow (since permeability typically decreases rapidly with depth: Brooks et al., 2004; Bishop et al., 2011) and to originate relatively close to flowing channels. One would expect F_{yw} to increase under wetter conditions, as the water table rises into more permeable near-surface zones and as the flowing channel network extends to more finely dissect the landscape (Godsey and Kirchner, 2014), thus shortening the path length of subsurface flows as well as multiplying the wetted catch-

ment area in riparian zones. In a companion paper (Kirchner, 2016), I show that young water fractions can be estimated separately for individual flow regimes, allowing one to infer how shifts in hydraulic forcing alter the fraction of streamflow that is generated via fast flowpaths. I further demonstrate how one can estimate the chemistry of "young water" and "old water" end-members, based on comparisons of F_{yw} and solute concentrations across different flow regimes.

Because one can estimate F_{yw} from irregularly and sparsely sampled tracer time series, it can be used to facilitate intercomparisons among many catchments that lack more detailed tracer data. For example, Jasechko et al. (2016) have recently used the approach outlined here to calculate young water fractions for hundreds of catchments around the globe, ranging from small research watersheds to continental-scale river basins, and to examine how they respond to variations in catchment characteristics.

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

6 Summary and conclusions

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

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

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

References

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

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

Barnes, C. J. and Bonell, M.: Application of unit hydrograph techniques to solute transport in catchments, Hydrol. Process., 10, 793–802, 1996.

Bethke, C. M. and Johnson, T. M.: Groundwater age and groundwater age dating, Annu. Rev. Earth Planet. Sci., 36, 121–152, doi:10.1146/annurev.earth.36.031207.124210, 2008.

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

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

Bishop, K., Seibert, J., Nyberg, L., and Rodhe, A.: Water storage in a till catchment II: Implications of transmissivity feedback for flow paths and turnover times, Hydrol. Process. 25, 3950–3959, doi:10.1002/hyp.8355, 2011.

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

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

Brooks, E. S., Boll, J., and McDaniel, P. A.: A hillslope-scale experiment to measure lateral saturated hydraulic conductivity, Water Resour. Res., 40, W04208, doi:10.1029/2003WR002858, 2004.

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

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

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

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

Feng, X. H., Faiia, A. M., and Posmentier, E. S.: Seasonality of isotopes in precipitation: A global perspective, J. Geophys. Res.-Atmos., 114, D08116, doi:10.1029/2008jd011279, 2009.

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

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

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

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

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

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

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

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

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

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

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

Jasechko, S., Kirchner, J. W., Welker, J. M., and McDonnell, J. J.: Substantial proportion of global streamflow less than three months old, Nat. Geosci., in press, 2016.

Kirchner, J. W.: A double paradox in catchment hydrology and geochemistry, Hydrol. Process., 17, 871–874, 2003.

Kirchner, J. W.: Aggregation in environmental systems – Part 2: Catchment mean transit times and young water fractions under hydrologic nonstationarity, Hydrol. Earth Syst. Sci., 20, 299–328, doi:10.5194/hess-20-299-2016, 2016.

Kirchner, J. W. and Neal, C.: Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection, P. Natl. Acad. Sci. USA, 110, 12213–12218, doi:10.1073/pnas.1304328110, 2013.

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

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

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

Kirchner, J. W., Tetzlaff, D., and Soulsby, C.: Comparing chloride and water isotopes as hydrological tracers in two Scottish catch-

ments, Hydrol. Process., 24, 1631–1645, doi:10.1002/hyp.7676, 2010.

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

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

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

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

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

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

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

McGuire, K. J., McDonnell, J. J., Weiler, M., Kendall, C., McGlynn, B. L., Welker, J. M., and Seibert, J.: The role of topography on catchment-scale water residence time, Water Resour. Res., 41, W05002, doi:10.1029/2004WR003657, 2005.

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

Niemi, A. J.: Residence time distributions of variable flow processes, Int. J. Appl. Radiat. Isotop., 28, 855–860, 1977.

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

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

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

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

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

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

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

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

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

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