Tandem use of transit time distribution and fraction of young water reveals the dynamic flow paths supporting streamflow at a mountain headwater catchment

Ravindra Dwivedi¹, Christopher Eastoe², John F. Knowles³, Jennifer McIntosh¹, Thomas Meixner¹, Ty P.A. Ferre¹, Rebecca Minor³, Greg Barron-Gafford³,⁴, Nathan Abramson¹, Michael Stanley⁵, and Jon Chorover⁶

¹Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, AZ, U.S.A.
²Department of Geosciences (retired), The University of Arizona, Tucson, AZ, U.S.A.
³School of Geography, Development & Environment, The University of Arizona, Tucson, AZ, U.S.A.
⁴Biosphere 2, The University of Arizona, Tucson, AZ, U.S.A.
⁵Mt. Lemmon Water District, Tucson, AZ, U.S.A.
⁶Department of Environmental Science, The University of Arizona, Tucson, AZ, U.S.A.

Correspondence to: Ravindra Dwivedi (ravindradwivedi@email.arizona.edu)
Abstract. Current understanding of the dynamic flow paths and subsurface water storages that support streamflow in mountain catchments is inhibited by the lack of long-term hydrologic data and the frequent use of single age tracers that are not applicable to older groundwater reservoirs. To address this, the current study used both multiple metrics and tracers to characterize the transient nature of flow paths with respect to change in catchment storage at Marshall Gulch, a sub-humid headwater catchment in the Santa Catalina Mountains, Arizona, USA. The fraction of streamflow that was untraceable using stable water isotope tracers was also estimated. A Gamma-type transit time distribution (TTD) was appropriate for deep groundwater analysis, but there were errors in the TTD shape parameters arising from the short record length of $^3$H in deep groundwater and stream water, and inconsistent seasonal cyclicity of the precipitation $^3$H time series data. Overall, the mean transit time calculated from $^3$H data was more than two decades greater than the mean transit time based on $\delta^{18}$O at the same site. The fraction of young water ($F_{yw}$) in shallow groundwater was estimated from $\delta^{18}$O time series data using weighted wavelet transform (WWT), iteratively re-weighted least squares (IRLS), and TTD-based methods. Estimates of $F_{yw}$ depended on sampling frequency, the method of estimation, bedrock geology, hydroclimate, and factors affecting streamflow generation processes. The coupled use of $F_{yw}$ and discharge sensitivity indicated highly dynamic flow paths that reorganized with changes in shallow catchment storage. The utility of $^3$H to determining $F_{yw}$ in deeper groundwater was limited by data quality. Given that $F_{yw}$, discharge sensitivity, and mean transit time all yield unique information, this work demonstrates how co-application of multiple methods can yield a more complete understanding of the transient flow paths and observable storage volumes that contribute to streamflow in mountain headwater catchments.
1 Introduction

Mountain headwater catchments are critical sources of water to downstream valley-fill aquifers (Viviroli et al., 2007; Viviroli et al., 2003; Kohler and Marselli, 2009; Carroll et al., 2019; Milly and Dunne, 2020; Harpold et al., 2012; Eppolito and Fonseca, 2021; Bryan, 2021). However, the conditions under which shallow and deep subsurface water storages and transient flow paths support streamflow in complex mountain terrain are still incompletely understood (McDonnell et al., 2018; McDonnell, 2017). Sustainable Development Goal (SDG) #15 of the 2030 Agenda for Sustainable Development (United Nations, 2015) specifically lists this shortcoming as a major hurdle to attaining sustainable development (United Nations, 2021; Creed and Noordwijk, 2018). Therefore, additional research is needed to thoroughly characterize dynamic relationships between storage, flow paths, and streamflow in the mountains.

Dual tracer approaches using stable water isotopes and tritium ($^3$H) are recommended to determine the contribution of deep groundwater to streamflow in mountain sites characterized by fractured bedrock aquifers (Stewart et al., 2010; Stewart et al., 2012). This recommendation originates from previous work that shows how mean transit time (mTT) estimates based on stable water isotopes alone may be underestimated because the tails of the transit time distributions (TTDs) that correspond to longer transit times can become truncated (Stewart et al., 2010; Stewart et al., 2012; Frisbee et al., 2013; DeWalle et al., 1997), and additionally takes into consideration that certain model performance criteria (e.g., Nash-Sutcliffe Efficiency) are insensitive to longer transit times (Seeger and Weiler (2014). Underestimated transit times can have cascading impacts on subsurface weathering rates, leading to incorrect understanding of stream water chemistry (Frisbee et al., 2013; Clow et al., 2018). As a result, the current
study leverages the use of $^3$H as a tracer with longer-period variations as a means to more completely characterize deeper flow paths that contribute to streamflow.

Using a virtual or synthetic experimental model setup and stable water isotope data, Kirchner (2016b, 2016a) identified significant aggregation errors in mTT estimates for heterogeneous catchments. As an alternative, Kirchner (2016b, 2016a) proposed the fraction of young water metric ($F_{yw}$), i.e., the fraction of water that has resided in a catchment for less than a threshold period of time, that is largely insensitive to spatial and temporal aggregation errors when evaluated for annual tracer cycles in inflow and outflow. Using a similar virtual experimental model setup but with $^3$H as a tracer, Stewart et al. (2017) also noted spatial aggregation errors in mTT but not $F_{yw}$. Together, these results suggest that catchment storage estimates based on $F_{yw}$ should be robust. Importantly, these and other $F_{yw}$-based studies (Jasechko et al. (2016); von Freyberg et al. (2018); Gallart et al. (2020a); Clow et al. (2018); Gallart et al. (2020a); Jacobs et al. (2018); Stockinger et al. (2016); Jasechko et al. (2017); Table S6; Table 2) considered annual or seasonal cycles in stable water isotope data or only one period when using only tritium (Stewart et al., 2017) or both tracers (Rodriguez et al., 2021); the authors know of no previous studies that have considered both $^3$H and $\delta^{18}$O tracers and multiple periods.

When $F_{yw}$ is related to the discharge flux, it contributes to a more thorough understanding of groundwater flow path dynamics and thus water quantity and quality in headwater catchments. Accordingly, the resulting discharge sensitivity of $F_{yw}$ can be considered the “diagnostic finger print” for streamflow generation processes. In this way, von Freyberg et al. (2018) distinguished between sites in terms of dominant flowpaths and flowpath changes during large and small storms, and suggested evaluation of their framework in different climatic and geological settings. Gallart et al. (2020b) applied
an alternative formulation for discharge sensitivity to a sub-humid Mediterranean field site, again concluding that flow paths responded dynamically to precipitation flux. Similar conclusions resulted from studies of alluvial and lower-elevation mountain-block aquifers in Arizona, USA (Eastoe and Wright, 2019; Eastoe and Towne, 2018) where recharge only occurred during the wettest ~30% of months. In this study, $F_{yw}$ is considered along with discharge sensitivity to better understand the degree to which catchment storage and flow paths are interrelated.

The current study addresses the following research questions at a high-elevation, sub-humid mountain site: (i) what is the appropriate TTD type and mTT for the deep groundwater system that supports streamflow? (ii) What are the $F_{yw}$ and resulting $F_{yw}$-based catchment storage estimates calculated from age tracers applicable to younger and older groundwater and stable water isotope and $^3$H time series data, respectively? (iii) What is the discharge sensitivity of $F_{yw}$ as determined by stable water isotope tracers? Following a description of the field site and data, we describe theoretical models and estimation methods for $^3$H-based TTD and mTT estimation and $F_{yw}$ (using both stable water isotopes and $^3$H) and its discharge sensitivity (stable water isotopes only). We discuss potential reasons for bias in the $F_{yw}$ literature toward both stable isotope tracers and annual cyclic variations and broaden the $F_{yw}$ approach by using both stable water isotope and $^3$H tracers simultaneously. The present study complements Dwivedi et al. (2021) who estimated TTD and mTT for shallow groundwater storages at the same study site using gamma distributions fitted to stable water isotope time series data in precipitation and streamflow.
2 Study site and data

2.1 Study site

The study site was the Marshall Gulch catchment (MGC), a 1.55 km² headwater catchment located within the Santa Catalina Mountains ~26 km northeast of Tucson in southeast Arizona, USA (Figure 1). The elevation at MGC ranges from 2285 to 2632 m above sea level (asl) with a mean of 2428 m asl and a mean topographic slope of 22° (or ~40%). Bedrock at the field site is mostly granite at upper elevations and micaceous schist at lower elevations (Dickinson et al., 2002). The prevailing soil type is sandy loam (Holleran, 2013b) with soil depth varying from 0 m to 1.5 m (Pelletier and Rasmussen, 2009). Soils overlying micaceous schist are generally deeper and have a higher clay content than soils overlying granite (Heidbüchel et al., 2013; Holleran, 2013a). Based on a 30-year (1981-2010) record, the long-term average annual precipitation at MGC is 920 mm (PRISM Climate Group, 2018). The catchment received an average of 654 mm (±158 mm) of precipitation per year between water years (WY) 2008 through 2017; the mean annual streamflow for the same period was 247 mm (±138 mm). WY \( n \) is defined here as the period from July 1 of year \( n-1 \) through June 30 of year \( n \). Instrumentation relevant to this study within and around the field site is shown in Figure 1.

2.2 Data

2.2.1 Hydrologic fluxes

The MGC-scale daily precipitation (P) and streamflow (Q) data were calculated between WY 2008 and 2017 (Figure 2A; Dwivedi et al. (2019a); Dwivedi et al. (2020)). Precipitation was observed at 15-minute
intervals at eight measurement sites equipped with tipping bucket precipitation gages at seven locations and a heated precipitation gage at the remaining site (Figure 1). From the precipitation time series, Thiessen polygon-derived weights were used to estimate daily catchment-scale mean precipitation (Dwivedi et al., 2019a). Streamflow was measured at 30-minute intervals at the MG-Weir site (Figure 1) using a pressure transducer (U20-001-01; Onset) with maximum error of 0.62 kPa and accuracy of 0.02 kPa and a previously derived stage-discharge relationship (Heidbüchel et al., 2012).

### 2.2.2 Stable water isotope data in precipitation and streamflow

#### 2.2.2.1 Precipitation

The precipitation bulk samples at MGC were collected using bulk samplers at the Schist, Fern Valley, and Granite station and using ISCO autosamplers at the MG-Weir and Mt. Lemmon stations (Figure 1). At the Fern Valley, Granite, and Schist stations, two collectors were installed at each station and samples were collected every 5 to 7 days (Heidbüchel et al. (2012); Lyon et al. (2009). At the Mt. Lemmon and MG-Weir stations, daily bulk precipitation samples were collected. At the Mt. Lemmon station, sampling mainly focused on summer monsoons (Heidbüchel et al., 2012), whereas continuous samples were collected beginning in December 2009 at the MG-Weir station. At all stations, the data density decreased after 2012 (see Figure 2B). The catchment-scale time series of $\delta^{18}$O in precipitation was calculated as the unweighted mean of results from all stations and was characterized by irregular time intervals between WY 2008 through WY 2012. (Heidbüchel et al. (2012); Dwivedi et al. (2021).
2.2.2.2 Streamflow

Stream water samples were collected using an autosampler installed at the MG-Weir site prior to 2012 and by grab sampling after 2012 (Figure 2B). While the stream water autosampler collected daily samples, sub-daily samples were also collected on the rising and falling limbs of the hydrograph during large runoff events (Heidbüchel et al., 2012). In the current study, sub-daily samples are volume-weighted to daily resolution (Dwivedi et al., 2021).

2.2.3 Tritium in precipitation, streamflow and deep groundwater

We calculated the amount-weighted time series of $^3$H in Tucson precipitation since 1992 following (i) Eastoe et al. (2004), (ii) The Environmental Isotope Laboratory, The University of Arizona, (access date: September 26, 2017); and (iii) Dr. C. Eastoe, unpublished data. At low levels of $^3$H concentration in precipitation, the data have a 1σ precision of ± 0.5 TU (tritium units) or less. Annual concentration cycles in the $^3$H data are best captured by the post-2001 period, which corresponds to semi-annual aggregates representing all precipitation events; between 1992 and 2001, the data represent large precipitation events only. Prior to 1992, mean annual $^3$H concentrations in Tucson precipitation were modeled from Doney et al. (1992) or interpolated (hollow squares) between observed or modeled values to obtain amount-weighted precipitation $^3$H concentrations at the half-yearly time scale (Figure S1A). All data were corrected for the elevation difference between the Tucson (747 m asl) and MGC (mean elevation 2428 m asl) by using the $^3$H concentrations in three simultaneous precipitation samples (Figure S1B) from Tucson and the Palisades Ranger Station, Santa Catalina Mountains. Input $^3$H time series data are shown in Figures 3 and S1C. The $^3$H time series data for deep groundwater was composed of observations in streamflow (baseflow conditions; n = 9) and groundwater from fractured bedrock (n = 5; Table S1; top
right inset). Additionally, the five discharge measurements from Pigeon Spring (Figure 1) were considered representative of deep groundwater (Dwivedi et al., 2019b). Data are grouped into half-yearly brackets using the following criteria: (i) sampling months 6 to 10 of year n, and (ii) sampling months 11 of year n-1 to 5 of year n. For groups with three or more measurements, the data are expressed as a mean ± 1σ (Fig. 3 inset).

3 Method description and previous results

3.1 Multi-tracer based TTD and mTT estimation methods

3.1.1 Stable water isotope-based TTD and mTT estimates

Dwivedi et al. (2021) proposed an improved practical approach for estimating the TTD of stream water using long-term measurements of hydrologic fluxes and δ¹⁸O in precipitation and streamflow. Evaluating multiple TTD types and using the weighted wavelet spectral analysis method of Kirchner and Neal (2013), they determined that a combined Piston Flow and Gamma TTD was applicable for periods of up to one month, and that a Gamma TTD was applicable thereafter. The resulting Gamma TTD shape parameter (α) was 0.42±0.001 (dimensionless) and the mTT was 0.82±0.03 years.

3.1.2 Tritium-based TTD and mTT estimation

Previous work determined that deep groundwater at MGC was recharged at a time scale of 0.5 years or less on the basis of end-member mixing analysis (Ajami et al. (2011); Dwivedi et al. (2019b). Here, we use time series ³H data to expand on these results. Given that these data are only available at seasonal
resolution (Section 2.2.3; Figure 3), the Stewart et al. (2016) method, i.e., Equation (1) below, was used to estimate the mean transit time and best fitting TTDs:

\[ C_Q(t) = \int_0^\infty C_{Recharge}(t - \tau) e^{-\frac{\log_2\tau}{t_{1/2}}} h(\tau) d\tau \quad (1) \]

where \( C_{Recharge}(t) \) and \( C_Q(t) \) are ³H concentration in recharge and stream water at time \( t \), \( \tau \) is the transit time in years, \( t_{1/2} \) is the ³H half-life, and “\( e \)” is the exponential function. Since the decay in ³H concentration during recharge is insignificant in relation to the precision of the analysis (± 0.5 TU), the precipitation ³H concentration, i.e., the input function in Figure 3, is used as \( C_{Recharge}(t) \). Following Maloszewski and Zuber (1993), only TTD types that require at most two fitting parameters were evaluated. Thus, parallel exponential models (Seeger and Weiler, 2014; Hrachowitz et al., 2009) and exponential piston flow models (Georgek et al., 2017) are excluded here. The specific TTD types evaluated by the current study were Piston Flow (PF), Exponential (Exp), Gamma (Gam), Fixed path one-dimensional advection dispersion (ADE-1x), and Multiple path one-dimensional advection dispersion (ADE-nx) (Dwivedi et al., 2021).

### 3.1.2.1 Optimization of model parameters using the Downhill simplex method in conjunction with a performance criterion

The Downhill Simplex method (Nelder and Mead, 1965; Gupta, 2016) was used to evaluate the performance of each TTD (Dwivedi et al., 2021). The modified Kling Gupta efficiency or KGE’ (Gupta et al., 2009; Kling et al., 2012) was used as the model performance criterion:

\[ KGE’ = \sqrt{\left(\frac{Cov_{m,o}}{\sigma_m \sigma_o} - 1\right)^2 + \left(\frac{\sigma_m / \mu_m}{\sigma_o / \mu_o} - 1\right)^2 + \left(\frac{\mu_m}{\mu_o} - 1\right)^2} \quad (2) \]
In Equation (2), $Cov_{m,o}$ is the covariance between the modeled (subscript $m$) and observed (subscript $o$) time series, $\sigma_m$ and $\sigma_o$ represent one standard deviation, and $\mu_m$ and $\mu_o$ are the mean of the modeled and observed time series, respectively. A perfectly fitting model will have a KGE’ value of zero and the worst fitting model will have a KGE’ value of $\infty$. Following Godsey et al. (2010), the KGE’ criterion was estimated in a log-transformed space. Both KGE’ and the characteristics of the criterion response surface were utilized to search for the optimum model parameters (Dwivedi et al., 2021). The following ranges of model parameters were considered: Mean transit time: 1 to 50 years when using tritium, which serves as a groundwater age tracer at a time scale of 1 to 50 years (Suckow, 2014; Aggarwal, 2013; Gleeson et al., 2015). Shape parameter ($\alpha$) for the Gamma TTD: 0.1 to 15 (Stewart et al., 2017). Average catchment scale Péclet number (Pe): 0.1 to 100 (Kirchner et al., 2001; Kirchner and Neal, 2013).

3.2 Multi-tracer based $F_{yw}$ estimation

3.2.1 Mathematical development of the flux-weighted ($F_{yw}^*$) and unweighted ($F_{yw}$) fraction of young water

The fraction of young water ($F_{yw}$) can be estimated from the amplitude ratio of tracer concentrations in outflow and inflow for any tracer (Kirchner, 2016b; von Freyberg et al., 2018). Thus, if the amplitudes of the tracer concentrations in outflow and inflow for any period $\lambda$ are $A_Q(\lambda)$ and $A_P(\lambda)$, respectively, then:

$$F_{yw}(\lambda) = \frac{A_Q(\lambda)}{A_P(\lambda)}$$

(3)

This method is preferred when sufficient long-term tracer data are available and can be applied without a priori knowledge of the TTD type (Kirchner, 2016b). If the TTD type and its parameters are known, then $F_{yw}$ can also be estimated (Kirchner, 2016b; von Freyberg et al., 2018; Stewart et al., 2017):
\[ F_{yw}(\lambda) = \int_{0}^{T_{yw}} h(\tau) d\tau \]  \quad (4)

where \( T_{yw} \) is the threshold age for the young water, defined as the upper limit in Equation (4) for which both Equations (3) and (4) provide the equivalent value of \( F_{yw}(\lambda) \). Note that \( T_{yw} \) depends not only on the periods of sinusoidal cycles in tracer concentrations, but also on the TTD parameters.

In contrast to \( \delta^{18}O \)-based \( F_{yw} \) calculations that are based on tracer fluxes in precipitation and streamflow, the \( ^3H \)-based \( F_{yw} \) in the current study is based on tracer concentrations in recharging and discharging deep groundwater. For consistency with the literature, we express tracer flux-based fraction of young water values as \( F_{yw}^{*} \) and tracer concentration-based fraction of young water values as \( F_{yw} \).

However, the mathematics in both cases generally remain the same, and we provide only the mathematical derivation for \( F_{yw} \) below, unless otherwise noted. It is assumed that the precipitation tracer flux can be represented by Equation (5), which denotes an input function of sinusoidal type with period \( \lambda \):

\[ PC(t) = A_{P}(\lambda) \sin \left( \frac{2\pi}{\lambda} t - \phi_{P}(\lambda) \right) + K_{P}(\lambda) \]  \quad (5)

where \( A_{P}(\lambda) \), \( \phi_{P}(\lambda) \), and \( K_{P}(\lambda) \) are the amplitude, phase angle and a constant for a given period. Similarly, the stream water tracer flux can be represented by Equation (6), which denotes an output function of tracer fluxes with period \( \lambda \):

\[ QC(t) = A_{Q}(\lambda) \sin \left( \frac{2\pi}{\lambda} t - \phi_{Q}(\lambda) \right) + K_{Q}(\lambda) \]  \quad (6)

in which \( A_{Q}(\lambda) \), \( \phi_{Q}(\lambda) \), and \( K_{Q}(\lambda) \) are analogous to the variables of Equation (5). Note that \( A_{Q} \) is also used to refer to the \( ^3H \) tracer cycle in deep groundwater.
The stream water tracer flux (or concentration in the case of \( ^3\)H) can be conceptualized as the time-convolution of the input tracer flux weighted by TTD (Equation 1 when using a conservative tracer). For a tracer subject to radioactive decay, we use Equation 1 of Stewart et al. (2017), which represents the contribution to streamflow of different flow paths in time and space, in addition to a decay factor \( e^{-\kappa \tau} \) where \( 1/\kappa \) is the mean life of the isotope of interest. Thus,

\[
QC(t) = \int_0^\infty PC(t - \tau) e^{-\kappa \tau} h(\tau) d\tau
\]

(7)

where \( PC(t) \) and \( QC(t) \) are the transient tracer fluxes in precipitation and stream water, respectively, and \( h(\tau) \) is the transit time distribution or TTD (e.g., Equation 4 for the Gamma TTD).

Substitution of Equation (5) into Equation (7) yields:

\[
QC(t) = \int_0^\infty \left( A_P(\lambda) \sin(\omega (t - \tau) - \phi_P(\lambda)) + K_p(\lambda) \right) e^{-\kappa \tau} h(\tau) d\tau
\]

(8)

If \( h(\tau) \) is a Gamma TTD, Kirchner (2016b) has shown that the amplitude damping for the outflow relative to inflow tracer signals can be estimated from the Fourier transform of \( h(\tau) \). For example, for a Gamma TTD with shape parameter \( \alpha \) and scale parameter \( \beta \) (=mTT/\alpha), its power spectrum can be expressed as the following (Bain, 1982):

\[
(H(\omega))^2 = \frac{1}{(1+(\omega \beta)^2)^\alpha}
\]

(9)

Thus, the amplitude damping for any period can be expressed as:

\[
\frac{A_Q}{A_P} = \frac{1}{(1+(\omega \beta)^2)^{\frac{\alpha}{\tau}}}
\]

(10)
The decay factor \( \exp(-\kappa \tau) \) is not considered in Equation 10. If the decay factor is considered, then the scale parameter in the Gamma TTD can be transformed to yield an expression akin to the standard Gamma TTD equation with a power spectrum equation similar to Equation 10. The subsequent expression for \( A_Q / A_P(\omega) \) is:

\[
\frac{A_Q}{A_P}(\omega) = F_{yw}(\omega) = \frac{1}{(1 + \kappa \beta)^\alpha (1 + \frac{(\omega \beta \tau)^2}{1 + \kappa \beta})^{\frac{\alpha}{2}}} \tag{11}
\]

For a conservative tracer (e.g., \( \delta^{18}O \)), \( \kappa = 0 \) in Equation 11, yielding Equation 10. Consequently, the amplitude ratio of output to input tracer fluxes can be obtained analytically for any period if the TTD parameters are already known using Expression 11. To relate the amplitude ratio to TTD through the parameter \( T_{yw} \) (Equation 4), \( T_{yw} \) can be estimated by:

\[
T_{yw}(\lambda) = gaminv\left(\frac{A_Q}{A_P}(\lambda), \alpha, \beta\right) \tag{12}
\]

where \( gaminv \) is the standard MATLAB® inverse Gamma function. In section S1, Equation 11 is used to compare the output:input amplitude ratios for various periods with respect to the Exponential TTD, which is a special case of the Gamma TTD. In section S2, the formulation of \( T_{yw} \) (Equation 12) is compared with the alternative recommendations of Kirchner (2016b) and Stewart et al. (2017). For \( \delta^{18}O \), the applicable tracer period ranged from 2 days (due to the median daily sampling interval; see Section 2.2.2) to 5 years (the length of the high-density data record, WY 2008 to WY 2012), whereas the applicable \( ^3H \) tracer period ranged from 1 year (due to half-yearly sampling of \( ^3H \) in Tucson precipitation) to 27 years (the length of the data record).
3.2.2 Estimation of $F_{yw}$ and $T_{yw}$ using various tracers

3.2.2.1 Estimation of $F_{yw}^*$ and $T_{yw}^*$ using stable water isotope tracers

Multiple methods were utilized to compare $F_{yw}^*$ and $T_{yw}^*$ estimates from long-term stable water isotope data during the same period. In all cases, estimates of $F_{yw}^*$ were obtained using tracer flux data, i.e., as a product of tracer concentration and hydrologic flux. Daily precipitation and streamflow were used to calculate the fraction of precipitation contributing to streamflow where daily precipitation was aggregated to time steps corresponding to the availability of stable water isotope data in precipitation. The iteratively re-weighted least square (IRLS) method was used to estimate $F_{yw}^*$ by fitting sinusoidal functions with periods ranging from 2 days to 5 years to tracer flux data (Kirchner, 2016b; von Freyberg et al., 2018). Additional estimates of $F_{yw}^*$ were obtained using the weighted wavelet transform (WWT) method (Kirchner and Neal (2013) and the TTD method (Equation 11), also with periods of 2 days to 5 years.

With respect to $\delta^{18}$O, method- and period-based $F_{yw}^*$ estimates were coupled to previously established TTD parameters (Dwivedi et al., 2021), and multi-year average values of $F_{yw}^*$ were used to compare between different methods (Stockinger et al. (2019); Gallart et al. (2020a). The $T_{yw}^*$ was estimated as a function of period using $F_{yw}^*$ estimates obtained from application of each of the three methods.

The temporal variability of $\delta^{18}$O in precipitation was addressed by means of uncertainty analyses of $F_{yw}^*$ and $T_{yw}^*$. For $F_{yw}^*$, the temporal variability of $\delta^{18}$O was expressed as three statistics: daily mean, mean + 1 standard deviation ($\sigma$), and mean - 1$\sigma$ calculated for both precipitation (P) and stream water (Q); consideration of all pair combinations resulted in total nine scenarios. For each period, the minimum, mean (referred to as the ensemble mean below), maximum, and 1$\sigma$ of the $F_{yw}^*$ results were computed for
all nine scenarios. The $T^*_yw$ process was similar but included additional uncertainty associated with the Gamma TTD parameter such that there were 27 total scenarios for each period.

### 3.2.2.2 Estimation of $F_{yw}$ and $T_{yw}$ using $^3$H

The $^3$H tracer data in precipitation and deep groundwater at MGC are sparse relative to $\delta^{18}$O, especially for deep groundwater (Figure 3). As a result, application of the IRLS method to deep groundwater $^3$H was unsatisfactory i.e., fitting a sinusoidal function to the data resulted in significant amplitude uncertainty at a period of one year. Therefore, the $^3$H $F_{yw}$ and $T_{yw}$ estimates are based on the TTD method (Equations 16 and 17). Because the estimated $^3$H-based TTD parameters demonstrated significant variability (Section 4.1.1), $F_{yw}$ and $T_{yw}$ uncertainty analyses followed the TTD-based $F^*_yw$ and $T^*_yw$ calculations derived from $\delta^{18}$O (Section 3.2.2.1).

### 3.3 Estimation of discharge sensitivity of $F_{yw}$ from $\delta^{18}$O data

When there is a linear relationship between $F_{yw}$ and discharge (this linear relationship has been observed at MGC up to a threshold discharge; points in Figure 7A), von Freyberg et al. (2018) suggested that the discharge sensitivity of $F_{yw}$ for $\lambda=1$ year can be estimated by approximating the amplitude of the tracer cycle in stream water as a function of discharge alone, i.e., $A_Q=n+m*Q$, where $n$ and $m$ are constants. It is important to note here that what has been termed “discharge sensitivity” of $F_{yw}$ in previous studies is in fact the response of discharge to $F_{yw}$. Nonetheless, if tracer cycle in precipitation has amplitude $A_P$, then $F_{yw}$ can be expressed as:

$$F_{yw} = \frac{A_Q}{A_P} = \frac{n}{A_P} + \left(\frac{m}{A_P}\right)Q$$  \hspace{1cm} (13)
Using the expression for tracer cycle amplitude (Equation 13) and an equation similar to Equation (6) above for the sinusoidal tracer concentration cycle, the unweighted tracer cycle in stream water can be expressed as:

\[ C_Q(t) = (n + mQ) \sin \left( \frac{2\pi}{\lambda} t - \phi_Q \right) + K_Q \quad (14) \]

By fitting \( C_Q(t) \) from Equation (14) to the observed tracer time series (e.g., blue points in Figure 2B), the discharge sensitivity of \( F_{yw} \), which is equal to \( m/A_P \) and has units of 1/units of discharge, can be estimated for the annual tracer cycle. Note that von Freyberg et al. (2018) did not consider the phase angle, \( \phi_Q \), to be a function of discharge, because changes in \( \phi_Q \) between high and low flows would have only a minor influence on the discharge sensitivity of \( F_{yw} \).

Gallart et al. (2020b) noted limitations of the von Freyberg et al. (2018) approach, especially under high discharge conditions. They proposed an alternative method for discharge sensitivity estimation that is robust under high flow conditions. In the Gallart et al. (2020b) approach, \( F_{yw} \) is expressed as an exponential-type equation:

\[ F_{yw} = 1 - (1 - F_o) e^{-SQ} \quad (15) \]

where \( F_o \) (dimensionless) is the virtual \( F_{yw} \), i.e., \( F_{yw} \) for \( Q=0 \), and \( S \) is the discharge sensitivity of \( F_{yw} \) in units of 1/units of discharge. While the approach of von Freyberg et al. (2018) assumes that \( A_Q \) is a linear function of \( Q \), \( A_Q \) in the Gallart et al. (2020b) approach assumes a more complex function of \( Q \). Nonetheless, using Equation (15), the tracer concentration in stream water can be expressed as:

\[ C_Q(t) = A_P \left( 1 - (1 - F_o) e^{-SQ} \right) \sin \left( \frac{2\pi}{\lambda} t - \phi_Q \right) + K_Q \quad (16) \]
By fitting Equation 16 to the observed tracer cycle in stream water for $\lambda = 1$ year, $F_0$ and $S$ can be estimated. In contrast to Equation 14, the solution of Equation (16) for $F_0$ and $S$ is not straightforward because they appear in a single term of Equation 16. To estimate the two parameters, Gallart et al. (2020b) used a non-linear technique that may lead to parameter estimates unrelated to reality, for instance negative values of $F_0$.

Instead of a non-linear technique, the current study presents a practical, simplified approach. If $S^*Q \ll 1$, then $\exp(-SQ)$ can be approximated by the first two terms in its Maclaurin series expansion, so that Equation 16 becomes:

$$C_Q(t) = A_P F_0 \sin \left( \frac{2\pi}{\lambda} t - \phi_Q \right) + A_P (1 - F_0) SQ \sin \left( \frac{2\pi}{\lambda} t - \phi_Q \right) + K_Q$$

(17)

Thus, the IRLS method can be used for estimating $F_{yw}$ discharge sensitivity and its associated uncertainty. If $S^*Q \ll 1$, Gallart et al. (2020b) have shown that the discharge sensitivity of $F_{yw}$ should be similar to that estimated from the approach of von Freyberg et al. (2018). In the current work, we have used both the Gallart et al. (2020b) and von Freyberg et al. (2018) methods for estimating the discharge sensitivity of $F_{yw}$, in order to better understand the nature of transient flow paths that are activated as catchment storage changes.
4 Results

4.1 $^3$H-based TTD type, parameters, and modeled outflow tracer concentrations

4.1.1 $^3$H-based TTD type and mTT at Marshall Gulch

Both Piston Flow (PF) and Gamma (Gam) TTDs performed adequately and yielded TTD parameters within the permissible parameter space (Table 1). However, model performance expressed in terms of KGE’ was slightly better for the PF relative to the Gam TTD type (PF KGE’ was 29% lower). Comparison of the PF and Gam TTDs further suggested “approximate equifinality” (Kirchner, 2016a) in the PF TTD results (Figure 4E vs. 4B). For example, three separate PF TTD model runs yielded mTTs of 32.5, 29.5, and 35.5 years (Figure 4E; Case 5 in Table S3) with very similar KGE’ values (~0.4; Case 5 in Table S3). In contrast, three separate Gam TTD runs yielded similar mTTs (mTT ~ 26 years) and $\alpha$ parameters (5.23). Overall, the PF TTD mTT varied between 4 and 33 years with a coefficient of variation of 0.57 (Table 1), and the Gam TTD mTT and $\alpha$ parameters varied between 26 and 30 years (mean mTT = 27 yrs; coefficient of variation = 0.05) and 2.17 to 14.58 (unitless) (mean $\alpha$ = 6.53; coefficient of variation = 0.64), respectively.

4.1.2 Modeled tracer concentrations in deeper groundwater

Modeled $^3$H concentrations in deep groundwater were generally within their observed ranges and remained within $1\sigma$ of their simulated means for both PF and Gam TTDs, but modeled $^3$H concentration variability was lower for the Gam TTD type (Figure 5). The error bars in Figure 5 were determined by considering the analytical uncertainty in $^3$H concentrations and are based on the first set of model runs.
Although the ADE-1x TTD-based model produced $^3$H concentrations that were within their observed ranges, the estimated TTD parameters were sometimes at the edge of the allowable parameter space; simulated $^3$H concentrations resultant from the ADE-nx TTD-based model were far from observed concentrations and suggest that this TTD type is not applicable at MGC.

4.2 $F_{yw}^*$ and $T_{yw}^*$ based on $\delta^{18}$O

4.2.1 Comparison of $\delta^{18}$O tracer-based $F_{yw}^*$ estimates using various methods

Considering $F_{yw}^*$ variability due to $\delta^{18}$O variability in precipitation and stream water, the IRLS method yielded more variable results than the WWT or TTD-based methods, particularly for periods less than one year (Figure 6A). Note that both the IRLS and WWT methods are based on sinusoidal curve fitting to the observed tracer data, but the $F_{yw}^*$ estimates resultant from the WWT method are less scattered because it involves spectral smoothing of data noise (Dwivedi et al., 2020; Kirchner and Neal, 2013). For a period of 1 year, $F_{yw}^*$ estimated from the TTD-based method was higher than the corresponding estimates obtained from the IRLS and WWT methods (Figure 6A). Comparison of ensemble means (due to significant variability in $F_{yw}^*$ estimates) for $\lambda = 1$ year indicated a $F_{yw}^*$ ensemble mean $\pm 1\sigma$ of 34.9 $\pm$ 0.5% using the TTD-based method, compared to 11.4 $\pm$ 0.7% and 7.9 $\pm$ 0.2% for the IRLS and WWT methods, respectively.

4.2.2 Comparison of $\delta^{18}$O-based $T_{yw}^*$ estimates

As with $F_{yw}^*$, the $T_{yw}^*$ results showed significantly greater variability when estimated using the IRLS method compared to the WWT or TTD-based methods, especially for periods below 0.5 years (Figure
Estimation of $T_{yw}^*$ with any method requires \textit{a priori} knowledge of the TTD parameters (Equation 385) that are identical to those used to calculate $F_{yw}^*$. For $\lambda = 1$, the ensemble $T_{yw}^*$ means ± 1σ were 0.125 ± 0.0058 yrs (TTD), 0.008 ± 0.0013 yrs (IRLS), and 0.004 ± 0.0003 yrs (WWT).

\subsection*{4.3 $F_{yw}$ and $T_{yw}$ based on $^3$H}

\subsubsection*{4.3.1 $F_{yw}$}

The time series’ of $^3$H in groundwater and streamflow (Fig. 3, inset) were too sparse and coarse for reliable estimation of $A_Q/A_P$ using the IRLS or WWT methods (Equation 3). Instead, we used the TTD method to calculate $F_{yw}$ with model parameters drawn from section 4.1.1. The $F_{yw}$ was characterized by a gradual increase in $dF_{yw}/d\lambda$ (Figure 6C). For $\lambda = 1$, the ensemble mean-based $F_{yw}$ was $(1.6 \pm 2.40) \times 10^{-3}$ % (blue triangle in Figure 6C).

\subsubsection*{4.3.2 $T_{yw}$}

Although $T_{yw}$ estimated using the TTD method gradually increased with period, $dT_{yw}/d\lambda$ gradually declined (Figure 6D). As with the $F_{yw}$ estimates, large error bars reflect variability in the TTD parameters. For an annual cycle, the ensemble mean-based $T_{yw}$ was $2.03 \pm 2.22$ yr (blue triangle in Figure 6D).

\subsection*{4.4 Discharge sensitivity of annual $F_{yw}$ estimated from $\delta^{18}$O data}

The discharge sensitivity of $F_{yw}$ is the slope of $F_{yw}$ vs. discharge, $Q$ (Figure 7A). Following the methods of von Freyberg et al. (2018) and Gallart et al. (2020b) and for $\lambda = 1$ year, calculated discharge sensitivities were $0.09 \pm 0.02$ day/mm and $0.11 \pm 0.02$ day/mm at Marshall Gulch, respectively. However, discharge sensitivity depended on whether tracer data were weighted by streamflow. Discharge sensitivities were
0.09 ± 0.02 (mean ± standard error) and 0.11 ± 0.02 without weighting but decreased to 0.03 ± 0.01 day/mm and 0.04 ± 0.01 day/mm when the tracer data were weighted by streamflow. These estimates were computed by fitting a sinusoidal cycle of an annual period to the observed stream water δ¹⁸O data. Analysis of $F_{yw}$ vs. $Q$ suggests that $F_{yw}$ initially decreases with increasing $Q$ (Figure 7A). This pattern may be due to an evaporative increase in stream water δ¹⁸O under low-flow conditions, leading to an increase in $A_Q$, and thus an increased $A_Q/A_P$ ratio as $Q$ decreases (Jasechko (2019); Stockinger et al. (2017)). At higher flows ($Q > ~3.2 \text{mm/day}$), $F_{yw}$ remained constant, but there is lower confidence in these estimates because the high-$Q$ brackets of data contain fewer observations with significant variability. Given that the high flow observations correspond to periods immediately following high intensity precipitation, the results suggest an asymptotic nature of discharge sensitivity at higher flows.

5 Discussion

5.1 $^3$H-based TTD type and mTT estimates

Previous estimates of TTD type and mTT at Marshall Gulch (MGC) were based on single stable isotope tracers (Heidbüchel et al., 2012; Dwivedi et al., 2021). The current work compliments these studies by using a tracer ($^3$H) that is applicable over decades rather than years. For both $^3$H and δ¹⁸O tracers, a Gamma TTD type was appropriate for MGC with mTT ~0.82 yrs ($\alpha = 0.42$, unitless) using δ¹⁸O and ~27 yrs ($\alpha = 6.53$, unitless) using $^3$H. These composite results are consistent with Stewart et al. (2010) that also noted differences in mTTs using a combination of $^3$H and δ¹⁸O. The $^3$H-based mTT estimate was close to the 26-year interval over which the amount-weighted $^3$H data were available and was consistent with ages of bedrock-hosted groundwater from Dwivedi et al. (2019b). We therefore conclude that the
large difference between mTT calculated from $^3$H versus $\delta^{18}$O can be attributed to the range of applicability of each tracer. Stable water isotopes are generally considered applicable to determine groundwater ages up to 5 years (DeWalle et al., 1997; Dwivedi et al., 2021) and are therefore appropriate for subsurface storages with faster flow (Stewart et al., 2010). In contrast, $^3$H is generally considered applicable up to a period of 50 years (Suckow, 2014; Aggarwal, 2013) and is thus appropriate for estimating “hidden” or deep groundwater contributions to streamflow (Stewart et al., 2012). The current study that uses both tracers can separate the contributions of both quick and slow groundwater flow to streamflow in a headwater mountain catchment.

5.2 Subsurface storages

The $\delta^{18}$O tracer is ostensibly applicable to soil water storage at MGC because the residence time of soil water is expected to be low due to high hydraulic conductivity (Heidbüchel et al., 2013; Heidbüchel et al., 2012; van der Velde et al., 2014). As a result, the $\delta^{18}$O tracer-based TTD calculated by the current study is likely associated with soil water storage. The current work also suggests that short-term storage estimates depend on the method used to estimate $F_{yw}$ and $T_{yw}$. If the short-term storage in a catchment is defined as the upper limit of water storage with age $\leq T_{yw}^*$ (or $\leq T_{yw}$), and calculated as $F_{yw}^* \times T_{yw}^* \times Q$ (Jasechko et al., 2016), estimates for short-term storage at MGC vary between 0.08 mm (WWT method), 0.22 mm (IRLS method), and 10.7 (TTD method). Using a method akin to IRLS, previous reported global short-term storages ranged between 1 and 55 mm (median 14 mm) (Jasechko et al. (2016)). For catchments comparable in size to MGC and for which mean annual discharge data are reported, the range narrows to between 6 mm (Rietholzbach site) and 30 mm (McDonalds B site) (Jasechko et al. (2016)). Thus, the
short-term storage estimates determined using various methods at MGC generally fall within the global range, albeit at the lower end for sites with subhumid climate, this may reflect the thin soils at MGC. No similar attempts were made to estimate long-term storage from $^3$H data because of the shortcomings of the dataset.

In the traditional approach (see Rodriguez et al. (2021) and references therein), storage is a function of the tracer-based mTT multiplied by the long-term mean discharge. Application of this approach to MGC results in storage estimate of 6.7 m using an $^3$H-based mTT of 27 years and the observed long-term streamflow (WY 2008 to WY 2017). In contrast, an equivalent modern groundwater depth of 3 m on the land surface was determined by Gleeson et al. (2015). Acknowledging such issues, Kirchner (2016a) - using stable water isotopes and virtual experiments - suggested the use of volume-weighted rather than time-weighted mTTs. However, Peters et al. (2013) and Dwivedi et al. (2021) have shown that volume-weighted mTTs and time-weighted mTTs differ by only a factor of ~2 as opposed to orders of magnitude. Volume weighted mTTs are also difficult to obtain via $^3$H data due to the lack of multi-decade observations of both streamflow and tracer concentrations in outflow. Further, the use of mean long-term discharge is problematic as the contribution of deep groundwater to streamflow is likely to be lower than contributions from soil water or other near surface storages. Using end-member mixing analysis, Dwivedi et al. (2019b) reported a deep groundwater contribution to streamflow of 4.5% of the long-term streamflow at MGC. If this fraction is included in the storage calculation, then the storage estimate decreases to a more plausible 0.3 m. Taken together, the results of the current work suggest that the tracer-based mTT can be used to estimate storage volumes, provided they are interpreted with appropriate caution.
5.3 Comparison to previous estimates of $T_{yw}$ and $F_{yw}$

Previous work has proposed $F_{yw}$ as a metric that can be used to compare hydrologic characteristics among catchments (Jasechko, 2016; Kirchner, 2016b; von Freyberg et al., 2018). Here, $F_{yw}$ and $T_{yw}$ estimates for MGC are compared to other study sites; in all cases, the estimates correspond to $\lambda = 1$ year.

5.3.1 Estimates from stable water isotopes

The TTD-based $T^*_{yw}$ (0.125 ± 0.0058 yrs) was within the range reported by Kirchner (2016b) i.e., between 0.11 and 0.25 years for $\alpha$ ranging between 0.2 and 2, but $T^*_{yw}$ and $F^*_{yw}$ estimates from the IRLS and WWT methods were lower than the corresponding TTD-based metrics (Section 4.2). Gallart et al. (2020a) reported that $F^*_{yw}$ estimates for $\lambda = 1$ year increased with higher sampling resolution (this finding was also corroborated by Stockinger et al. (2016) with $F^*_{yw}$ values of 10.3%, 22.6%, and 30.4% resultant from weekly, high-resolution (30-minute as well as flow-dependent sampling), and “virtual thorough” sampling that involved using 5-minute discharge along with the $F_{yw}$ vs. Q relationship to estimate $F_{yw}$ (Table S6). The results of the current work support Gallart et al. (2020a) insofar as the TTD-based $F^*_{yw}$ represents a thorough sampling of flowpaths with transit time between 0 and $T^*_{yw}$ years. In this way, the TTD based $F^*_{yw}$ results may be more reliable than estimates derived from the IRLS or WWT methods that potentially lack thorough sampling of flowpaths between the transit times of 0 and $T^*_{yw}$ years. However, the literature on $F^*_{yw}$ is mostly based on IRLS or similar methods with few studies reporting TTD-based results (Table 2).

In comparing the results for MGC with those of other studies (Table 2), a problem arises because most previous studies used only the IRLS method or similar method to estimate $F_{yw}$. At MGC, the IRLS
method appears to underestimate $F_{yw}$ relative to the TTD method as explained above. It is probably not useful to compare the TTD results from MGC with IRLS results from elsewhere; therefore, we begin by comparing the IRLS result at MGC with those from other studies.

Comparison of the IRLS-based $F_{yw}^*$ estimates at MGC to the literature (Table 2) suggests that streamflow generation processes, hydroclimate, bedrock geology, sampling frequency of stream water, and the estimation method used all influence values of $F_{yw}$ in ways that have not been previously reported. For example, von Freyberg et al. (2018) characterized a site with $F_{yw}^* = 49\%$ that was characterized by fast shallow groundwater flow paths during both small and large precipitation events, but another site with $F_{yw}^* = 20\%$ where dynamic flow paths changed with precipitation events of different sizes. While the discharge sensitivity of MGC (section 4.4) also suggests dynamic flow paths, these flow paths principally appear to co-evolve with changes in catchment storage. At a site with a Mediterranean sub-humid climate, Gallart et al. (2020a) reported an $F_{yw}^*$ value of 22.6, and a meta-analysis of catchments comparable in area to MGC and located mostly in humid climates determined that $F_{yw}^*$ (estimated as 26\% greater than reported $F_{yw}$; von Freyberg et al. (2018)) ranged between 6 and 33\% (Jasechko et al. (2016) Table S6; Table 2); however, the fractured bedrock at MGC is functionally distinct than the "watertight" bedrock characterized by Gallart et al. (2020a) and the majority of humid sites in Jasechko et al. (2016) that are comparable in size to MGC. Similarly, $F_{yw}^*$ values reported by Clow et al. (2018), Zhang et al. (2018), and Bansah and Ali (2019) can be attributed to differences in the climate and/or bedrock characteristics between MGC and their field sites. $F_{yw}^*$ discrepancies between MGC and mountain sites in Lutz et al. (2018) and a headwater catchment in Stockinger et al. (2019) are likely due to their relatively coarser stream water sampling frequency relative to MGC (Table 2).
Comparison of the $F_{yw}^*$ at MGC to TTD-based $F_{yw}$ estimates in the literature shows that MGC is at lower end of the range reported by Wilusz et al. (2017) for humid Plynlimon catchments in the U.K. (Table 2). We attribute these differences to methodological inconsistencies including the use of different $T_{yw}^*$ values. Specifically, Wilusz et al. (2017) conducted rainfall-runoff modeling in conjunction with rank StorAge Selection (rSAS) function-based transit time modeling (more details in Harman, 2015), in order to estimate $F_{yw}^*$ when $T_{yw}^* = 0.25$ years, as opposed to $T_{yw}^* = 0.12$ years at MGC. In other cases, differences between TTD-based $F_{yw}^*$ estimates in the current study and those reported in the literature are the result of using a single period for TTD parameter estimation (e.g., Song et al. (2017); Stockinger et al. (2017)), as opposed to various periods and wavelet analysis to determine the appropriate TTD type and its parameters at MGC (Dwivedi et al., 2021). A TTD-based $F_{yw}^*$ estimate (1.5%) for an oceanic, forested catchment was significantly lower than MGC and was likely due to gently sloping topography at that site versus the steep topography at MGC (Rodriguez et al. (2021) Table 2). Remondi et al. (2019) applied integrated hydrological flow and transport models to determine $F_{yw}^*$ at MGC using a synthetic topography, but the results spanned almost the entire range of permissible $F_{yw}$ values and the model performed poorly during the winter season that is the dominant recharge period (Dwivedi et al., 2021).

### 5.3.2 Estimates from $^3$H

The ensemble mean $F_{yw}$ based on annual $^3$H cycles at MGC was $1.6 \times 10^{-3}$ %, or effectively 0% (Figure 6C). For comparison, the lowest $F_{yw}$ value in Stewart et al. (2017) is ~8% for a system composed of two homogeneous sub-systems, each having an mTT of 25 years with the Gamma TTD shape parameter $\alpha = 10$ (unitless). We attribute this difference to either: (i) the constant $T_{yw}$ value that Stewart et al. (2017)
used for both $\alpha$ and tracer cycle period; or (ii) differences in the hydrogeologic settings between MGC and the Stewart et al. (2017) New Zealand catchments, or (iii) the use of multiple lumped parameter models by Stewart et al. (2017), as opposed to the $F_{yw}$ values in the current work that are based on fitting a single TTD to the whole $^3$H dataset for deep groundwater. If Equation (11) is applied to the MGC data using a Gamma TTD with parameters $\alpha = 10$ (unitless) and $mTT = 25$ years (e.g., from Stewart et al., 2017), the resulting value of $F_{yw}$ would be $1.1 \times 10^{-10}\%$, still effectively 0%, indicating that parameter choice is not responsible for the difference. Using TTD parameter estimates from rSAS (rank StorAge Selection) functions, Rodriguez et al. (2021) reported a $F_{yw}$ estimate of 1.8% for a forested headwater catchment that is closer to the near-zero $F_{yw}$ estimate at MGC but may also be subject to differences in sampling protocols and/or calculation methods i.e., Rodriguez et al. (2021) sampled stream water under varying flow conditions, in contrast to baseflow sampling in this work, and used a $T_{yw}$ value of 0.2 years.

A negligible $F_{yw}$ at MGC calls into question of the suitability of the $^3$H-based $F_{yw}$ approach for deeper groundwater. The fractured bedrock storage at MGC has a large $mTT$ (~27 years), and the annual tracer cycle will be highly damped as a result ($F_{yw} \sim 0$; section 4.3.1). From the standpoint of the IRLS method (Equation 3), the $^3$H data in precipitation or deep groundwater, if useful, should have an amplitude $A_P$ or $A_Q$ greater than the $^3$H measurement precision (0.5 TU, Section 2.2.3) for some period between 1 and 27 years. At MGC, this is true for $A_P$ at periods greater than 19 years, but not for $A_Q$ at any period up to 27 years (Section S3); consequently, the available data are inadequate for calculating $F_{yw}$. This is apparent in the lack of consistent annual periodicity in the Tucson Basin precipitation data (Figure 3), which may not be possible to overcome even with a much larger $^3$H dataset.
5.4 Dynamic catchment behavior revealed by the discharge sensitivity

The discharge sensitivity of $F_{yw}$ at MGC suggests that flowpaths in shallow storages restructure and reorganize dynamically as catchment storage changes (Figure 7B). Using the discharge sensitivity of $F_{yw}$ for various Swiss catchments, von Freyberg et al. (2018) suggested three cases (Figure 7B) for dynamic evolution of flowpaths: Case 1: catchments with high $F_{yw}$ but low discharge sensitivity in which fast flowpaths dominate and persist during both large and small events; Case 2: catchments with low $F_{yw}$ but high discharge sensitivity in which different flowpaths dominate during large and small events; and Case 3: catchments with low $F_{yw}$ and low discharge sensitivity in which slow flowpaths dominate and persist during both large and small events. When evaluated without any flow-weighting, the discharge sensitivity of $F_{yw}$ at MGC suggests that it falls under Case 2 i.e., reorganization of flowpaths with change in catchment storage, as recently noted in the more humid Plynlimon, U.K. catchments (Wilusz et al., 2017). The linear-regression approach to calculating $F_{yw}$ specifically indicates a threshold of $\sim 3.2$ mm/day above which the mean $F_{yw}$ does not increase with Q at MGC (Figure 7A). Using a method similar to Jasechko et al. (2016) with estimates for $F_{yw}$ and $T_{yw}$ based on Gamma TTD parameters from Dwivedi et al. (2021) using their “Method 1” and Equation (17) in this work, discharge of $3.2$ mm/day results in a short-term storage estimate of $0.05$ m. Thus, after a threshold of $0.05$ m short-term near-surface storage at MGC, the current study supports that infiltration may activate deeper groundwater flowpaths (Dwivedi et al. (2019b)).

There is evidence for a global inverse relationship between topographic slope and $F_{yw}$ (Jasechko et al., 2016; 2017). Topographic roughness and fractured bedrock permeability may also play roles in promoting infiltration to fractured-bedrock aquifers in steep mountainous catchments once shallow
storage is exceeded. A case in point is the difference between recharge seasonality in two mountain blocks of similar lithology: the Santa Catalina Mountains, Arizona where MGC is located (mainly winter recharge) and the neighboring Rincon Mountains, Arizona (both summer and winter recharge). Eastoe and Wright (2019) attributed the difference in infiltration of summer rainwater to topographic control by the orientation of pegmatite sheets that are steeply dipping in the Santa Catalina Mountains versus sub-horizontal in the Rincon Mountains that is more conducive to infiltration during high intensity summer precipitation events. This current work suggests the presence of an additional factor, threshold storage, that contributes to observed inverse relationships between $F_{yw}$ and topographic slope.

5.5 Limitations of the proposed approach

Considered together, $F_{yw}$ and its discharge sensitivity provide complimentary information about transient flowpaths (Figure 8A). Increasing catchment storage specifically results in reorganization of transient flow paths and activation of deep flow paths rather than simple acceleration of flow along any given path. Importantly, the magnitude of the $F_{yw}$ metric depends on the method used to estimate it. Recent literature has suggested limitations of the $\delta^{18}$O-based $F_{yw}$ metric (e.g., Jacobs et al. (2018) Stockinger et al. (2017) Jasechko et al. (2016). However, the literature has mainly reported $F_{yw}$ estimates based on either IRLS or similar sinusoidal curve fitting methods for annual or seasonal tracer cycles (Table 2; Table S6). The current work contributes to the growing body of $F_{yw}$ research by quantifying the variability of IRLS-based results at a mountain headwater catchment (Figure 6A). For an annual tracer cycle, $F_{yw}$ from the IRLS method was one-third of $F_{yw}$ from the TTD method, and it is therefore likely that previously reported $F_{yw}$ estimates may be underestimated. This would have significant implications for $F_{yw}$-based understanding.
of contaminant and nutrient transport, surface water quality (Kirchner, 2016a; Jasechko et al., 2016), and estimation of TTD parameters (Lutz et al., 2018). As a result, future studies utilizing IRLS or similar methods may wish to report $F_{yw}$ for various periods, in addition to the annual period, in order to better constrain the variability of the results. Future studies that reported TTD-based results would also be useful to characterize the methodological sensitivity of $F_{yw}$ across a broader range of natural systems.

The use of a $^3$H-based $F_{yw}$ metric has been recommended toward an improved understanding of deep and/or slow flowpaths contributing to streamflow (Jacobs et al., 2018; Jasechko, 2019). However, the current study highlights that the $^3$H-based $F_{yw}$ metric may be inappropriate when there are insufficient deep groundwater data, which is a general limitation in groundwater aquifers including MGC (Rodriguez et al., 2021; Gleeson et al., 2015). This limitation can also lead to significant variability in the estimated Gamma TTD parameters when the $^3$H tracer is applied to the question of “hidden streamflow” (Stewart et al., 2010; Stewart et al., 2012; Seeger and Weiler, 2014; Jacobs et al., 2018). In contrast to $F_{yw}$, the $^3$H-based mTT metric does not depend on any particular period of tracer cycles in inflow and outflow, but aggregation errors may lead to estimates of mTT that are low by several orders of magnitude relative to known mTTs from virtual experiments, especially in heterogeneous catchments such as MGC (Kirchner, 2016b; Stewart et al., 2017). The current work also demonstrates that the $^3$H-based mTT can lead to greatly over-estimated total deep groundwater storage estimates. To address this issue, appropriate long-term discharge estimates, not including storm runoff, are critical to accurate storage calculations (Section 5.2). Finally, the current results support the use of multiple (“lumped”) parameter models, qualified by site-specific hydrogeological information to reduce aggregation errors in real catchments, but
acknowledge that model parameters may be difficult to constrain in the multiple parameter approach (Stewart et al. (2017); Jacobs et al. (2018); Hrachowitz et al. (2009)).

**6 Conclusions**

This study supports concurrent application of multiple metrics for a more complete understanding of the transient flow paths and storage volumes that contribute to streamflow in a sub-humid mountain headwater catchment. Among the various combinations of tracers and metrics that were tested, the most appropriate metrics at MGC included δ\(^{18}\)O-based \(F_{yw}\), discharge sensitivity, and shallow storage m\(\text{TT}\), in addition to \(^{3}\)H-based deeper storage m\(\text{TT}\). Application of the weighted wavelet transform (WWT), iteratively re-weighted least square (IRLS), and transit time distribution (TTD) methods to annual cycles of δ\(^{18}\)O in stream water resulted in flux-based \(F_{yw}\) values of 7.9 ± 0.2%, 11.4 ± 0.7%, and 34.9 ± 0.5%, respectively. The current study therefore constrains the degree to which \(F_{yw}\) depends on the method of estimation. At MGC, the Gamma TTD was preferred on the basis that it thoroughly sampled flow paths between a transit time of zero and a threshold age for young water. In comparison, the IRLS results were scattered over the periods of interest and were only approximately one-third of the TTD-based estimate an annual period; WWT results were similar to IRLS but showed much less scatter owing to spectral smoothing of the data.

The Gamma TTD-based m\(\text{TT}\) using \(^{3}\)H data was 27 years. The same methodology yielded an MTT of 0.82 years when based on δ\(^{18}\)O (Dwivedi et al., 2021); hence, we conclude that the former m\(\text{TT}\) may correspond to groundwater stored in fractured bedrock, whereas the latter applies to shallow storages in the soil profile. The shape parameters of the \(^{3}\)H-based Gamma TTD at MGC demonstrated
significant variability arising from the short length and inconsistent seasonal cyclicity of the available 
$^3$H time series data that precluded adequate estimation of $F_{yw}$ in fractured-bedrock groundwater. 

Although data quality could be addressed by longer-term observation and attention to precision, 
variations of $^3$H in precipitation at some locations may restrict the applicability of this approach. In 
summary, using $\delta^{18}$O-based $F_{yw}$ together with its discharge sensitivity was an effective method with 
which to quantify the dynamic nature of shallow groundwater flowpaths at MGC. Beyond a threshold 
$F_{yw}$ in short-term storage, additional infiltration is likely to activate deeper groundwater flow paths. 

**Author contributions:**

R. Dwivedi: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization

C. Eastoe: Writing – original draft preparation, Writing – review & editing

J. F. Knowles: Writing – review & editing

J. McIntosh: Funding acquisition, Project administration, Resources, Supervision, Writing – original draft 
preparation, Writing – review & editing

T. Meixner: Funding acquisition, Project administration, Resources, Supervision, Writing – original draft 
preparation, Writing – review & editing

P. A. “Ty” Ferre: Funding acquisition, Project administration, Resources, Supervision, Writing – original 
draft preparation, Writing – review & editing

R. Minor: Writing – review & editing

G. Barron-Gafford: Writing – review & editing
N. Abramson: Data curation, Writing – review & editing

M. Stanley: Resources, Writing – review & editing

J. Chorover: Funding acquisition, Project administration, Resources, Supervision, Writing – original draft preparation, Writing – review & editing

Acknowledgments

We would like to thank Drs. J. Kirchner and J. von Freyberg for guiding the fraction of young water analysis, and Drs. J. Kirchner and M. Stewart for helpful comments on a preliminary version of this paper. The authors acknowledge support from National Science Foundation Grants EAR 1331408 and 0724958 to the Santa Catalina Mountains and Jemez River Basin Critical Zone Observatory. RD also acknowledges the Geological Society of America for Graduate Student Research and Horton Research Grants, and the The University of Arizona Graduate and Professional Student Council for Research and Travel grants; TM, JM, TF, and RD acknowledge additional support from the Water Resources Research Center 104(b). The corresponding author would like to thank his wife, Jessie Dwivedi, and their children, Darcy and Noah Dwivedi, for their patience and support during this project. Some aspects of this study are based on previously published data. The authors declare no conflicts of interest.

References


Kohler, T., and Marselli, D.: Mountains and Climate Change - From Understanding to Action, 75 p., 2009.


USGS NED 1 arc-second n33w111 1 x 1 degree ArcGrid 2018: https://viewer.nationalmap.gov/basic/, access: May 15, 2018, 2018.
Figure 1: Marshall Gulch Catchment (MGC; the catchment boundary is shown in green), located within the Santa Catalina Mountains Critical Zone Observatory (SCM-CZO) in southeast Arizona, USA (inset map), along with the precipitation, stream water, and deep groundwater collection sites. The digital elevation model data are from the U.S. Geological Survey (2018).
Figure 2. (A) Timeseries plots of daily precipitation (P) and streamflow (Q) and (B) $\delta^{18}O$ in P and Q from water year (WY) 2008 through WY 2017. The error bars in (B) show one standard deviation.
Figure 3. Amount-weighted $^3$H concentration in precipitation (data from Eastoe et al. (2004); The Environmental Isotope Laboratory, The University of Arizona; Eastoe, unpublished data; Figure S1), and $^3$H concentrations in deep fractured bedrock groundwater (blue points; a combination of MG-Weir site and Pigeon Spring; see Figure 1).
Figure 4. Response surfaces for various TTD types when using $^3$H input and output functions as shown in Figure 3. Results for the three separate model runs are shown as different symbols.
Figure 5. Observed (gray points and error bars) and modeled (blue points and error bars) $^3$H concentration in deeper groundwater for (A) Exp, (B) Gam, (C) ADE-1x, (D) ADE-nx, and (E) PF TTD types. Error bars for the modeled concentrations represent one standard deviation of all the modeled concentrations for the first model run (i.e., run #1 in Table S3), based on uncertainty in amount-weighted $^3$H concentrations in precipitation and deeper groundwater.
Figure 6. (A and B): $F_{yw}$ (ensemble mean) and $T_{yw}$ (ensemble mean) vs. period ($\lambda$), based on $\delta^{18}$O data and (C and D): $F_{yw}$ (ensemble mean) and $T_{yw}$ (ensemble mean) vs. period ($\lambda$), based on $^3$H data. The blue triangles in each plot show the ensemble mean of $F_{yw}$ (or $F_{yw}$) and $T_{yw}$ (or $T_{yw}$) values for $\lambda = 1$ year, calculated using the TTD method.
Figure 7. (A) Relationship between $F_{yw}$ with discharge for various flow regimes (total number of observations for each flow regime shown in plot legend) computed using the methods of von Freyberg et al. (2018) and Gallart et al. (2020b) with (solid lines with square markers) and without (solid lines) flow-weighting when estimating the discharge sensitivity parameters. (B) Relationship between $F_{yw}$ with $Q$ with respect to the framework proposed by von Freyberg et al. (2018) (see section 5.4 for more details). Note, the blue line in (B) is the inferred $F_{yw}$ from discharge at MGC using the discharge sensitivity of $F_{yw}$ calculated following von Freyberg et al. (2018) without considering flow-weighting. The horizontal and vertical error bars in (A) correspond to standard errors in $Q$ and $F_{yw}$, respectively.
Figure 8. Novel contributions of this work in terms of inferred transient nature of flowpaths that restructure with catchment storage and threshold short-term storage after which the propensity for precipitation to infiltrate and activate deep groundwater increases. (A) Transit time distribution or TTD (h(τ)) and TTD parameters estimated using tritium, which compliments previous stable water isotope-based TTD estimates at the same site (Dwivedi et al., 2021). (B) The nine visually indistinguishable h(τ) curves are based on stable water isotope data and show TTDs with mean transit time (mTT) of 0.82±0.03 (mean ± one standard deviation) and α=0.42±0.001 (unitless). (C) The nine h(τ) curves are based on tritium and correspond to TTDs with mean transit time (mTT) of 27±1.4 and α=6.53±4.21 (unitless). The black curves in (B) and (C) show Gamma TTDs with TTD parameters of mTT=0.82 yr and α=0.42 (unitless) and mTT=27 yr and α=6.53 (unitless), respectively.
Table 1. Estimated TTD parameters for various TTD types. The TTD parameters in columns (2) and (3) are based on the input and output functions shown in Figure 3. The TTD parameter statistics in columns (5) through (14) are based on the first set of model runs that consider amount-weighted $^3$H concentration uncertainty in precipitation and concentration uncertainty in deep groundwater (Table S3). Parameter 1 is the mean transit time (in years). Parameter 2 is not applicable for the PF TTD type and is the scale parameter $\alpha$ (unitless) for the Exp and Gam TTD and the Pe parameter for the ADE-1x and ADE-nx TTD types. The parameter $\alpha$ is set to 1 for the Exp TTD type. KGE’ is the modified Kling-Gupta Efficiency, which ranges between 0 (for the best fitting model) and infinity (for the worst fitting model).

<table>
<thead>
<tr>
<th>TTD type</th>
<th>Parameter 1</th>
<th>Parameter 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parame</td>
<td>Parame</td>
</tr>
<tr>
<td></td>
<td>ter 1</td>
<td>ter 2</td>
</tr>
<tr>
<td>Piston flow</td>
<td>32.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Exponential</td>
<td>50.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gamma</td>
<td>26.34</td>
<td>5.23</td>
</tr>
<tr>
<td>ADE-1x</td>
<td>3.63</td>
<td>100.00</td>
</tr>
<tr>
<td>ADE-nx</td>
<td>24.44</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NA: not applicable; NE: not estimated; Min: minimum; One std: one standard deviation; CV: coefficient of variation; and Max: maximum.
Table 2. Synthesis of $F_{yw}^*$ estimates reported in literature including the current study. The $F_{yw}^*$ values for studies shown in bold were obtained by scaling $F_{yw}$ estimates by a factor of 26% from von Freyberg et al. (2018).

<table>
<thead>
<tr>
<th>Date source</th>
<th>Study site</th>
<th>$F_{yw}^*$ for annual tracer cycle (method used)</th>
<th>Mean topographic gradient (%)</th>
<th>Sampling frequency</th>
<th>Climate</th>
</tr>
</thead>
<tbody>
<tr>
<td>This study</td>
<td>MGC</td>
<td>7.9 (WWT), 11.4 (IRLS), and 34.9 (TTD)</td>
<td>40</td>
<td>median daily data</td>
<td>Sub-humid</td>
</tr>
<tr>
<td>von Freyberg et al. (2018)</td>
<td>Erlenbach, Lumpenenbach, Reitholzbach, and Vogelbach out of 22 catchments</td>
<td>20 to 49 (IRLS)</td>
<td>14.6 to 28.9</td>
<td>Biweekly</td>
<td>Humid to temperate continental climate</td>
</tr>
<tr>
<td>Jaseckho et al. (2016)</td>
<td>Brugga - Oberried, DE; Botorpstrommen - Gunnebo, SE; Rietholzbach - Mosnang (Pegel), CH; McDonalds B</td>
<td>6 to 33 (periodic regression method)</td>
<td>Not provided</td>
<td>~ monthly</td>
<td>Humid (Brugga – Oberried, McDonalds B); warm and temperate (Botorpstrommen - Gunnebo);</td>
</tr>
<tr>
<td>Study</td>
<td>Location</td>
<td>Duration</td>
<td>Frequency</td>
<td>Climate/Region</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------</td>
<td>-------------</td>
<td>----------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Gallart et al. (2020a)</td>
<td>Lebanon State Forest, USA</td>
<td>22.6 (least square fitting)</td>
<td>25.6</td>
<td>temperate humid (Rietholzbach - Mosnang Pegel);</td>
<td></td>
</tr>
<tr>
<td>Song et al. (2017)</td>
<td>Zuomaokong watershed, Qinghai-Tibet Plateau</td>
<td>26 (TTD)</td>
<td>4.5</td>
<td>Mediterranean sub-humid</td>
<td></td>
</tr>
<tr>
<td>Wilusz et al. (2017)</td>
<td>Lower Hafren and Tanllwyth, UK</td>
<td>30 to 55 (TTD)</td>
<td>Not provided</td>
<td>Weekly humid sites</td>
<td></td>
</tr>
<tr>
<td>Zhang et al. (2018)</td>
<td>Boulder Creek watershed, USA</td>
<td>8 to 28</td>
<td>Not provided</td>
<td>Weekly Alpine, subalpine, and Montane</td>
<td></td>
</tr>
<tr>
<td>Lutz et al. (2018)</td>
<td>Bode catchment, Germany</td>
<td>1.3 to 19 (IRLS and ordinary least square)</td>
<td>16.5 (sites in mountains)</td>
<td>Monthly Cold and wet in high elevations</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Location</td>
<td>Flow Duration</td>
<td>Flow Frequency</td>
<td>Climate</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------------</td>
<td>---------------</td>
<td>----------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>Bansah and Ali (2019)</td>
<td>South Tobacco Creek watershed, Canada</td>
<td>42 to 91 (sinusoidal curve fitting)</td>
<td>Not provided</td>
<td>Sub-humid</td>
<td></td>
</tr>
<tr>
<td>Clow et al. (2018)</td>
<td>Andrews Creek, CO, and Andrews spring, CO, USA out of 11 catchments</td>
<td>19 to 21 (multiple regression method)</td>
<td>85 to 102</td>
<td>Monthly</td>
<td>Cold and wet during winter and warm and dry during summer</td>
</tr>
<tr>
<td>Stockinger et al. (2017)</td>
<td>Wusteback headwater catchment, Germany</td>
<td>13 to 16 (multiple linear regression) and 14 to 16 (TTD)</td>
<td>Not provided</td>
<td>Weekly</td>
<td>humid</td>
</tr>
<tr>
<td>Stockinger et al. (2019)</td>
<td>Wusteback headwater catchment, Germany</td>
<td>5 to ~16 (IRLS)</td>
<td>Not provided</td>
<td>Weekly</td>
<td>humid</td>
</tr>
<tr>
<td>Rodriguez et al. (2021)</td>
<td>Weierbach Catchment, Luxembourg</td>
<td>1.5 (TTD-based method that uses StorAge Selection functions (more details in Rodriguez and Klaus (2019)))</td>
<td>71% area with slope between 0-9% and 29% area with slope between 9 to 96% (Rodriguez and Klaus (2019))</td>
<td>Sub-daily sampling for stable water isotopes and bi-weekly for tritium</td>
<td>Temperate and semi-oceanic</td>
</tr>
</tbody>
</table>