

Aggregation effects on tritium-based mean transit times and young water fractions in spatially heterogeneous catchments and groundwater systems

Michael K. Stewart¹, Uwe Morgenstern², Maksym A. Gusyev^{3,4}, Piotr Maloszewski⁵

5 ¹Aquifer Dynamics & GNS Science, P.O. Box 30368, Lower Hutt, 5040, New Zealand

²GNS Science, Tritium & Water Dating Laboratory, Avalon, Lower Hutt, 5040, New Zealand

³International Centre for Water Hazard and Risk Management (ICHARM), Public Works Research Institute (PWRI), Tsukuba, Japan

⁴National Graduate Institute for Policy Studies (GRIPS), Tokyo, Japan

10 ⁵AGH University of Science and Technology Cracow, Department of Hydrogeology and Engineering Geology, Al. Mickiewicza 30, 30-059 Cracow, Poland

Correspondence to: Michael K. Stewart (m.stewart@gns.cri.nz)

Abstract. Kirchner (2016a) demonstrated that aggregation errors due to spatial heterogeneity, represented by two homogeneous subcatchments, could cause severe underestimation of the mean transit times (MTTs) of water travelling through catchments when simple lumped parameter models were applied to interpret seasonal tracer cycle data. Here we examine the effects of such errors on the MTTs and young water fractions estimated using tritium concentrations in two-part hydrological systems. We find that MTTs derived from tritium concentrations in streamflow are just as susceptible to aggregation bias as those from seasonal tracer cycles. Likewise, groundwater wells or springs fed by two or more water sources with different MTTs will also have aggregation bias. However, the transit times over which the biases are manifested are different because the two methods are applicable over different time ranges, up to 5 years for seasonal tracer cycles and up to 200 years for tritium concentrations. Our virtual experiments with two water components show that the aggregation errors are larger when the MTT differences between the components are larger and the amounts of the components are close to 50% of the mixture. We also find that young water fractions derived from tritium (based on a young water threshold of 18 years) are almost immune to aggregation errors as were those derived from seasonal tracer cycles with a threshold of about two months.

25 **1 Introduction**

Environmental tracers are commonly used to obtain transit time distributions (TTDs) in groundwater systems (Maloszewski and Zuber, 1982) or catchments (McDonnell et al., 2010). Transit time is the time it takes for rainfall to travel through a system from recharge to emergence in a well, spring or stream. TTDs provide important information about transport, mixing and storage of water in systems and therefore on the retention and release of pollutants. In addition, mean transit times (MTTs) determined from these distributions provide practical information for various aspects of water resources management. For example, MTTs have been used to estimate the volume of groundwater storage providing baseflow in catchments (Morgenstern

et al., 2010; Gusyev et al., 2016) and to predict lag times and life expectancies of contaminants in the subsurface (Hrachowitz et al., 2016). The drinking water securities of wells in New Zealand are partly assessed by an absence of water with less than one-year travel time by the NZ drinking water quality standard (Ministry of Health, 2008). As useful as they are, TTDs cannot be measured directly in the field and have to be inferred from age-dependent tracer concentrations with the use of lumped parameter models (LPMs).

Catchments are inherently heterogeneous on various scales. Point-scale properties vary greatly from place to place, while streams integrate the various catchment outputs. The top-down approach uses catchment outputs, such as streamflow and stream chemistry, to infer or predict catchment TTDs. The hope is that these average out local heterogeneities allowing one simple LPM to provide a good fit and its parameters to be representative of the catchment. But individual areas within catchments can vary greatly because of geology, geography, aspect, etc. Groundwater systems also show heterogeneity. Kirchner (2016a) showed by means of virtual experiments that aggregating subcatchments with different TTDs can lead to severe underestimation of the composite MTT when simple LPMs were applied to interpret seasonal tracer cycles. This is because the smoothing out of the seasonal cycles is a non-linear process which acts more rapidly on the younger water components thereby causing underestimation of the composite MTT. He also found that the young water fraction was a much more robust metric than the MTT against aggregation error. These results raise an important question: Are tritium-derived MTTs also susceptible to aggregation error due to spatial heterogeneity? This work aims to answer this question.

Seasonal tracer cycle and tritium-based MTTs are determined by different methods and have given very different results in catchments. The seasonal tracer cycle method depends on damping of input cycles on passing through a system into the output, whereas the tritium method depends on radioactive decay of tritium between input and output (with half-life of 12.32 yr). Effects of mixing within systems need to be accounted for in both cases (Maloszewski and Zuber, 1982). Results from seasonal tracer cycles have given MTTs up to about 5 years at which point the input cycles in homogeneous systems are completely damped within tracer measurement errors, while results from tritium measurements show that large proportions of the flow in many streams have MTTs of one to two decades or more (Stewart et al., 2010; Seegar and Weiler, 2014; Michel et al., 2015). Aggregation errors due to the non-linearity of the damping of the seasonal tracer cycles in time (noted above) add to this loss of signal in seasonal tracer cycles, thereby increasing the underestimation of the real MTTs in streams. Similarly, radioactive decay of tritium is a non-linear process and therefore spatial aggregation errors are expected when water components with different MTTs are combined (Bethke and Johnson, 2008).

Calibration of LPMs using environmental radioisotope and stable isotope data has been the subject of study for many years (see Maloszewski and Zuber (1982) and early work summarised therein). If a catchment outflow is a mixture of two or more components of different water ages, it can be difficult to calibrate an LPM uniquely when we only have data for tracers. For example, for springs in Czatkowice, Poland, only when the proportion in which the water components (water fluxes) were mixed was known could the unique answer based on tritium measurements be found (Grabczak et al., 1984; Maloszewski and Zuber, 1993). In heterogeneous catchments, it is always helpful to (i) measure a variable tracer periodically, and (ii) to combine

those data with water fluxes in the inputs and outputs to separate “fast” and “slow” components; see for example studies at Lainbach Valley, Germany (Maloszewski et al., 1983), and Schneealpe, Austria (Maloszewski et al., 2002). The choice of LPM, or equivalently the TTD function, must be based more on the hydrogeological situation and not on artificial mathematical (fitting) considerations. Consideration of hydrological parameters known independently (e.g. mean thickness of the water bearing layers in the catchment) is required for model validation in order to examine if the model is likely to be applicable to the real situation. We can have a very well-calibrated model in terms of tracer data being fitted by an LPM, but the MTT can be far from the hydrological reality.

The aim of this paper is to examine the aggregation effects of spatially heterogeneous catchments and groundwater systems on MTTs and young water fractions determined using tritium concentrations. We conducted our investigation by combining two dissimilar water components in virtual experiments and comparing the true mixed MTTs with the tritium-inferred apparent MTTs, as Kirchner (2016a) did with seasonal tracer cycles. Our experiments did not include examination of non-stationary hydrological systems, in which Kirchner (2016b) had found similar underestimation of MTTs with seasonal tracer cycles. We also examined aggregation effects for young water fractions estimated using tritium. Our calculations are based on the gamma LPM with shape factors (α) between 1 and 10, which is also representative of other frequently-used simple LPMs such as the exponential, exponential piston flow and dispersion models. The different tritium input functions for Northern and Southern Hemisphere locations were also tested.

2 Methods

2.1 Transit time determination: Simple and compound lumped parameter models

The varied flow paths of water through the subsurface of catchments imply that outflows contain mixtures of water with different transit times. That is, the water in the stream does not have a discrete transit time, but has a distribution of transit times (TTD). This distribution is often described by a conceptual flow or mixing model, which reflects the average (steady-state) conditions in the catchment or groundwater system.

Rainfall incident on a catchment is affected by immediate surface/near surface runoff and longer-term evapotranspiration loss. The remainder constitutes recharge to the subsurface water stores. Tracer inputs to the subsurface water stores (i.e. seasonal tracer cycles and tritium concentrations in the recharge water) are modified during passage through the hydrological system by mixing of water with different transit times (represented by the flow model) and radioactive decay in the case of tritium before appearing in the output. The convolution integral and an appropriate flow model are used to relate the tracer input and output. The convolution integral is given by

$$C_{out}(t) = \int_0^{\infty} C_{in}(t - \tau)h(\tau) \exp(-\lambda\tau) d\tau \quad (1)$$

where C_{in} and C_{out} are the input and output concentrations in the recharge and baseflow respectively. t is calendar time and the integration is carried out over the transit times τ . $h(\tau)$ is the flow model of the hydrological system based on the distribution

of water fluxes in the catchment. The exponential term accounts for radioactive decay of tritium. λ is the tritium decay constant ($= \ln 2/T_{1/2}$), where $T_{1/2}$ is the half-life of tritium (12.32 years).

Tritium concentrations in precipitation were different in each hemisphere, and are proxies for tritium recharge concentrations (C_{in}). Input functions (tritium concentrations in monthly samples of precipitation) at Kaitoke, New Zealand in the Southern Hemisphere (Morgenstern and Taylor, 2009) and Trier, Germany in the Northern Hemisphere (IAEA/WMO, 2016) are given in Fig. 1. Tritium data for Trier before 1978 was calculated by regression from data for Vienna, Austria. Both curves have pronounced bomb peaks due to nuclear weapons testing mainly in the Northern Hemisphere during the 1950s and 1960s, but the peak was much larger in the Northern Hemisphere than in the Southern Hemisphere. Since then there have been steady declines due to leakage of tritium from the stratosphere into the troposphere followed by removal by rainout and radioactive decay. However, the tritium concentrations in the troposphere are now reaching the background cosmogenic levels which they had before the dawn of the nuclear age (conventionally taken as 1950). The levelling-out process occurred about 20 years ago in the Southern Hemisphere and 5-10 years ago in the Northern Hemisphere. The bomb peaks have been good markers of 1960s precipitation in past tritium studies, but the steady declines which mimic radioactive decay of tritium have caused problems with ambiguous (i.e. multiple) age estimations for given tritium values (Stewart et al., 2010).

The curves also show smaller variations due to annual peaks in tritium concentrations caused by increased stratospheric leakage during spring in each hemisphere, and possibly small longer-term variations related to sunspot cycles. Tritium concentrations are expected to remain at the present cosmogenic levels for the foreseeable future, and this means that multiple age solutions are becoming less of a problem (Stewart et al., 2012; Stewart and Morgenstern, 2016; Gusyev et al., 2016). However, the minimal variation will mean that tritium will not be very effective for identifying flow models in the future.

Several simple flow models are commonly used in tracer studies. The piston flow model (PFM) describes systems in which all of the water in the output has the same transit time (MTT or τ_m). Its TTD is

$$h(\tau) = \delta(\tau - \tau_m) \quad (2)$$

where the single parameter is τ_m [yr], and $\delta(\tau - \tau_m)$ is a δ -function that gives a spike when $\tau = \tau_m$ (see Fig 2a), The output tritium concentration is

$$C_{out}(t) = C_{in}(t - \tau_m)\exp(-\lambda\tau_m) \quad (3)$$

and the output concentration equals the input concentration delayed in time by τ_m and for tritium decayed by radioactive decay during the delay.

The exponential model (EM) is given by

$$h(\tau) = \frac{1}{\tau_m} \exp\left(-\frac{\tau}{\tau_m}\right) \quad (4)$$

where again the single parameter is τ_m [yr]. In this model, water parcels with different transit times combine in the outflow to approximate the exponential TTD. It is mathematically equivalent to the well-mixed model (also called the linear reservoir), but it does not imply that full mixing occurs within real systems.

The gamma model (GM) has TTDs based on the gamma distribution

$$5 \quad h(\tau) = \frac{\tau^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\tau/\beta} \quad (5)$$

where the two parameters α [-] and β [yr⁻¹] are shape and scale factors respectively, and $\tau_m = \alpha\beta$ (Kirchner et al., 2000). The gamma distribution reduces to the exponential distribution for the special case of $\alpha = 1$.

The exponential piston flow model (EPM) combines a volume with exponential transit times followed by a piston flow volume to give a model with two parameters (Maloszewski and Zuber, 1982). The TTD is given by

$$10 \quad h(\tau) = 0 \quad \text{for } \tau < \tau_m(1-f) \quad (6a)$$

$$h(\tau) = \frac{1}{f\tau_m} \exp\left(-\frac{\tau}{f\tau_m} + \frac{1}{f} - 1\right) \quad \text{for } \tau \geq \tau_m(1-f) \quad (6b)$$

where f is the ratio of the exponential volume to the total volume. Maloszewski and Zuber (1982) used the parameter η instead of f , where $\eta = 1/f$. $f\tau_m$ is the time required for water to flow through the exponential volume, while $\tau_m(1-f)$ is the time in the piston flow section,

15 The dispersion model (DM) assumes a tracer transport which is controlled by advection and dispersion processes (Maloszewski and Zuber, 1982), with a TTD of

$$h(\tau) = \frac{1}{\tau\sqrt{4\pi(P_D)\tau/\tau_m}} \exp\left[-\frac{(1-\frac{\tau}{\tau_m})^2}{4(P_D)\tau/\tau_m}\right] \quad (7)$$

where P_D [-] is the dispersion parameter (being the measure of the variance of the transit time distribution, i.e. the sum of the variance resulting from the space distribution of the infiltration through the catchment surface and variance resulting from the dispersive flow through the underground). The two parameters are τ_m and P_D .

This paper makes a particular distinction between *simple LPMs* (meaning specifically the gamma model, the exponential piston flow model with end members piston flow and exponential models, and the dispersion model) and *compound LPMs* (binary or other parallel combinations of simple LPMs). Simple LPMs describe homogeneous systems, while compound LPMs can accommodate heterogeneity in the system.

25 Compound LPMs have generally only been explored for more complicated systems or when simple LPMs have given poor fits to data (such as seasonal tracer cycles or tritium concentrations) (e.g. Maloszewski et al., 1993; Stewart and Thomas, 2008; Blavoux et al., 2013; Morgenstern et al., 2015). The binary parallel LPM is given by

$$LPM = bLPM_1 + (1 - b)LPM_2 \quad (8)$$

where LPM_1 and LPM_2 are simple LPMs with individual PDFs representing two water components contributing to the system output, and $b[-]$ is the fraction of the first component in the combined output. The overall combined MTT (τ_m) is

$$\tau_m = b\tau_1 + (1 - b)\tau_2 \quad (9)$$

- 5 An example of a compound LPM is the parallel combination of two exponential models describing a system with young and old water components. This is called the ‘double exponential model’ when applied to tritium (Michel, 1992; Taylor et al., 1992) and the ‘two parallel linear reservoirs’ (TPLR) model when applied to seasonal tracer cycles (Weiler et al., 2003). The PDF is given by

$$h(\tau) = \frac{b}{\tau_f} \exp\left(-\frac{\tau}{\tau_f}\right) + \frac{(1-b)}{\tau_s} \exp\left(-\frac{\tau}{\tau_s}\right) \quad (10)$$

- 10 τ_f and τ_s are the MTTs of the fast and slow reservoirs respectively. The model has three parameters with the overall combined MTT (τ_m) being

$$\tau_m = b\tau_f + (1 - b)\tau_s \quad (11)$$

- Other compound LPMs referred to in this work are the double gamma model (DGM), double exponential piston flow model (DEPM) and the double dispersion model (DDM), which are binary parallel combinations of the respective models. They each
15 have five parameters.

2.2 Estimation of spatial aggregation effects on mean transit times

- To estimate the effects of spatial aggregation on mean transit times (MTTs), we perform virtual experiments by combining two homogeneous subsystems. Each subsystem or water component is described by a simple LPM (a gamma model with assumed parameters α and β). The combined or mixed system is then describable by a compound LPM (Eq. 8), which yields
20 the “true” MTT via Eq. 9 using the assumed MTTs of the components.

- To determine the “apparent” MTT, the tritium concentrations of the water components from 1940 to the present are calculated from the gamma models applying to each component using the convolution process described above (Eq. 1). The input function was first assumed to be constant at 2 TU for the calculations given in Section 3.1.1, then the Kaitoke or Trier input functions (Fig. 1) were used for the calculations in Section 3.1.2 and 3.1.3. In all cases, the tritium concentrations of the mixed system
25 (C_m) are given by

$$C_m = bC_1 + (1 - b)C_2 \quad (12)$$

where C_1 [TU] and C_2 [TU] are the tritium concentrations in components 1 and 2 respectively. The mixed system is then treated as if it is homogeneous to produce the “apparent” MTT by fitting a simple LPM (a gamma model) to the tritium concentrations of the mixture (C_m). The true and apparent MTTs of the mixture are compared for different assumed values of the MTTs of

the components. b is assumed to be 0.5 for simplicity in what follows. Following Kirchner (2016a), we did not consider evapotranspiration in our analysis of tritium aggregation effects.

2.3 Determination of young water fractions

The young water fraction (Y_f) is the fraction of water with transit times between zero and a young water threshold (t_y), i.e.

$$5 \quad Y_f = \int_0^{t_y} h(\tau) \cdot d\tau \quad (13)$$

The young water threshold for tritium was estimated by trial and error using the gamma model with parameter α in the range 1 to 10. It was found that a constant threshold value of 18 years gave agreement between the apparent and true young water fractions to within about 10%. This included the case with the greatest difference in ages between the two water components (i.e. waters with MTTs of 3 and 397 years respectively in this study). Accordingly, the young water threshold has been taken

15 as 18 years in what follows. The "true" Y_f is determined by mixing the two waters according to the equation

$$Y_{ftrue} = bY_{f1} + (1 - b)Y_{f2} \quad (14)$$

in analogy with Eq. (9). b is the fraction of component 1 in the mixture, and Y_{f1} are Y_{f2} are the young water fractions of the two components. The "apparent" Y_f is determined by fitting a simple LPM to the tritium concentrations of the mixture (Eq. 12). b is assumed to be 0.5.

15 2.4 Comparison of transit time distributions of different flow models

The transit time distributions of the three cases of the gamma model investigated in this work are illustrated in Fig. 2a, as normalised probability density functions (i.e. $h(\tau) \times \tau_m$) versus normalised transit times (τ/τ_m). These cover the range of shapes observed in streams and groundwater using tritium concentrations. They are also approximately representative of the other simple flow models described above. The gamma model case with $\alpha=1$ is the exponential distribution (linear storage); the same as the exponential piston flow model with $f = 1$. Gamma model cases with $\alpha=3$ and 10 are more peaked and have smaller tails (short and long transit times are reduced compared to transit times close to the mean). The piston flow model is the end member of the series, being all peak and no tail (see Fig. 2a).

20 The other simple flow models are compared with the gamma models in Table 1 and Figs. 2b-c. The standard deviation (sd) and Nash Sutcliffe efficiency (NSE) are used to quantify the goodness-of-fit between the gamma model (GM_i) and the best-fitting version of each of the other models (LPM_i), where

$$sd = \sqrt{\frac{\sum_1^N (GM_i - LPM_i)^2}{N}} \quad (15)$$

$$\text{and } NSE = 1 - \frac{\sum_1^N (GM_i - LPM_i)^2}{\sum_1^N (LPM_i - \bar{LPM}_i)^2} \quad (16)$$

The NSE efficiency can vary between $-\infty$ and 1. $NSE = 1$ indicates a perfect fit between the gamma model and the other model, while $NSE = 0$ means that the variation between the models is the same as the variation about the mean of the other model. The standard deviation and Nash Sutcliffe efficiency gave the same results in terms of identifying the most similar shapes of the gamma, exponential piston flow and dispersion models (Table 1). TTD shapes for the gamma model with α between 1 and 10 are equivalent to exponential piston flow model shapes with exponential fractions (f) between 1.0 and 0.44 (Table 1), which have been found suitable for interpreting tritium concentrations in baseflow and groundwater (e.g. Maloszewski et al., 1983; Stewart et al., 2007; Morgenstern and Stewart, 2004). The useful range of the dispersion model has dispersion parameters (P_D) between about 1.3 and 0.05 corresponding to the gamma model with α between 1 and 10 (Table 1). The gamma and exponential piston flow model shapes become less similar to each other as α increases to 10, while the gamma and dispersion model shapes become more similar.

3 Results

3.1 Aggregation effects on mean transit times determined using tritium

3.1.1 Relationships between mean transit time and tritium concentration

We first demonstrate the relationships between mean transit time and tritium concentration for mixed systems (Fig. 3) by assuming constant annual input tritium concentration of 2 TU over time, i.e. without the bomb pulse during the nuclear age and only natural background concentrations are present. This simplifying assumption is necessary to allow for the analysis shown in Fig. 3; with the real peaked input the figures would be much more complicated. The assumption of a constant tritium input function is however becoming increasingly realistic in the Southern Hemisphere, with the bomb tritium from 50 years ago now fading away and assuming no more large-scale releases of tritium to the atmosphere. This assumption is not limited to tritium but would also be valid for all radioactive tracers with constant input such as carbon-14 and argon-39.

Fig. 3a shows the relationship for the gamma model with shape factor $\alpha = 1$. The red points indicate the assumed water components (with MTTs of 3 and 197 years respectively) and the red dashed line is the mixing relationship between them described by Eqs. 9 and 12. The “true” MTT (100 years) of a 50:50 mixture of the components (i.e. $b = 0.5$) is shown on the red dashed line. The black curve is the result of applying the gamma model with $\alpha = 1$ to the mixed tritium concentrations (Eq. 12). A 50:50 mixture of the components gives the “apparent” MTT shown (20.5 years), which is much less than the “true” MTT. This results from the strongly non-linear character of the black curve (Fig. 3a) and therefore combining two dissimilar subsystems causes aggregation bias in a similar way to that demonstrated for seasonal tracer cycles by Kirchner (2016a) in his Fig. 5 (and also for radioactive decay by Bethke and Johnson (2008) in their Fig. 3a).

Figs. 3b-d show the same calculations applied to the gamma models with $\alpha = 3$ and 10 and the piston flow model (PFM). The different shape factors describe different fractional contributions of past water inputs to the present water output as illustrated by the transit time distributions in Fig. 2a. The gamma models with $\alpha = 3$ and 10 have slightly greater differences between the true and apparent MTTs than the gamma model with $\alpha = 1$. The piston flow model is the most sharply peaked of all, and has

the greatest true-apparent MTT difference of 100 to 15 years. Since there is no mixing, the non-linearity of the black curve is solely due to radioactive decay of tritium (Fig. 3d).

3.1.2 Effect of young component fraction (b) on aggregation

Fig. 4 shows the effect of changing the fraction of the young component (b) on the aggregation error for a mixture of two components with MTTs of 10 and 90 years. As b increases from zero, the aggregation effect increases from zero reaching a maximum near $b = 0.5$ and then decreasing to zero again at $b = 1$. This is an important factor in the aggregation error. The virtual experiments below were carried out with $b = 0.5$ to show the maximum effects.

3.1.3 True versus apparent mean transit times

The true versus the apparent MTTs calculated using the real tritium input function from Kaitoke (expressed as annual values) are given in Fig. 5. The calculations were structured in that the two water components were initially assumed to have the same MTTs (i.e. $\tau_1 = \tau_2$) and therefore the mixture had the same true and apparent MTTs and plotted on the 1:1 lines. The second component (MTT2) was then allowed to become older in 50 year steps so that the difference in MTTs between the two components increased. This caused the apparent MTTs to become younger than the true MTTs and the points to move further and further away from the 1:1 line as shown by the curves in Fig. 5. The dots show the effects of the step changes in MTT2. As expected, the greatest age differences caused the biggest deviations from the 1:1 lines.

The different values of α cause differences to the patterns observed, but the patterns are similar overall. They are tighter around the 1:1 line for $\alpha = 1$ showing smaller aggregation effects, and are most divergent for $\alpha = 10$. Errors of fitting for determining the apparent MTTs (expressed as standard deviations (Eq. 15)) are greatest when component 1 is youngest, these are shown by fine dashed lines above and below the curves. The errors are largest with $\alpha = 10$. The fitting errors are important because big errors would lead researchers to apply more complicated and therefore more realistic LPMs (such as binary LPMs), as many have in the past (e.g. Maloszewski et al., 1983; Uhlenbrook et al., 2002; Stewart and Thomas, 2008; Morgenstern et al., 2015).

Using the Trier (Northern Hemisphere) tritium input function (Fig. 1) results in very similar aggregation biases for tritium MTTs (Fig. 6) compared to those obtained with the Kaitoke input (Fig. 5). Using Northern or Southern Hemisphere tritium input functions makes only slight differences to the curves. Note that the problem of multiple age solutions often experienced using tritium with the Northern Hemispheric input function (e.g. Stewart et al., 2012) does not arise here because we calculate around 75 tritium values (one for each year) and this constrains the final 'apparent' fitting to a single unique solution. However, the fitting errors for the apparent MTTs with the Trier input function are much larger than those determined with the Kaitoke input function.

Some of the calculation results are replotted in Fig. 7 to compare results for the Northern and Southern Hemispheres. This figure shows the possible aggregation error (expressed as percentage difference of the apparent from the true MTTs) versus

the MTT of component 2 (MTT2) for the gamma model with $\alpha = 1$. The solid curves show results for MTT1 = 10 years, these are restatements of the curves in Figs. 5a and 6a for $\alpha = 1$ and MTT1 = 10 years. Aggregation errors are 8% for Southern Hemisphere and 15% for Northern Hemisphere locations by the young water threshold (18 years).

3.2 Aggregation effects on young water fractions

5 The effect of combining two different water components on the true and apparent young water fractions (Y_f) of a mixture are examined in this section using the same procedure as before (i.e. testing mixtures with MTT1 at 10 years, 25 years, etc.). The two water components were initially assumed to have the same MTTs and young water fractions (i.e. $Y_{f1} = Y_{f2}$) and therefore the mixture had the same true and apparent young water fractions and plotted on the 1:1 lines in Figs. 8 and 9. The second component (MTT2) was then allowed to become older in 50-year steps so that the differences in MTTs and young water
10 fractions between the two components increased. But now the true and apparent young water fractions did not diverge very much from each other (Figs. 8 and 9). The figures show the young water fractions decreasing as the mixtures become older, but the curves lie mostly along the 1:1 lines. There are only small divergences from an apparent to true young fraction ratio of one (up to about 10%). The maximum divergences from this ratio are affected by the choice of young water threshold (Eq. 13). The present calculations have been made using a young water threshold of 18 years. With higher values for the threshold,
15 the maximum divergences from the 1:1 line were found to become larger. Consequently, 18 years is taken as the recommended value for the young water threshold.

For stable isotopes, Kirchner (2016a) reported a young water threshold range from 0.1 to 0.25 years (or approximately two months) for the gamma model shape factor α ranging from 0.2 to 2. From our tritium evaluation with MTT1 at 10 years, the young water threshold of tritium-based transit times was 18 years for all values of the shape factor α between 1 and 10.

20 Young water fractions evaluated using tritium are of practical interest for various threshold ages, for example one year for assessing drinking water security of groundwater wells (water mixtures without any fraction of water of less than one year are regarded as secure in terms of potential for pathogen contamination (Close et al., 2000; Ministry of Health, 2008)), or 60 years to assess the fraction of water that has already been impacted by high-intensity industrial agriculture starting after WWII (e.g. Morgenstern et al., 2015).

25 3.3 Aggregation effects on MTTs for seasonal tracer cycles

Aggregation effects for seasonal tracer cycles have been determined by the methods of Kirchner (2016a) for comparison with the tritium effects. The rainfall input variation has been approximated as a sine wave with a one-year period to imitate the seasonal tracer cycle, and the sine wave has been traced through the convolution using the gamma distribution. Fig. 10 shows the aggregation effects for the gamma model with $\alpha = 1$. The pattern is very like those observed using tritium concentrations
30 (Fig. 5), so it is clear that the effects are effectively the same whether seasonal tracer cycles or radioactive isotopes are being used. Although our methodology was the same as Kirchner's in that two components were combined, we followed the process of starting with the same MTTs and then allowing the second component to become older. For this reason, the results show

the dependence of the aggregation error on the difference in MTTs more explicitly than the random sampling of non-similar MTT components method of Kirchner.

4 Discussion

4.1 Implications of tritium MTT aggregation bias

5 The analysis of Sections 3.1 and 3.2 has shown that tritium-derived MTTs are just as susceptible to aggregation bias as seasonal tracer cycles when flows from dissimilar parts of catchments are combined using simple LPMs. Likewise, groundwater wells or springs fed by two or more water sources with different MTTs will also show aggregation bias. However, the transit times over which the biases are manifested are different because the two methods are applicable to different time ranges, up to 5 years for seasonal tracer cycles and up to 200 years for tritium concentrations (based on appropriate mixing models). Note particularly that the bias *not only* applies to samples at the limits of the methods (i.e. with very small tracer cycles or near-zero tritium concentrations), *but also* applies to MTTs far below these limits.

The calculations have been made for extreme cases to highlight the aggregation bias. Firstly, the heterogeneity is assumed to be represented by just two homogeneous but different areas of hydrological systems. This is the worst type of heterogeneity for aggregation bias. Secondly, the water components from these areas are assumed to combine in the proportions of 1:1 in the outlet. This causes close to the maximum aggregation bias for a given pair of waters, which ranges from zero for $b = 0$ or 1 to a maximum around $b = 0.5$ (Fig. 4). Obviously there is no aggregation bias when the MTTs of the two components are the same, and the bias increases as the difference between the MTTs increases (Figs. 5 and 6). The bigger the span of MTTs between the two components, the bigger the aggregation error. When the old component is so old that it has essentially no tritium and could have any age (hundreds, thousands or even millions of years), the aggregation error could be very large. This is a well-recognised problem with the use of many radioactive isotopes and chemicals for dating water (e.g. Cook and Böhlke, 2012; Stewart, 2012). The analogous problem with seasonal tracer cycles is when the old component is too old to have any seasonal variation at all and the age is effectively truncated at around 5 years (Stewart et al., 2010).

4.2 How can aggregation error be detected in tritium-based MTTs?

Both simple and compound LPMs can be free of aggregation error or conversely be affected by aggregation error depending on whether or not they capture the nature of the heterogeneity in the catchment or groundwater system relevant to the error. Simple LPMs have fewer parameters, but have no ability to capture heterogeneity because of their underlying perceptual model (i.e. the assumption of homogeneity), and therefore would be expected to underestimate MTTs because of aggregation error if there is heterogeneity producing flows with different MTTs in the system. (However, note that a highly skewed simple LPM (in that case a gamma model with $\alpha = 0.24$) was able to mimic the MTT of a specific case of binary flow in Sect. 3.1.1 (see Reviewer's Report #1). This did not mean that the model gave a good representation of the TTD of the binary system. The dispersion model has the same ability to be highly skewed if the dispersion parameter (DP) is allowed to become very much larger than normal.) Compound LPMs have more parameters and therefore more flexibility to capture heterogeneity, but the

model structure must be based on the underlying perceptual model or the parameters extracted could misrepresent the system and lead to aggregation error. Calibration of the parameters of compound LPMs requires that there be sufficient quantity and quality of data and with enough variation to enable retrieval of the increased number of parameter values with compound LPMs. In the past, the bomb tritium pulse introduced strong variation in tritium concentrations of precipitation, but variation in the future will be very much less because of the passing of the bomb pulse from the atmosphere.

We therefore suggest that the answer to the question in the title of this section may be what has often been practised in the past, even though the term “aggregation error” was not used (e.g. Maloszewski et al., 1993; Uhlenbrook et al., 2002; Stewart and Thomas, 2008; Morgenstern and Taylor, 2009; Stewart, 2010; Blavoux et al., 2013; Morgenstern et al., 2015). This ideally involves evaluation of many types of information about a hydrological system (geological, hydrological, hydrochemical, tritium and other isotopes) to establish a perceptual model, and experiments with simple and compound LPMs in harmony with the derived perceptual model to fit tritium data (and if available other types of chemical or isotopic data). Compound LPMs in harmony with the perceptual model would be expected to yield MTTs with less aggregation error than simple LPMs, because the former have the ability to separate young and old water components while the latter do not. Comparison of MTTs from simple and compound models should then show if there is much aggregation error. Parameters yielded by best-fitting models have been used in the past, but may not be the most appropriate ones if the parameters are to be used in other contexts. There is also risk of missing less apparent (alternative) parameter solutions if there are any elsewhere in the parameter space. Gallart et al. (2016) applied a GLUE-based uncertainty assessment method which used Monte Carlo searching of the parameter space of the exponential piston flow model to estimate MTTs from tritium. This allowed the uncertainties of the parameters to be quantified.

20 **5 Summary**

MTT estimations based on tritium concentrations show very similar aggregation effects to those for seasonal tracer variations. Our virtual experiments with two water components show that the aggregation errors are largest when the MTT differences between the two components are largest and the amounts of the components subequal. We also find that young water fractions derived from tritium based on a young water threshold of 18 years are almost immune to aggregation errors as were those derived from seasonal tracer cycles with a threshold of about two months. We conclude with comments on the implications of aggregation bias on tritium MTTs and detection of aggregation errors in past studies.

References

Blavoux, B., Lachassagne, P., Henriot, A., Ladouche, B., Marc, V., Beley, J.-J., Nicoud, G., and Olive, P.: A fifty-year chronicle of tritium data for characterising the functioning of the Evian and Thonon (France) glacial aquifers, *J. Hydrol.*, 494, 116–133, doi.org/10.1016/j.jhydrol.2013.04.029, 2013.

- Bethke, C.M. and Johnson, T.M.: Groundwater age and groundwater age dating. *Ann. Review Earth and Planet. Sci.*, 36, 121–152, doi: 10.1146/annurev.earth.36.031207.124210, 2008.
- Close, M., Stewart, M.K., Rosen, M., Morgenstern, U., and Nokes, C.: Investigation into secure groundwater supplies, ESR Client Report FW0034, for the Ministry of Health. 30p. 2000.
- 5 Cook, P.G. and Böhlke, J-K.: Determining Timescales for Groundwater Flow and Solute Transport. Chapter 1 in Eds. P.G. Cook and A.L.[Herczeg](#). *Environmental Tracers in Subsurface Hydrology*. pp 1-30, 2012.
- Gallart, F., Roig-Planasdemunt, M., Stewart, M.K., Llorens, P., Morgenstern, U., Stichler, W., Pfister, P., and Latron, J.: A GLUE-based uncertainty assessment framework for tritium-inferred transit time estimations under baseflow conditions. *Hydrol. Process. (Keith Beven Tribute)*, 30, 4741–4760, DOI: 10.1002/hyp.10991, 2016.
- 10 Grabczak, J., Maloszewski, P., Rozanski, K., and Zuber, A.: Estimation of the tritium input function with the aid of stable isotopes, *Catena*, 11, 105-114, 1984.
- Gusyev, M.A., Morgenstern, U., Stewart, M.K., Yamazaki, Y., Kashiwaya, K., Nishihara, T., Kuribayashi, D., Sawano, H., and Iwami, Y.: Application of tritium in precipitation and baseflow in Japan: a case study of groundwater transit times and storage in Hokkaido watersheds. *Hydrology and Earth System Sciences* 20, 3043-3058. doi:10.5194/hess-20-15 3043-2016, 2016.
- Hrachowitz, M., Benettin, P., van Breukelen, B.M., Fovet, O., Howden, N.J.K., Ruiz, L., van der Velde, Y., and Wade, A.J.: Transit times—the link between hydrology and water quality at the catchment scale. *WIREs Water*, 3, 629–657. doi: 10.1002/wat2.1155, 2016.
- IAEA/WMO Global Network of Isotopes in Precipitation. The GNIP Database. Accessible at: <http://www.iaea.org/water>, 20 2016.
- Kirchner, J.W.: Aggregation in environmental systems – Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments, *Hydrol. Earth Syst. Sci.*, 20, 279-297, doi:10.5194/hess-20-279-2016, 2016a.
- Kirchner, J.W.: Aggregation in environmental systems – Part 2: Catchment mean transit times and young water fractions under 25 hydrologic nonstationarity. *Hydrol. Earth Syst. Sci.*, 20, 299-328, doi:10.5194/hess-20-299-2016, 2016b.
- Kirchner, J.W., Feng, X., and Neal, C.: Fractal stream chemistry and its implications for contaminant transport in catchments, *Nature*, 403, 524-527, doi:10.1038/35000537, 2000.
- 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.
- 30 Małoszewski, P., Stichler, W., Zuber, A., and Rank, D.: Identifying the flow systems in a karstic-fissured-porous aquifer, the Schneealpe, Austria, by modelling of environmental ^{18}O and ^3H isotopes, *J. Hydrol.*, 256, 48-59. 2002
- Małoszewski, P., and Zuber, A.: Determining the turnover time of groundwater systems with the aid of environmental tracers: 1. Models and their applicability, *J. Hydrol.*, 57, 207-231, doi:10.1016/0022-1694(82)90147-0, 1982.

- Małozzewski, P., and Zuber, A.: Principles and practice of calibration and validation of mathematical models for the interpretation of environmental tracer data in aquifers. *Adv. Wat. Resour.*, 16, 173-190, 1993.
- McDonnell, J.J., McGuire, K., Aggarwal, P., Beven, K., Biondi, D., Destouni, G., Dunn, S., James, A., Kirchner, J., Kraft, P., Lyon, S., Malozzewski, 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., Wörman, A., and Wrede, S.: How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis. *Hydrol. Process.*, 24, 1745-1754, 2010.
- Michel, R.L., Aggarwal, P., Araguas-Araguas, L., Kurttas, T., Newman, BD, and Vitvar, T.: A simplified approach to analysing historical and recent tritium data in surface waters. *Hydrol. Process.* 29, 572–578, DOI: 10.1002/hyp.10174, 2015.
- Michel, R.L.: Residence times in river basins as determined by analysis of long-term tritium records, *J. Hydrol.* 130, 367–378, 1992.
- Ministry of Health: Drinking-water Standards for New Zealand 2005 (revised 2008). Ministry of Health, PO Box 5013, Wellington, New Zealand, ISBN 978-0-478-31809-8, 2008.
- Morgenstern, U., Daughney, C. J., Leonard, G., Gordon, D., Donath, F. M., and Reeves, R.: Using groundwater age and hydrochemistry to understand sources and dynamics of nutrient contamination through the catchment into Lake Rotorua, New Zealand, *Hydrol. Earth Syst. Sci.*, 19, 803–822, 2015.
- Morgenstern, U., Stewart, M.K., and Stenger, R.: Dating of streamwater using tritium in a post nuclear bomb pulse world: continuous variation of mean transit time with streamflow. *Hydrol. Earth Syst. Sci.*, 14, 2289-2301, 2010.
- Morgenstern, U., and Stewart, M.K.: Stream water dating and future nutrient load to Lake Taupo, New Zealand,. In International Workshop on the Application of Isotope Techniques in Hydrological and Environmental Studies, UNESCO, Paris, 6–8 September. 2004.
- Morgenstern, U., and Taylor, C.B.: Ultra low-level tritium measurement using electrolytic enrichment and LSC, *Isotopes Environ. Health Stud.*, 45, 96–117. doi:10.1080/10256010902931194, 2009.
- Seeger, S. and Weiler, M.: Re-evaluation 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.
- Stewart, M.K.: A 40-year record of carbon-14 and tritium in the Christchurch groundwater system, New Zealand: Dating of young samples with carbon-14, *J. Hydrol.*, 430-431, 50-68, DOI: 10.1016/j.jhydrol.2012.01.046, 2012.
- Stewart, M.K., Mehlhorn, J., and Elliott, S.: Hydrometric and natural tracer (^{18}O , silica, ^3H and SF_6) evidence for a dominant groundwater contribution to Pukemanga Stream, New Zealand, *Hydrol. Process.*, 21, 3340–3356, DOI: 10.1002/hyp.6557, 2007.
- Stewart, M.K. and Morgenstern, U.: Importance of tritium-based transit times in hydrological systems, *WIREs Water* 3, 145–154, doi: 10.1002/wat2.1134, 2016.
- 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, 2010.

- Stewart, M.K., Morgenstern, U., McDonnell, J.J., and Pfister, L.: The 'hidden streamflow' challenge in catchment hydrology: A call to action for streamwater transit time analysis, *Hydrol. Process.*, 26, 2061-2066, doi: 10.1002/hyp.9262. 2012.
- 5 Stewart, M.K. and Thomas, J.T.: A conceptual model of flow to the Waikoropupu Springs, NW Nelson, New Zealand, based on hydrometric and tracer (^{18}O , Cl, ^3H and CFC) evidence, *Hydrol. Earth Syst. Sci.*, 12, 1-19, <http://www.hydrol-earth-syst-sci.net/12/1/2008/hess-12-1-2008.html>, 2008.
- Taylor, C.B., Brown, L.J., Cunliffe, J.J, and Davidson, P.W.: Environmental isotope and ^{18}O applied in a hydrological study of the Wairau Plain and its contributing mountain catchments, Marlborough, New Zealand, *J. Hydrol.* 138, 269–319, 1992.
- 10 Uhlenbrook, S., Frey, M., Liebundgut, C., Maloszewski, P.: Hydrograph separations in a mesoscale mountainous basin at event and seasonal timescales, *Water Resour. Res.*, 38, doi:10.1029/2001WR000938, 2002.
- Weiler, M., McGlynn, B. L., McGuire, K. J., and McDonnell, J. J.: How does rainfall become runoff? A combined tracer and runoff transfer function approach, *Water Resour. Res.*, 39, 1315, doi:10.1029/2003WR002331, 2003.

Table 1: Comparison of the shapes of the gamma (GM), exponential piston flow (EPM) and dispersion models (DM) transit time distributions. The shape parameters of the best-fitting versions of the other models and the goodnesses-of-fit (standard deviations (sd) and Nash-Sutcliffe efficiencies (NSE)) between them and the GM are given.

GM α	1.0	3.0	10.0
EPM f	1.00	0.74	0.44
sd	0.00	0.15	0.24
NSE	1.00	0.70	0.61
DM P_D	1.36	0.22	0.05
sd	0.12	0.05	0.03
NSE	0.78	0.97	0.99

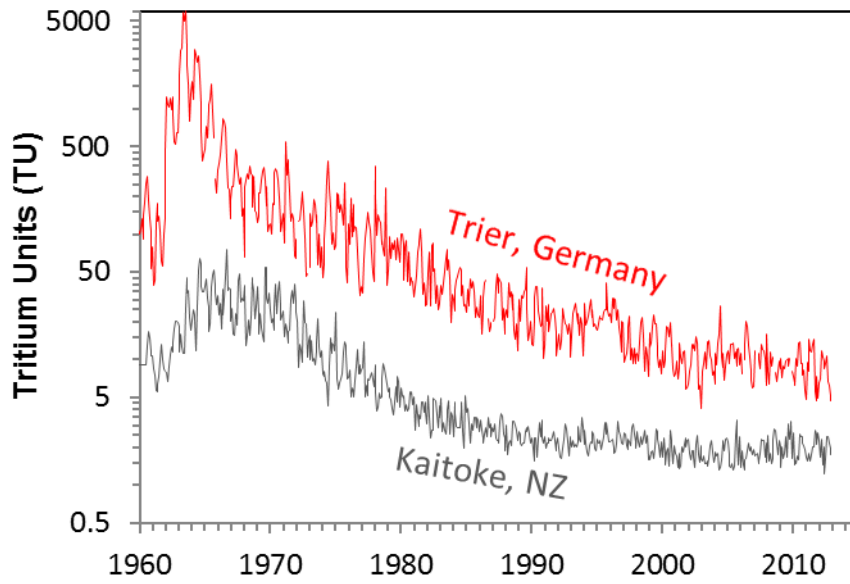
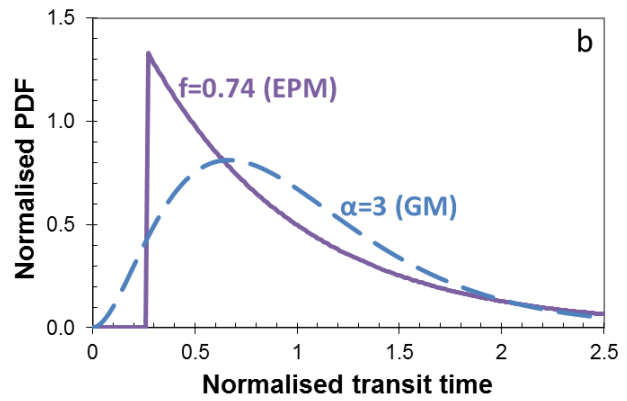
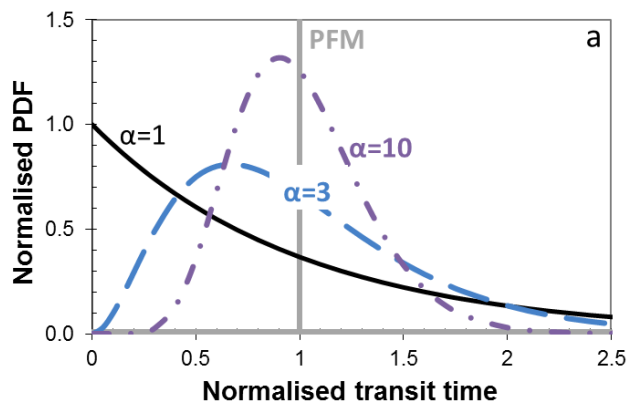


Figure 1. Tritium concentrations (TU) in monthly precipitation samples at Kaitoke, New Zealand in the Southern Hemisphere, and Trier, Germany in the Northern Hemisphere.



5

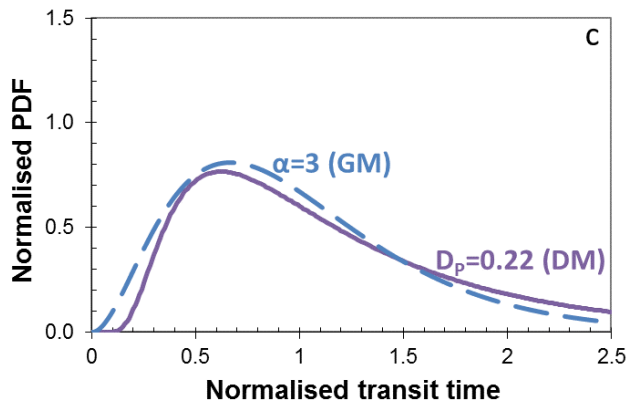


Figure 2. (a) Gamma distributions (GM) for shape factors α between 1 and 10. The axes show normalised transit time (τ/τ_m) and normalised probability density function PDF ($h(\tau) \times \tau_m$). Note that the distribution for GM ($\alpha = 1$) is the same as that for the exponential model (EM). (b-c) Comparison between the gamma model with $\alpha = 3$ and the best fitting exponential piston flow and dispersion models.

5

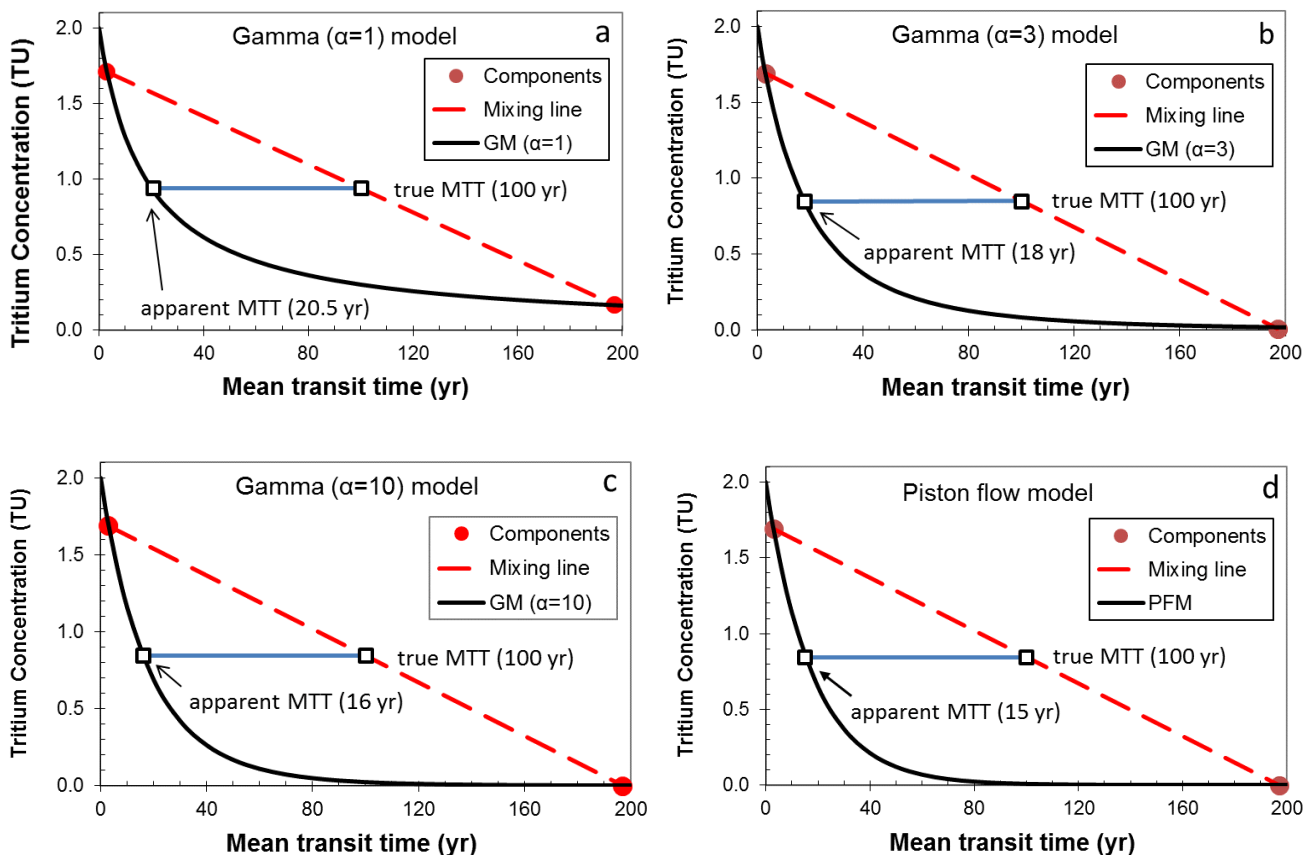
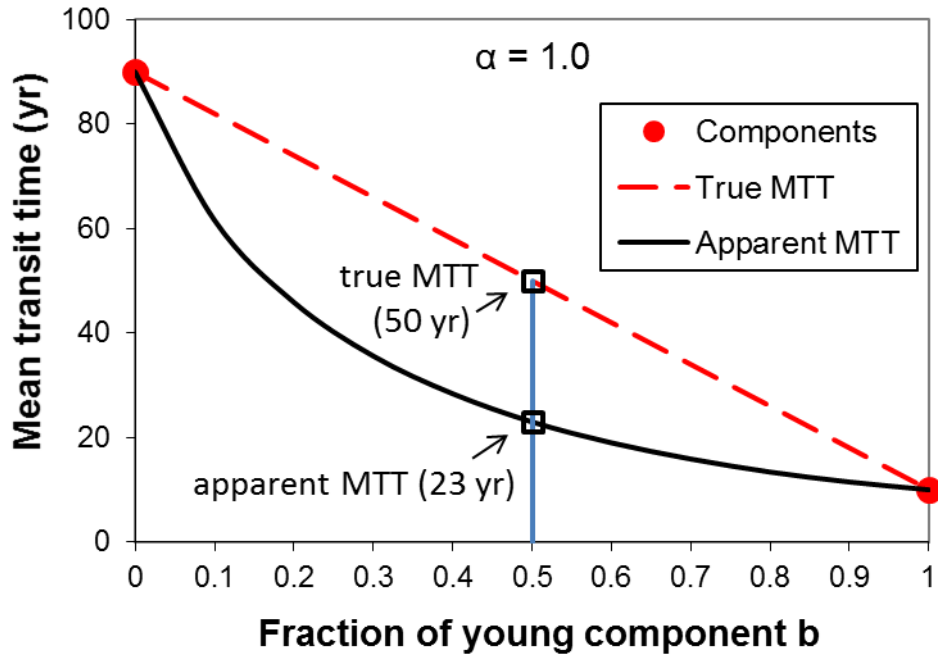
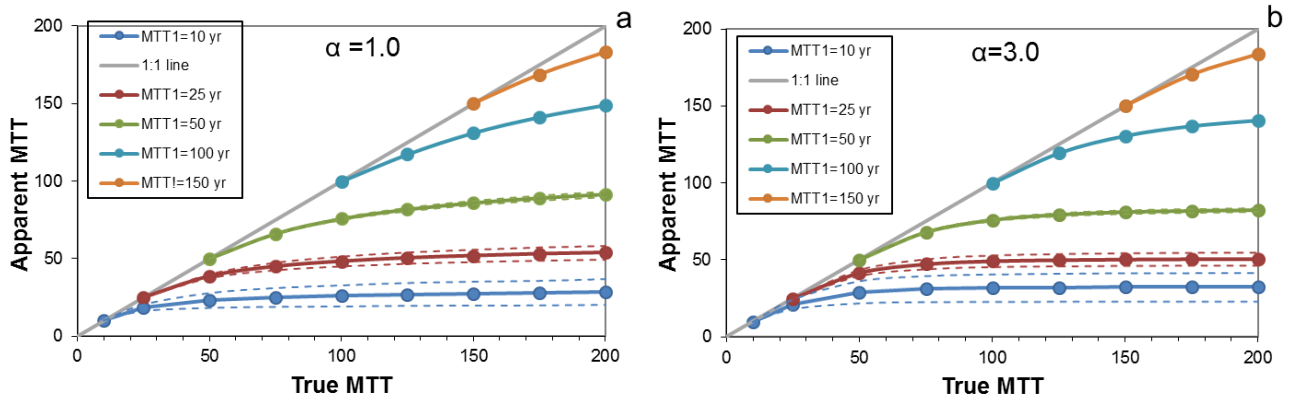


Figure 3(a-d). Aggregation errors when the tritium input concentration is assumed to be constant at 2 TU. Mean transit times (MTTs) are inferred from tritium concentrations in mixed runoff from two subcatchments with different tritium concentrations and

MTTs (shown by red dots) using a range of gamma and the piston flow models. The relationships between MTTs and tritium concentrations given by the simple models (black curves) are strongly non-linear causing marked differences between the true and apparent MTTs.



5 Figure 4. Effect of changing b (the young component fraction) on the aggregation error with the gamma model with $\alpha = 1$ for mixing of two components with MTTs of 10 and 90 years. The blue line shows the effects at $b = 0.5$.



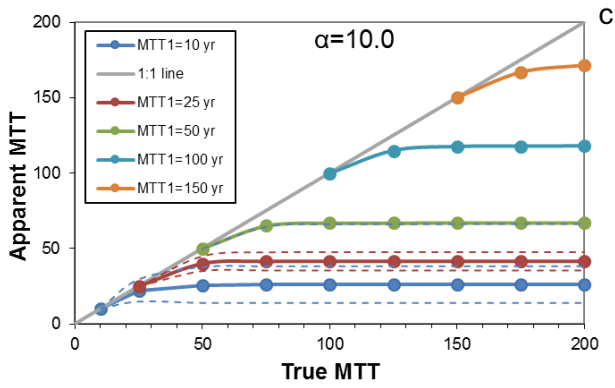


Figure 5(a-c): Aggregation effects for tritium MTTs for gamma models with different values of α using the Kaitoke input function. Curves show changes as component 2 (MTT2) becomes older in 50-year steps and therefore the mixtures older in 25-year steps (shown by dots). The first step of the MTT1=10 yr curve is 15 years. Fitting errors in the apparent MTTs are shown by fine dashed lines.

5

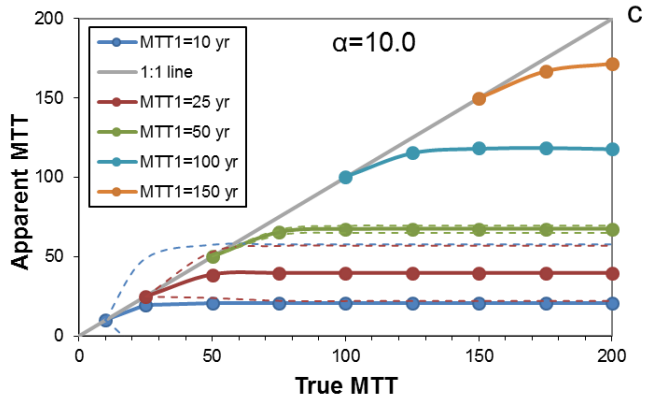
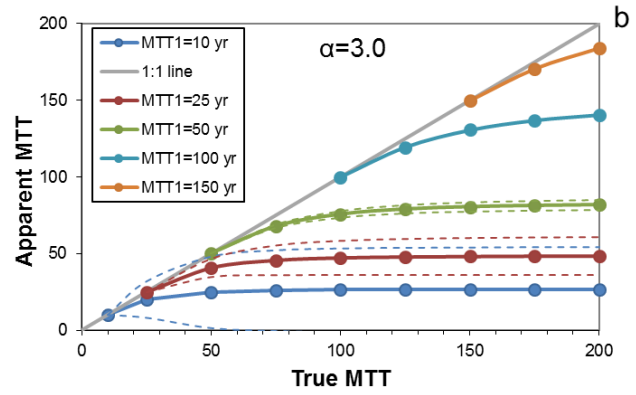
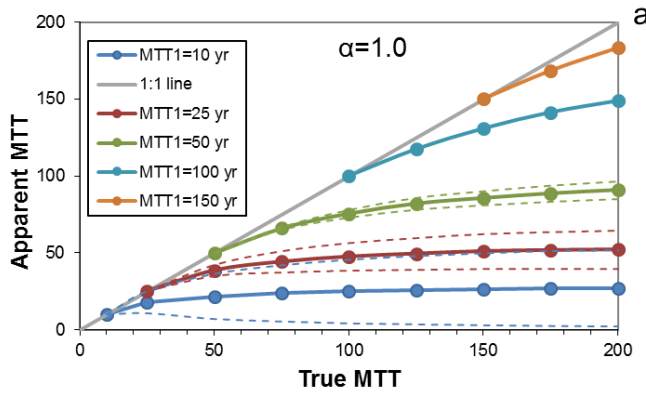


Figure 6(a-c). Aggregation effects for tritium MTTs using the Trier input function. Symbols as in Fig. 5.

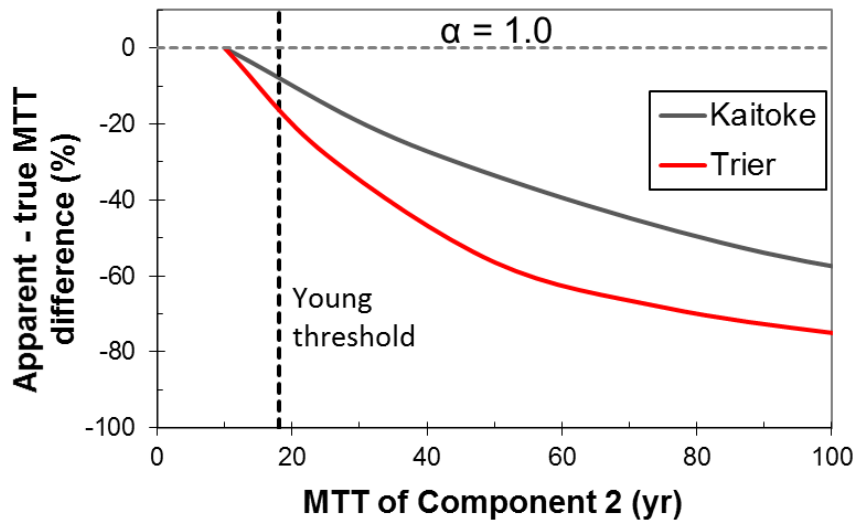
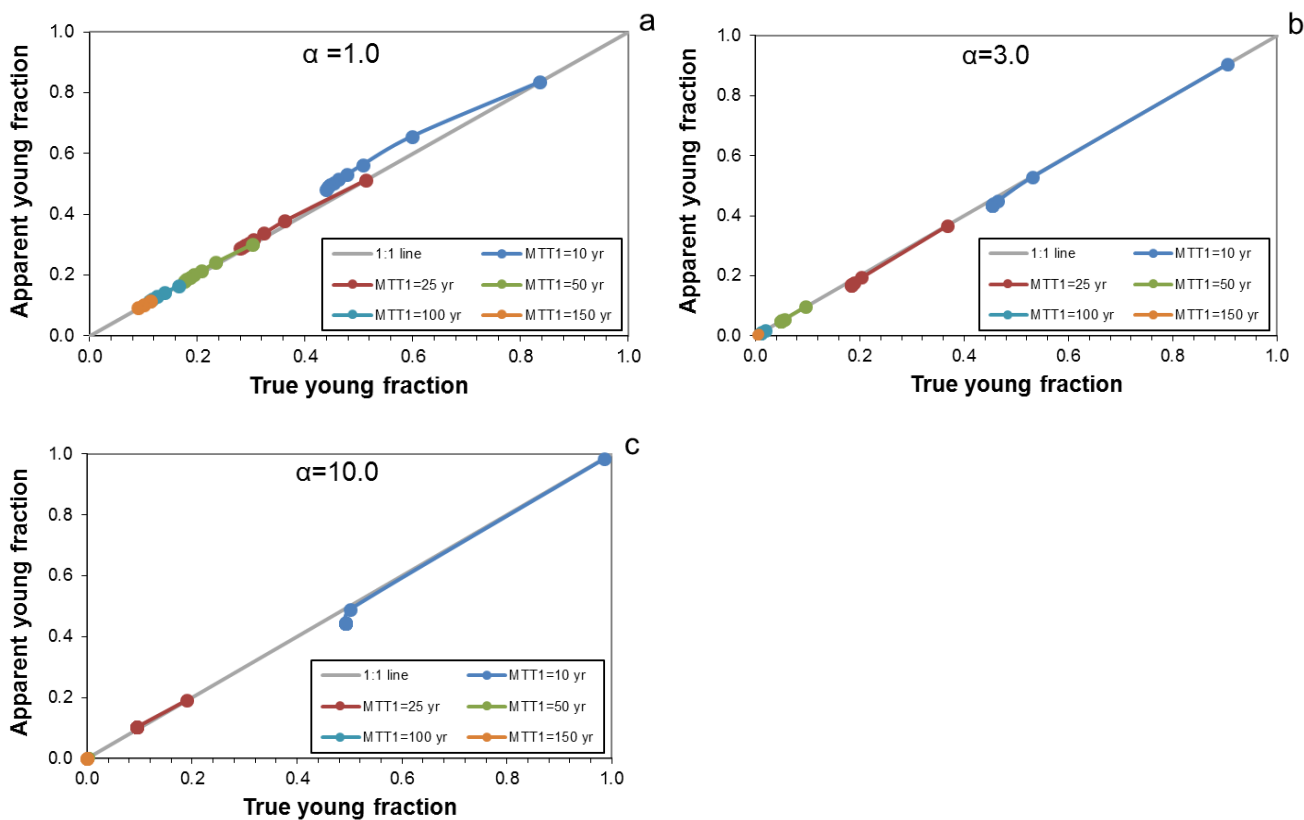
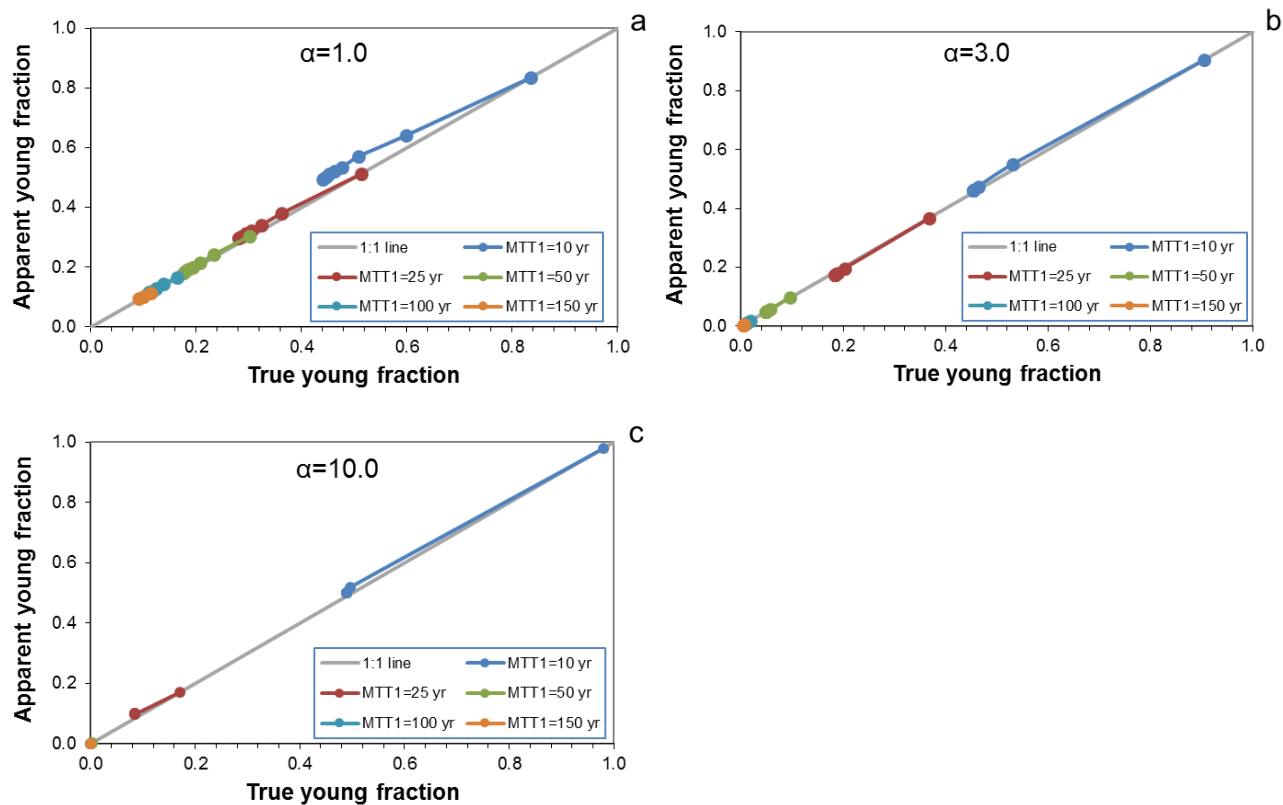


Figure 7. Comparison of maximum aggregation effects for Southern (Kaitoke) and Northern (Trier) Hemispheres for the gamma model with $\alpha = 1$. The curves show results with component 1 having mean transit time of 10 years (MTT1 = 10 yr).



5

Fig. 8(a-c). True versus apparent tritium young water fractions for gamma models with different values of α using the Kaitoke input function. Curves show changes as component 2 (MTT2) becomes older in 50-year steps and therefore the mixtures older in 25-year steps (shown by dots).



5

Figure 9(a-c). Tritium young water fractions using the Trier, Germany tritium input function. Symbols as in Fig. 8.

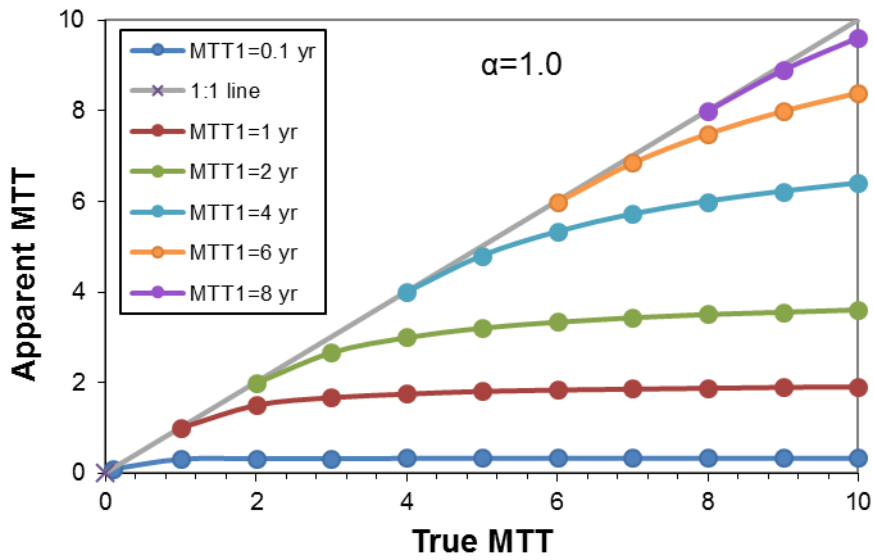


Figure 10. Aggregation effects on MTTs determined using seasonal tracer cycles for the gamma model with $\alpha = 1.0$. Curves show changes as component 2 (MTT2) becomes older in 2-year steps and therefore the mixtures older in 1-year steps (shown by dots)