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

Michael K. Stewart[1], Uwe Morgenstern[2], Maksym A. Gusyev[3,4], Piotr Maloszewski[5]

[1]Aquifer Dynamics & GNS Science, P.O. Box 30368, Lower Hutt, 5040, New Zealand
[2]GNS Science, Tritium & Water Dating Laboratory, Avalon, Lower Hutt, 5040, New Zealand
[3]International Centre for Water Hazard and Risk Management (ICHARM), Public Works Research Institute (PWRI), Tsukuba, Japan
[4]National Graduate Institute for Policy Studies (GRIPS), Tokyo, Japan
[5]AGH 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 show 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. We also find that young water fractions derived from tritium are almost immune to aggregation errors as were those derived from seasonal tracer cycles.

To investigate the implications of these findings for past and future use of tritium for estimating MTTs in catchments and groundwater systems, we examined case studies from the literature which had long series of tritium measurements and in which simple and more complicated lumped parameter models had been used. We find that MTT aggregation errors are small when the component waters have similar MTTs to each other. On the other hand, aggregation errors can be large when very young water components are mixed with older components. In general, well-chosen compound lumped parameter models should be used as they will reduce potential aggregation errors due to the application of simple lumped parameter models. An opportunity to determine a realistic compound lumped parameter model is given by matching simulations to time series of tritium measurements (underlining the value of long series of past tritium measurements), but such results should be validated by reference to the characteristics of the hydrological system to ensure that the parameters found by modelling correspond to reality. In addition, the range of tritium concentrations in hydrological systems in the future is likely to be limited because of

the passing of the bomb peak. This will make model identification more difficult and highlights the fact that such time series should be built now as long as the tail of the bomb peak is still available.

1 Introduction

Water can take very complex flow pathways in catchments through shallow and deep aquifers. 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 and emerge 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. 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 (a well-known relatively successful application of an LPM to model catchment acidification is described by Cosby et al., 1985). 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 using seasonal tracer cycles. 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 also 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 possibly more) components of different water ages, it is 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; Małoszewski 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. Calculation 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 used is 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 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 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 simple LPMs frequently used for interpreting tritium data, such as the exponential, exponential piston flow and dispersion models. The different tritium input functions for Northern and Southern Hemisphere locations were tested. We also surveyed some applications of tritium dating from the literature to see if they had sufficient data to constrain the parameters of mixtures of young and old waters. MTTs from simple and compound LPMs applied to the data were compared to examine the aggregation errors for these real examples. This has allowed us to consider the practical implications of our findings and provide guidance for tritium sampling and interpretation in heterogeneous catchments and groundwater systems.

2 Methods

2.1 Transit time determination: Simple and compound lumped parameter models

The different flow paths of water through the subsurface of catchments imply that outflows contain mixtures of water with different transit times. i.e. The water in the stream does not have a discrete age, but has a distribution of ages. 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 of different ages (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

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

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 transit time distribution (TTD) function of the hydrological system constructed based on the distribution of the water fluxes in the catchment (flow model). The exponential term accounts for radioactive decay of tritium. (λ is the tritium decay constant ($= \ln2/T_{1/2}$, where $T_{1/2}$ is the half-life of tritium (12.32 years).)

Tritium in precipitation was different in each hemisphere, and is a proxy 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. 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 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 tritium is becoming increasingly useful for dating because multiple age solutions are now less of a problem (Stewart et al., 2012; Stewart and Morgenstern, 2016; Gusyev et al., 2016). Effective use of tritium does require highly sensitive and accurate tritium measurements, however, because the natural cosmogenic tritium concentrations and variations are very low.

5 Several simple flow models are commonly used in tracer studies, and are represented by probability density functions (PDFs). 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

$$h(\tau) = \delta(\tau - \tau_m) \tag{2}$$

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
10 concentration is

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

The exponential model (EM) is given by

$$h(\tau) = \frac{1}{\tau_m}\exp(-\frac{\tau}{\tau_m}) \tag{4}$$

where again the single parameter is τ_m[yr]. In this model, water parcels with different transit times combine in the outflow to
15 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

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

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
20 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

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

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

25 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 f=1/η. τ_m(1-f) is the time required for water to flow through the piston flow section, while fτ_m is the mean transit time through the exponential volume.

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

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

where P_D[-] is the dispersion parameter (being the measure of the variance of the transit time distribution, i.e. the sum of the

5   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 LPMs) and *compound LPMs* (binary or other parallel combinations of simple LPMs). Simple LPMs describe systems as homogeneous, but have been widely

10   applied to any and all systems, no matter how heterogeneous they may have been.

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

$$LPM = bLPM_1 + (1-b)LPM_2 \tag{8}$$

15   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

$$\tau_m = b\tau_1 + (1-b)\tau_2 \tag{9}$$

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.,

20   1992) and the 'two parallel linear reservoirs' (TPLR) model when applied to seasonal tracer cycles (Weiler et al., 2003). The PDF is given by

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

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

25   $$\tau_m = b\tau_f + (1-b)\tau_s \tag{11}$$

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 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 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 calculations in Section 3.1.2. In all cases, the tritium concentrations of the mixed system (C_m) are given by

$$C_m = bC_1 + (1-b)C_2 \tag{12}$$

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. Note that the tritium concentrations in recharge was used for the case studies.

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.

$$Y_f = \int_0^{t_y} h(\tau).d\tau \tag{13}$$

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

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

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.

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).

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

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

$$\text{and } NSE = 1 - \frac{\sum_1^N (GM_i - LPM_i)^2}{\sum_1^N (LPM_i - \overline{LPM_i})^2} \tag{16}$$

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 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 due to radioactive decay of tritium (Fig. 3d).

3.1.2 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. 4. 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. 4. 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. 5) compared to those obtained with the Kaitoke input (Fig. 4). 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., 2010) 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. 6 to compare results for the Northern and Southern Hemispheres and to investigate aggregation errors for young components. 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. 4a and 5a 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). Note that this value for the young water threshold is essentially based on the calculations for MTT1 = 10 yr, because these have the greatest effect compared with the calculations for MTT1 = 25 yr etc. (Figs. 7, 8 below).

Further calculations were made for MTT1 = 0.01 yr (dashed curves in Fig. 6). The aggregation errors are very large and occur at relatively young MTT2 values. Aggregation errors could be up to 50% for Southern Hemisphere locations and up to 75% for Northern Hemisphere locations by the young water threshold. Despite the large MTT aggregation errors at the young water threshold, the young water fraction remains unaffected for both tritium input curves as demonstrated in the following section.

3.2 Aggregation effects on young water fractions

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. 7 and 8. The second component (MTT2) was then allowed to become older in 50-year steps so that the differences in MTTs and young water fractions between the two components increased. But now the true and apparent young water fractions did not diverge very much from each other (Figs. 7 and 8). 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, the maximum divergences from the ratio were found to become larger. Consequently, 18 years is taken as the recommended value for the young water threshold below.

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, the young water threshold of tritium-based transit times was 18 years with shape factors from 1 to 10 (and MTT1 starting at 10 years).

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.

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. 9 shows the aggregation effects for the gamma model with α = 1. The pattern is very similar to those observed using tritium concentrations (Fig. 4), 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 Case studies from the literature

The calculations above have shown that fitting simple LPMs to tritium data can potentially cause significant (or severe) underestimation of the MTT when two or more dissimilar tributaries or a multi-aquifer system feed a sampled outlet. On the other hand, the young water fraction is not likely to be affected by such errors. A number of studies using tritium to determine MTTs have been reported in the literature. How significant have such errors been in actual practice? This section describes some case studies from the literature to explore this question in different hydrogeologic settings of New Zealand, where relatively long tritium records are available. The case studies have been chosen to cover the age dating range of the tritium method in well, spring or stream flows.

Our method of investigation is to compare the results obtained with the simple and compound LPMs fitted to the tritium data from the case studies in combination with other information on the catchment/groundwater systems. The compound LPMs are binary parallel combinations of the exponential piston flow or dispersion models. We contend (see discussion below) that the binary models have very much less potential for aggregation bias than the simple models because the former allow young water components to be treated separately from old water components while the latter do not. In addition, the particular geohydrological characteristics of the cases studied led to the identification of two predominant types of water in their outflows.

The MTTs derived with the binary models are taken as the "true" MTTs and those from the simple models as the "apparent" MTTs for this comparison.

4.1 Two water components at Waikoropupu Springs (karstic springs fed by Arthur Marble overlain by Tertiary sediments)

5   Tritium measurements at the Waikoropupu Springs began in 1966 and cover almost the rise and fall of the tritium bomb peak in precipitation (Stewart and Thomas, 2008). Fig. 10a shows the tritium concentrations of the recharge, the Main Spring, and the best-fitting model simulations of the data. The mixing models used were two simple LPMs (exponential piston flow and dispersion models) and two compound LPMs (double exponential piston flow (DEPM) and dispersion (DDM) models).

The compound models were used because flows, $\delta^{18}O$ and Cl measurements showed that there were two separate water systems
10   contributing to the Main Spring (a shallow system and a deep system, see Stewart and Thomas, 2008). The fraction of the shallow system (b = 0.24) contributing to the Main Spring was determined from a balance model based on the flows, $\delta^{18}O$ and Cl concentrations. The transit time distributions of the models have similar shapes with peaks of very young water and long tails of much older water (Fig. 10b). All models gave good fits to the data, and the mean transit times (MTTs) were constrained close to 8 years (Table 2). The best-fitting exponential piston flow and dispersion models had MTTs of 7.9 and 8.2 yr
15   respectively. The DEPM model fitted well and gave MTTs of 0.1 and 10.2 years for the two components. $\delta^{18}O$ measurements had shown that the shallow system had MTT of 1.1 years (Stewart and Thomas, 2008), so the DEPM did not reproduce this well. The DDM fitted better with an overall MTT of 7.9 years and identified the MTTs of the two components well (1.1 and 10.0 years). A full uncertainty analysis using multivariable parameter estimation methods for fitting the models to the data (e.g. Gallart et al., 2016) would be valuable (especially for the four-parameter double models, the fifth parameter (b) being
20   estimated separately).

There is little aggregation error in the MTTs in this case (i.e. the 'apparent' MTTs of the exponential piston flow and dispersion models are very similar to the 'true' overall MTTs of the DEPM and DDM model). This is because both systems are young in relation to the young water threshold for tritium (18 years). The young fractions are also similar to each other at about 0.7.

4.2 Kuratau River (volcanic ash deposits and andesite)

25   Kuratau River flows into Lake Taupo in the North Island of New Zealand. Samples from the river were analysed for tritium from 1960 to the present making it the longest tritium time-series in New Zealand (Morgenstern and Taylor, 2009). Geological evidence strongly supports the presence of two subsystems within the catchment. The very impermeable Whakamaru Group ignimbrites and andesitic and basaltic lavas produce very young water, while the area with the highly permeable Taupo/Oruanui ignimbrites and tephras produces much older water (Morgenstern, 2007). The highly contrasting permeabilities
30   of these rocks is corroborated by observations in adjacent catchments. Distributed groundwater models calibrated with groundwater levels, river discharges and tritium concentrations substantiated these flows (Gusyev et al., 2013; 2014).

The best EPM and DM models (Table 2) had MTTs of 4 years but fitted relatively poorly to samples collected around 1970 (Fig. 11a), while compound mixing models (double EPM and DM models) with overall MTTs of 15 and 19 years fitted much better. The compound models comprised about 50% of a component with MTT of less than 1 year and 50% of a component with MTT of 35 years. Strong aggregation bias is shown by the marked difference in MTTs between the simple models and the compound models (giving apparent and true MTTs of 4 and 15-19 years respectively), due to the dominance of the young component. Note that the samples collected after 2000 are not capable of distinguishing between the models in the Southern Hemisphere using annual tritium input values, although this is still possible in the Northern Hemisphere. The young fractions are about 0.7 (Table 2).

4.3 Hangarua Spring and Hamurana Stream (volcanic ash deposits, Mamaku Ignimbrite)

Hangarua Spring drains from the Mamaku Plateau and flows into Lake Rotorua via the Hamurana Stream, which also gains water from other springs. Tritium samples were collected from about 1970 to the present. For Hangarua Spring and Hamurana Stream, and for many other streams and springs drawing from the Mamaku Ignimbrite plateau, two different flow contributions are demonstrated by the tritium measurements (Morgenstern et al., 2015). These contributions are (relatively) young water from shallow aquifers seen in minor streams maintained by shallow aquifers, and old water from deep aquifers seen in aquifers with very deep groundwater tables in the area (Rosen et al., 1998).

The best fitting exponential piston flow model (EPM) for Hangarua Spring has MTT of 56 years but fits poorly to the measured data, while the compound model (DEPM) with MTT of 116 years fits well (Fig. 12a, Table 2). It consists of 35% 18-year-old water and 65% 167-year-old water (f parameters listed in Table 2). A moderate aggregation bias is demonstrated by the difference in MTTs (Table 2). Application of the dispersion model (DM) and double version (DDM) gives somewhat different results, but still shows a moderate aggregation effect. The DM had MTT of 109 years but fitted very poorly, while the DDM had MTT of 179 years. This suggests that when assessing an aggregation effect, one should not use the DM and the DEPM together, or the EPM and DDM.

Similarly, Hamurana Stream shows moderate aggregation bias (apparent to true MTTs of 61 and 144 years with the EPM and 76 to 160 years with the DM) (Table 2, Fig. 13a,b). The compound models consist of about 35% of 12-year-old water and 65% 220-year-old water. The young fractions are all about 0.23 based on t_y of 18 years (Table 2).

4.4 Reconciliation of tritium and carbon-14 results: Christchurch groundwater system (interleaved alluvial gravel and marine sediments)

Samples from a deep groundwater well in Christchurch, New Zealand, demonstrate possible effects of two water feeds with different mean transit times to the well (Stewart, 2012). The well (M35/3637) taps the Wainoni Aquifer (Aq. 4), where it is unconfined in west Christchurch. The first tritium measurement was in 1986 and five subsequent measurements showed a steady rise from near-zero tritium in 1986 as the bomb tritium peak passed through the site (Fig. 14a). The best-fitting

exponential piston flow model (EPM) simulation to all points had f = 0.75 and MTT = 105 years. EPM fits to each individual point gave mean ages close to 105 yr.

A double EPM (DEPM) model simulation which included an old water component with zero tritium concentration was also applied to the M35/3637 data, in order to investigate whether the mean age could really be older than the 105 years given by the EPM simulation (Stewart, 2012). Addition of the old water component did not improve the fit to the tritium data, but did not make it worse for addition of a small proportion (up to about 20%) of tritium-free water (Fig 14b). (The DEPM curve shown in Fig. 14a has two water components with mean ages of 100 and 1200 years respectively, with f=0.75 for each. The older water makes up 15% of the mixture.) So the mean water age could easily be older than the 105 years given by the EPM model (e.g. 265 years with the DEPM parameter values above), because of the aggregation error due to input of two water components with different ages.

Carbon-14 measurements collected at the same time as the tritium measurements (1986 to 2006) gave mean ages of 94, 283, 190 and 324 years according to the EPM model with f = 0.75 (Stewart, 2012). These show that the water feeding the well became older on average after 1986. The later tritium samples were not able to show this increase in mean age because the extra old water added had very little tritium and therefore was 'invisible' to the tritium method.

5 Discussion

5.1 Implications of tritium MTT aggregation bias

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 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. 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 case 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 has a mild heightening effect on the aggregation bias. Both of these assumptions maximise the aggregation bias. 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. 4 and 5). The bias is particularly severe when one component is especially young. Fig. 6 shows the bias for the extreme case when the young component (MTT1) is 0.01 years. The bias rapidly increases as the old component (MTT2) becomes older. By the time the MTT difference between the components is equal to the young water threshold (18 years), the bias is 50% for Southern Hemisphere locations and 75% for Northern Hemisphere locations. On the

14

other hand, when MTT1 = 10 years (dashed curves), maximum errors are much smaller (8% for Southern Hemisphere and 15% for Northern Hemisphere locations by the young water threshold).

5.2 Does the compound MTT represent the "true" MTT?

The type of heterogeneity leading particularly to aggregation bias is heterogeneity that produces flows with very different
5   MTTs. The extreme case is when the catchment or groundwater system is divided into two parts, as noted above. Luther and Haitjema (1998) observed that cases of "few and distinct" heterogeneity have marked effects on the bulk hydrogeological quantities necessary for digital models. We looked for evidence of real-world aggregation error by examining some tritium dating case studies from the literature (Section 4). These were cases where we knew there were "few and distinct" heterogeneities in the catchments. It was not difficult to find such cases, but others with no such heterogeneities could also
10   have been chosen. Many catchments have outflows composed of quickflow and baseflow, which could also have very different MTTs leading to large aggregation errors. However, tritium dating studies are often carried out on the baseflow (i.e. low flow) portions of streamflows.

We believe that the use of compound LPMs could strongly reduce aggregation errors in hydrological systems with "significant and distinct" heterogeneity. For example, we consider a simple case of a catchment split into two parts by two
15   very different rock types that produce waters with very different MTTs; i.e. the most extreme "significant and distinct" heterogeneity one can imagine. A binary LPM describes this type of system, and when optimised with suitable data would very effectively separate the young MTT from the old MTT waters in the catchment outflow, and therefore minimise aggregation errors in MTT. If we now consider a catchment split into four parts with two areas of each rock type, the binary LPM when optimised is still very effective for separating the two types of water, while the potential for aggregation errors is
20   smaller. In systems which are split into eight, sixteen, etc. parts the binary LPM retains its effectiveness, but the potential for aggregation errors becomes very much smaller because the system starts to look homogeneous at larger scales. There is, of course, a wide range of different types of hydrological systems, but the binary LPM is likely to remain effective in cases of "significant and distinct" heterogeneity, which are the ones of concern for aggregation error.

5.3 How much have aggregation effects affected tritium MTTs in past studies?

25   Seasonal tracer cycles have been far more widely used to determine MTTs in streams than tritium concentrations. It is clear that many (if not most) studies using seasonal tracer cycles *interpreted with simple LPMs* will have been affected by aggregation bias. But we contend that tritium studies will have been affected less, despite aggregation bias also applying to tritium-derived MTTs, because many of the tritium studies in the literature applied compound models calibrated by fitting to time series of tritium measurements rather than or as well as using simple LPMs. Provided the compound LPMs were well-
30   chosen based on the characteristics of the catchments, they will produce more accurate TTDs than the simple LPMs and therefore will reduce aggregation bias on MTTs.

A good example is the study of Blavoux et al. (2013) describing the interpretation of an exceptionally long and very detailed record of tritium concentrations from the Evian-Cachat Spring in France. The tritium record was much too complicated to be fitted by a simple LPM. Instead, the detailed records of input and output allowed accurate specification of a combined model comprising of exponential (τ_m = 8 yr) and dispersion (τ_m = 60 yr) models in series, with a small bypass flow in parallel with them, followed by a piston flow model (τ_m = 2.5 yr) in series giving an overall τ_m of 70 yr. The combined model was closely related to the hydrogeology of the area and produced an accurate TTD for the average stationary state of the system, so there is little possibility of aggregation bias.

Four key examples of such studies were also described by Stewart et al. (2010) when comparing stable isotope and tritium estimations of MTTs. These studies nicely showed truncation of stable isotope TTDs compared to tritium TTDs. The studies were of Lainbach Valley streamflow in Germany (Maloszewski et al., 1983), Brugga Basin streamflow in Germany (Uhlenbrook et al., 2002), Waikoropupu Springs flow in New Zealand (the first case study described above) and Pukemanga streamflow in New Zealand (Stewart et al., 2007). Both simple and compound LPMs were applied in the original studies to interpret the MTTs in these streams. The compound LPMs were based on streamflow characteristics and gave better fits to the tritium data, but more importantly separated young and old water flows.

5.4 Considerations for future use of tritium MTTs

This work is cautionary for applications of simple LPMs to hydrological systems (at least for estimation of MTTs) because of the risk of underestimation of MTTs due to aggregation bias. Compound LPMs are to be preferred, but often there can be considerable difficulty in uniquely quantifying the parameters especially if the output data is limited. In the future, the range of input tritium variation will be much more limited because of the passing of bomb tritium from hydrological systems in the Southern Hemisphere and from their gradual waning in Northern Hemisphere systems (Stewart and Morgenstern, 2016).

However, there can be good reasons for choosing realistic compound LPMs. These reasons include simulation of the data if there is a good record in time and use of other contrasting time tracers such as gases (CFCs and SF_6). Hydrological reasons can be based on baseflow separation methods (Stewart, 2015; Duvert et al., 2016) or conceptual models of catchments (Hale et al., 2016). For example, Maloszewski et al. (1983) tested three LPMs of increasing complexity at Lainbach Valley with the most complex including a bypass flow representing direct runoff (30% of total flow) and shallow and deep reservoirs (52.5 and 17.5% of flow respectively) representing indirect runoff. Deuterium and tritium measurements were used to calibrate the LPMs. Other reasons can be hydrogeological (two rock types in catchments, illustrated by the Kuratau River case study) or chemical (mixtures of water types, illustrated by the Waikoropupu Spring case study).

6 Summary and Conclusions

MTT estimations based on tritium concentrations show very similar aggregation effects to those for seasonal tracer variations. Kirchner (2016a) recently demonstrated that aggregation errors due to heterogeneity in catchments could cause severe underestimation of the mean transit times (MTTs) of water travelling through catchments when simple lumped parameter

models (LPMs) were applied to interpret seasonal tracer cycles. Here we examine the effects of such errors on the MTTs and young water fractions estimated using tritium concentrations. 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 show aggregation bias. However, the transit times

5 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. We also find that young water fractions derived from tritium are almost immune to aggregation errors as were those derived from seasonal tracer cycles.

To investigate the implications of these findings for past and future use of tritium for estimating mean transit times in catchments and groundwater systems, we examined case studies from the literature in which simple and compound LPMs had

10 been used. We find that MTT aggregation errors are small when the component waters have similar MTTs to each other. On the other hand, aggregation errors are large when very young water components are mixed with old components. In general, well-chosen compound LPMs should be used as they will significantly reduce potential aggregation errors due to the application of simple LPMs. Well-chosen means that the (compound) LPM is based on hydrologically and geologically validated information. The choice of a suitable LPM can be assisted by matching simulations to time series of tritium

15 measurements (underlining the value of long series of past tritium measurements), but such results should also be validated to ensure that the parameters won from modelling correspond to reality (since we nearly always have sufficient hydrological/geological data to examine the modelling results).

References

Blavoux, B., Lachassagne, P., Henriot, A., Ladouche, B., Marc, V., Beley, J.-J., Nicoud, G., and Olive, P.: A fifty-year

20 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.

Botter, G., Bertuzzo, E., and Rinaldo, A.: Transport in the hydrologic response: Travel time distributions, soil moisture

25 dynamics, and the old water paradox, Water Resour. Res., 46, W03514, doi:10.1029/2009WR008371, 2010.

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.

Cosby, A.G., Hornberger, G.M., Galloway, J.N., and Wright, R.F.: Modeling the effects of acid deposition: assessment of a lumped parameter model of soil water and streamwater chemistry. Water Resour. Res., 21, 51-63, 1985.

30 Duvert, C., Stewart, M.K., Cendón, D.I., and Raiber, M.: Time-series of tritium, stable isotopes and chloride reveal short-term variations in groundwater contribution to a stream, Hydrol. Earth Syst. Sci., 20, 257-277, doi:10.5194/hess-20-257-2016, 2016.

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.

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., Abrams, D., Toews, M.W., Morgenstern, U., and Stewart, M.K.: A comparison of particle-tracking and solute transport methods for simulation of tritium concentrations and groundwater transit times in river water. Hydrol. Earth Syst. Sci., 18, 3109-3119, doi:10.5194/hess-18-3109-2014, 2014.

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-20-3043-2016, 2016.

Gusyev, M.A., Toews, M., Morgenstern, U., Stewart, M.K., White, P., Daughney, C., and Hadfield, J.: Calibration of a transient transport model to tritium data in streams and simulation of groundwater ages in the western Lake Taupo catchment, New Zealand, Hydrol. Earth Syst. Sci., 17, 1217-1227, 2013.

Hale, V.C., McDonnell, J.J., Stewart, M.K., Solomon, D.K., Pack, R.T., Ice, G.G., and Doolitte, J.: Effect of bedrock permeability on stream base flow mean transit time scaling relationships: 2. Process study of storage and release. Water Resour. Res., 52, 1375–1397**,** doi**:**10.1002/2015WR017660, 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.

Hrachowitz, M., Savenije, H., Bogaard, T. A., Tetzlaff, D., and Soulsby, C.: What can flux tracking teach us about water age distribution patterns and their temporal dynamics?, Hydrol. Earth Syst. Sci., 17, 533–564, doi:10.5194/hess-17-533-2013, 2013.

IAEA/WMO Global Network of Isotopes in Precipitation. The GNIP Database. Accessible at: http://www.iaea.org/water, 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 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.

Luther, K.H., and Haitjema, H.M,: Numerical experiments on the residence time distributions of heterogeneous groundwatersheds, J. Hydrol. 207, 1-17, 1998.

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.

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 ^{3}H 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.

Małoszewski, 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., Maloszewski, P., Newman, B., Pfister, L., Rinaldo, A., Rodhe, A., Sayama, T., Seibert, J., Solomon, K., Soulsby, C., Stewart, M., Tetzlaff, D., Tobin, C., Troch, P., Weiler, M., Western, A., 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.: Lake Taupo streams –Water age distribution, fraction of landuse impacted water, and future nitrogen load. Environment Waikato Technical Report TR 2007/26, Document #1194728, p. 28, 2007.

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.

Rinaldo, A., Beven, K. J., Bertuzzo, E., Nicotina, L., Davies, J., Fiori, A., Russo, D., and Botter, G.: Catchment travel time distributions and water flow in soils, Water Resour. Res., 47, W07537, doi:10.1029/2011WR010478, 2011.

Rosen, M. R., Milner, D.,Wood, C. P., Graham, D., and Reeves, R. Hydrogeologic investigation of groundwater flow in the Taniwha Springs area. Institute of Geological and Nuclear Sciences Client Report 72779C.10, GNS Science, Lower Hutt, New Zealand, 1998.

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.: Promising new baseflow separation and recession analysis methods applied to streamflow at Glendhu Catchment, New Zealand, Hydrol. Earth Syst. Sci., 19, 2587-2603, doi:10.5194/hess-19-2587-2015, 2015.

Stewart, M.K., Mehlhorn, J., and Elliott, S.: Hydrometric and natural tracer (18O, silica, 3H and SF6) 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.

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, ^{3}H 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 18O applied in a hydrological study of the Wairau Plain and its contributing mountain catchments, Marlborough, New Zealand, J. Hydrol. 138, 269–319, 1992.

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

Table 2. **Parameters of simple and compound LPMs applied to tritium measurements from the Main Spring of the Waikoropupu Springs, Kuratau River, Hangarua Spring and Hamurana Stream. The young water fraction (Y_f) is the fraction of water with ages less than the young water threshold (18 yr).**

Feature	Model	MTT τ_m (yr)	b	Component 1 τ_1 (yr)	f_1	D_{P1}	Component 2 τ_2 (yr)	f_2	D_{P2}	sd (TU)	Y_f (yr)
	EPM	7.9	1.00	7.9	0.94	--	--	--	--	0.36	0.86
Waikoropupu	DEPM	7.8	0.24[1]	0.1	0.50	--	10.2	0.79	--	0.22	0.61
Main Spring	DM	8.3	1.00	8.3	--	1.8	--	--	--	0.30	0.66
	DDM	7.8	0.24[1]	1.1	--	0.09	10.0	--	0.63	0.09	0.77
	EPM	4.4	1.00	4.4	0.89	--	--	--	--	2.9	0.88
Kuratau	DEPM	15.2	0.55	0.8	1.00	--	33.0	0.50	--	1.6	0.66
River	DM	3.6	1.00	3.6	--	1.8	--	--	--	2.6	0.83
	DDM	19.2	0.46	0.3	--	0.07	35.5	--	0.92	1.6	0.53
	EPM	56	1.00	56	0.90	--	--	--	--	0.20	0.27
Hangarua	DEPM	116	0.35	17.9	0.69	--	167	0.89	--	0.07	0.24
Spring	DM	109	1.00	109	--	1.8	--	--	--	0.36	0..14
	DDM	179	0.40	18.8	--	0.31	286	--	0.17	0.05	0.19
	EPM	61	1.00	61	0.99	--	--	--	--	0.28	0.26
Hamurana	DEPM	144	0.32	13.3	0.80	--	204	0.82	--	0.02	0.26
Stream	DM	76	1.00	76	--	1.8	--	--	--	0.63	0.20
	DDM	160	0.38	12.0	--	0.25	250	--	0.20	0.05	0.27

-- Columns not applicable for these cases. [1]Value of b determined from mass balance (see text).
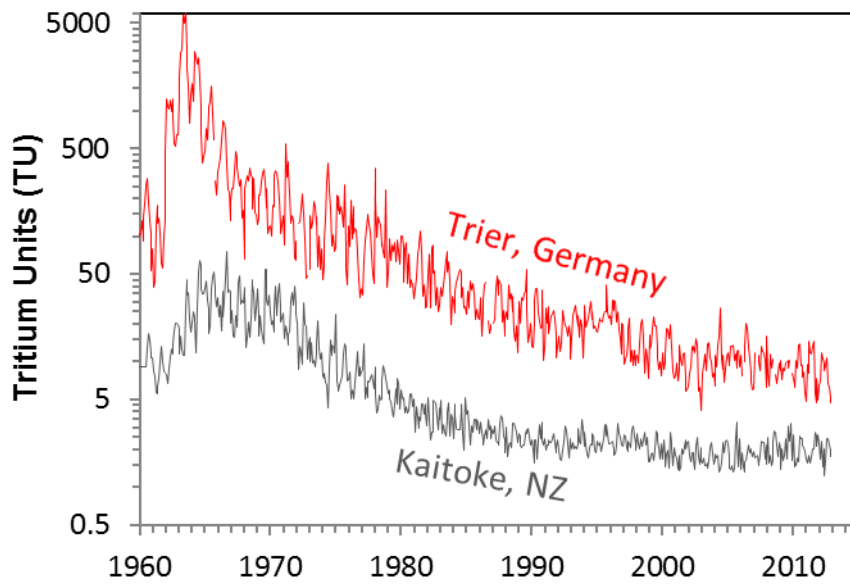
Figure 1. Tritium concentrations (TU) in monthly precipitation samples at Kaitoke, New Zealand in the Southern Hemisphere, and Trier, Germany in the Northern Hemisphere.

5
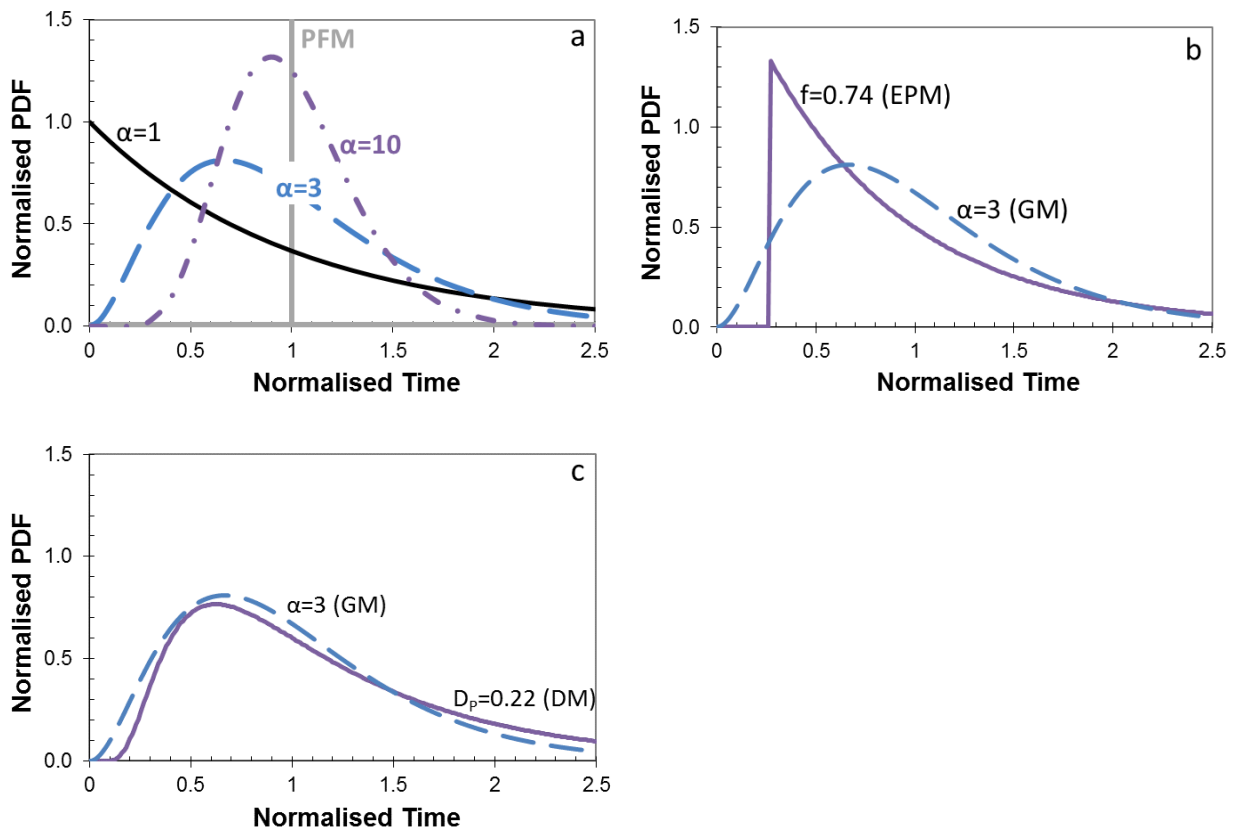
Figure 2. (a) Gamma distributions (GM) for shape factors α between 1 and 10. The axes show normalised transit time (τ/τ_m) and normalised probability density function PDF (h(τ) x τ_m). Note that the distribution for GM (α = 1) is the same as that for the exponential model (EM). (b-c) Comparison between the gamma model with α = 3 and the best fitting exponential piston flow and dispersion models.

5
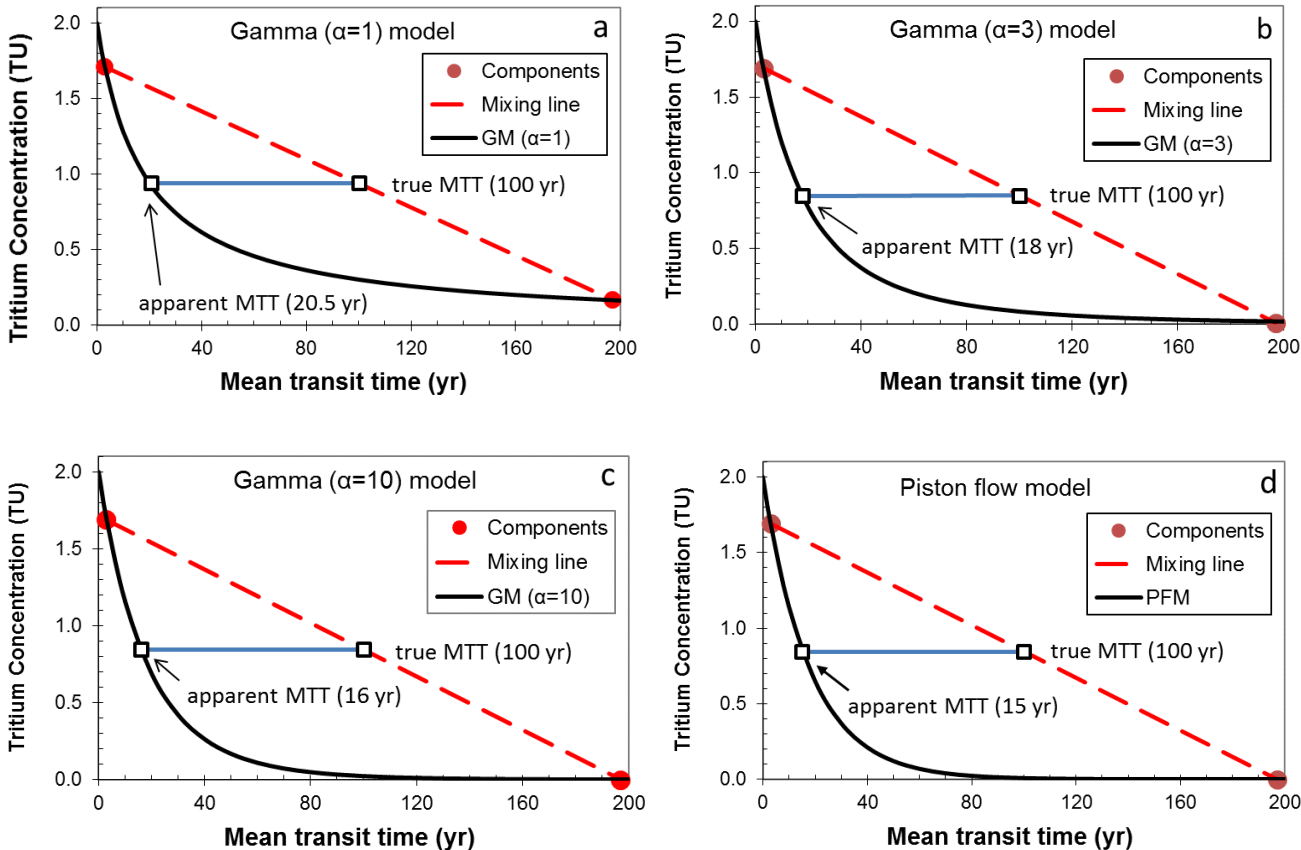
Figure 3(a-d). Aggregation errors when the tritium input concentration is assumed to be constant at 2 TU. Mean transit times (MTTs) are inferred from tritium concentrations in mixed runoff from two subcatchments with different tritium concentrations and MTTs (shown by red dots) using a range of gamma 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.

10
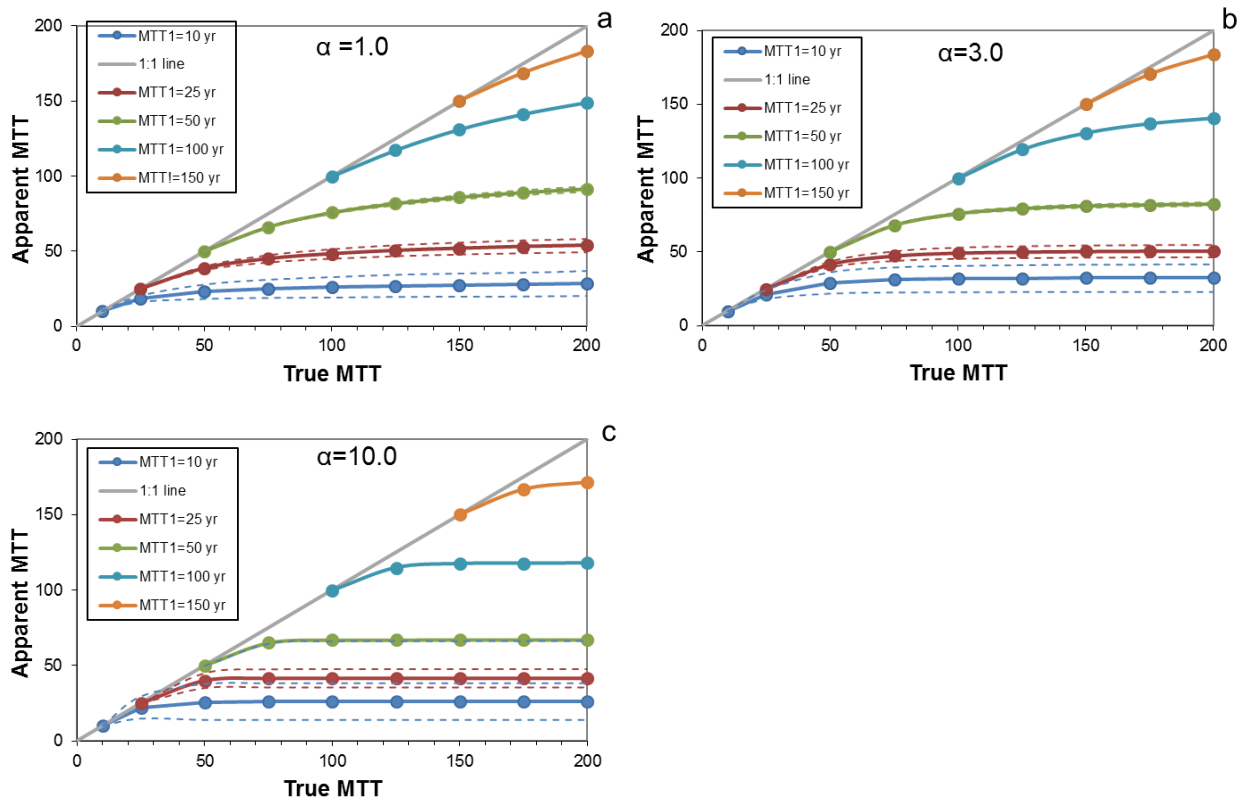
Figure 4(a-c): Aggregation effects for tritium MTTs for gamma models with different values of α using the Kaitoke input function. Curves show changes as component 2 (MTT2) becomes older in 50-year steps and therefore the mi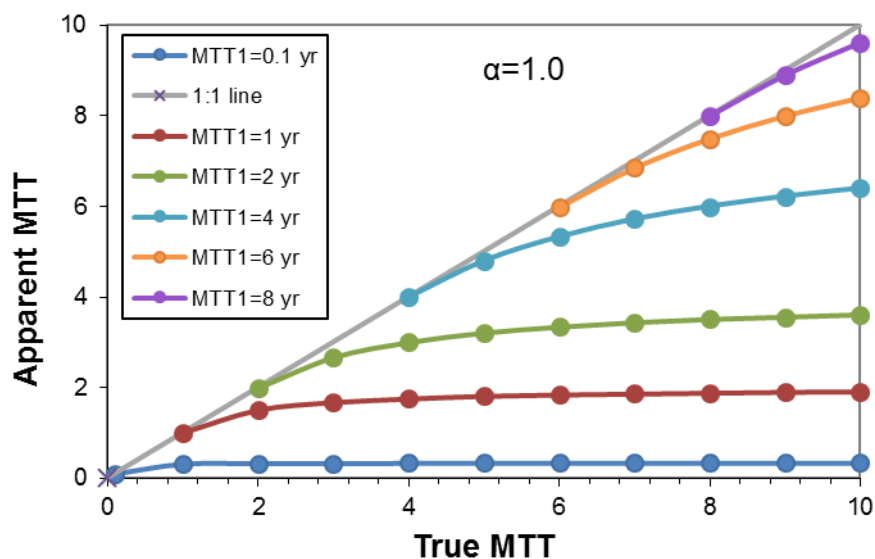xtures older in 25-year steps (shown by dots). The first step of the MTT1=10 yr curve is 15 years. Fitting errors in the apparent MTTs are shown by fine dashed lines.
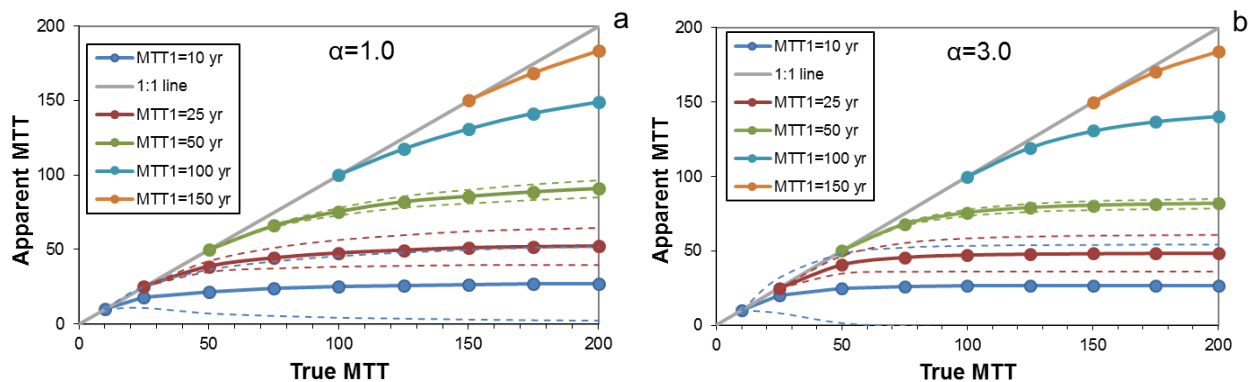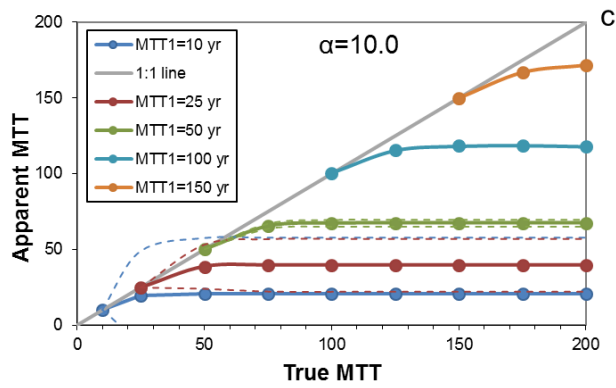
Figure 5(a-c). Aggregation effects for tritium MTTs using the Trier input function. Symbols as in Fig. 4.
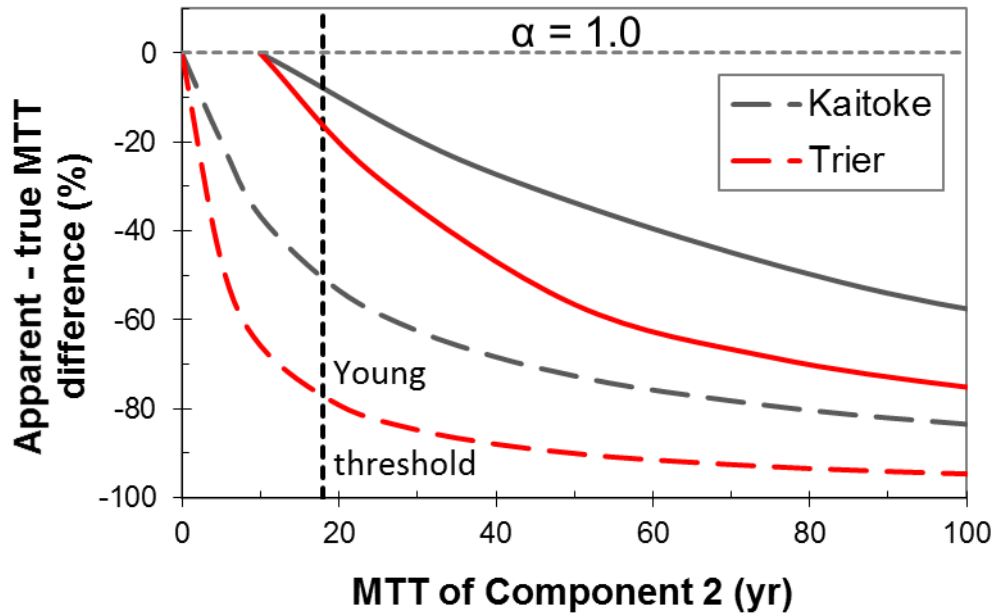
5 **Figure 6. Comparison of maximum aggregation effects for Northern (Trier) and Southern (Kaitoke) Hemispheres for the gamma model with α = 1. The solid curves show results with component 1 having mean transit time of 10 years (MTT1 = 10 yr) and the dashed curves with MTT1 = 0.01 yr.**
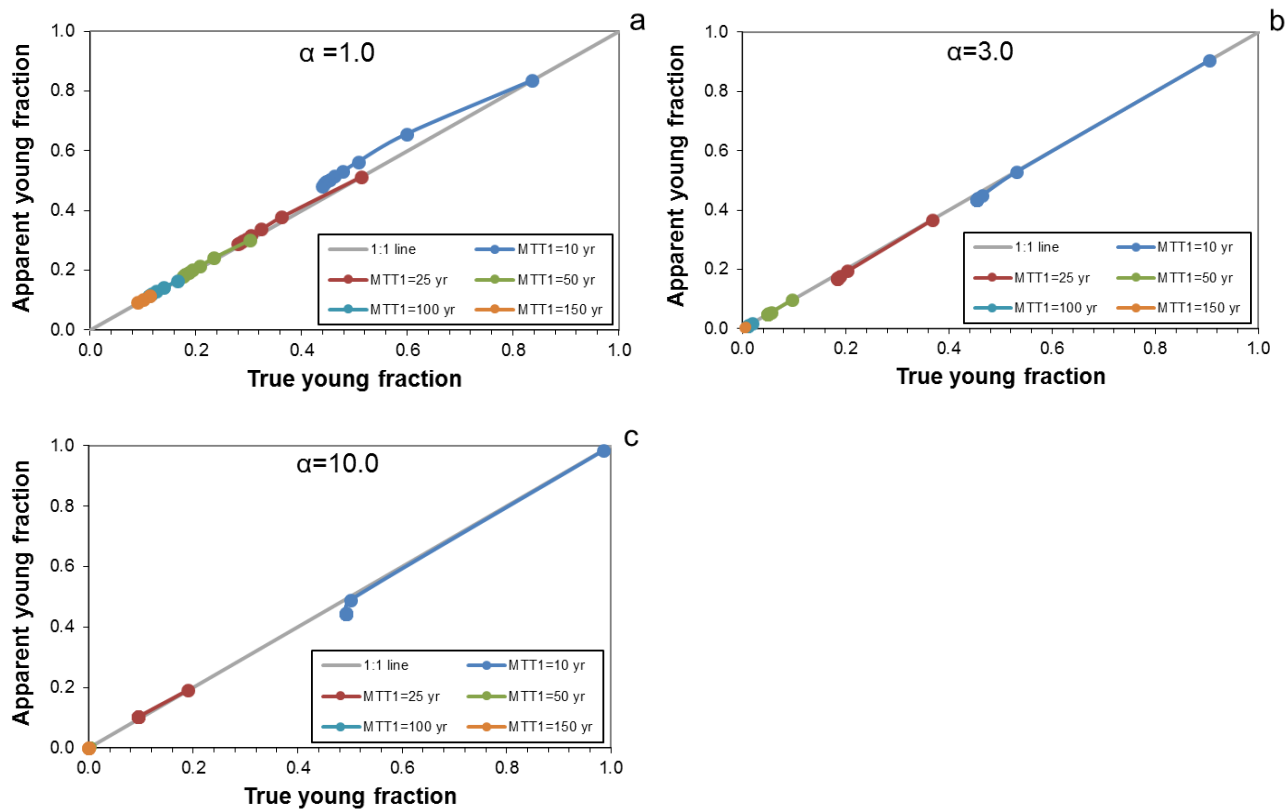
Fig. 7(a-c). True versus apparent tritium young water fractions for gamma models with different values of α using the Kaitoke input function. Curves show changes as component 2 (MTT2) becomes older in 50-year steps and therefore the mixtures older in 25-year steps (shown by dots).
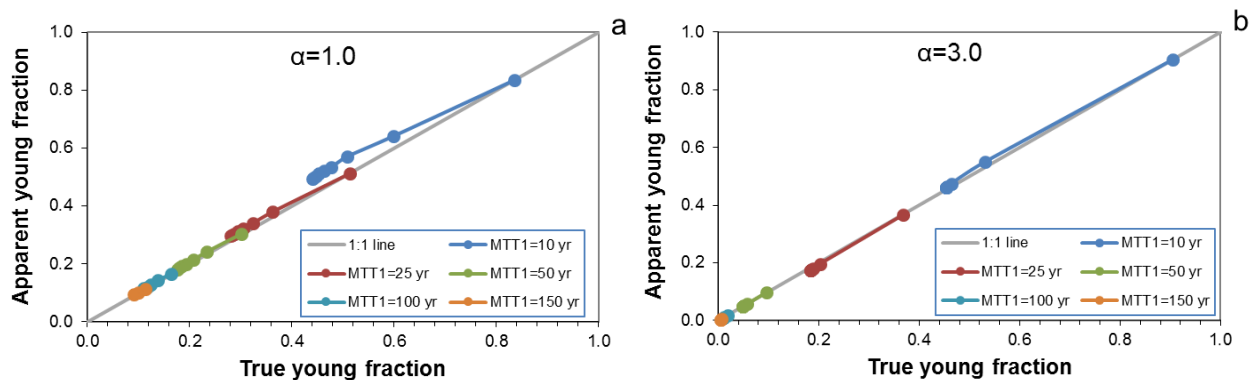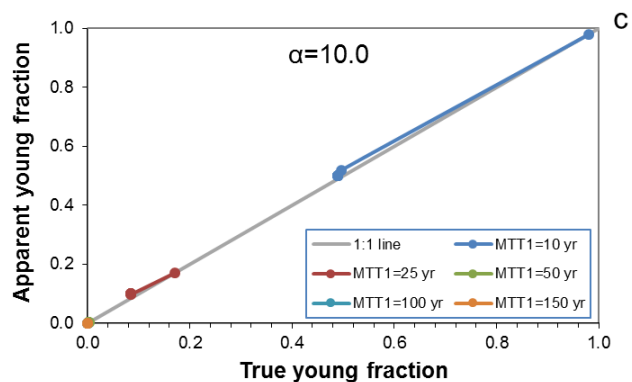
Figure 8(a-c). Tritium young water fractions using the Trier, Germany tritium input function. Symbols as in Fig. 7.

Figure 9. Aggregation effects on MTTs determined using seasonal tracer cycles for the gamma model with α = 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)
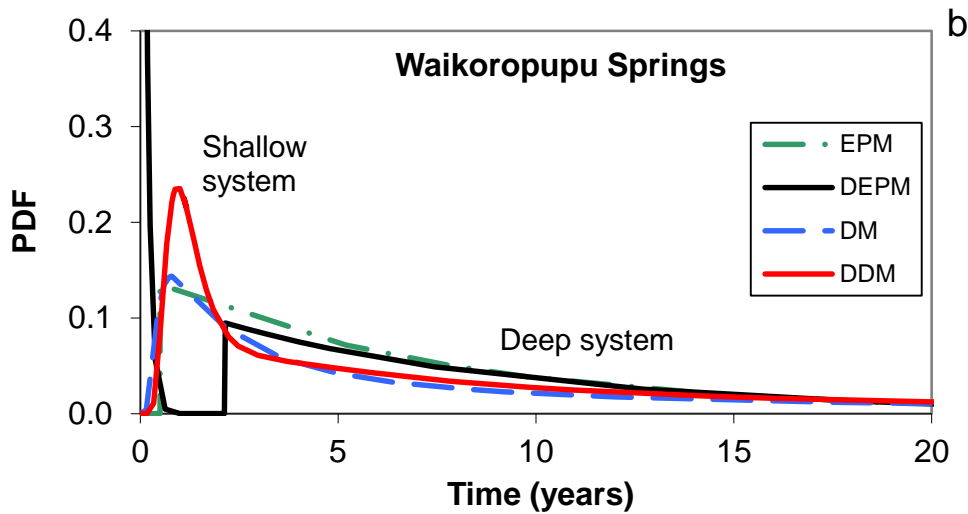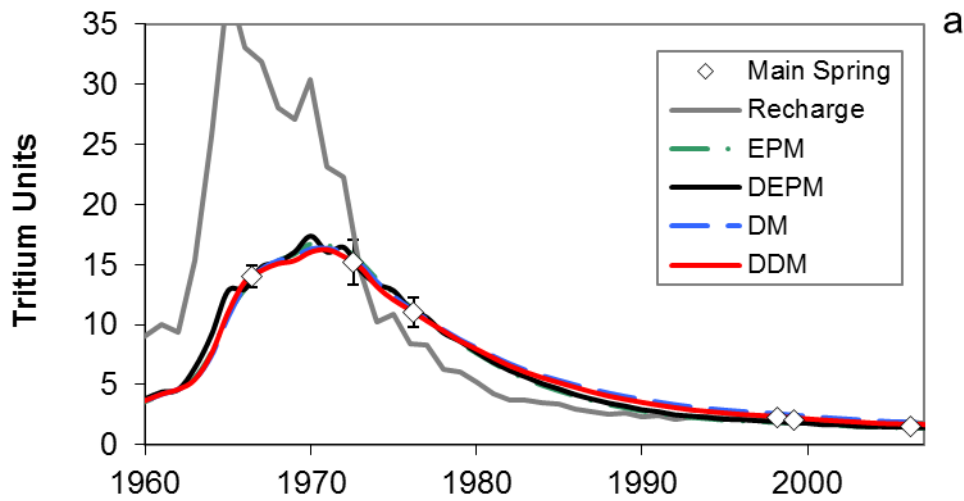
Figure 10. (a) Tritium measurements of the Main Spring of the Waikoropupu Springs, New Zealand, and model fits to the data. EPM and DM are exponential piston flow and dispersion models, and DEPM and DDM are double (i.e. binary parallel) EPM and DM models. (b) Transit time distributions of the best-fit simulations.
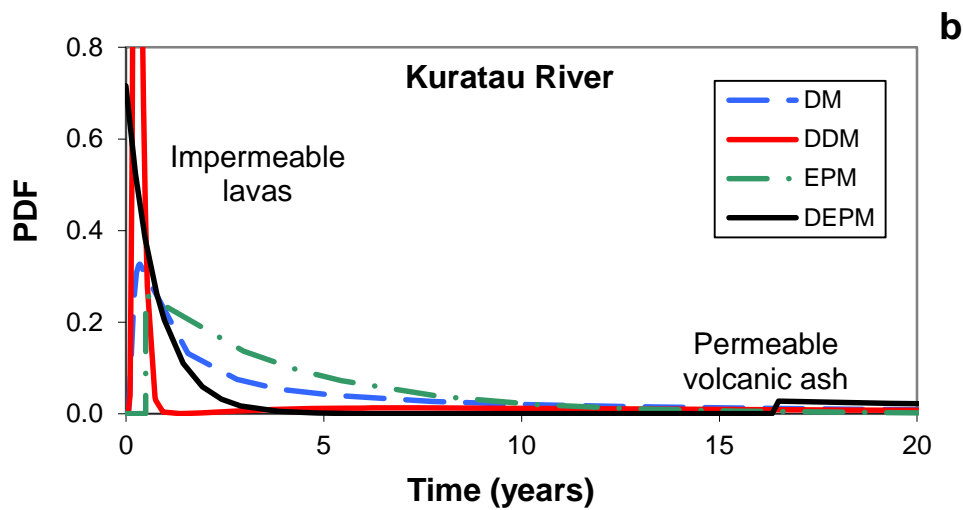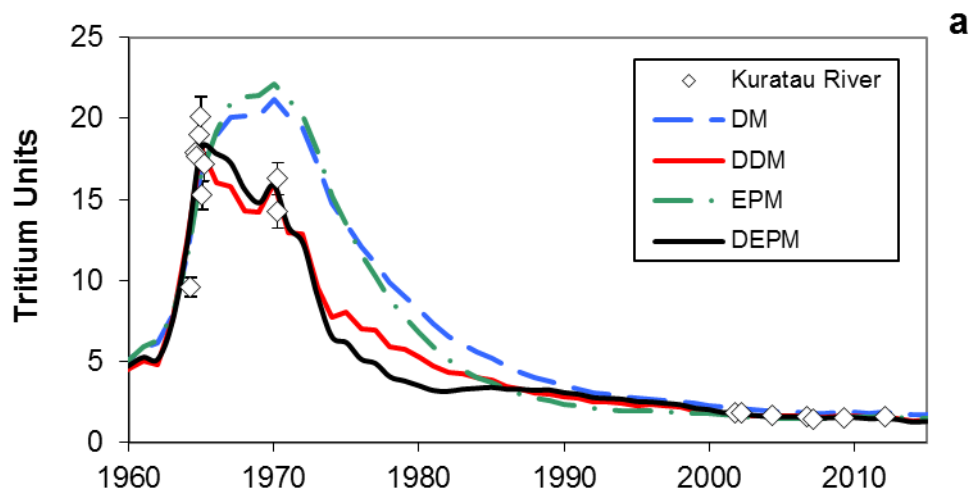
Figure 11. (a) Tritium measurements and model fits for Kuratau River, New Zealand. (b) Transit time distributions of the best-fit simulations.
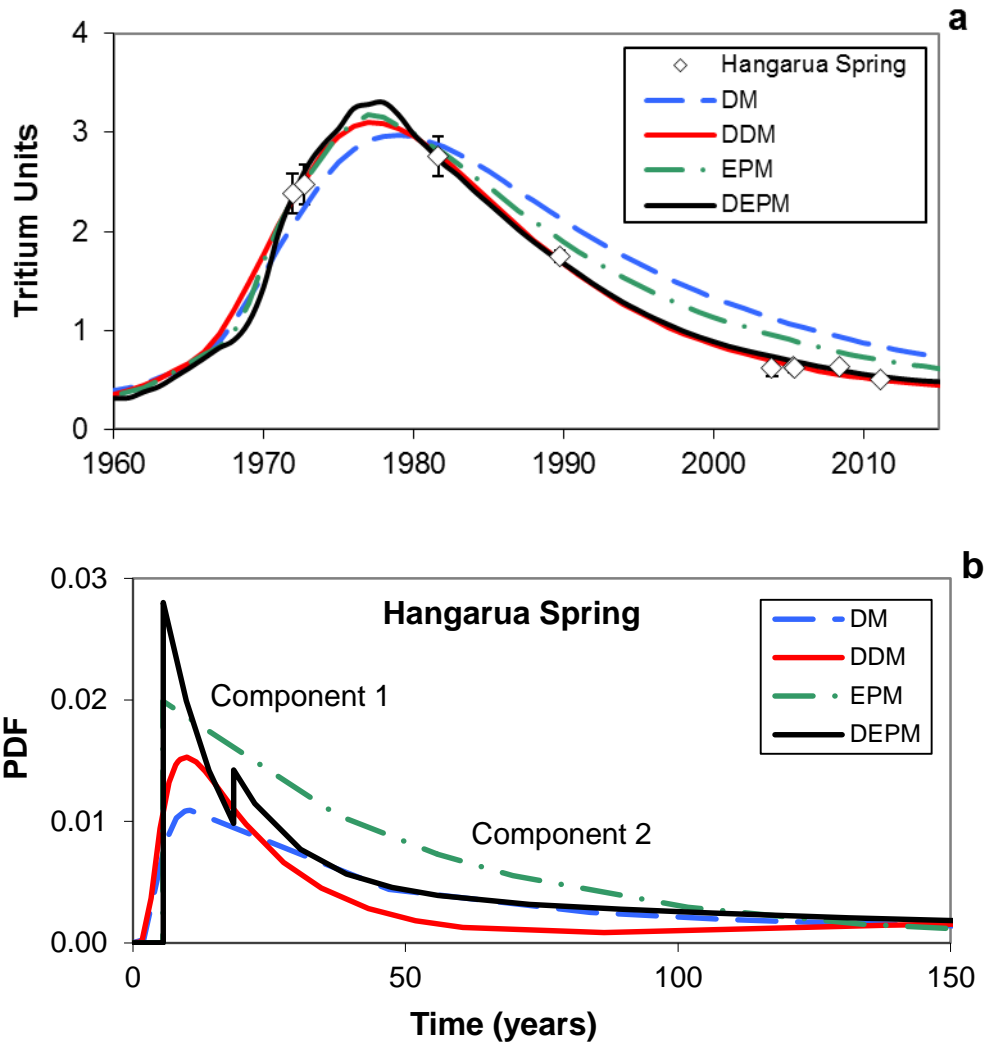
5

Figure 12. (a) Tritium measurements and model fits for Hangarua Spring, Rotorua, New Zealand. (b) Transit time distributions of the best-fit simulations.
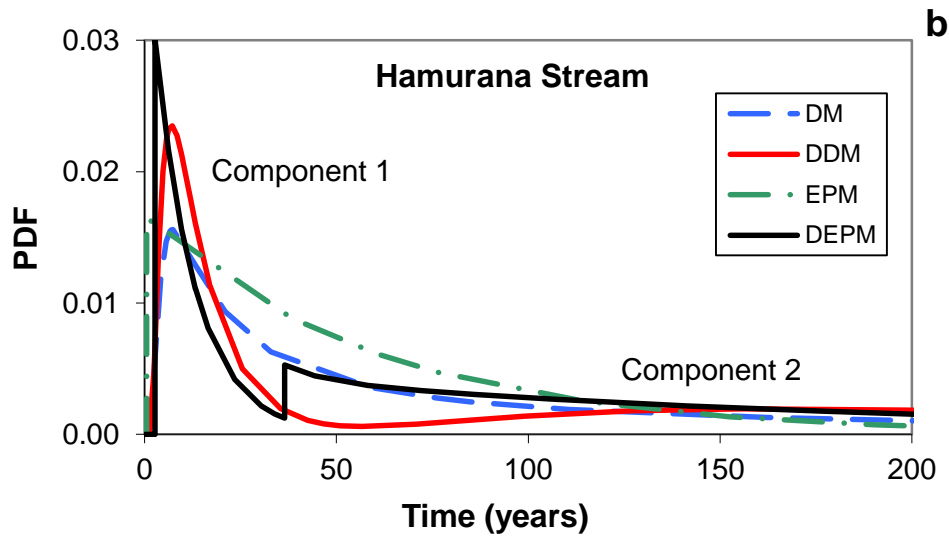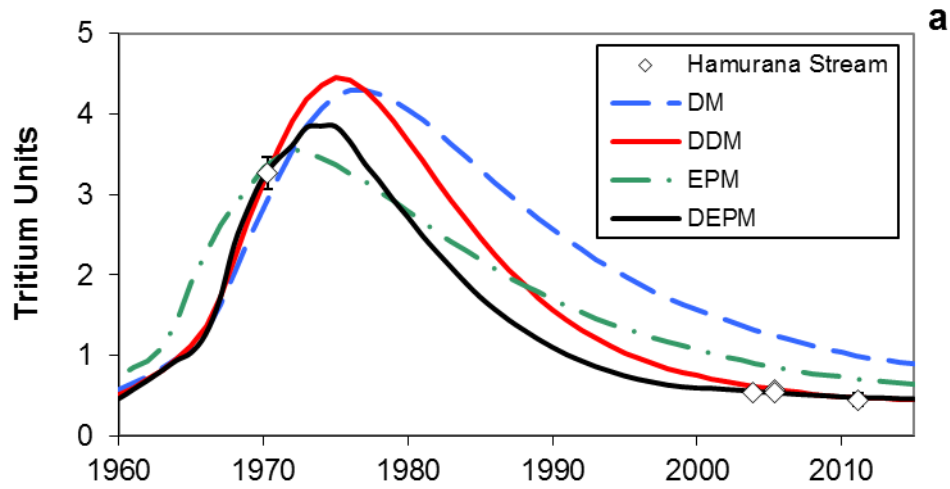
Figure 13. (a) Tritium measurements and model fits for Hamurana Stream, Rotorua, New Zealand. (b) Transit time distributions of the best-fit simulations.
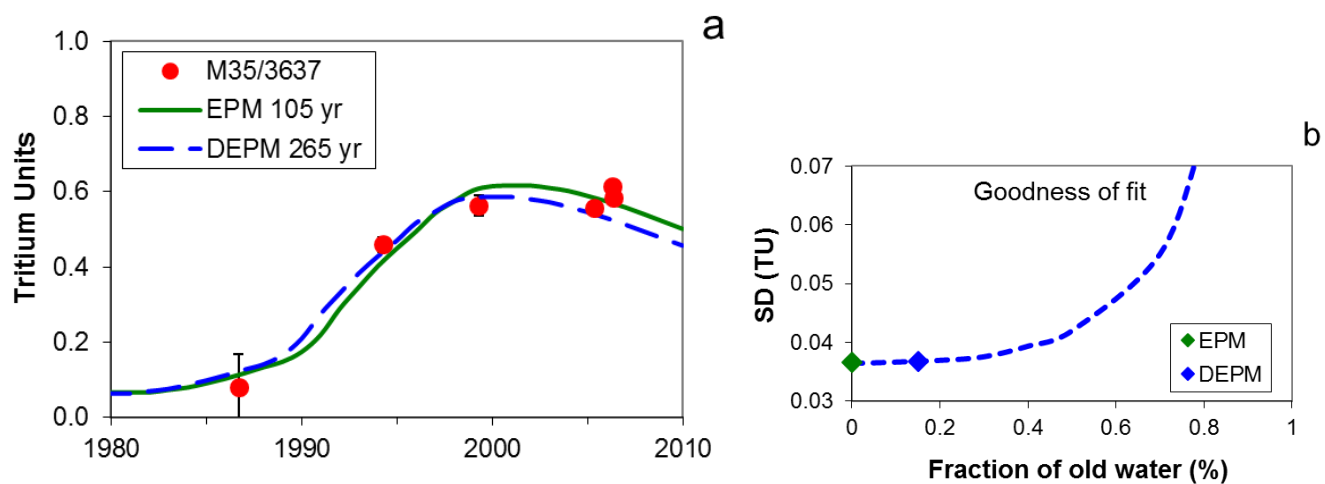
5

Figure 14. (a) Tritium measurements and simulations for groundwater well M35/3637 in Christchurch, New Zealand. (b) Variation of the goodness-of-fit criterion (sd) with fraction of old water.