Editor Initial Decision: Reconsider after major revisions (25 Oct 2016) by Markus
Hrachowitz: Comments to the Author

Editor Comment As you have seen, the three reviewers do, in principle, appreciate the topic you
address in your manuscript but they also raise a few, quite critical issues.

Reply We appreciate the work of the Editor and Reviewers on this manuscript.

Editor Comment 1) The first is the quality of the presentation of your work. The manuscript is, in
places, written in quite a confusing way with quite some inconsistencies in terminology and the
symbols used, making it difficult to follow. In addition, some of the methods of your analysis are
not explained in sufficient detail and clarity so as to allow the reader to repeat your experiments
and to (ideally) reproduce your results. I think this first issue can (and needs to) be addressed and
solved quite easily.

Reply Major changes have been made to the manuscript in response to the Reviewers' comments.
We have changed the structure (it is now structured as: 1. Introduction, 2. Methods, 3. Results, 4.
Case studies from the literature, 5. Discussion, and 6. Summary and Conclusions), but more
importantly we have expanded the description of the calculations in the Methods with two new
sections (Sects. 2.2 and 2.3), and expanded the description in the Results with section 3.1 now
divided into subsections 3.1.1 and 3.1.2. In response to the Reviewers' comments we have also
eliminated inconsistencies in terminology, and have reduced jargon by trying to spell out lumped
parameter model names (exponential model not EM, etc.). The tables have been revamped and
figures improved.

Editor Comment 2) The second point is, in my opinion, somewhat more problematic, as it
touches on the core assumptions of the experiment. The reviewers emphasized, and I fully agree
with that, that real systems and their heterogeneity are in general not known. Neither are their
real TTDs and MTTs. While it is plausible that compound TTDs will reduce the effect of natural
heterogeneity on the aggregation bias, it is, at this point and in the absence of the necessary
information, impossible to say what a "well-chosen" compound TTD is. If, for example, a *real*
system can be represented by a hyper-exponential TTD, it is still not known, how many
components , i.e. parallel, exponential processes active at different (unknown) time scales, are
necessary to correctly describe this system. Is two components enough? If not, how do we know
and how many do we need? What indeed has been shown in your manuscript is that more
flexibility (i.e. more parameters, and thus potentially more processes) can reduce the problem.
However, in real world systems it is far from obvious how relevant this reduction is, as the
absolute aggregation bias is not known. The bottom line is the question: "What is the real world
reference, which we can compare our models to?"

Reply This is one of those propositions that could go round and round without getting anywhere.

1

To make progress we think one needs to ask "What is the kind of spatial heterogeneity that leads particularly to aggregation bias when simple LPMs are used? We think this is captured in the phrase by Luther and Haitjema (1998), that is that "significant and distinct" heterogeneity has particularly marked effects on the bulk hydrogeological quantities (see also the comments by Farlin and Kirchner, and our response). The calculation procedure uses the most extreme form of this type of heterogeneity (two homogeneous subsystems) and this is fully described by a compound (binary) LPM (or hyper-exponential if the subsystems are exponential models).

When it comes to real-world systems, the optimum compound LPM could in theory be very complicated. But binary LPMs have the capacity of separating young water from old water, which is the crux of the aggregation problem. More complicated LPMs would not add much, and would have more parameters which would be difficult to quantify. This has been the basis for our approach, which is explained in various places (such as Sects. 2.1, 2.2, 5.1, 5.2).

Editor Comment 3) In addition, it is stated that, for example a gamma distribution as TTD will result in considerable aggregation error (fig.3). I would argue that such a generalizing statement is not necessarily justified, as the choice of the gamma distribution parameters very much depends on the performance metric and "calibration" strategy used. Fitting a distribution to another one is always a difficult task and many factors will considerable influence the results: how many points are used for fitting (e.g. only the points at the actual time steps used in the model or also some sub-steps?)? How long a tail is considered for fitting (i.e. the longer the tail, the more the performance metric will tend to get the tails right at the cost of a worse representation of the early phases)? Another aspect that is not fully clear, is if we should rather fit the PDFs or the CDFs of the distributions (depends on the type of sample we use)? All these aspects will strongly influence, which TTD is chosen as meaningful.

Reply We assume that this comment refers to the Fig. 3 caption "… 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." This refers to mixing of two waters with MTTs of 3 and 197 years respectively and resultant tritium concentrations determined by convolution with constant tritium input. The black curve results from fitting a gamma model with the specified α to the tritium concentrations determined by mixing the two waters. The tritium concentrations of the mixture vary as the proportions of the two waters are varied (shown by the dashed red line). The difference between the true and apparent MTTs is displayed for a 50:50 mixture. This is an idealised system and we have only adjusted the β ($=\tau_m/\alpha$) of the gamma model to fit it to the tritium concentrations of the mixture.

Editor Comment 4) For example, experimenting a bit with a gamma distribution over 500 time steps, with the squared errors computed from these 500 time steps, I found that a gamma distribution with alpha=0.4 and beta=250 (i.e. MTT=100) gives a sum of squared errors (compared to the compound exponential with MTTs of 3 and 197, as in the manuscript) much lower than for the gamma distribution used in the manuscript (alpha~0.3, beta~150, MTT~45),

while exhibiting essentially *NO* aggregation error (MTTs are almost equal at 100). And, as also pointed out by the reviewers, there are many more different parameter combinations that fit better than the combination given in the manuscript, while exhibiting partly considerably lower aggregation bias.

5

Reply Our calculations for Figs. 3b and 3c in the original ms. had errors which we have now corrected. We have also omitted the calculations for the gamma model with α = 0.3 from the revised paper (Figs. 3, 4, 5, etc.) because they are not applicable for tritium.

10 **Editor Comment 5)** These points are not yet sufficiently well explained and clarified in the author replies to the reviewers. Given their central importance for the analysis and its interpretation, these points need particular attention when revising your manuscript.

Reply OK

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

Aggregation effects on tritium-based mean transit times and young water fractions in spatially heterogeneous catchments and groundwater systems~~, and implications for past and future applications of tritium~~

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. Applications of simple lumped parameter models to describe aspects of hydrological systems rest on assumptions of homogeneity that are rarely valid. The lumped parameters are supposed to represent the quantities within the system as well as those of the overall system, but such quantities will obviously vary greatly from place to place within heterogeneous systems. Less appreciated is the fact that aggregation errors will affect overall system parameters as well. Kirchner (2016a) recently demonstrated that aggregation errors due to heterogeneity in catchments could cause severe underestimation of the mean transit times (MTTs) of wat~f~er travelling through catchments when simple lumped parameter models 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 over which the biases are manifested are very different; for seasonal tracer cycles it is ~~2-3 months~~ up to about 5 years, while for tritium concentrations it is ~~6-12 years~~ up to about 200 years. 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 <u>which had long series of tritium measurements and</u> in which simple and more complicated lumped parameter models had been used. We find that MTT aggregation errors are small when either <u>the</u> component waters are young (~~less than 6-12 years,~~ as found in many catchments), or component waters have similar MTTs to each other. On the other hand, aggregation errors are large when very young water components are mixed with <u>very</u> old components. In general, well-chosen compound lumped parameter models should be used as they will ~~eliminate~~ <u>reduce</u> potential aggregation errors due to the application of simple lumped parameter models. The choice of a suitable lumped

parameter model can be assisted by matching simulations to time series of tritium measurements (underlining the value of long series of past tritium measurements), but such results should also be finally 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.

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 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?

Hence, examination of the effects of aggregating differing flows on tritium-based MTTs is needed to test the applicability of tritium-based MTTs in catchments and groundwater systems. 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

5

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 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;
5 Seegar and Weiler, 2014; Michel et al., 2015). This difference makes it clear that TTDs determined from seasonal tracer cycles often underestimate the real MTTs in streams. Such underestimation is due to the non-linearity of the damping of the seasonal tracer cycles as noted above. Similarly, radioactive decay of tritium is a non-linear process and therefore spatial aggregation errors are expected when water components with different ages 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
10 (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 ~~impossible~~ difficult to calibrate an LPM uniquely when we only have data for ~~the one~~ 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 ~~required~~ helpful to (i)
15 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
20 validation in order to examine if the model used is applicable to the real ~~natural~~ 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 dissimilar water components and comparing the true mixed MTTs with the tritium-inferred apparent MTTs, as Kirchner
25 (2016a) did 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 ~~0.3~~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 tested. We also surveyed some applications of tritium dating from the literature~~, which~~to see if they had sufficient data to constrain the parameters of mixtures of young and
30 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.

6

Our experiments did not include examination of non-stationary hydrological systems. Kirchner (2016b) had found similar underestimation of MTTs and robustness of young water fractions based on seasonal tracer cycles in a non-stationary system (a two-box model). We briefly consider non-stationary aggregation effects in the light of the differences in the tritium and seasonal tracer cycles methods.

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 (LPM), which reflects the average (steady-state) conditions in the catchment.

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 a 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

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

where C_{in} and C_{out} are the input and output concentrations in the precipitation 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 input (C_{in}) was different in each hemisphere. 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 as examples in Fig. 1. 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

7

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 ~~to~~ 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 ~~much~~ 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.

Several simple flow models (LPMs) are commonly used in tracer studies. The piston flow model (PFM) describes systems ~~for~~ in which ~~all outputs have the same transit time~~all of the water in the output has the same transit time (MTT or τ_m). ~~, i.e. the outputs are not combinations of different component fluxes with different transit times.~~ The TTD is

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

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), and~~The~~ output tritium 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 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 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

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

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

8

where f is the ratio of the exponential volume to the total volume. (Maloszewski and Zuber (1982) used the parameter η; 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 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 ~~GM~~gamma model, ~~, EPM~~the exponential piston flow model with end members ~~PFM~~piston flow and ~~EM~~exponential models, and the ~~DM~~ dispersion model LPMs) and *compound LPMs* (binary or other parallel combinations of simple LPMs). Simple LPMs describe homogeneous systems, but have been widely 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) (~~Compound LPMs have often been used to represent more complicated systems~~ (e.g. Maloszewski et al., 1993; Stewart and Thomas, 2008; ~~Morgenstern et al., 2015~~ Blavoux et al., 2013; Morgenstern et al., 2015). The binary parallel LPM is given by

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

where LPM_1 and LPM_2 are simple LPMs representing two water components contributing to the output. 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 t~~The~~ parallel combination of two ~~EM~~ exponential models ~~in parallel.~~ 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)~~.~~ It is given by

$$h(\tau) = \frac{b}{\tau_f}\exp\left(-\frac{\tau}{\tau_f}\right) + \frac{(1-b)}{\tau_s}\exp\left(-\frac{\tau}{\tau_s}\right) \tag{810}$$

where b is the fraction of the young component (fast reservoir), and 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

$$\tau_m = b\tau_f + (1-b)\tau_s \tag{911}$$

9

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, we perform a virtual experiment 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 tritium concentrations of the mixed system (C_m) are given by This section examines the effect of combining two different water components on the MTT of the mixture. The 'true' MTT is determined by mixing the two waters according to Eq. (9) with b being the fraction of the first component in the mixture ((1-b) is the fraction of the second component), and τ_1 and τ_2 are the MTTs of the two components. The 'apparent' MTT is determined by fitting a simple LPM (the GM) to the tritium concentrations of the mixture (C_m) given by

$$C_m = b.C_1 + (1 - b).C_2 \hspace{3cm} (12\text{1})$$

where C_1 and C_2 are the tritium concentrations in components 1 and 2 respectively 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., i.e equal fractions of both components. 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 ages between zero and a young water threshold (t_y), i.e.

$$Y_f = \int_0^{t_y} h(\tau).d\tau \hspace{3cm} (13\text{2})$$

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 has been estimated as the value that gives agreement between the apparent and true young water fractions to within about 10%. This included the case of the apparent and true young water fractions for 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

10

what follows. This case gives the greatest difference between the apparent and true MTTs. As the threshold increases, the apparent Y_f increases relative to the true Y_f. The "true" Y_f is determined by mixing the two waters according to the equation

$$Y_{ftrue} = b \cdot Y_{f1} + (1-b) \cdot Y_{f2} \hspace{3cm} (14)$$

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 from an LPM fitted to the tritium concentrations of the mixture (Eq. 12). B is assumed to be 0.5. b is assumed to be 0.5. Tritium aggregation biases are much smaller for young water fractions than for MTTs (i.e. the points plot much closer to the 1:1 line) as observed by Kirchner (2016a) for seasonal tracer cycles. Fig. 6 gives the results calculated with the GM with different shape parameters

2.2 4 Comparison of transit time distributions of different flow models

The TTDs transit time distributions of the four three cases of the GM gamma model investigated in this work are illustrated in Fig. 2a, as normalised probability density functions (i.e. h(τ) x τ_m) versus normalised transit times (τ/τ_m). These cover a wide the range of TTDs shapes observed in streams using seasonal tracer cycles, and streams and groundwater using tritium concentrations. They are also approximately representative of the other simple flow models described above. The GM gamma model case with α=0.3 has 'heavy' tails (short and long transit times are strongly emphasised compared with transit times close to the mean). Case α=1 is the exponential distribution (linear storage); the same as the exponential piston flow model with f = 1. Gamma model Ccases with α=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).

Distribution shapes described by the range of α values between 0.3 and 1.0 for the GM have been found useful for interpreting seasonal tracer cycles in streams (Kirchner, 2016a), but have generally not been as effective for interpreting tritium concentrations in stream baseflow or groundwater (see below). Of the simple LPMs, the EPM has no equivalent shapes to the GM for α less than 1 and the DM has none for α less than 0.8 (Table 1). The standard deviation (sd) and Nash Sutcliffe efficiency (NSE) are is used to quantify the goodness-of-fit between the GM gamma model (GM_i) and the best-fitting version of each of the other models (LPM_i), where

$$sd = \sqrt{\frac{\sum_1^N (GM_i - LPM_i)^2}{N}} \hspace{3cm} (15)$$

The most basic of compound LPMs (the TPLR model) gives TTD shapes which are comparable with the GM shapes in this range of α (Fig. 1d). and $$NSE = 1 - \frac{\sum_1^N (GM_i - LPM_i)^2}{\sum_1^N (LPM_i - \overline{LPM_i})^2} \hspace{3cm} (16)$$

The NSE efficiency can vary between -∞ 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 GM, EPM and TPLR models give the exponential model when they have the appropriate parameters ($\alpha = 1$ for GM, $f = 1$ for EPM, and both τ_f and τ_s equal to τ_m for TPLR). 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 GM gamma model with α between 1 and 10 are equivalent to EPM exponential piston flow model shapes with exponential fractions (f) between 1.0 and 0.4 (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 DM dispersion model has dispersion parameters (P_D) between about 2.5 and 0.05 corresponding to the GM gamma model with α between 0.8 and 10 (Table 1). The GM gamma and EPM exponential piston flow model shapes become less similar to each other as α increases to 10, while the GM gamma and DM dispersion model shapes become more similar. The TPLR does not mimic the GM in this range (the EM version ($\tau_f = \tau_s = \tau_m$) gives the best fits to the GM with α greater than 1), and probably gives more realistic shapes than the simple LPMs when there are two major water components present in streams.

3 Results

3.1 Aggregation effects on mean transit times determined using tritium

This section examines the effect of combining two different water components on the MTT of the mixture. The 'true' MTT is determined by mixing the two waters according to Eq. (9) with b being the fraction of the first component in the mixture ((1-b) is the fraction of the second component), and τ_1 and τ_2 are the MTTs of the two components. The 'apparent' MTT is determined by fitting a simple LPM (the GM) to the tritium concentrations of the mixture (C_m) given by

$$C_m = b.C_1 + (1-b).C_2 \hspace{3cm} (11)$$

where C_1 and C_2 are the tritium concentrations in components 1 and 2 respectively. b is assumed to be 0.5 for simplicity in what follows, i.e equal fractions of both components. Following Kirchner (2016a), we did not consider evapotranspiration in our analysis of tritium aggregation effects.

3.1.1 Relationships between mean transit time and tritium concentration

We first demonstrate the relationships between The relationship between mean transit time and tritium concentration for mixed systems (is illustrated in Fig. 3) for the assumption of by assuming constant input tritium concentration of 2 TU over time, i.e. without the bomb pulse during the nuclear age. This simplifying assumption is necessary to allow for the analysis shown in Fig. 3; with the real peaked input the figures would be scrambled 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 ~~piston flow model~~gamma model with shape factor α = 1. The red points indicate the assumed water components (with MTTs of 3 and 197 year~~s~~ respectively) and the red dashed line is the mixing relationship
5   between them (described by Eqs. 9 and 12~~1~~). The "true" MTT (100 year~~s~~) of a 50:50 mixture of the components is shown on the red dashed line. The black curve is the result of applying the ~~piston flow model (PFM)~~gamma model with α=1 to the mixed tritium concentrations (Eq. 12~~0~~). ~~The PFM describes a flow system in which tritium decays radioactively with a half-life of 12.32 yr, but there is no mixing so that all of the water has the same age.~~ A 50:50 mixture of the components gives the 'apparent' MTT shown (20.5~~15~~ year~~s~~), which is much less than the "true" MTT. This results from the strongly non-linear
10   character of the black curve (Fig. 3a) and therefore combining two dissimilar subcatchments 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~~-c~~ d show the same calculations applied to the ~~GM~~ gamma models with α = 3 and 10, ~~different shape factors~~ and the piston flow model (PFM). The different shape factors describe different fractional contributions of past water inputs to the
15   present water output as illustrated by the transit time distributions~~TTDs~~ in Fig. 2a. ~~Application of the GM takes account of input of new water (with tritium concentration of 2 TU) mixing with old water already present in the system with lower tritium concentrations because of radioactive decay. The heavy-tailed TTD (α = 0.3, Fig. 3b) causes a flattening of the curves but there is still a considerable difference between the true and apparent MTTs (100 to 44 yr). The exponential TTD (α = 1) shows a larger difference (100 to 26 yr), and t~~The ~~more peaked TTDs~~ gamma models with ~~(~~α = 3 and 10~~)~~ have greater differences
20   between true and apparent MTTs.~~still.~~ The piston flow model is the most sharply peaked of all, and describes a flow system in which tritium decays radioactively with a half-life of 12.32 yr, but there is no mixing so that all of the water has the same age (Fig. 2a). It ~~TTD (α = 10)~~ ~~approaches the result for the PFM~~has a true-apparent MTT difference of 100 to 15 years. ~~The PFM describes a flow system in which tritium decays radioactively with a half-life of 12.32 yr, but there is no mixing so that all of the water has the same age.~~

25

3.1.2 Aggregation effects on mean transit times determined using tritium

~~Fig. 4 compares t~~The true versus the apparent MTTs calculated using the real ~~peaked~~ tritium input function from Kaitoke (Fig.1) are given in Fig. 4. The calculations were structured as virtual experiments in which 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 ~~(Eqs. 9~~
30   ~~and 11)~~ 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 MTT~~s~~ to become younger than the true MTT~~s~~ 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

13

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 α = 0.3 1 showing smaller aggregation effects, and are most divergent for α = 10. Errors of fitting for determining
5 the apparent MTTs (expressed as standard deviations (Eq. 15)) are greatest when component 1 is youngest, these are shown by fine dashed lines for 3 and 25 years above and below the curves. The errors are largest with α = 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).

10 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 almost no 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 50 75 tritium values (one for each year) and this constrains the final 'apparent' fitting to a single unique
15 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.

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 is are examined in this section. As before, the two water components were initially assumed to have the same MTTs and young water
20 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. 6 and 7. 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. 6 and 7). 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
25 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.

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

30 $$Y_f = \int_0^{t_y} h(\tau).d\tau \qquad (12)$$

14

The young water threshold for tritium has been estimated as the value that gives agreement within 10% of the apparent and true young water fractions for 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). This case gives the greatest difference between the apparent and true MTTs. As the threshold increases, the apparent Y_f increases relative to the true Y_f. The 'true' Y_f is determined by mixing the two waters according to the equation

$$Y_{ftrue} = b.Y_{f1} + (1-b).Y_{f2} \qquad (13)$$

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 from an LPM fitted to the tritium concentrations of the mixture (Eq. 11). b is assumed to be 0.5. Tritium aggregation biases are much smaller for young water fractions than for MTTs (i.e. the points plot much closer to the 1:1 line) as observed by Kirchner (2016a) for seasonal tracer cycles. Fig. 6 gives the results calculated with the GM with different shape parameters

The young fractions are dependent on the choice of the young water thresholds used in Eq. (12) because water with transit times longer than the thresholds are omitted from the calculation. However, the result would hold with any reasonable and consistent choice of the young water thresholds, because the normally underestimated water with old ages (i.e. the long tail portions of the TTDs) would be cut off in all cases. They would then not be causing the apparent young fractions to deviate from the true young fractions. On the other hand, the MTTs include the underestimated long tails and therefore give discrepant results.

The young water thresholds (t_y) used for these calculations are plotted against α in Fig 8. The method used for determination of t_y is described earlier in this paper. The points fit a power law given by

$$t_y = 11.6\alpha^{-0.274} \qquad (14)$$

The reason for this relationship, which is similar to that found by Kirchner for seasonal tracer cycles, is to be found in the gamma distribution. It is important to note the differences in resulting thresholds between tritium and stable isotope tracers. For stable isotopes, Kirchner (2016a) reported a young water threshold range from 0.1 to 0.25 years (or approximately two months) for the ~~GM~~ gamma model shape factor α ranging from 0.2 to 2~~, respectively~~. From our tritium evaluation, the young water threshold of tritium-based transit times ~~ranged from 16 to 6~~was 18 years with shape factors from ~~0.3~~1 to 10. ~~In addition, this relationship is also applicable to estimate the young water fractions of other LPMs with parameters corresponding to α in Table 1.~~

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.

15

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 ~~(GM)~~. Fig. ~~9~~ 8

5 shows the aggregation effects for the ~~effective range of GM shapes~~gamma model ~~for streams (~~with α ~~=~~ 1~~between 0.3 and 1.0)~~. The~~se~~ pattern~~s~~ ~~are~~ 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.

~~3.~~4~~4~~ Case studies from the literature

The calculations above have shown that fitting simple LPMs to tritium data can potentially cause significant (or severe)

10 underestimation of the MTT when two or more dissimilar tributaries feed a sampled outlet. On the other hand, the young water fraction is not likely to be affected by such errors ~~(as also found for seasonal tracer cycles by Kirchner, 2016a)~~. 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

15 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 parallel binary 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

20 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.

~~3.~~4.1 Two water components at Waikoropupu Springs (karstic springs fed by Arthur Marble overlain by Tertiary

25 **~~fc~~sediments)**

Tritium measurements at the Waikoropupu Springs began in 1966 and cover almost the rise and fall of the tritium bomb ~~spike~~ peak in precipitation (Stewart and Thomas, 2008). Fig. ~~10a~~ 9a 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 (~~EPM~~exponential piston flow and ~~DM~~dispersion models) and ~~a~~ two compound LPMs (~~the~~ double exponential piston flow (DEPM) and ~~DM~~

30 dispersion ~~or~~ (DDM) models).

The ~~DDM~~ compound models w~~ere~~as used because <u>flows,</u> δ^{18}O and Cl measurements showed that there were two separate water systems contributing to the Main Spring (a shallow system and a deep system, see Stewart and Thomas<u>,</u> ~~(~~2008~~)~~). <u>The fraction of the shallow system (b = 0.24) contributing to the Main Spring was determined from a balance model based on the flows, δ^{18}O and Cl concentrations.</u> The ~~residence~~ <u>transit</u> time distributions of the models have similar shapes with peaks of very young water and long tails of much older water (Fig. ~~10b~~<u>9b</u>). ~~The variation of the quality of the fits with MRT is shown in Fig. 9c, with the goodness-of-fit being expressed as the standard deviation (sd) of the simulations about the measurements.~~

All ~~three~~ models gave good fits to the data, and the mean residence times were ~~sharply~~ constrained close to 8 years (Table 2~~, Fig. 10c~~). The best-fitting ~~EPM~~ <u>exponential piston flow</u> and ~~DM~~ <u>dispersion</u> models had MTTs of 7.9 and 8.2 yr respectively. <u>The DEPM model fitted well and gave MTTs of 0.1 and 10.2 years for the two components. δ^{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 MRT 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 estimated separately).</u> ~~The DDM model fitted very well indeed (sd 0.08 TU) with overall MRT of 7.9 yr from Eq. 9. This model had 26% water from the shallow system with MRT = 1.2 yr, and 74% from the deep system with MRT = 10.2 yr.~~

There is little aggregation error in the MTTs in this case (i.e. the 'apparent' MTTs of the ~~EPM~~ <u>exponential piston flow</u> and ~~DM~~ <u>dispersion</u> models are very similar to the 'true' overall MTT<u>s</u> of the <u>DEPM and</u> DDM model). This is because both systems are young in relation to the young water threshold for tritium ~~applying to this case~~<u>(18 years).</u> ~~(t_y is 11 yr).~~ The young fractions are also similar to each other at about 0.7.

~~3.~~4.2 Kuratau River (volcanic ash deposits and andesite)

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). The best EPM <u>and DM</u> model<u>s</u> (~~almost an EM,~~ Table 2<u>3</u>) had ~~an~~ MTT<u>s</u> of 4 year<u>s</u> but fitted relatively poorly to samples collected around 1970 (Fig. 10a<u>1</u>), while ~~a~~ compound mixing model<u>s</u> (double EPM <u>and DM models</u>) with overall MTT<u>s</u> of 15<u>1</u> <u>and 19</u> years fitted much better. This comprised ~~65~~<u>about 50</u>% of a <u>simple model</u>~~n EM~~ with MTT<u>s</u> of <u>less than</u> 1 year and 3<u>50</u>% of a~~n EPM~~ <u>model</u> with MTT of 3<u>50</u> years. The young~~er~~ component is believed to be derived from drainage from the ~~almost~~ <u>very</u> impermeable <u>lavas (Whakamaru Group ignimbrites and</u> andesite<u>s and basaltic lavas)</u> ~~in the catchment~~, while the older component comes from very porous <u>volcanic ash deposits (Taupo/Oruanui ignimbrites and tephras) in the catchment (Morgenstern, 2007).</u> ~~volcanic ash deposits; distributed~~ <u>Distributed</u> groundwater models calibrated with groundwater levels, river discharges and tritium concentrations substantiated these flows (Gusyev et al., 2013; 2014). Strong aggregation bias is shown by the marked difference in MTTs between the simple ~~EPM~~ model<u>s</u> and the compound model<u>s</u> (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, although this is still possible in the Northern Hemisphere. The young fractions ~~of 0.93 and 0.65 were determined from the TTDs with threshold t_y of 10 yr~~ are about 0.7 (Table 3~~2~~).

~~3.~~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. The best fitting ~~EPM~~ exponential piston flow model (EPM) for Hangarua Spring has MTT of 5~~6~~8 years but fits poorly to the measured data, while the compound model (DEPM) with MTT of ~~90~~116 years fits well (Fig. ~~12a~~11a, Table 2). It consists of 35% 1~~8~~6-year-old water and 65% 1~~67~~30-year-old water (f parameters listed in Table 3~~2~~). A moderate aggregation bias is demonstrated by the difference in MTTs (Table 3~~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 ~~a~~ moderate aggregation bias (apparent to true MTTs of ~~70~~61 and 1~~4~~24 years with the EPM and 76 to 160 years with the DM) (Table 2, Fig. 12a,b). The compound model~~s~~ consist~~s~~ of about 35% of 12-year-old water and 65% ~~185~~220-year-old water. The young fractions are all about 0.2~~3~~15 based on t_y of 1~~8~~0 years (Table 3~~2~~).

~~3.~~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. 13a). 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 13b). (The DEPM curve shown in Fig. 13a 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.

4 5 Discussion

45.1 Implications of tritium MTT aggregation bias

The analysis of Sections 3.1 and 3.2 haves 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, there is an important difference between the biases of these methods and that is the time periods over which the biases are manifested. The bias applies for transit times greater than 2-3 months (i.e. the young water threshold)up to about five years for seasonal tracer cycles (Kirchner, 2016a), whereas it applies for times greater than 6 16up to about 200 years for tritium as demonstrated here, with the values of both depending on the GM shape factor. 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. In addition, the bias is particularly severe when one component is especially young relative to the young water threshold (2-3 months for seasonal tracer cycles and 18 years for tritium), unless both components are especially young.

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

The answer to this depends on the nature of the heterogeneity. The type of heterogeneity leading particularly to aggregation bias is heterogeneity that produces flows with very different MTTs. The extreme case is when the catchment or groundwater system is divided into two parts and the flows from each part have very different transit times. This is the case examined by the virtual experiments in Sections 3.1 and 3.2. As noted by Luther and Haitjema (1998), such 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 have been chosen. Many catchments have outflows composed of quickflow and baseflow, which could also have very different MTTs leading to large aggregation errors.

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

19

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 is ideally suited for 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 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 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 most concern for aggregation error.

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

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. because the values of the MTTs determined were between 2-3 months and 4-5 years. But, we contend that tritium studies will have been affected far less, despite aggregation bias also applying to tritium-derived MTTs, because:

(1) Tritium-derived MTTs can only have been affected by aggregation bias if they are greater than 6-16 years (or 6-12 years for α in the range 1 to 10) and were determined using simple LPMs. Many stream studies have shown that the waters are young. Stewart et al. (2010) surveyed most of the tritium studies on streamflow up to then. They found that MTTs determined in baseflow averaged about 10 ± 8 years for headwater catchments (with two outlying volcanic ash flow catchments omitted) and 10 ± 5 years for large rivers. Baseflow averaged about 50% of the total flows from the catchments in both groups. In addition, Michel et al. (2015) concluded that "the range of residence times for active water in the majority of [surface] hydrologic systems throughout the world is on the order of one to two decades" based on a surface water database (GNIR) of over 6500 measurements of tritium assembled by the IAEA from measurements made between the late 1940s and the present. These suggest that aggregation bias is not important for many streamflow studies because of the transit times involved.

(2) 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-chosen based on the characteristics of the catchments, they will produce more accurate TTDs than the simple LPMs and therefore will eliminate or reduce aggregation bias on MTTs.

A very 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 an EM ($\tau_m = 8$ yr) and DM ($\tau_m = 60$ yr) in series, with a small bypass flow in parallel with them, followed

20

by a PFM (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. The studies also illustrate the first point above, in that not much aggregation error is expected in the tritium MTTs for these streams because all of the identified water components are young relative to the young water threshold age (6-16 yr) for tritium.

45.3 Aggregation effects due to non-stationarity in systems

This study has not looked specifically at aggregation effects due to the non-stationary nature of hydrological systems. These will be different for catchments and groundwater systems, with streamflow variations often being far more dynamic than those in flows from wells and springs.

Methods of determining TTDs from tritium concentrations are quite different from those used for seasonal tracer cycles. The latter method in principle requires a series of samples from both input and output of a hydrological system in order to determine the reduction in their variation during transport through the system. Using assumptions about mixing, this damping is then interpreted to give a TTD that is characteristic of a stationary system. But flows through hydrological systems such as catchments are never stationary because they are driven by intrinsically variable rainfall. Consequently, seasonal tracer cycle methods produce TTDs which are averages of the TTDs during the period of sampling. However, methods have now been developed for stable isotope/chloride variations which allow determination of time-variant TTDs (Botter et al., 2010; Rinaldo et al., 2011; Hrachowitz et al., 2013).

On the other hand, use of tritium for determining TTDs depends on its radioactive decay rate which is applicable to single samples. Hence a series of tritium measurements can in principle yield a series of TTDs and the behaviour of a catchment during a range of hydrological conditions can be investigated. Interpretation of single tritium measurements to yield TTDs is of course not necessarily straight-forward, because:

(1) The type of LPM and parameter value to apply to a single sample needs to be assumed (e.g. the shape parameter pre-determined). The specification of the LPM has often been determined by fitting LPMs to time series of tritium samples (as described in the case studies, Section 3.4), which obviously cannot be done with single samples. This work shows that it would be dangerous in terms of aggregation error to use a simple LPM for this. Hence the recommended procedure would be to

21

sample several tritium samples separated in time at each of a number of different streamflows and experiment with fitting different LPMs at each of the flows.

(2)   Several values of the age parameter may allow the simulation to fit the measurement (i.e. there can be multiple solutions or ambiguous ages) because of input variations resulting from nuclear weapons testing in the past. This still applies for sites in the Northern Hemisphere, but is now largely past for the Southern Hemisphere (Stewart et al., 2012; Stewart and Morgenstern, 2016; Gallart et al., 2016).

45.4   Considerations for future use of tritium MTTs

This work sounds a death knell 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, unless the simple LPMs are based on long series of tritium measurements. Seasonal tracer cycles should probably not be interpreted at all using simple LPMs. Fortunately, there can be good reasons for choosing compound LPMs which are more realistic than simple LPMs. 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).

5 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 over which the biases are manifested are very different; for seasonal tracer cycles it is 2-3 months up to about 5 years, while for tritium concentrations it is 6-12 years up to about 200 years. 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 more complicated LPMs had been used. We find that MTT aggregation errors are small when either all of the component waters are young (less than 6-12 years, as found in many catchments), or they have similar MTTs to each other. On the other hand, aggregation errors

22

are large when very young water components are mixed with very old components. In general, well-chosen compound LPMs should be used as they will ~~eliminate or~~markedly 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 measurements (underlining the value of

5   long series of past tritium measurements), but such results should also be ~~finally~~ 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
10   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
15   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.

20 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
25   conditions, In press, Hydrol. Process. (Keith Beven Tribute), 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
30   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

23

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., 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. 2004: 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., 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.

25

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) and two parallel linear reservoirs (TPLR) 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. Blank cells indicate α values for which no fits with sd < 0.25 can be found.

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
TPLR τ_f/τ_m	1.00	1.00	
TPLR b	0 – 1	0 – 1	
sd	0.00	0.21	

5

Table 2: Results of model simulations for the Main Spring of the Waikoropupu Springs. The young water fraction (Y_f) is the fraction of water with ages less than the young water threshold age.

Model (parameters)	Shallow fraction b	Mean transit time τ_m (yr)	Young fraction Y_f	Standard deviation sd (TU)
EPM (f=0.95)	--	7.9	0.75	0.40
DM (P_D=1.8)	--	8.2	0.71	0.29
DDM (τ_s=1.2 yr, P_{Ds}=0.12) (τ_d=10.2 yr, P_{Dx}=0.6)	0.26	7.9	0.65	0.08

10

Table 3: Parameters of simple and compound LPMs applied to tritium measurements from the Kuratau River, Hangarua Spring and Hamurana Stream.

Feature	Model	b	Overall MTT τ_m (yr)	Y_f	Component 1 τ_1 (yr)	f_1	Component 2 τ_2 (yr)	f_2
Kuratau River	EPM	1.00	4	0.93	4	0.99	--	--
	DEPM	0.65	11	0.65	1.0	1.00	30	0.50
Hangaroa Spring	EPM	1.00	58	0.10	58	0.91	--	--
	DEPM	0.35	90	0.17	16	0.63	130	0.87
Hamurana Stream	EM	1.00	70	0.14	70	1.00	--	--
	DEPM	0.35	124	0.21	12	0.77	185	0.82

15

20

27

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 age.**

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)
Waikoropupu Main Spring	EPM	7.9	1.00	7.9	0.94	--	--	--	--	0.36	0.86
	DEPM	7.8	0.24[1]	0.1	0.50	--	10.2	0.79	--	0.22	0.61
	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
Kuratau River	EPM	4.4	1.00	4.4	0.89	--	--	--	--	2.9	0.88
	DEPM	15.2	0.55	0.8	1.00	--	33.0	0.50	--	1.6	0.66
	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
Hangarua Spring	EPM	56	1.00	56	0.90	--	--	--	--	0.20	0.27
	DEPM	116	0.35	17.9	0.69	--	167	0.89	--	0.07	0.24
	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
Hamurana Stream	EPM	61	1.00	61	0.99	--	--	--	--	0.28	0.26
	DEPM	144	0.32	13.3	0.80	--	204	0.82	--	0.02	0.26
	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
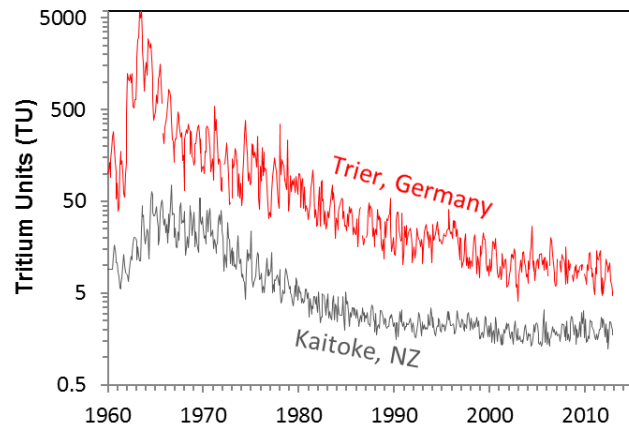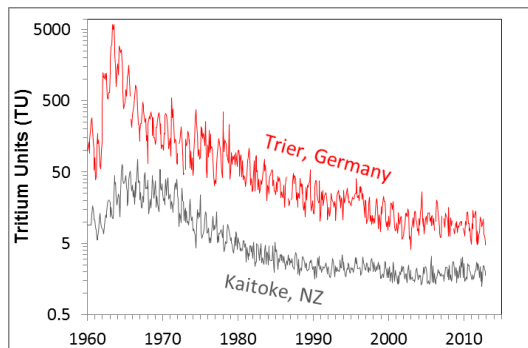
Figure 1. Tritium concentrations (TU) in monthly precipitation samples at Kaitoke, New Zealand in the Southern Hemisphere, and Trier, Germany in the Northern Hemisphere.
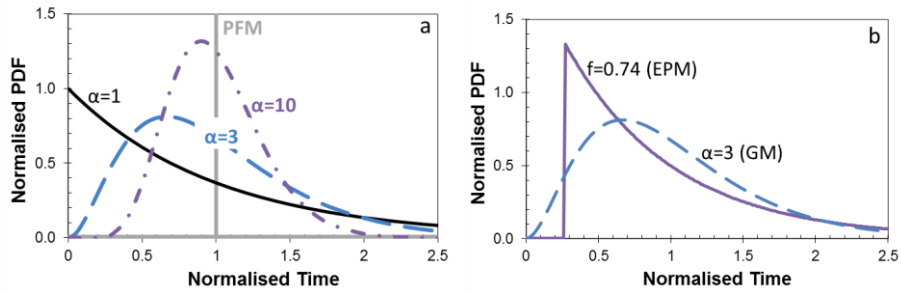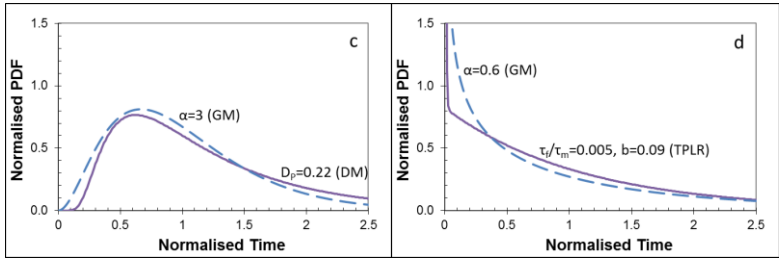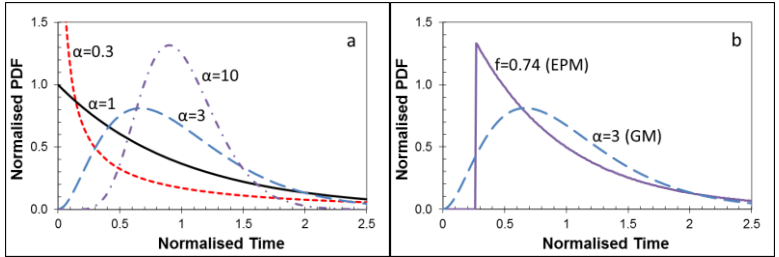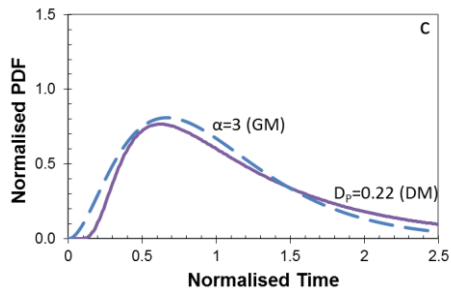
Figure 2. (a) Gamma distributions (GM) for shape factors α between ~~10,~~3 and 10. The axes show normalised transit time (τ/τ_m) and normalised ~~probability density function~~ PDF ($h(\tau)$ x τ_m). Note that the distribution for GM ($\alpha = 1$) is the same as that for the exponential model (EM). **(b-~~cd~~)** Comparison between ~~particular GMs~~ the gamma model with $\alpha = 3$ and the best fitting ~~EPM~~exponential piston flow and~~,~~ ~~DM~~ dispersion ~~and TPLR flow~~ models.
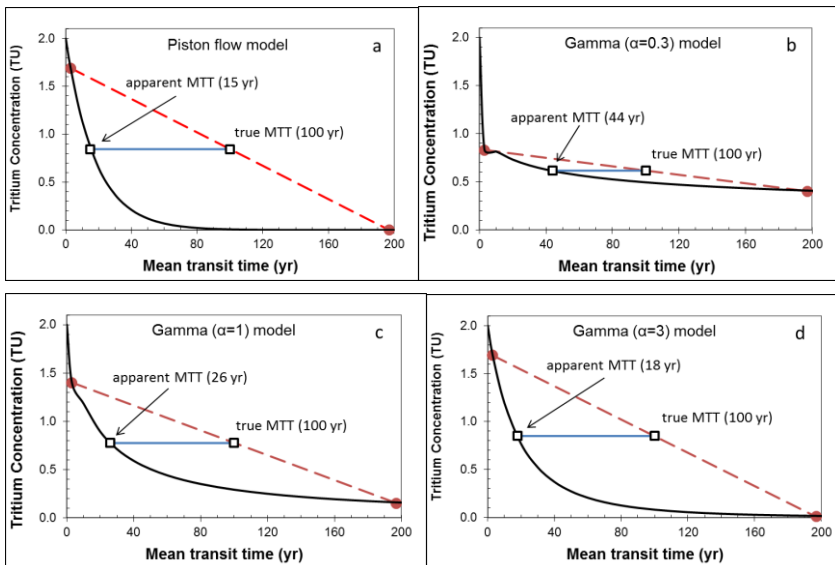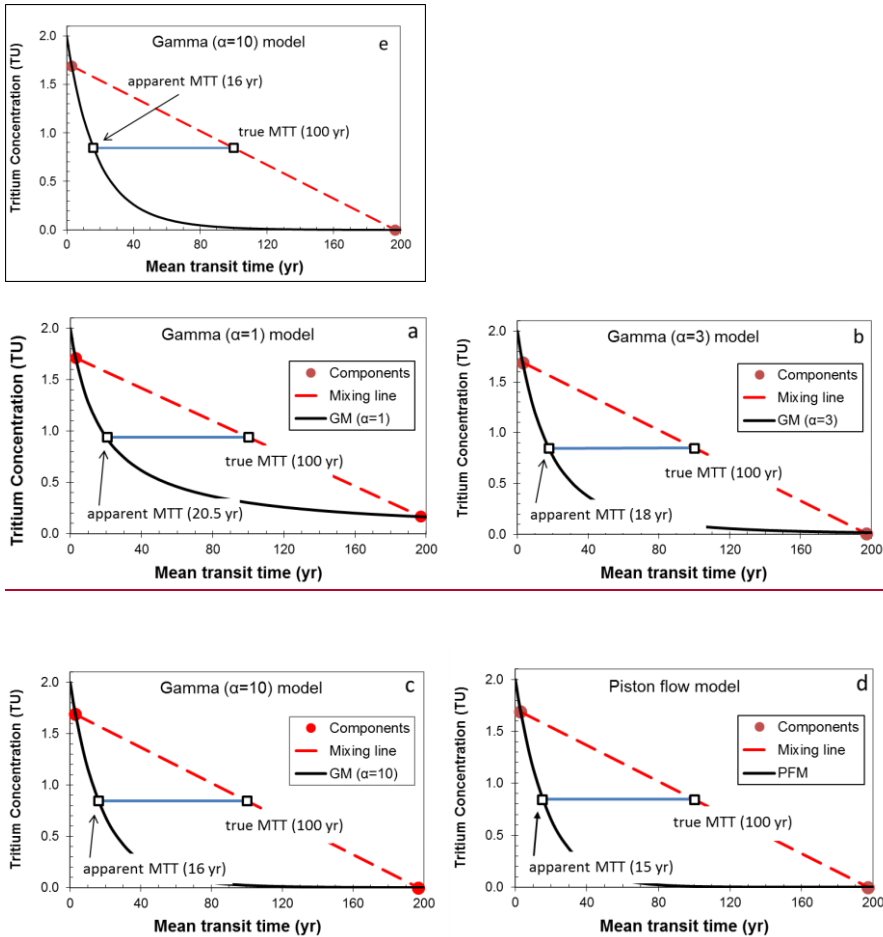
Figure 3(a-ed). Aggregation errors when the tritium input concentration is assumed to be constant at 2 TU. when mMean transit times (MTTs) is are inferred from tritium concentrations in mixed runoff from two subcatchments with different tritium concentrations and MTTs (shown by red dots) using a range of gamma and the piston flow and a range of gamma models. The tritium input concentration is assumed to be constant at 2 TU for clarity. 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.
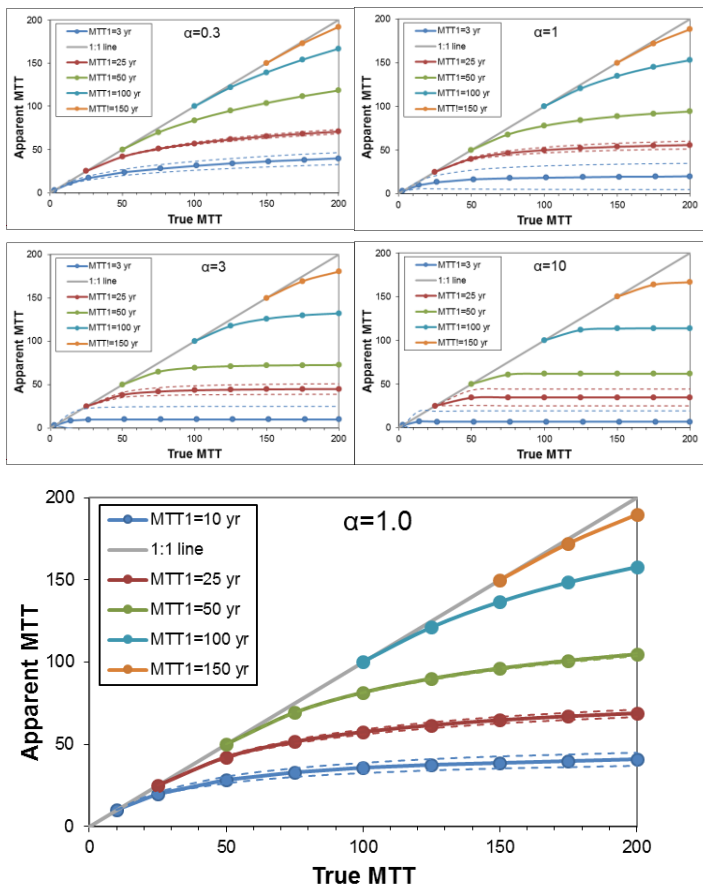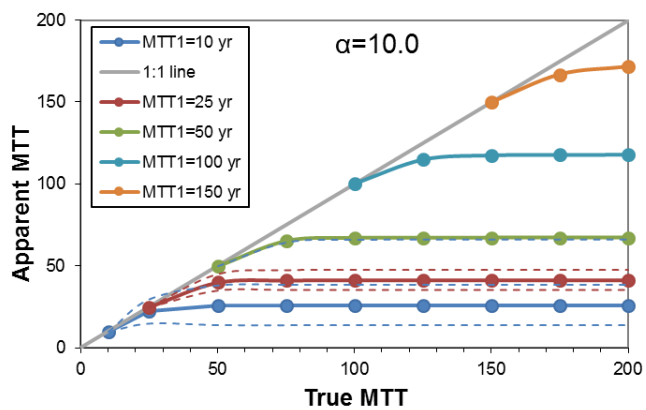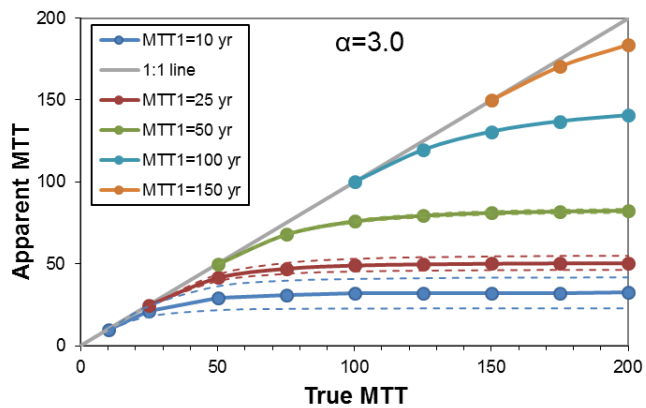
Figure 4: Aggregation effects for tritium MTTs ~~for gamma models with different values of α~~ using the Kaitoke input function. ~~Curves show changes as component 2 (MTT2) becomes older in 50-year steps (shown by dots).~~ Fitting errors in the apparent MTTs are shown ~~for MTT1 = 3 and 25 yr~~ by ~~fine~~ dashed lines.
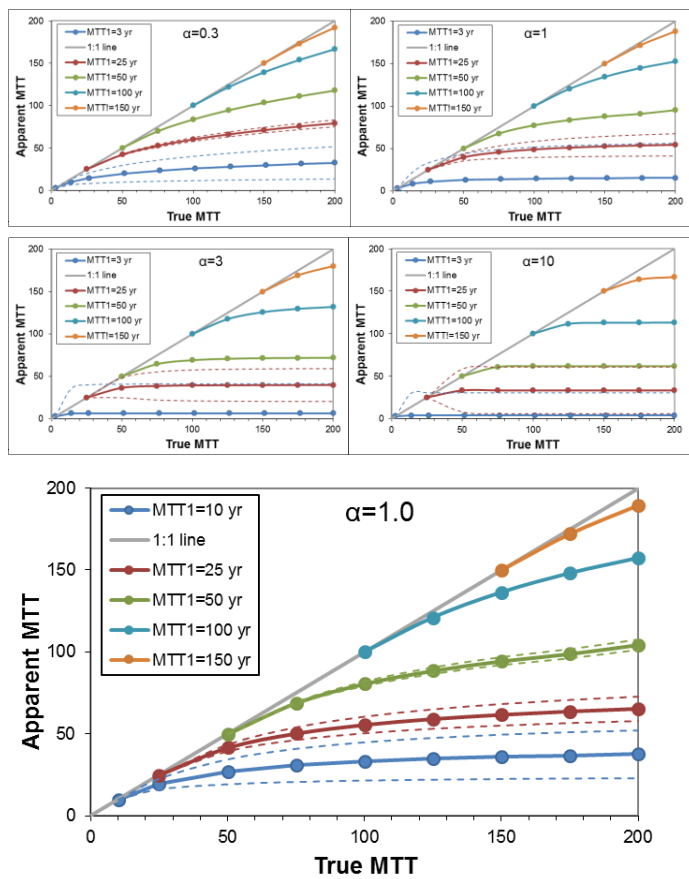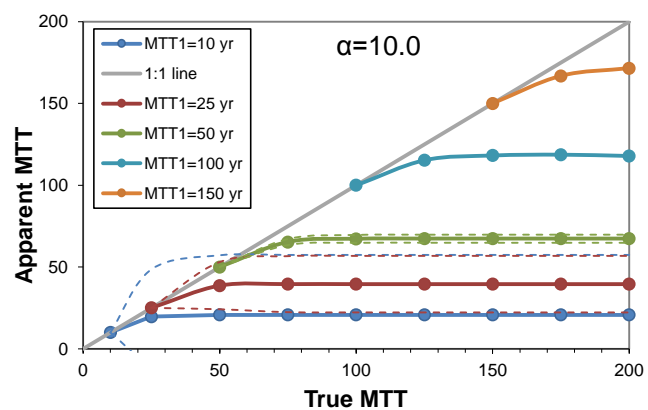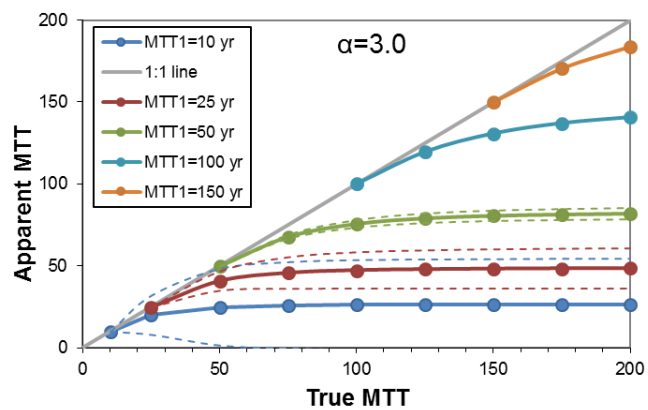
Figure 5. Aggregation effects for tritium MTTs using the Trier input function. ~~Symbols as in Fig. 4.Fitting errors in the apparent MTTs are shown for MTT1 = 3 and 25 yr by dashed lighter lines.~~
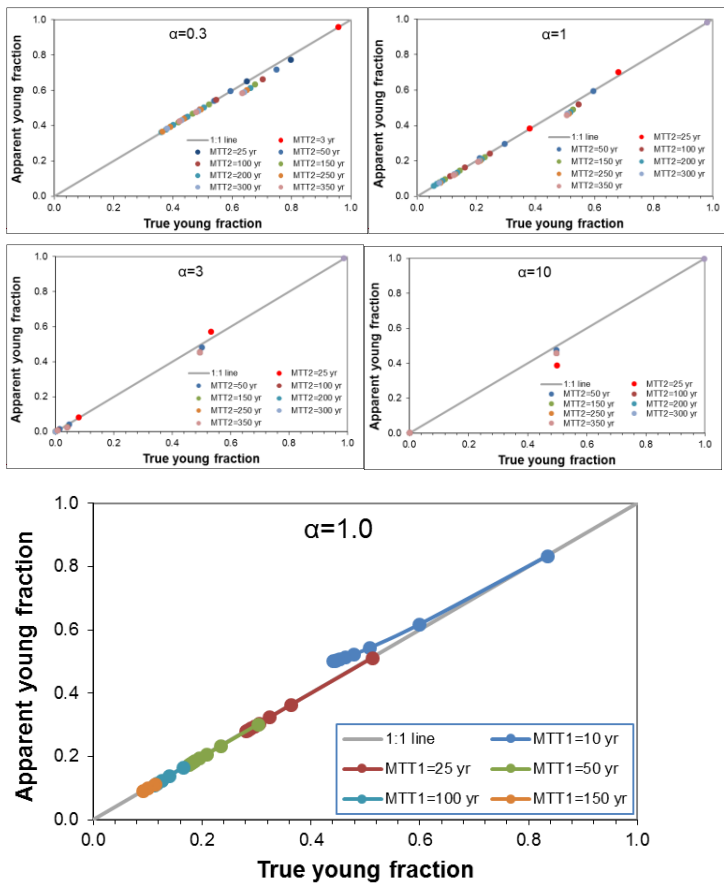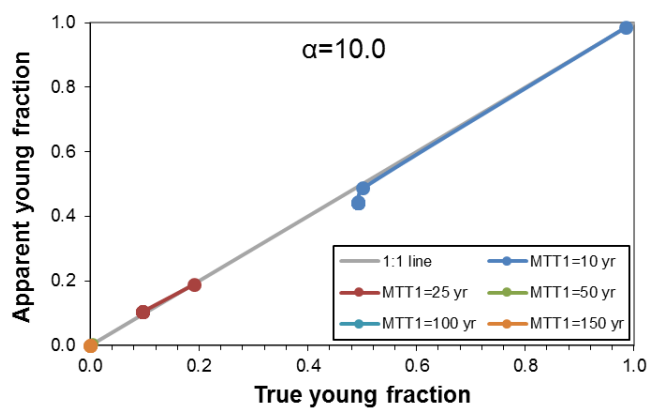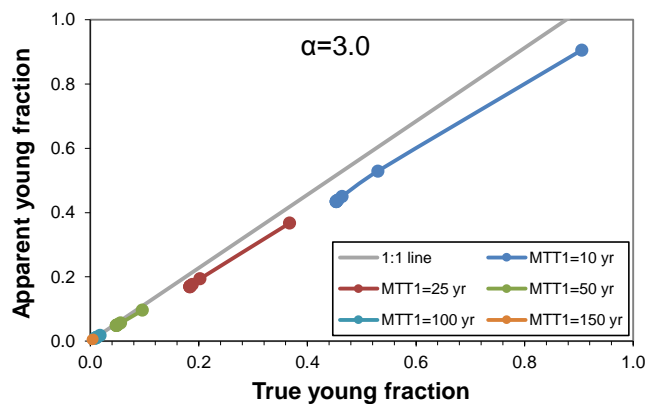
Fig. 6. True versus apparent tritium young water fractions for ~~gamma models with~~ different values of α ~~with the GM for~~ using the Kaitoke input function. Curves show changes as component 2 (MTT2) becomes older in 50-year steps (shown by dots).
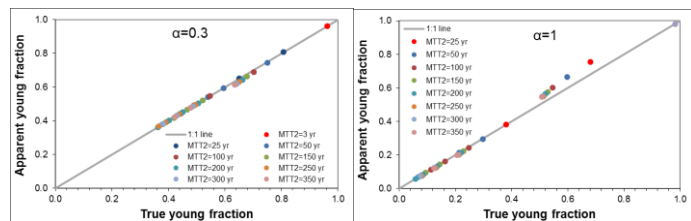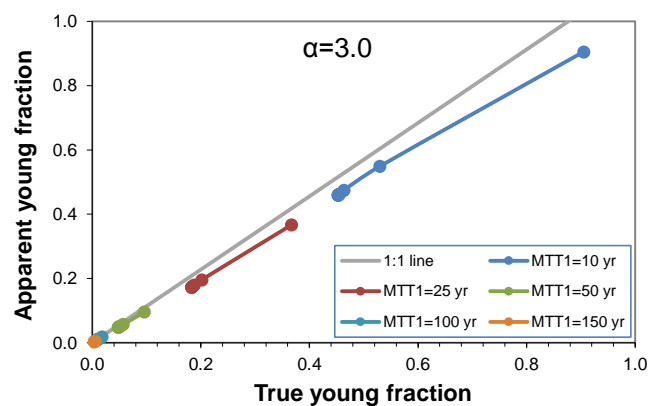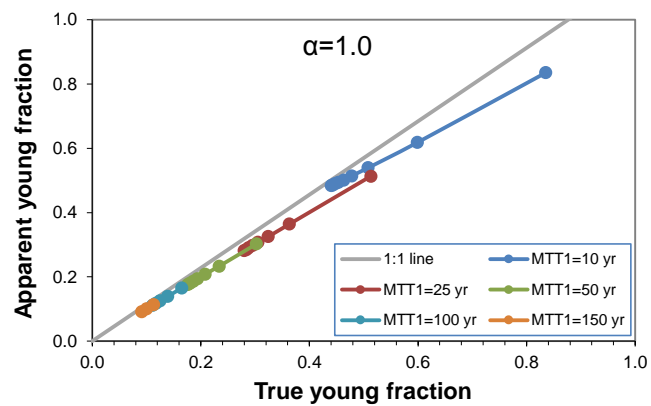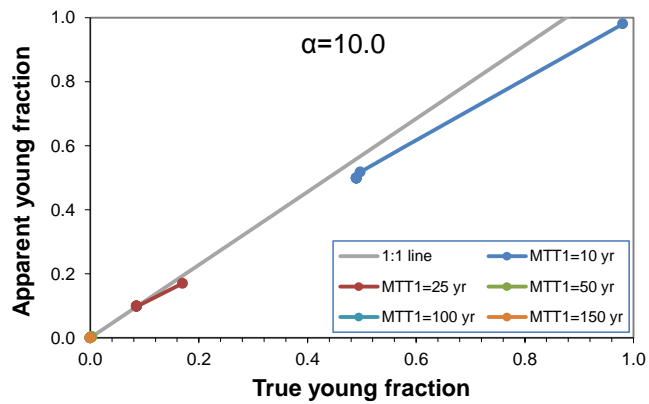
Figure 7. Tritium young water fractions using the Trier, Germany tritium input function. Symbols as in Fig. 6.
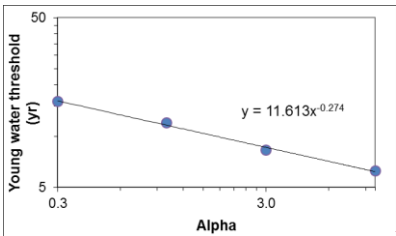
Figure 8. Plot of young water threshold versus α for tritium. Note the log scales.

5

40
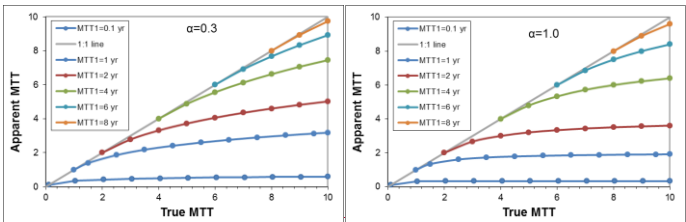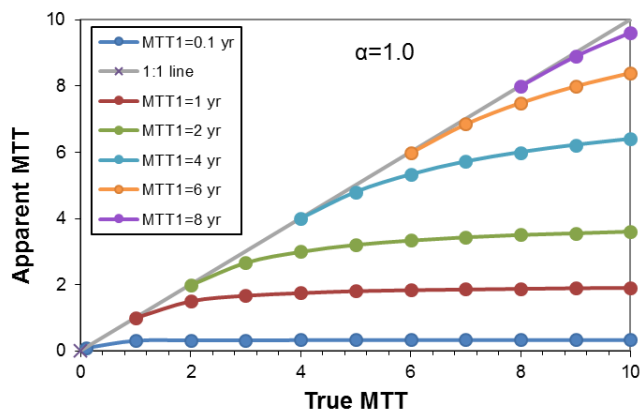
Figure 98. Aggregation effects on MTTs determined using seasonal tracer cycles for the ~~gamma model effective range of GM shapes (with~~ α = ~~0.3 – 1.0~~). Curves show changes as component 2 (MTT2) becomes older in 2-year steps (shown by dots).
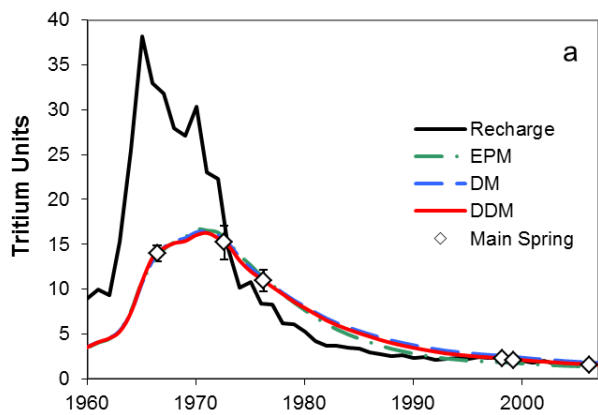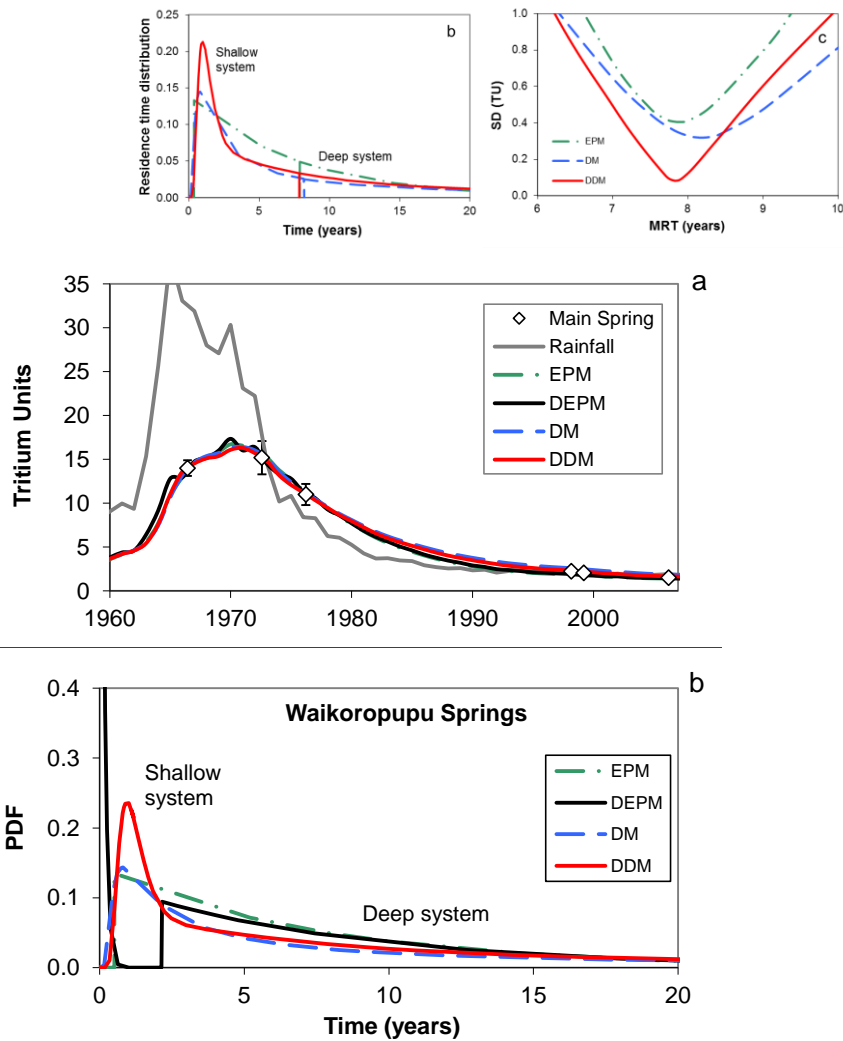
Figure ~~109~~. (a) Tritium measurements of the Main Spring of the Waikoropupu Springs, New Zealand**, and model fits to the data**. **EPM and DM are exponential piston flow and dispersion models, and DEPM and DDM are double (i.e. binary parallel) EPM and DM models. (b) Transit ~~Residence~~ time distributions of the best-fit simulations. ~~(c) Variation of goodness-of-fit criterion with MRT.~~**
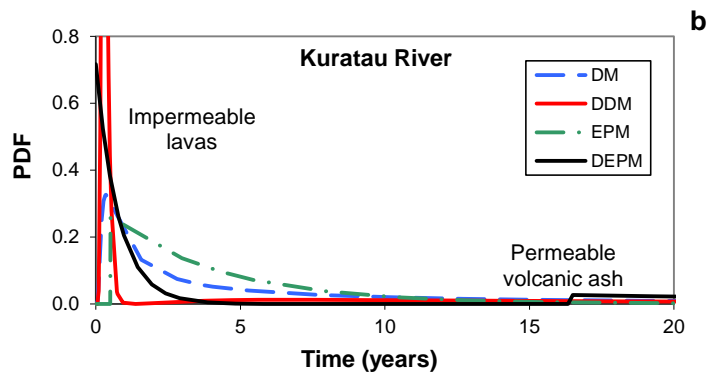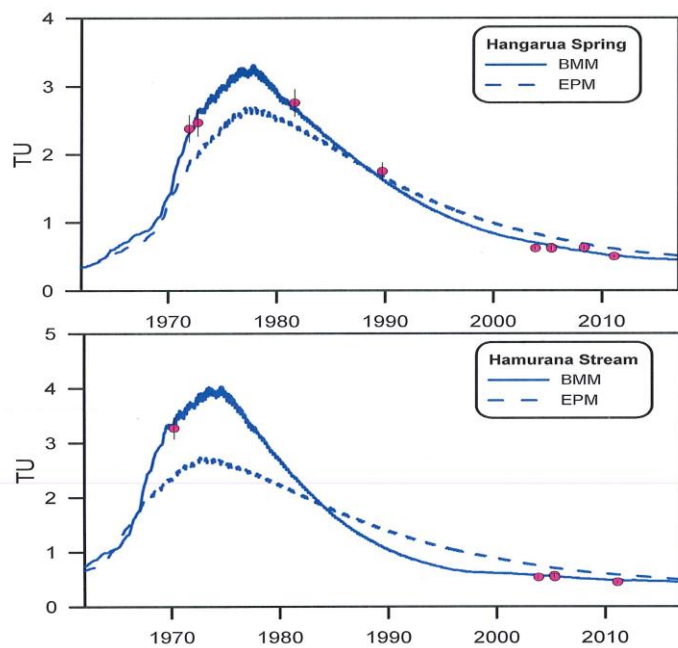
a

b

Kuratau River

Impermeable lavas

Permeable volcanic ash

Legend:
- DM
- DDM
- EPM
- DEPM
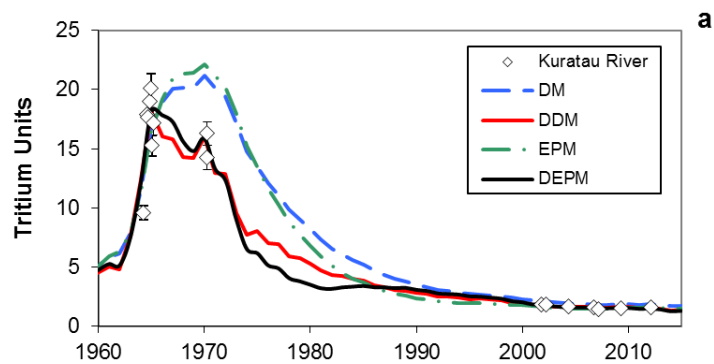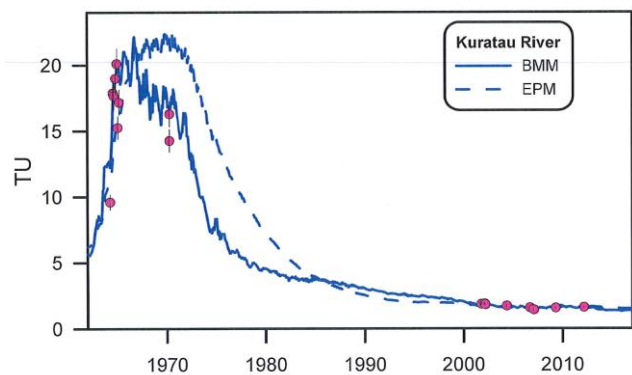
Figure 10~~1~~. (a) Tritium measurements and model fits for Kuratau River, New Zealand. (b) Transit time distributions of the best-fit simulations.
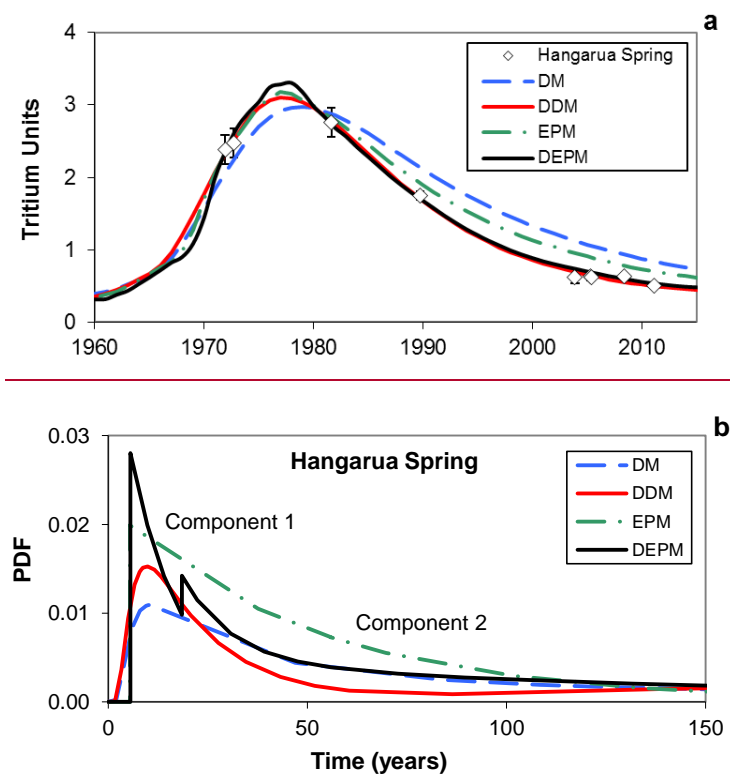

Hangarua Spring
- BMM
- EPM

Hamurana Stream
- BMM
- EPM

44

Figure 1~~1~~<ins>2</ins>. <ins>(a)</ins> Tritium measurements and model fits for ~~(a)~~ Hangarua Spring ~~and (b) Hamurana Stream~~, Rotorua, New Zealand. <ins>(b) Transit time distributions of the best-fit simulations.</ins>
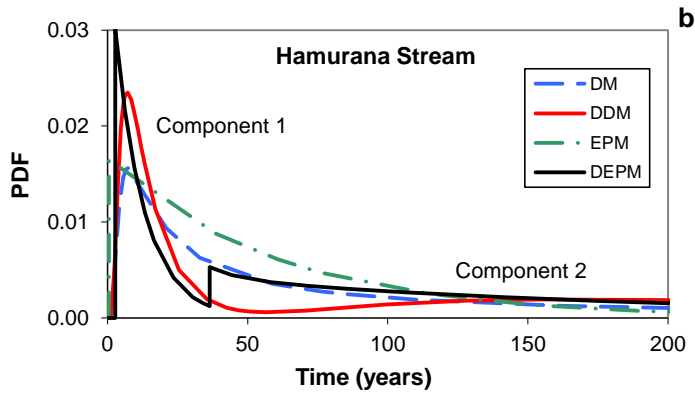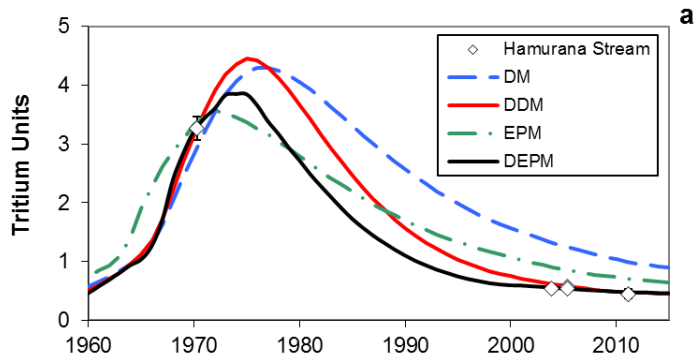
Figure 12. (a) Tritium measurements and model fits for Hamurana Stream, Rotorua, New Zealand. (b) Transit time distributions of the best-fit simulations.
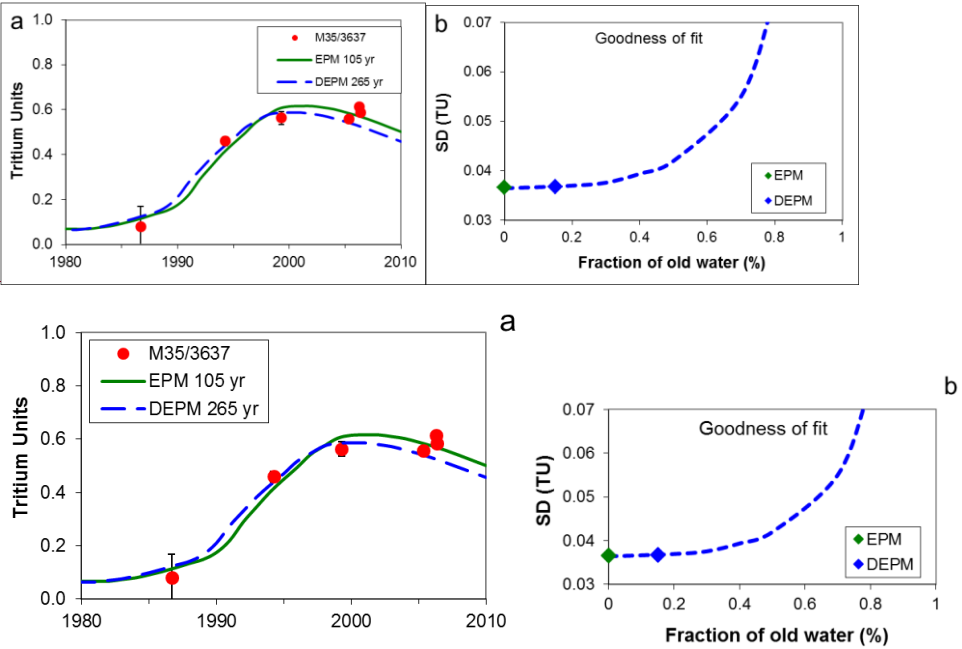
5

Figure 13. (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 component.