Response to Editor and Reviewers

Authors' responses in red

Editor Report

Dear authors,

thank you for the revised manuscript. The reviewers and myself appreciate the efforts made. Besides appreciating the general objective and direction of the paper, we think that the revised version is strongly improved with respect to organization, structure and clarity.

Unfortunately, the most critical issue is not yet adequately resolved, as concisely pointed out by reviewer #3 (reflecting the other previous reviewers as well as reviewer #5):
"[...] IF the heterogeneity in the catchment actually coincides with the chosen compound model (that is, if the catchment actually consists of two compartments that each behave according to their chosen age distributions), and IF the parameter estimation procedures actually retrieve parameter values that match the true values for the two compartments, then that model will, indeed, subsume the heterogeneity in the real-world system and eliminate the aggregation bias. But this only says that IF the model is actually the correct model, then it will be correct. The problem is, how can we know whether the model (including the underlying perceptual model) and its parameters are correct? [...]"

As large parts of the interpretation of the manuscript critically hinge on the assumption of a "well chosen" model, it will be necessary to unambiguously substantiate this claim with data.
Another way forward could be, as also pointed out by reviewer #3, to "[...] restrict the claims that are made to ones that actually demonstrated [...]".

Response: We have omitted contentious material and restricted the claims to ones that are actually demonstrated.

[Comment: Of course nobody knows the hydrogeologic flow situation, and never will know it because once one digs into the aquifer to make observations, it is being changed through the process. Therefore, we make observations as to how a tracer moves through the system and deduce a flow model from those observations, also taking into account other hydrogeologic information. With increasing complexity of the tracer input(s), increasingly complex flow processes can be resolved. That is the whole nature of the work. Tritium input and output records that cover the full range of the bomb pulse (like those for the case studies previously given in this paper, and for example by Blavoux et al., 2012) are excellent in terms of their large ranges of variation, and superior to those used for rainfall/runoff and stable isotope modelling. Hence, kneejerk application of lessons from those latter types of modelling are probably not justified. This will not apply to tritium samples collected in places where there is not already a record of past tritium concentrations.]

In addition, the question concerning parameter uncertainty issue (or "how sure are you that the chosen parameters are indeed the most appropriate ones?") has not at all been addressed, raising questions about the robustness of the chosen parameters.

Response: We have removed the entire second part of the paper and plan to submit it somewhere when we have done a full GLUE analysis (a la Gallart et al., 2016) of the simple and compound models.

[Comment: We have applied the GLUE analysis for the EPM model to the case studies previously included in the paper and found that the best-fit and likelihood-weighted MTTs and their uncertainty ranges are in good correspondence.]

The above points need to be addressed in detail before the manuscript can be considered for publication. I would thus be glad if the authors invested some more effort to develop the manuscript to the point.

Best regards,

Markus Hrachowitz

Report#1

This paper has already been through a number of thoughtful reviews. Here I will focus on a few contentious elements rather than provide a complete review.

This paper aims first to extend the results from Kirchner (2016) on the effects of 'aggregation errors' on transit times estimated from seasonal cycles of stable water isotopes (or other passive tracer) to water ages estimated from tritium observations. I understand the "aggregation error" (as Kirchner used the term) as error in data interpretation arising from a poor choice of probability distribution for the transit time distribution (or 'Lumped Parameter Model' – LPM – as the groundwater community prefers) in certain circumstances. Specifically, it is the case where the sampled water contains contributions from sources whose individual transit time distributions have very different means. If the chosen distribution for the combined sample is unable to represent the breadth (variance and skewness) of the combined transit time distribution it will provide biased estimates of the mean transit time.

Response: Agreed

If the aims of the paper were simply to point out that spatially variable transit time distributions have similar effects on Tritium observations as Kirchner found they did on stable water isotopes, the paper could be a useful contribution – particularly

if the robustness of the <18 year old fraction were more convincingly established. However the paper aims to have further-reaching conclusions regarding the superiority of 'compound' LPMs, and I have some slight issues with these as they currently stand – but I believe these can be remedied with some revision.

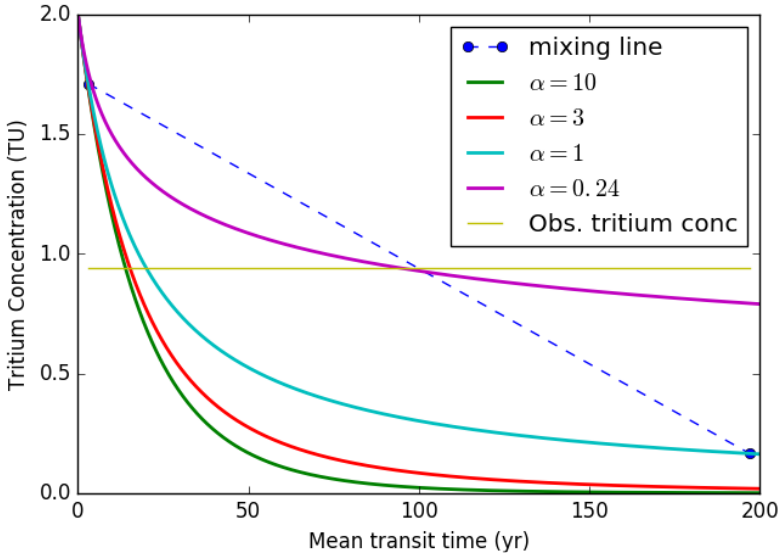
Response: OK

The heart of the analysis is the set of virtual experiments described in section 2.2. I think there is a limitation with the virtual experiments described, and which has perhaps led to some of the contentious reviews in the previous round. Specifically, the experiments are conducted by combining two 'simple' LPMs to predict a stream tritium concentration. This concentration is then analyzed by assuming a single 'simple' LPM again. The results show that this leads to significant bias. The authors conclude that 'simple' LPMs are susceptible to aggregation errors, and that a 'compound' LPM is required.

The problem with this is that their choices for the assumed 'simple' LPM are unnecessary constrained. As far as I can tell, the analysis is restricted to cases where both the 'true' LPM used to construct the data, and the 'assumed' LPM used to analyze it, are identical in form: both piston (Dirac delta distributions) or both gamma distributions with identical shape parameters (and always shape parameters greater than 1). The authors do not consider the case where (for example) the 'true' LPMs are exponential, but the 'assumed' LPM is a gamma distribution with a shape parameter less than 1. There is no reason not to do so that I can think of, since in practice we will not know the underlying distributions of the contributing parts. If we do not know the 'true' LPM, we are at liberty to adopt any physically reasonable distribution for the 'assumed' LPM.

Response: The procedure described above was followed in order to produce Figs. 3a-d (Sect 3.1.1) which were for explanatory purposes. For determining the actual (virtual) aggregation effects in Figs. 5, 6 (Sect. 3.1.3) the shape factor (α) was allowed to change along with the MTT to produce the best-fit of the simple LPM to the tritium data.

In this case, it is possible to choose a shape parameter for which there is zero aggregation error (at least in terms of mean age). The figure below is identical to Figure 3a (where alpha is 1) but with the curves for alpha=3 and 10 (like figures 3 b and 3c) included, along with the curve for alpha=0.24, which was not considered by the authors. For the latter case there is zero aggregation error. The predicted tritium concentration and MTT (100 years) of the 'simple' assumed LPM are both almost exactly that of the 'compound' true LPM.



Response: This is an interesting observation. The dispersion model also has large skewness when the dispersion parameter has high values. This is referred to in the Discussion (Sect. 4.1).

This seems to contradict the conclusion of the paper that suggests that 'simple' LPM should not be used because they have higher aggregation errors. Here a simple LPM perfectly reproduces the MTT of the aggregate. It is able to do so because the small alpha gives it a large variance and skewness, enabling it to capture the influence of

the young and old components. Note that, of course, it is not an accurate representation of the true TTD – however the data (a single tritium observation) is insufficient to determine that. Also, I was only able to choose the 'right' value of alpha because I know what I was aiming for.

As the authors point out a compound distributions is a sensible choice where distinct sources can be identified, and the partitioning of flow between them is known. It is useful to be able to incorporate such information into the model used to interpret the data. However in the absence of such auxillary information there is nothing fundamentally different about 'compound' LPMs that makes them immune to aggregation errors in some way that other type of distribution are not. There are many other distributions that are also reasonable choices even in the case of large heterogeneity, so long as they have enough flexibility to accommodate larger amounts of skew than an exponential or other type of distribution tested by the authors can.

Response: We agree that both simple and compound LPMs can be free of aggregation error or conversely be affected by aggregation error. We have removed this part of the paper.

Of course, in the absence of such auxillary information the compound LPMs and other flexible distributions will have multiple free parameters that must be estimated. As the number of free parameters increase, the ability of the model to reproduce calibration data inevitably increases, without necessarily increasing the physical realism of the calibrated parameters, or any metrics derived from the model (like the mean). This brute fact must always be acknowledged and dealt with by those urging us to adopt more complex models, and is not dealt with here (as several reviewers of the previous version pointed out, to no avail).

Response: We agree and have stated this in the revised ms (Sect. 4.2).

In conclusion, this paper makes a useful contribution, and I believe it warrants publication if the authors are able to clarify the issues raised above, and pay a little more attention to the trade-offs associated with increasing the number of free

parameters.

Response: We thank the Referee for his constructive input and addressed his comments in the revised manuscript.

Report#2

The revised version of this manuscript is an improvement. It resolves several of the issues that were raised in the first round of reviews.

Unfortunately, there are also important areas where the manuscript has hardly changed in response to major issues identified by the reviewers as well as the editor. To the extent that aspects of the paper haven't changed, my assessments of those aspects also remain unchanged.

The paper's central claim is that "well-chosen compound lumped parameter models should be used as they will reduce potential aggregation errors due to the application of simple lumped parameter models". This claim, from the abstract, is stated much more strongly in the text, e.g., "... the binary models have very much less potential for aggregation bias than the simple models".

These claims must be demonstrated with evidence, and not simply asserted.

Response: This is not the central claim of the paper. The central claim is "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." (Quoted from the abstract – first paragraph.) This is also what is given in the title. The revised paper deals only with this topic.

The reviewer's asserted "central claim" above was always subsidiary and has now been removed.

To the extent that evidence is offered, it is mostly true by definition, since: a) a "well-chosen" model is operationally defined in the paper as one that is consistent with the perceptual geological model, which is assumed to be correct b) the authors define the "true" MTT as the MTT of the compound model for the purposes of the results in Section 4, and c) the authors measure the aggregation error as the deviation of the simple model MTT from the compound model MTT (which is considered to be the "true" MTT).

Under this system of definitions, of course the compound models are better than the simple models... because the compound models are defined from the start as being correct and therefore having no aggregation error!

Response: This is not relevant now.

Of course IF the heterogeneity in the catchment actually coincides with the chosen compound model (that is, if the catchment actually consists of two compartments that each behave according to their chosen age distributions), and IF the parameter estimation procedures actually retrieve parameter values that match the true values for the two compartments, then that model will, indeed, subsume the heterogeneity in the real-world system and eliminate the aggregation bias. But this only says that IF the model is actually the correct model, then it will be correct.

Response: Agreed

The problem is, how can we know whether the model (including the underlying perceptual model) and its parameters are correct?

The authors' position seems to be that as long as the model fits the perceptual model and the data, it is correct. But long experience in hydrological modeling has shown that models with many parameters can often fit the available data, whether or not those models are structurally correct. In many cases, the models will also fit well for multiple, widely differing parameter sets – in other words, the parameter values will not be identifiable, even though the models fit the data nicely.

Response: Not relevant now.

To justify the statements made in the manuscript, one would need to present evidence that if the model fits the perceptual model and the data, then the model structure and the parameter values are guaranteed to be correct – that is, that it is not possible for the real-world structure, or its parameter values, to be significantly in error. No evidence is presented to meet this burden of proof.

There is another way forward, and that is to restrict the claims that are made to ones that actually demonstrated. For example, the authors have analyzed several cases where there is geological evidence for a multicomponent catchment system, and in those cases, compound LPM's that mirror these perceptual models fit the tritium data better than simple LPM's do. So that is a claim that could be made in the paper.

Response: Not relevant now.

But if the manuscript is to make broader claims (for example, about whether compound LPM's will reduce or eliminate aggregation errors), then those claims need to be substantiated with evidence. That evidence also needs to be nontrivial; it is not enough to posit a system of two exponential distributions, and then assert that a compound LPM with two exponential distributions would fit this system, and would have no aggregation bias. That is simply irrelevant to the problem we face in the real world, which is that we do not already know the "right answer", including the structure of the system we are trying to analyze.

The manuscript demonstrates an important point, which is that where we already have geological information about aquifer partitioning, then that information can be very useful in constraining transit time models. Where we have that information, we don't need to rely on tracers alone. Another way to say it is, if the transit time models agree with the geological information and the tracers, then they strengthen our confidence in both of them. That is a useful point to make.

Response: Not relevant now.

But the generalization that "well-chosen" models are always preferred is either true by definition (of course "well-chosen" is better than "poorly chosen"), or else it leaves open the critical issue: how do we know when our models are "well chosen"? Is two compartments "well chosen", or just one? Or six? How should we decide?

Response: Not relevant now.

More complex models, with more adjustable parameters, will almost always fit data sets better, but the parameters themselves may be highly uncertain. The critical question that the manuscript must come to grips with is: how well can the parameters be constrained? What are their uncertainties? Holding all the other parameter values constant and just varying one of them is not a valid way to estimate parameter uncertainty, unless those other values are actually known to have their assigned values. But as far as I can tell (the manuscript doesn't say, and that is itself a problem), either this, or nothing at all, is what has been done.

Response: Not relevant now.

My previous review said that there was no excuse for not properly analyzing parameter uncertainty, and the response was that this had not been done in tritium papers until very recently. The fact that it has rarely been done before is no excuse for not doing it now. The only alternative is to remove all claims about whether parameters (including the MTT's that are derived from them) are well constrained.

Response: Not relevant now.

Some particularly problematic passages are quoted below (this is an incomplete list):

"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 very different rock types that produce waters with very different MTTs; i.e. the most extreme “significant and distinct” heterogeneity one can imagine. A binary LPM describes this type of system, and when optimised with suitable data would very effectively separate the young MTT from the old MTT waters in the catchment outflow, and therefore minimise aggregation errors in MTT. "

This assumes the existence of "suitable data", but gives no criteria by which to determine whether data are "suitable". No analysis is presented to support any of the statements in the passage. They may be true, or they may not, but unless they can be supported by evidence they should be removed.

Response: Not relevant now.

"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."

No evidence is presented to substantiate these statements. In Kirshner [2016], splitting the model system into more components does not reduce the aggregation bias, which suggests that the statements made here are not correct.

Response: Not relevant now.

"There is, of course, a wide range of different types of hydrological systems, but the binary LPM is likely to remain effective in cases of “significant and distinct” heterogeneity, which are the ones of concern for aggregation error."

What do "remain effective" and "significant and distinct" mean?

Response: Not relevant now.

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

bias on MTTs.

These statements may or may not be true, but they are not supported by any clear evidence that I can find in the manuscript.

Response: Not relevant now.

"A good example is the study of Blavoux et al. (2013) describing the interpretation of an exceptionally long and very detailed record of tritium concentrations from the Evian-Cachat Spring in France. The tritium record was much too complicated to be fitted by a simple LPM. Instead, the detailed records of input and output allowed accurate specification of a combined model comprising of exponential (τm = 8 yr) and dispersion (τm = 60 yr) models in series, with a small bypass flow in parallel with them, followed by a piston flow model (τm = 2.5 yr) in series giving an overall τm 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."

The Blavoux et al. study relied on samples taken over decades, including ones that nicely traced out the bomb pulse. It is a nice study, but is completely misleading as an indication of what will be possible with tritium data, except in the very few places where records go back that far.

Response: The quoted paragraph has now been omitted. We agree that the Blavoux et al. approach can only be applied to places where there is sufficient past tritium data (such as a number of sites in New Zealand). [But we are puzzled, was the Blavoux et al., 2012, approach also "inexcusable" (in the reviewer's terminology) or is this word only applied to our approach although we do the same thing?]

A central problem with the paper is that it gives the reader the impression that fitting more complex LPM's will solve all kinds of problems. But now and into the future, with the bomb pulse gone, there will be very little tritium variation available to fit even simple models to, let alone more complex ones.

Response: It was not our intention to give such an impression. The paper states in several places that identifying parameters of LPMs will be difficult in the future.

"Compound LPMs are to be preferred, but often there can be considerable difficulty in uniquely quantifying the parameters especially if the output data is limited."

The statement that compound LPM's "are to be preferred" is unsubstantiated, except perhaps as a statement of the authors' preferences. And the second part of the sentence effectively cancels the first, since (although this is not stated), the "difficulty in uniquely quantifying the parameters" becomes geometrically more difficult as the number of parameters grows. The manuscript needs to show that the "considerable difficulty" is not a problem in the cases presented here, and also in the types of situations where tritium is likely to be used in the future, when there will be no bomb pulse to work with.

Response: Not relevant now

"We find that MTT aggregation errors are small when the component waters have similar MTTs to each other. On the other hand, aggregation errors can be large when very young water components are mixed with older components."

Unlike in the case of seasonal tracer cycles, with tritium the absolute ages (such as "very young") are not relevant because tritium decays exponentially. Aggregation errors can arise in tritium whenever the component ages span

a large range, regardless of where that range is centered. The exponential decay curve implies that the aggregation error depends on the ratios of the ages, not their absolute values.

Response: Not relevant now. This argument is mostly correct, but ignores the effect of mixing models.

Even if we have no "very young" water at all, for example, adding some much older (tritium-depleted) water can potentially lead to a large aggregation error, depending on how much older that water is. If, for example, we added 50% zero-tritium water to our mixture, we would decrease the average 3H concentration by half and therefore increase the model age by 12.3 years... but we could increase the ACTUAL mean age by hundreds or thousands of years, depending on how old the zero-tritium water actually is.

The authors point out in section 4.4 that additions of tritium-free water will often have a very small effect on model fits to tritium time series, but (although they don't emphasize this), this could have a potentially huge effect on what the true mean transit time is (because the tritium-free water could be 200, or 2000, or 20,000 years old). Thus the true mean transit time can become decoupled from the mean transit time estimated from tritium. This is a clear aggregation error, and it should be identified as such. The same potential aggregation error exists (but is not mentioned) in the other case examples as well. Every TTD model in the paper, both simple and compound, is vulnerable to it.

Response: This is a well-known problem mentioned in Sect. 4.1 (and that was illustrated in the now-removed section 4.5).

"In general, well-chosen compound lumped parameter models should be used as they will reduce potential aggregation errors due to the application of simple lumped parameter models. An opportunity to determine a realistic compound lumped parameter model is given by matching simulations to time series of tritium measurements (underlining the value of long series of past tritium measurements), but such results should be validated by reference to the characteristics of the hydrological system to ensure that the parameters found by modelling correspond to reality."

Again, the value of comparing model assumptions to "the characteristics of the hydrological system" (presumably determined independently?) is clear, but claims that any model is "realistic" – and, more generally, that the approach outlined here is a reliable guide to model "realism" – need to be supported with evidence.

Response: Not relevant now.

Minor items:

The claim that tritium measurements are usable over a 200-year time range should be substantiated. 200 years is almost 17 half-lives, during which tritium concentrations will decay to 10^-5 of their original values. Even assuming we start with 5,000 TU (the peak of the bomb pulse), this will decay to about 0.05 TU after 200 years. Even if such a measurement is analytically feasible, it is hard to see how it is useful in practice, because contamination with even a tiny bit of younger water would obscure the tiny traces of the 200-year-old water. (And remember, this is starting from the strongest possible tritium signal, the peak of the bomb pulse.)

Response: This is very easy to substantiate. To adopt the rhetoric of the reviewer, there is no excuse for the calculation made by the reviewer here. His/her calculation is based on the piston flow mixing model (an assumption that has dogged the use of all radioactive isotopes and some chemicals (e.g. tritium, carbon-14, CFCs, etc.) for water dating over much of their history. The PFM does not apply to very many water flow situations at

all (if any) and particularly not when older waters are involved. Waters with MTTs of 200 years will often have measurable tritium concentrations when appropriate mixing models are used (see table). For example, using a mixing model which has been found to apply to many situations (EPM with parameter f = 0.75), 200-year-old MTT water would have tritium concentrations of 0.07 TU in the Southern Hemisphere (measurement error about ±0.02 TU) and 2.65 TU in the Northern Hemisphere (measurement error about ±0.1 TU). Note that such ages are mixing model-dependent (as are all ages calculated using tritium and other isotopes/chemicals), and that the tritium observed in the sample would have come almost entirely from the younger water in the mixture which would have contained bomb pulse tritium.

Table 1. Tritium concentrations in 2017 for water with mean transit times of 200 years calculated using an exponential piston flow mixing model. Pre-bomb tritium inputs of 2 TU in the Southern Hemisphere, and 8 TU in the Northern Hemisphere were assumed.

Exponential piston flow model (EPM)	Alternative model name	Tritium concentration (TU) in 2017 of 200-year MTT water	
Parameter f		Southern Hemisphere	Northern Hemisphere
0	PFM	0.00002	0.0001
0.5		0.0008	0.004
0.75		0.07	2.65
1.0	EM	0.25	3.12

The statement that there is no aggregation bias when the MTT's of the two components are the same (see the abstract, and Section 5.1, for example) are rather trivial. If there is no heterogeneity, then of course no bias can result from it.

Response: Not relevant.

On page 5, line 19, the dimensions of beta are wrong.

Response: Agreed.

The conclusions are mostly a word-for-word repetition of the abstract. If the authors don't have anything further to say, then there is no need for the conclusions. There is need to print the same sentences in two different places.

Response: Both conclusions and abstract have been revised.

Report#3

General Comments:
I was not a reviewer in the initial round of reviews and therefore I cannot judge exactly how much improvement has been made in the revision process so far. Looking at the initial reviews and the revised manuscript I can however see that the authors did make some efforts to resolve problems that were brought forward.
I can also say that the manuscript is well-written in terms of style and structure.

The part I am struggling with is more a general one. The authors conclude that using a well-chosen compound model solves the issue of spatial aggregation errors. Still, they never really show how to 'choose well'. They demonstrate which of the models causes the smallest aggregation error, but what does that really mean? If a gamma model with shape parameter 10 represents the dynamics of the real-world hydrology best, this model should still be used regardless of the fact that it causes more aggregation errors than a gamma model with shape parameter 1 (or is that a wrong assumption?).

Response: We agree that any simple LPM should be used if it represents the dynamics of the real-world hydrology. There would be no aggregation error in that case. Simple LPMs should be tried in any case to see if they can fit the data.

Also, the f parameter bothers me. How can it be constrained and how does it influence the results? How does it affect the young water fractions and the young water thresholds? The authors often state that they strongly believe that the use of compound LPMs could strongly reduce aggregation errors. What about a little demonstration of whether the model actually works. If you mix two waters of different ages in a specified ratio (not only 50:50), can you define both MTTs correctly using the binary model or do you get lost in equifinality problems when using 5 fitting parameters? You could set up a model scenario with known inputs and outputs to demonstrate that your method actually works.

Response: We think the reviewer means b (the fraction of the younger water component in the mixture) not f.

This is an interesting point, although not now applicable to our revised paper. We tried fitting a DEPM model to the tritium values resulting from the combination of two EPM models (EPM1 and EPM2) using the Kaitoke (Southern Hemisphere) tritium input function. This procedure is equivalent to a situation where one had a time sequence of 75 tritium samples (one for each year from 1940 to 2015), i.e. a very good record of the tritium in the stream or groundwater. The fitting procedure using the Solver Add-In of the Excel spreadsheet gave mixed results, but generally was unable to recapture uniquely all five parameters of the DEPM (MTT1, f1, MTT2, f2 and b). Likewise with fitting four parameters, i.e. after fixing b based on measurements in the catchment or groundwater system. Varying three parameters (the MTTs and b), on the other hand, gave unique solutions which recovered the original parameter values. This required fixing the two f values at their original values. Work by Gallart et al. (2016) on the uncertainties of the parameters of the exponential piston flow model had shown that the f parameter had remarkably low identifiability (i.e. almost any value would do), whereas the MTT was usually much more clearly identifiable.

The procedure we have usually adopted when fitting a time sequence of samples is to initially insert parameter values deemed likely from the hydrogeology of the system (and our previous experience) and try to get close to a good fit by manually tweaking the parameter values. Then apply the Solver fitting procedure to locate the best-fit position or positions. However, there can be several parameter combinations giving fits of various qualities (i.e. equifinality), so further experimentation to explore the full parameter spaces is often necessary.

Another issue is the exclusion of anything non-stationary from the discussion. How would the omnipresent time-variance influence the results? Are all the assumptions (and conclusions) still valid if the fast component (or both components) change velocity over time? How would aggregation in time affect the apparent mTTs?

Response: Our work did not include the effect of non-stationarity on aggregation error.

Is the determination of the young water fraction threshold simply done by minimizing the difference between apparent and true young water fractions? You do not specify this method. So what exactly is the young water fraction? Is it a random number fulfilling some mathematical/statistical requirements – like for example the value where the young water transit time distribution has its peak? What is it actually good for, if its value ranges from

0.1 to 18 years? You state that certain thresholds are important (1 year, 60 years) but you do not show/test whether they can be predicted accurately (I guess they cannot because they are far from the value of 18 years). Maybe you can discuss some of these issues in more detail.
In essence, what I want to say is that this manuscript would profit considerably from going even further beyond reproducing Kirchner's 2016 paper with tritium.

Response: The young water threshold of 18 years was determined by trial and error, with the requirement that the apparent and true young water fractions not deviate from each other by more than 10%. 18 years was about the highest value for which this requirement was met.

Specific Comments:
Page 13, Line 21: What do you mean by 'using the DM and DEPM together or the EPM and DDM'? This is confusing me. This sentence has been removed. (It was just a suggestion that the EPM and DM models should not be mixed for that particular case study. If one starts with the EPM then it should be compared with the DEPM not the DDM, when comparing results from single and binary LPMs.).
The young water fraction in Table 2 should not have dimensions of (yr). Agreed

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

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

[1]Aquifer Dynamics & GNS Science, P.O. Box 30368, Lower Hutt, 5040, New Zealand
[2]GNS Science, Tritium & Water Dating Laboratory, Avalon, Lower Hutt, 5040, New Zealand
[3]International Centre for Water Hazard and Risk Management (ICHARM), Public Works Research Institute (PWRI), Tsukuba, Japan
[4]National Graduate Institute for Policy Studies (GRIPS), Tokyo, Japan
[5]AGH University of Science and Technology Cracow, Department of Hydrogeology and Engineering Geology, Al. Mickiewicza 30, 30-059 Cracow, Poland

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

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

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

~~reality. In addition, the range of tritium concentrations in hydrological systems in the future is likely to be limited because of the passing of the bomb peak. This will make model identification more difficult and highlights the fact that such time series should be built now as long as the tail of the bomb peak is still available.~~

1 Introduction

~~Water can take very complex flow pathways in catchments through shallow and deep aquifers.~~ Environmental tracers are commonly used to obtain transit time distributions (TTDs) in groundwater systems (Maloszewski and Zuber, 1982) or catchments (McDonnell et al., 2010). Transit time is the time it takes for rainfall to travel through a system ~~and emerge~~from recharge to emergence in a well, spring or stream. TTDs provide important information about transport, mixing and storage of water in systems and therefore on the retention and release of pollutants. In addition, m~~M~~ean transit times (MTTs) determined from these distributions provide practical information for various aspects of water resources management. For example, MTTs have been used to estimate the volume of groundwater storage providing baseflow in catchments (Morgenstern et al., 2010; Gusyev et al., 2016) and to predict lag times and life expectancies of contaminants in the subsurface (Hrachowitz et al., 2016). The drinking water securities of wells in New Zealand are partly assessed by an absence of water with less than one-year travel time by the NZ drinking water quality standard (Ministry of Health, 2008). As useful as they are, TTDs cannot be measured directly in the field and have to be inferred from age-dependent tracer concentrations with the use of lumped parameter models (LPMs).

Catchments are inherently heterogeneous on various scales. Point-scale properties vary greatly from place to place, while streams integrate the various catchment outputs. The top-down approach uses catchment outputs, such as streamflow and stream chemistry, to infer or predict catchment TTDs. The hope is that these average out local heterogeneities allowing one simple LPM to provide a good fit and its parameters to be representative of the catchment ~~(a well-known relatively successful application of an LPM to model catchment acidification is described by Cosby et al., 1985)~~. But individual areas within catchments can vary greatly because of geology, geography, aspect, etc. Groundwater systems also show heterogeneity. Kirchner (2016a) showed by means of virtual experiments that aggregating subcatchments with different TTDs can lead to severe underestimation of the composite MTT when simple LPMs were applied to interpret seasonal tracer cycles. This is because the smoothing out of the seasonal cycles is a non-linear process which acts more rapidly on the younger water components thereby causing underestimation of the composite MTT. He also found that the young water fraction was a much more robust metric than the MTT against aggregation error ~~using seasonal tracer cycles~~. These results raise an important question: Are tritium-derived MTTs also susceptible to aggregation error due to spatial heterogeneity? This work aims to answer this question.

Seasonal tracer cycle and tritium-based MTTs are determined by different methods and have given very different results in catchments. The seasonal tracer cycle method depends on damping of input cycles on passing through a system into the output, whereas the tritium method depends on radioactive decay of tritium between input and output (with half-life of 12.32 yr).

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

Calibration of LPMs using environmental radioisotope and stable isotope data has been the subject of study for many years (see Maloszewski and Zuber (1982) and early work summarised therein). If a catchment outflow is a mixture of two ~~(~~or ~~possibly~~ more~~)~~ components of different water ages, it ~~is~~ can be difficult to calibrate an LPM uniquely when we only have data for tracers. For example, for springs in Czatkowice, Poland, only when the proportion in which the water components (water fluxes) were mixed was known could the unique answer based on tritium measurements be found (Grabczak et al., 1984; Małoszewski and Zuber, 1993). In heterogeneous catchments, it is always helpful to (i) measure a variable tracer periodically, and (ii) to combine those data with water fluxes in the inputs and outputs to separate "fast" and "slow" components; see for example studies at Lainbach Valley, Germany (Maloszewski et al., 1983), and Schneealpe, Austria (Maloszewski et al., 2002). The choice of LPM, or equivalently the TTD function, must be based more on the hydrogeological situation and not on artificial mathematical (fitting) considerations. ~~Calculation~~ Consideration of hydrological parameters known independently (e.g. mean thickness of the water bearing layers in the catchment) is required for model validation in order to examine if the model ~~used is~~is likely to be applicable to the real situation. We can have a very well-calibrated model in terms of tracer data being fitted by an LPM, but the MTT can be far from the hydrological reality.

The aim of this paper is to examine the aggregation effects of spatially heterogeneous catchments and groundwater systems on MTTs and young water fractions determined using tritium concentrations. We conducted our investigation by combining two dissimilar water components in virtual experiments and comparing the true mixed MTTs with the tritium-inferred apparent MTTs, as Kirchner (2016a) did with seasonal tracer cycles. Our experiments did not include examination of non-stationary hydrological systems, in which Kirchner (2016b) had found similar underestimation of MTTs with seasonal tracer cycles. We also examined aggregation effects for young water fractions estimated using tritium. Our calculations are based on the gamma LPM with shape factors (α) between 1 and 10, which is also representative of other frequently-used simple LPMs such as the exponential, exponential piston flow and dispersion models. The different tritium input functions for Northern and Southern Hemisphere locations were tested. ~~We also surveyed some applications of tritium dating from the literature to see if they had sufficient data to constrain the parameters of mixtures of young and old waters. MTTs from simple and compound LPMs applied to the data were compared to examine the aggregation errors for these real examples. This has allowed us to consider~~

~~the practical implications of our findings and provide guidance for tritium sampling and interpretation in heterogeneous catchments and groundwater systems.~~

2 Methods

2.1 Transit time determination: Simple and compound lumped parameter models

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

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

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

where C_{in} and C_{out} are the input and output concentrations in the recharge and baseflow respectively. t is calendar time and the integration is carried out over the transit times τ. h(τ) is the ~~transit time distribution (TTD) function~~flow model 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 (= $\ln 2/T_{½}$~~)~~, where $T_{½}$ is the half-life of tritium (12.32 years).~~)~~

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

markers of 1960s precipitation in past tritium studies, but the steady declines which mimic radioactive decay of tritium have caused problems with ambiguous (i.e. multiple) age estimations for given tritium values (Stewart et al., 2010).

The curves also show smaller variations due to annual peaks in tritium concentrations caused by increased stratospheric leakage during spring in each hemisphere, and possibly small longer-term variations related to sunspot cycles. Tritium concentrations are expected to remain at the present cosmogenic levels for the foreseeable future, and this means that multiple age solutions are becoming less of a problem ~~tritium is becoming increasingly useful for dating because multiple age solutions are now less of a problem~~ (Stewart et al., 2012; Stewart and Morgenstern, 2016; Gusyev et al., 2016). However, the minimal variation will mean that tritium will not be very effective for identifying flow models in the future.~~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 are commonly used in tracer studies.~~, and are represented by probability density functions (PDFs).~~ The piston flow model (PFM) describes systems in which all of the water in the output has the same transit time (MTT or τ_m). Its TTD is

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

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

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

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

The exponential model (EM) is given by

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

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

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

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

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

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

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

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

where f is the ratio of the exponential volume to the total volume. Maloszewski and Zuber (1982) used the parameter η instead of f, where ~~f=1/~~η = 1/f. fτ$_m$ is the time required for water to flow through the exponential volume, while τ$_m$(1-f) is the time ~~time required for~~in ~~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

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

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

This paper makes a particular distinction between *simple LPMs* (meaning specifically the gamma model, the exponential piston flow model with end members piston flow and exponential models, and the dispersion model ~~LPMs~~) and *compound LPMs* (binary or other parallel combinations of simple LPMs). Simple LPMs describe ~~systems as~~ homogeneous systems, while compound LPMs can accommodate heterogeneity in the system.~~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) (e.g. Maloszewski et al., 1993; Stewart and Thomas, 2008; Blavoux et al., 2013; Morgenstern et al., 2015). The binary parallel LPM is given by

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

where LPM$_1$ and LPM$_2$ are simple LPMs with individual PDFs representing two water components contributing to the system output, and b[-] is the fraction of the first component in the combined output. The overall combined MTT (τ$_m$) is

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

An example of a compound LPM is the parallel combination of two exponential models describing a system with young and old water components. This is called the 'double exponential model' when applied to tritium (Michel, 1992; Taylor et al.,

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

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

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

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

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

2.2 Estimation of spatial aggregation effects on mean transit times

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

To determine the "apparent" MTT, the tritium concentrations of the water components from 1940 to the present are calculated from the gamma models applying to each component using the convolution process described above (Eq. 1). The input function was first assumed to be constant at 2 TU for the calculations given in Section 3.1.1, then the Kaitoke or Trier input functions (Fig. 1) were used for the calculations in Section 3.1.2 and 3~~2~~.1.3. In all cases, the tritium concentrations of the mixed system (C_m) are given by

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

where C_1[TU] and C_2[TU] are the tritium concentrations in components 1 and 2 respectively. The mixed system is then treated as if it is homogeneous to produce the "apparent" MTT by fitting a simple LPM (a gamma model) to the tritium concentrations of the mixture (C_m). The true and apparent MTTs of the mixture are compared for different assumed values of the MTTs of the components. b is assumed to be 0.5 for simplicity in what follows. Following Kirchner (2016a), we did not consider evapotranspiration in our analysis of tritium aggregation effects. ~~Note that the tritium concentrations in recharge was used for the case studies.~~

2.3 Determination of young water fractions

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

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

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

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

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

2.4 Comparison of transit time distributions of different flow models

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

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

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

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

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 standard deviation and Nash Sutcliffe efficiency gave the same results in terms of identifying the most similar shapes of the gamma, exponential piston flow and dispersion models (Table 1). TTD shapes for the gamma model with α between 1 and 10 are equivalent to exponential piston flow model shapes with exponential fractions (f) between 1.0 and 0.44 (Table 1), which have been found suitable for interpreting tritium concentrations in baseflow and groundwater (e.g. Maloszewski et al., 1983;

Stewart et al., 2007; Morgenstern and Stewart, 2004). The useful range of the dispersion model has dispersion parameters (P_D) between about 1.3 and 0.05 corresponding to the gamma model with α between 1 and 10 (Table 1). The gamma and exponential piston flow model shapes become less similar to each other as α increases to 10, while the gamma and dispersion model shapes become more similar.

3 Results

3.1 Aggregation effects on mean transit times determined using tritium

3.1.1 Relationships between mean transit time and tritium concentration

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

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

Figs. 3b-d show the same calculations applied to the gamma models with α = 3 and 10 and the piston flow model (PFM). The different shape factors describe different fractional contributions of past water inputs to the present water output as illustrated by the transit time distributions in Fig. 2a. The gamma models with α = 3 and 10 have slightly greater differences between the true and apparent MTTs than the gamma model with α = 1. The piston flow model is the most sharply peaked of all, and has the greatest true-apparent MTT difference of 100 to 15 years. Since there is no mixing, the non-linearity of the black curve is solely due to radioactive decay of tritium (Fig. 3d).

3.1.2 Effect of young component fraction (b) on aggregation

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

3.1.32 True versus apparent mean transit times

The true versus the apparent MTTs calculated using the real tritium input function from Kaitoke (expressed as annual values) are given in Fig. 54. The calculations were structured in that the two water components were initially assumed to have the same MTTs (i.e. $\tau_1 = \tau_2$) and therefore the mixture had the same true and apparent MTTs and plotted on the 1:1 lines. The second component (MTT2) was then allowed to become older in 50 year steps so that the difference in MTTs between the two components increased. This caused the apparent MTTs to become younger than the true MTTs and the points to move further and further away from the 1:1 line as shown by the curves in Fig. 54. The dots show the effects of the step changes in MTT2. As expected, the greatest age differences caused the biggest deviations from the 1:1 lines.

The different values of α cause differences to the patterns observed, but the patterns are similar overall. They are tighter around the 1:1 line for α = 1 showing smaller aggregation effects, and are most divergent for α = 10. Errors of fitting for determining the apparent MTTs (expressed as standard deviations (Eq. 15)) are greatest when component 1 is youngest, these are shown by fine dashed lines above and below the curves. The errors are largest with 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).

Using the Trier (Northern Hemisphere) tritium input function (Fig. 1) results in very similar aggregation biases for tritium MTTs (Fig. 65) compared to those obtained with the Kaitoke input (Fig. 54). Using Northern or Southern Hemisphere tritium input functions makes only slight differences to the curves. Note that the problem of multiple age solutions often experienced using tritium with the Northern Hemispheric input function (e.g. Stewart et al., 20120) does not arise here because we calculate around 75 tritium values (one for each year) and this constrains the final 'apparent' fitting to a single unique solution. However, the fitting errors for the apparent MTTs with the Trier input function are much larger than those determined with the Kaitoke input function.

Some of the calculation results are replotted in Fig. 76 to compare results for the Northern and Southern Hemispheres. ~~and to investigate aggregation errors for young components.~~ This figure shows the possible aggregation error (expressed as percentage difference of the apparent from the true MTTs) versus the MTT of component 2 (MTT2) for the gamma model with α = 1. The solid curves show results for MTT1 = 10 years, these are restatements of the curves in Figs. 54a and 65a for α = 1 and MTT1 = 10 years. Aggregation errors are 8% for Southern Hemisphere and 15% for Northern Hemisphere locations by the young water threshold (18 years). ~~Note that this value for the young water threshold is essentially based on the~~

~~calculations for MTT1 = 10 yr, because these have the greatest effect compared with the calculations for MTT1 = 25 yr etc. (Figs. 7, 8 below).~~

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

3.2 Aggregation effects on young water fractions

The effect of combining two different water components on the true and apparent young water fractions (Y_f) of a mixture are examined in this section using the same procedure as before (i.e. testing mixtures with MTT1 at 10 years, 25 years, etc.). The two water components were initially assumed to have the same MTTs and young water fractions (i.e. $Y_{f1} = Y_{f2}$) and therefore the mixture had the same true and apparent young water fractions and plotted on the 1:1 lines in Figs. ~~8~~7 and ~~9~~8. The second component (MTT2) was then allowed to become older in 50-year steps so that the differences in MTTs and young water fractions between the two components increased. But now the true and apparent young water fractions did not diverge very much from each other (Figs. ~~8~~7 and ~~9~~8). The figures show the young water fractions decreasing as the mixtures become older, but the curves lie mostly along the 1:1 lines. There are only small divergences from an apparent to true young fraction ratio of one (up to about 10%). The maximum divergences from this ratio are affected by the choice of young water threshold (Eq. 13). The present calculations have been made using a young water threshold of 18 years. With higher values for the threshold, the maximum divergences from the ~~ratio~~ 1:1 line were found to become larger. Consequently, 18 years is taken as the recommended value for the young water threshold below.

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

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

3.3 Aggregation effects on MTTs for seasonal tracer cycles

Aggregation effects for seasonal tracer cycles have been determined by the methods of Kirchner (2016a) for comparison with the tritium effects. The rainfall input variation has been approximated as a sine wave with a one-year period to imitate the

seasonal tracer cycle, and the sine wave has been traced through the convolution using the gamma distribution. Fig. 109 shows the aggregation effects for the gamma model with α = 1. The pattern is very similar tolike those observed using tritium concentrations (Fig. 54), so it is clear that the effects are effectively the same whether seasonal tracer cycles or radioactive isotopes are being used. Although our methodology was the same as Kirchner's in that two components were combined, we followed the process of starting with the same MTTs and then allowing the second component to become older. For this reason, the results show the dependence of the aggregation error on the difference in MTTs more explicitly than the random sampling of non-similar MTT components method of Kirchner.

4 Case studies from the literature

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

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

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

Tritium measurements at the Waikoropupu Springs began in 1966 and cover almost the rise and fall of the tritium bomb peak in precipitation (Stewart and Thomas, 2008). Fig. 10a shows the tritium concentrations of the recharge, the Main Spring, and the best-fitting model simulations of the data. The mixing models used were two simple LPMs (exponential piston flow and dispersion models) and two compound LPMs (double exponential piston flow (DEPM) and dispersion (DDM) models).The compound models were used because flows, $\delta^{18}O$ and Cl measurements showed that there were two separate water systems contributing to the Main Spring (a shallow system and a deep system, see Stewart and Thomas, 2008). The fraction of the shallow system (b = 0.24) contributing to the Main Spring was determined from a balance model based on the flows, $\delta^{18}O$ and

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

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

4.2 Kuratau River (volcanic ash deposits and andesite)

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

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

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

Hangarua Spring drains from the Mamaku Plateau and flows into Lake Rotorua via the Hamurana Stream, which also gains water from other springs. Tritium samples were collected from about 1970 to the present. For Hangarua Spring and Hamurana Stream, and for many other streams and springs drawing from the Mamaku Ignimbrite plateau, two different flow contributions

are demonstrated by the tritium measurements (Morgenstern et al., 2015). These contributions are (relatively) young water from shallow aquifers seen in minor streams maintained by shallow aquifers, and old water from deep aquifers seen in aquifers with very deep groundwater tables in the area (Rosen et al., 1998).

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

Similarly, Hamurana Stream shows moderate aggregation bias (apparent to true MTTs of 61 and 144 years with the EPM and 76 to 160 years with the DM) (Table 2, Fig. 13a,b). The compound models consist of about 35% of 12-year-old water and 65% 220-year-old water. The young fractions are all about 0.23 based on t_5 of 18 years (Table 2).

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

Samples from a deep groundwater well in Christchurch, New Zealand, demonstrate possible effects of two water feeds with different mean transit times to the well (Stewart, 2012). The well (M35/3637) taps the Wainoni Aquifer (Aq. 4), where it is unconfined in west Christchurch. The first tritium measurement was in 1986 and five subsequent measurements showed a steady rise from near-zero tritium in 1986 as the bomb tritium peak passed through the site (Fig. 14a). The best-fitting exponential piston flow model (EPM) simulation to all points had f = 0.75 and MTT = 105 years. EPM fits to each individual point gave mean ages close to 105 yr.

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

Carbon-14 measurements collected at the same time as the tritium measurements (1986 to 2006) gave mean ages of 94, 283, 190 and 324 years according to the EPM model with f = 0.75 (Stewart, 2012). These show that the water feeding the well

~~became older on average after 1986. The later tritium samples were not able to show this increase in mean age because the extra old water added had very little tritium and therefore was 'invisible' to the tritium method.~~

~~5~~4 Discussion

~~4~~5.1 Implications of tritium MTT aggregation bias

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

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

~~**4The bias is particularly severe when one component is especially young. Fig. 6 shows the bias for the extreme case when the young component (MTT1) is 0.01 years. The bias rapidly increases as the old component (MTT2) becomes older. By the time the MTT difference between the components is equal to the young water threshold (18 years), the bias is 50% for Southern Hemisphere locations and 75% for Northern Hemisphere locations. On the other hand, when MTT1 = 10 years (dashed curves), maximum errors are much smaller (8% for Southern Hemisphere and 15% for Northern Hemisphere locations by the young water threshold).**~~

Formatted: Font: Bold

Formatted: Font: Bold, Highlight

Formatted: Font: Bold

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

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

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

Both simple and compound LPMs can be free of aggregation error or conversely be affected by aggregation error depending on whether or not they capture the nature of the heterogeneity in the catchment or groundwater system relevant to the error. Simple LPMs have fewer parameters, but have no ability to capture heterogeneity because of their underlying perceptual model (i.e. the assumption of homogeneity), and therefore would be expected to underestimate MTTs because of aggregation error if there is heterogeneity producing flows with different MTTs in the system. (However, note that a highly skewed simple LPM (in that case a gamma model with $\alpha = 0.24$) was able to mimic the MTT of a specific case of binary flow presented in Sect. 3.1.1 (see Reviewer's Report #1). This did not mean that the model gave a good representation of the TTD of the binary system. The dispersion model has the same ability to be highly skewed if the dispersion parameter (DP) is allowed to become very much larger than normal.) Compound LPMs have more parameters and therefore more flexibility to capture heterogeneity, but the model structure must be based on the underlying perceptual model or the parameters extracted could misrepresent the system and lead to aggregation error. Calibration of the parameters of compound LPMs requires that there be sufficient quantity and quality of data and with enough variation to enable retrieval of the increased number of parameter values with compound LPMs. In the past, the bomb pulse introduced strong variation in tritium concentrations of precipitation, but variation in the future will be very much less because of the passing of the bomb tritium from the atmosphere.

We therefore suggest that the answer to the question in the title of this section may be what has often been practised in the past, even though the term "aggregation error" was not used (e.g. Maloszewski et al., 1993; Uhlenbrook et al., 2002; Stewart and Thomas, 2008; Morgenstern and Taylor, 2009; Stewart, 2010; Blavoux et al., 2013; Morgenstern et al., 2015). This ideally involves evaluation of many types of information about a hydrological system (geological, hydrological, hydrochemical, tritium and other isotopes) to establish a perceptual model, and experiments with simple and compound LPMs in harmony with the derived perceptual model to fit tritium data (and if available other types of chemical or isotopic data). Compound LPMs in harmony with the perceptual model would be expected to yield MTTs with less aggregation error than simple LPMs, because the former have the ability to separate young and old water components while the latter do not. Comparison of MTTs from simple and compound models should then show if there is much aggregation error. Parameters yielded by best-fitting models have been used in the past, but may not be the most appropriate ones if the parameters are to be used in other contexts.

Formatted: Normal, Line spacing: single

Formatted: Underline

Formatted: Justified

Formatted: Font: (Default) +Body (Times New Roman)

Formatted: Font: (Default) +Body (Times New Roman), 10 pt

Formatted: Font: 10 pt

Formatted: Font: (Default) +Body (Times New Roman), 10 pt

Formatted: Font: (Default) +Body (Times New Roman)

Formatted: Font: 10 pt

Formatted: Font: (Default) +Body (Times New Roman)

Formatted: Font: (Default) +Body (Times New Roman)

There is also risk of missing less apparent (alternative) parameter solutions if there are any elsewhere in the parameter space. Gallart et al. (2016) applied a GLUE-based uncertainty assessment method which used Monte Carlo searching of the parameter space of the exponential piston flow model to estimate MTTs from tritium. This allowed the uncertainties of the parameters to be quantified.

5 Summary

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

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

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

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

Formatted: Space After: 6 pt

Formatted: Space Before: 0 pt, After: 6 pt

Formatted: Justified

Formatted: Font: Not Italic

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

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

5.4 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. But prior knowledge of the type of LPM and the values of the parameters are needed to determine TTDs for single samples. 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 sample several tritium samples separated in time at each of a number of different streamflows and experiment with fitting different LPMs (and possibly compound LPMs) at each of the flows.

This work is cautionary for applications of simple LPMs to hydrological systems (at least for estimation of MTTs) because of the risk of underestimation of MTTs due to aggregation bias. Compound LPMs are to be preferred, and there can be good reasons for choosing realistic compound 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).

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 different because the two methods are applicable to different time ranges, up to 5 years for seasonal tracer cycles and up to 200 years for tritium concentrations. We also find that young water fractions derived from tritium are almost immune to aggregation errors as were those derived from seasonal tracer cycles.

To investigate the implications of these findings for past and future use of tritium for estimating mean transit times in catchments and groundwater systems, we examined case studies from the literature in which simple and compound LPMs had been used. We find that MTT aggregation errors are small when the component waters have similar MTTs to each other. On the other hand, aggregation errors are large when very young water components are mixed with old components. In general, well-chosen compound LPMs should be used as they will significantly reduce potential aggregation errors due to the application of simple LPMs. Well-chosen means that the (compound) LPM is based on hydrologically and geologically validated information. The choice of a suitable LPM can be assisted by matching simulations to time series of tritium measurements (underlining the value of long series of past tritium measurements), but such results should also be validated to ensure that the parameters won from modelling correspond to reality (since we nearly always have sufficient hydrological/geological data to examine the modelling results).

References

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

Cook, P.G. and Böhlke, J-K.: Determining Timescales for Groundwater Flow and Solute Transport. Chapter 1 in Eds. P.G. Cook and A.L.Herczeg. Environmental Tracers in Subsurface Hydrology. pp 1-30, 2012.

~~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.~~

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

Gallart, F., Roig-Planasdemunt, M., Stewart, M.K., Llorens, P., Morgenstern, U., Stichler, W., Pfister, P., and Latron, J.: A GLUE-based uncertainty assessment framework for tritium-inferred transit time estimations under baseflow conditions. Hydrol. Process. (Keith Beven Tribute*), 30*, 4741-4760, DOI: 10.1002/hyp.10991, 2016.

Grabczak, J., Maloszewski, P., Rozanski, K., and Zuber, A.: Estimation of the tritium input function with the aid of stable isotopes, Catena, 11, 105-114, 1984.

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

Gusyev, M.A., Morgenstern, U., Stewart, M.K., Yamazaki, Y., Kashiwaya, K., Nishihara, T., Kuribayashi, D., Sawano, H., and Iwami, Y.: Application of tritium in precipitation and baseflow in Japan: a case study of groundwater transit times and storage in Hokkaido watersheds. *Hydrology and Earth System Sciences 20*, 3043-3058. doi:10.5194/hess-20-3043-2016, 2016.

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

Formatted: Space Before: 0 pt, After: 6 pt

Hale, V.C., McDonnell, J.J., Stewart, M.K., Solomon, D.K., Pack, R.T., Ice, G.G., and Doolitte, J.: Effect of bedrock permeability on stream base flow mean transit time scaling relationships: 2. Process study of storage and release. Water Resour. Res., 52, 1375–1397, doi:10.1002/2015WR017660, 2016.

Hrachowitz, M., Benettin, P., van Breukelen, B.M., Fovet, O., Howden, N.J.K., Ruiz, L., van der Velde, Y., and Wade, A.J.: Transit times—the link between hydrology and water quality at the catchment scale. WIREs Water, 3, 629–657. doi: 10.1002/wat2.1155, 2016.

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

IAEA/WMO Global Network of Isotopes in Precipitation. The GNIP Database. Accessible at: http://www.iaea.org/water, 2016.

Kirchner, J.W.: Aggregation in environmental systems – Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments, Hydrol. Earth Syst. Sci., 20, 279-297, doi:10.5194/hess-20-279-2016, 2016a.

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

Kirchner, J.W., Feng, X., and Neal, C.: Fractal stream chemistry and its implications for contaminant transport in catchments, Nature, 403, 524-527, doi:10.1038/35000537, 2000.

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

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

Małoszewski, P., Stichler, W., Zuber, A., and Rank, D.: Identifying the flow systems in a karstic-fissured-porous aquifer, the Schneealpe, Austria, by modelling of environmental ^{18}O and ^{3}H isotopes, J. Hydrol., 256, 48-59. 2002

Małoszewski, P., and Zuber, A.: Determining the turnover time of groundwater systems with the aid of environmental tracers: 1. Models and their applicability, J. Hydrol., 57, 207-231, doi:10.1016/0022-1694(82)90147-0, 1982.

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

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

Michel, R.L, Aggarwal, P., Araguas-Araguas, L., Kurttas, T., Newman, BD, and Vitvar, T.: A simplified approach to analysing historical and recent tritium data in surface waters. Hydrol. Process. 29, 572–578, DOI: 10.1002/hyp.10174, 2015.

Michel, R.L.: Residence times in river basins as determined by analysis of long-term tritium records, J. Hydrol. 130, 367–378, 1992.

Ministry of Health: Drinking-water Standards for New Zealand 2005 (revised 2008). Ministry of Health, PO Box 5013, Wellington, New Zealand, ISBN 978-0-478-31809-8, 2008.

~~Morgenstern, U.: Lake Taupo streams - Water age distribution, fraction of landuse impacted water, and future nitrogen load. Environment Waikato Technical Report TR 2007/26, Document #1194728, p. 28, 2007.~~

Morgenstern, U., Daughney, C. J., Leonard, G., Gordon, D., Donath, F. M., and Reeves, R.: Using groundwater age and hydrochemistry to understand sources and dynamics of nutrient contamination through the catchment into Lake Rotorua, New Zealand, Hydrol. Earth Syst. Sci., 19, 803–822, 2015.

Morgenstern, U., Stewart, M.K., and Stenger, R.: Dating of streamwater using tritium in a post nuclear bomb pulse world: continuous variation of mean transit time with streamflow. Hydrol. Earth Syst. Sci., 14, 2289-2301, 2010.

Morgenstern, U., and Stewart, M.K.: Stream water dating and future nutrient load to Lake Taupo, New Zealand,. In International Workshop on the Application of Isotope Techniques in Hydrological and Environmental Studies, UNESCO, Paris, 6–8 September. 2004.

Morgenstern, U., and Taylor, C.B.: Ultra low-level tritium measurement using electrolytic enrichment and LSC, Isotopes Environ. Health Stud., 45, 96–117. doi:10.1080/10256010902931194, 2009.

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

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

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

Stewart, M.K.: A 40-year record of carbon-14 and tritium in the Christchurch groundwater system, New Zealand: Dating of young samples with carbon-14, J. Hydrol., 430-431, 50-68, DOI: 10.1016/j.jhydrol.2012.01.046, 2012.

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

Stewart, M.K., Mehlhorn, J., and Elliott, S.: Hydrometric and natural tracer (18O, silica, 3H and SF6) evidence for a dominant groundwater contribution to Pukemanga Stream, New Zealand, Hydrol. Process., 21, 3340–3356, DOI: 10.1002/hyp.6557, 2007.

Stewart, M.K. and Morgenstern, U.: Importance of tritium-based transit times in hydrological systems, WIREs Water 3, 145–154, doi: 10.1002/wat2.1134, 2016.

Stewart, M.K., Morgenstern, U. and McDonnell, J.J.: Truncation of stream residence time: How the use of stable isotopes has skewed our concept of streamwater age and origin, Hydrol. Process., 24, 1646-1659, 2010.

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

Stewart, M.K. and Thomas, J.T.: A conceptual model of flow to the Waikoropupu Springs, NW Nelson, New Zealand, based on hydrometric and tracer (^{18}O, Cl, ^{3}H and CFC) evidence, Hydrol. Earth Syst. Sci., 12, 1-19, http://www.hydrol-earth-syst-sci.net/12/1/2008/hess-12-1-2008.html, 2008.

Taylor, C.B., Brown, L.J., Cunliffe, J.J, and Davidson, P.W.: Environmental isotope and 18O applied in a hydrological study of the Wairau Plain and its contributing mountain catchments, Marlborough, New Zealand, J. Hydrol. 138, 269–319, 1992.

Uhlenbrook, S., Frey, M., Liebundgut, C., Maloszewski, P.: Hydrograph separations in a mesoscale mountainous basin at event and seasonal timescales, Water Resour. Res., 38, doi:10.1029/2001WR000938, 2002.

Weiler, M., McGlynn, B. L., McGuire, K. J., and McDonnell, J. J.: How does rainfall become runoff? A combined tracer and runoff transfer function approach, Water Resour. Res., 39, 1315, doi:10.1029/2003WR002331, 2003.

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

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

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

Feature	Model	MTT τ_m (yr)	b	Component 1 τ_1 (yr)	f_1	D_{P1}	Component 2 τ_2 (yr)	f_2	D_{P2}	sd (TU)	Y_f (yr)
Waikoropupu Main Spring	EPM	7.9	1.00	7.9	0.94	--	--	--	--	0.36	0.86
	DEPM	7.8	0.24[‡]	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	--	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

-- Columns not applicable for these cases. [‡]Value of b determined from mass balance (see text).

Formatted: Indent: Left: 0 cm, First line: 0 cm, Line spacing: single

~~– Columns not applicable for these cases.~~ ~~[‡]Value of b determined from mass balance (see text).~~

Formatted: Indent: Left: 0 cm, Hanging: 1.27 cm, Line spacing: single

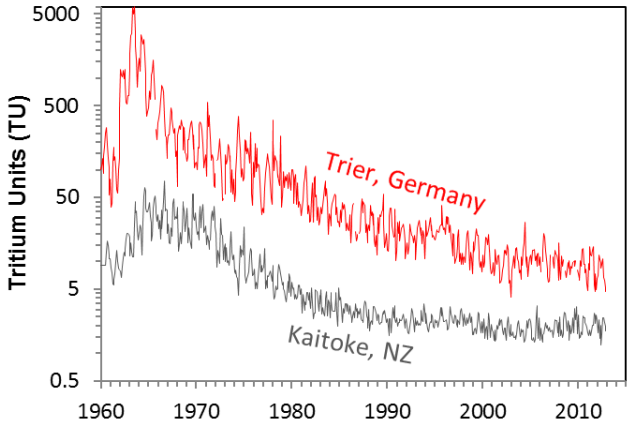
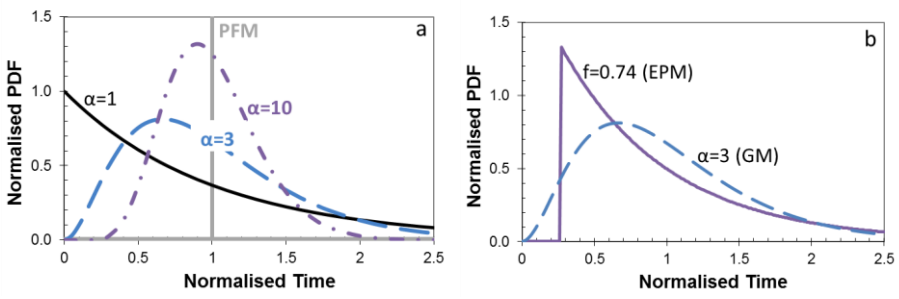


Figure 1. Tritium concentrations (TU) in monthly precipitation samples at Kaitoke, New Zealand in the Southern Hemisphere, and Trier, Germany in the Northern Hemisphere.



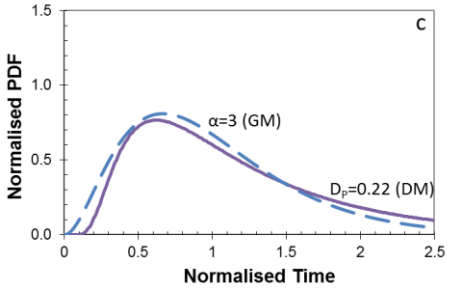


Figure 2. (a) Gamma distributions (GM) for shape factors α between 1 and 10. The axes show normalised transit time (τ/τ_m) and normalised probability density function PDF ($h(\tau) \times \tau_m$). Note that the distribution for GM ($\alpha = 1$) is the same as that for the exponential model (EM). (b-c)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

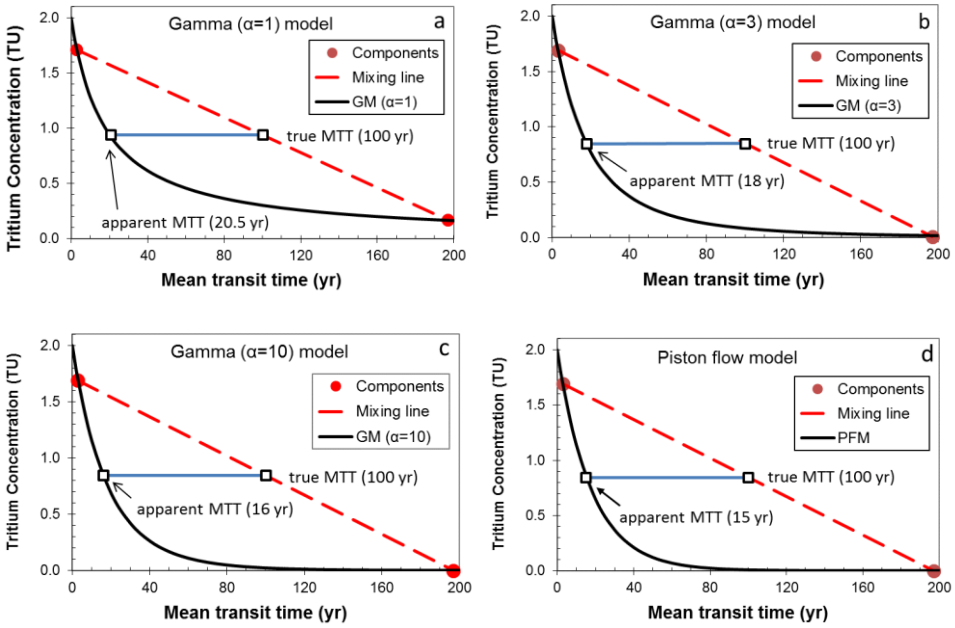


Figure 3(a-d). Aggregation errors when the tritium input concentration is assumed to be constant at 2 TU. Mean transit times (MTTs) are inferred from tritium concentrations in mixed runoff from two subcatchments with different tritium concentrations and

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

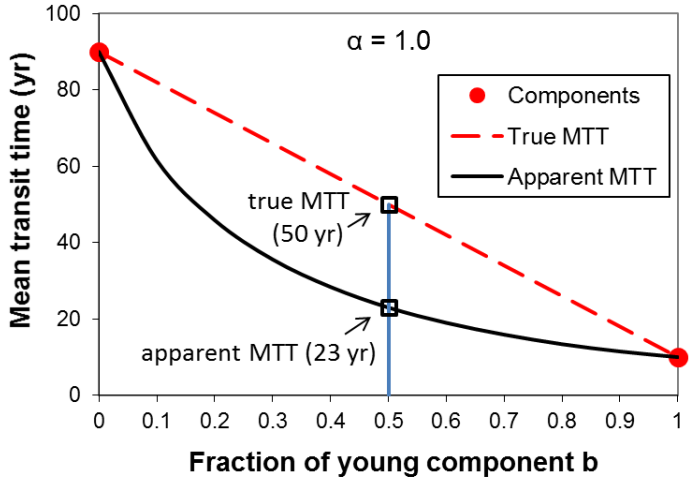
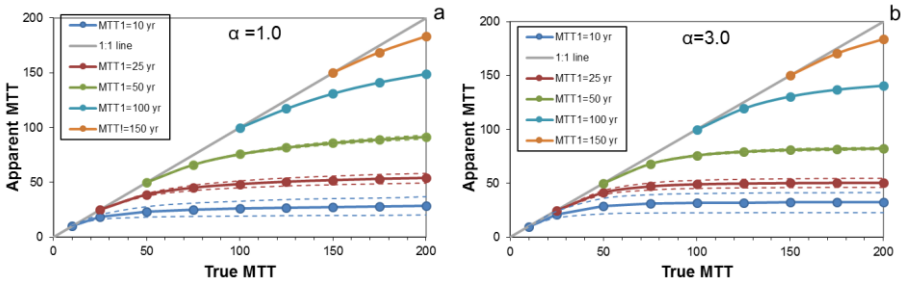


Figure 4. **Effect of changing b (the young component fraction) on the aggregation error with the gamma model with $\alpha = 1$ for mixing of two components with MTTs of 10 and 90 years. The blue line shows the effects at b = 0.5.**

Formatted: Font: 9 pt



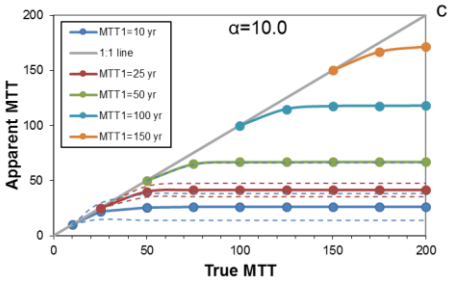


Figure 45(a-c): Aggregation effects for tritium MTTs for gamma models with different values of α using the Kaitoke input function. Curves show changes as component 2 (MTT2) becomes older in 50-year steps and therefore the mixtures older in 25-year steps (shown by dots). The first step of the MTT1=10 yr curve is 15 years. Fitting errors in the apparent MTTs are shown by fine dashed lines.

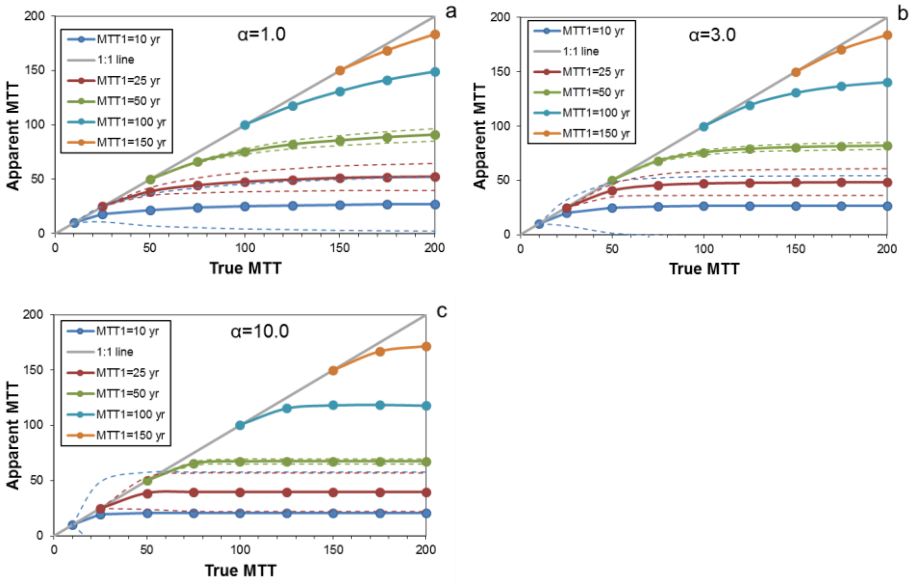


Figure 56(a-c). Aggregation effects for tritium MTTs using the Trier input function. Symbols as in Fig. 4.

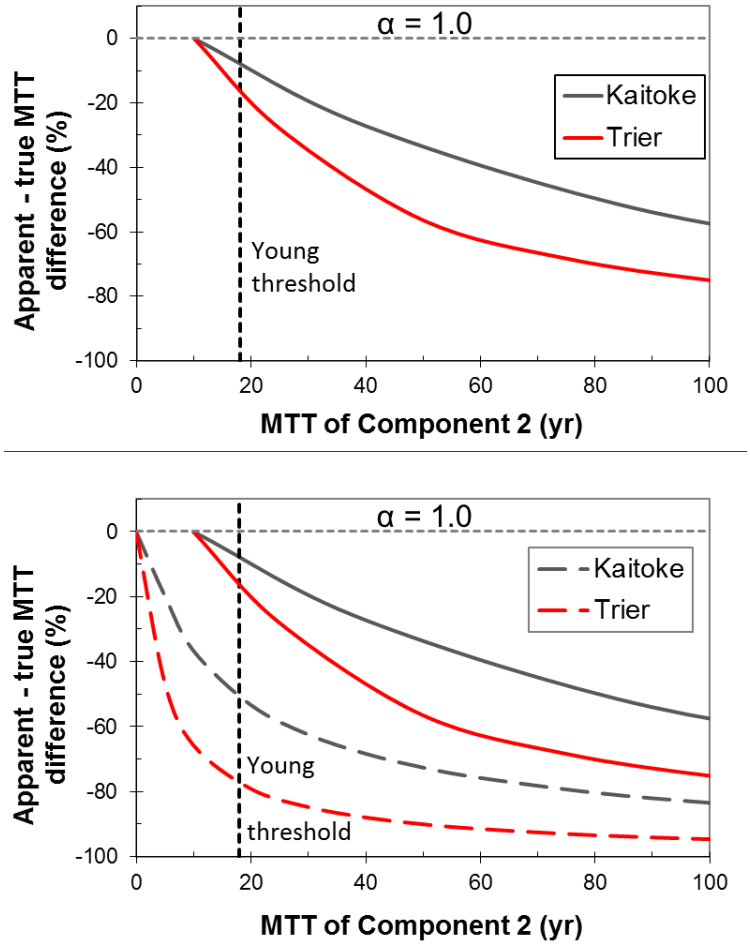


Figure ~~67.~~ Comparison of maximum aggregation effects for Northern (Trier) and Southern (Kaitoke) Hemispheres for the gamma model with α = 1. The ~~solid~~ curves show results with component 1 having mean transit time of 10 years (MTT1 = 10 yr)~~, and the dashed curves with MTT1 = 0.01 yr.~~

Formatted: Space After: 0 pt, Line spacing: 1.5 lines

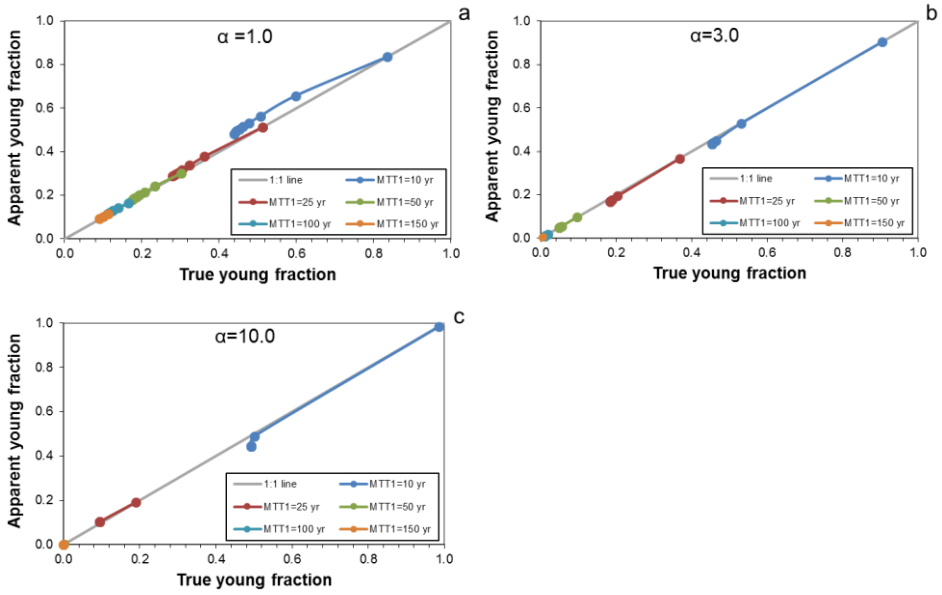
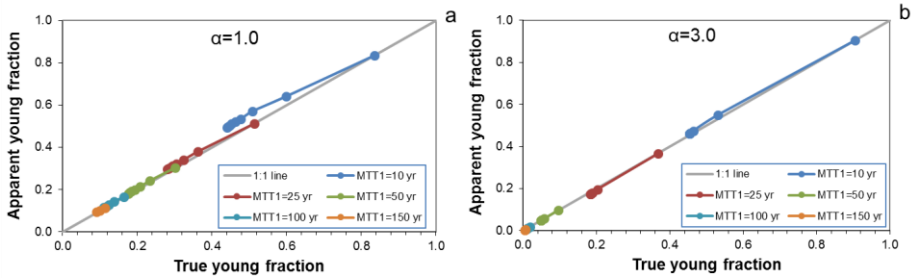


Fig. ~~7~~8(a-c). True versus apparent tritium young water fractions for gamma models with different values of *α* using the Kaitoke input function. Curves show changes as component 2 (MTT2) becomes older in 50-year steps and therefore the mixtures older in 25-year steps (shown by dots).



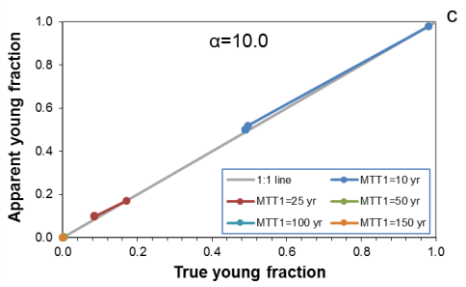


Figure ~~8~~9(a-c). Tritium young water fractions using the Trier, Germany tritium input function. Symbols as in Fig. 7.

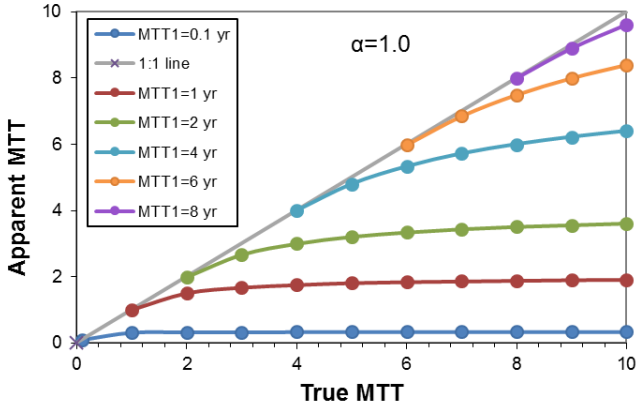
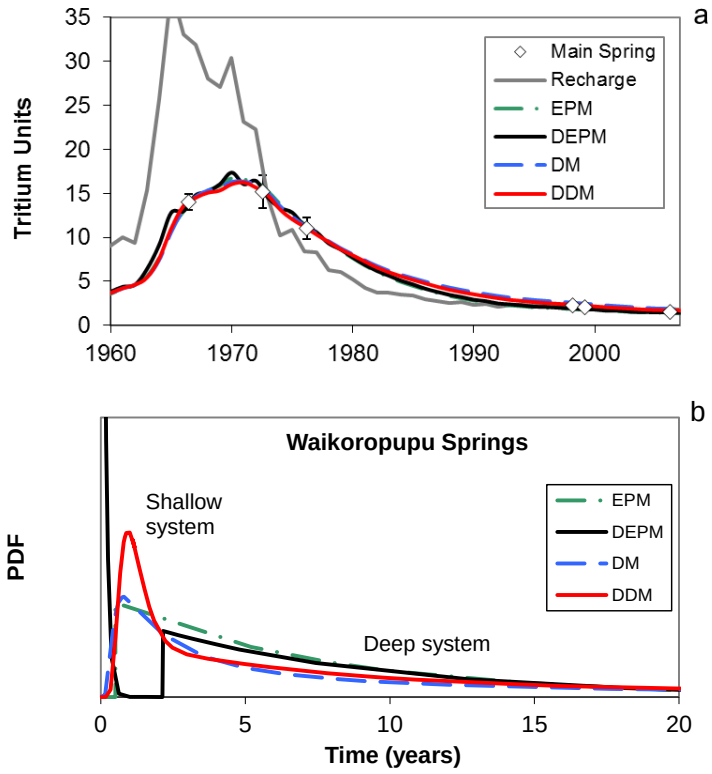
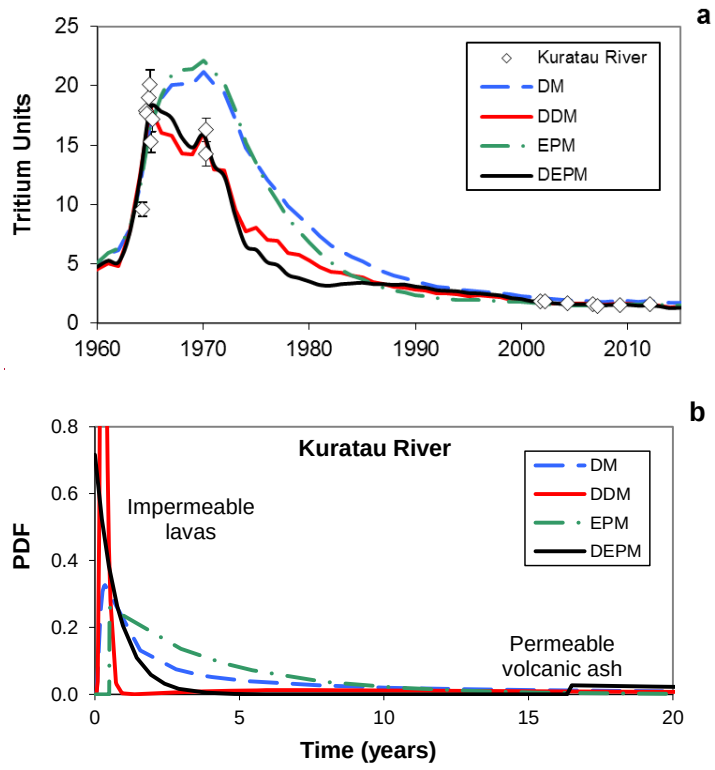


Figure ~~9~~10. Aggregation effects on MTTs determined using seasonal tracer cycles for the gamma model with α = 1.0. Curves show changes as component 2 (MTT2) becomes older in 2-year steps and therefore the mixtures older in 1-year steps (shown by dots)



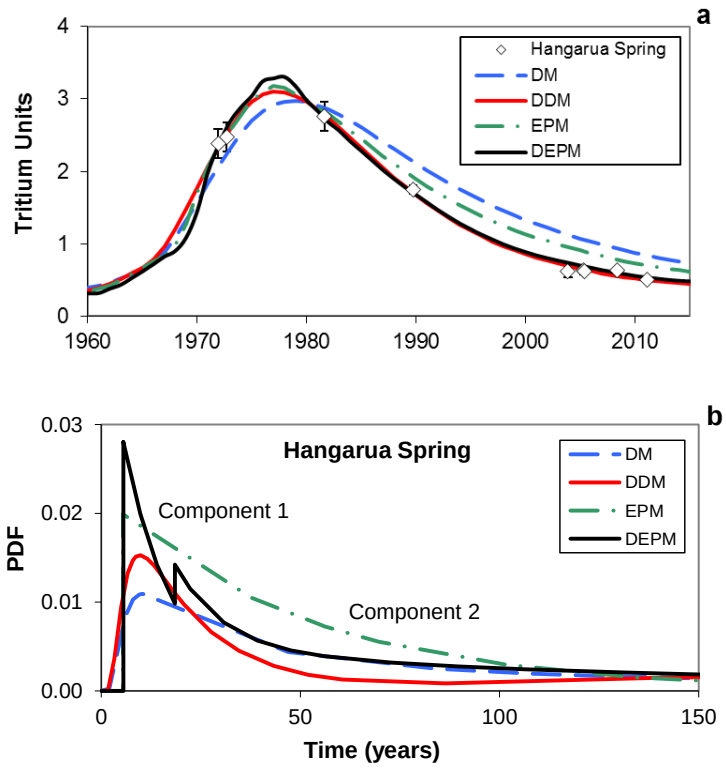
~~Figure 10. (a) Tritium measurements of the Main Spring of the Waikoropupu Springs, New Zealand, and model fits to the data. EPM and DM are exponential piston flow and dispersion models, and DEPM and DDM are double (i.e. binary parallel) EPM and DM models. (b) Transit time distributions of the best-fit simulations.~~

Formatted: Left

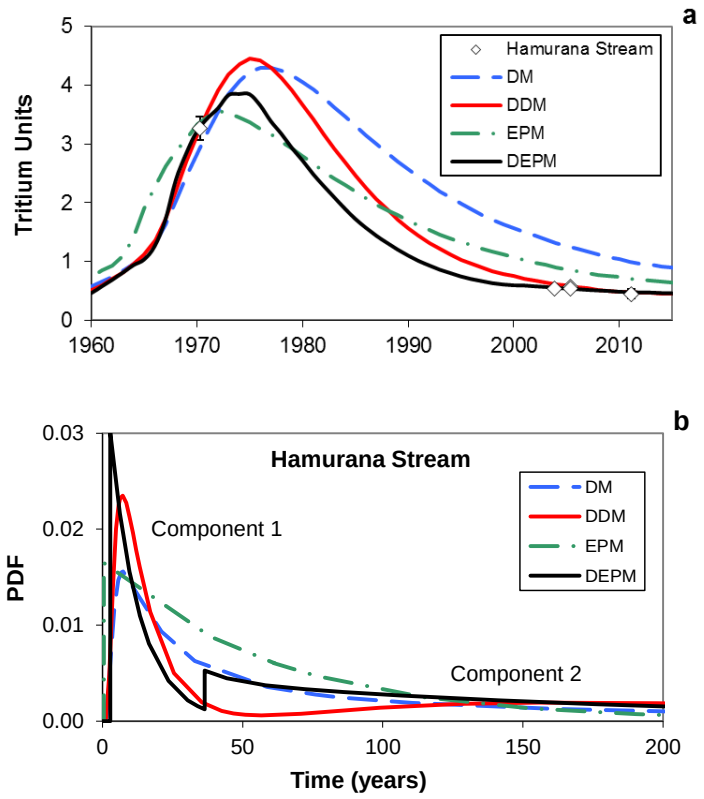


~~Figure 11. (a) Tritium measurements and model fits for Kuratau River, New Zealand. (b) Transit time distributions of the best-fit simulations.~~

Formatted: Space After: 0 pt, Line spacing: 1.5 lines

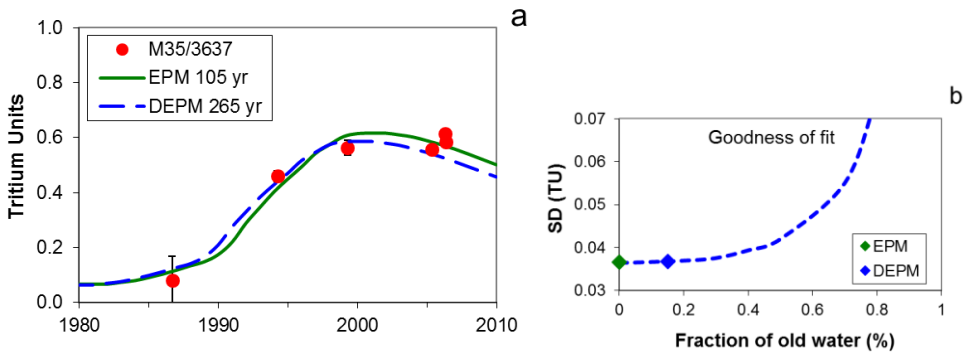


~~Figure 12. (a) Tritium measurements and model fits for Hangarua Spring, Rotorua, New Zealand. (b) Transit time distributions of the best-fit simulations.~~



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

5



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