We appreciate the many helpful comments of Anonymous Referee #3.

Ref #3: General comments
The results summarized in the first paragraph of the abstract and Figures 1-9 of the manuscript itself are, to the best of my knowledge, broadly correct. However, the results reported in the second paragraph (and the related discussion and analyses in section 3.4 and 4) need to be reconsidered.

Reply: We will reconsider the second paragraph of the abstract and related analyses and discussion as suggested by the Referee.

Ref #3: 1. The statement that "well-chosen compound lumped parameter models should be used as they will eliminate potential aggregation errors due to the application of simple lumped parameter models" directly implies that aggregation errors only arise with simple LPM's, but not with compound LPM's, or at least not with "well-chosen" ones (whatever that means).

Reply: We stand by this statement, except that we would now replace the word "eliminate" by "reduce or greatly reduce". In general, "well-chosen" means that the compound lumped parameter model (LPM) captures important aspects of the water flow in the catchment or groundwater system, as in a conceptual model. In particular, delineation between subsystems delivering young water and old water to the system outflow is very important in this context because the aggregation error is caused by the disproportionate effect of young water.

Ref #3: 1.a. This is not consistent with the analysis presented elsewhere in the paper. Figures 3-5 show clear aggregation errors from the use of simple lumped parameter models, but they would also show clear aggregation errors if more complex lumped parameter models were used. For example, a gamma model with alpha=0.3 closely approximates a compound LPM, but it is clearly vulnerable to aggregation errors, as shown in Figures 4 and 5.

Reply: As noted above, a compound model would have to have a transit time distribution (TTD) representing the relative contributions of young and old water subsystems to be able to reduce potential aggregation errors. This separation can be achieved approximately by a binary or more complicated LPM, but not by a simple LPM. If a simple LPM (such as a gamma model with alpha=0.3) accurately described the TTD of a system, it would not have much aggregation error. (This is shown by the increasing aggregation error as the contrast between the ages of the two water components increase in Fig. 4a).

Ref #3: 1.b. Since the analyses in Figures 3-5 have been used to demonstrate aggregation errors in simple LPM's, exactly the same analyses must be applied to compound LPM's to demonstrate that these aggregation errors disappear. Until this is done, the claims in the abstract have not been demonstrated, and must be removed.

Reply: We reject this demand, because the Referee has misunderstood the situation. It is obvious that compound (especially binary) models would work very well with the virtual experiments carried out here (i.e. the analyses in Figures 3-5), because the experiments are based on combining two subsystems with the same or different mean transit times. Optimised compound models would simply separate the components back out again and produce (near) zero aggregation error. Applying compound models to the results of the virtual experiments is of course not the same

thing as experimenting with real systems, which is why we considered some cases from the literature.

Ref #3: 2. In some of the examples that are presented, the compound LPM's clearly fit the data better, but of course they should, because they have more free parameters. Whether these parameters are fitted by formal calibration or by "expert judgment" and fitting by hand makes little practical difference; in either case they make the fitted curves more flexible and thus more conformable to the data. (This comes at the cost of greater parameter uncertainty; more about that below.)

Reply: What the Referee has not understood is that the improvement in fit to the data by an optimised compound LPM (due to there being more free parameters) means that the model more accurately represents the TTD of the system. This means that the compound LPM has less potential to produce aggregation error.

Ref #3: 2.a. In the case of the "DDM" in Figure 10 and Table 2, for example, there are FIVE adjustable parameters: b, tau_s, tau_d, P_Ds, and P_Dd (incorrectly labeled as a second "P_Ds" in the table). So Figure 10 shows a five-parameter fit to just six data points (which are themselves not fully independent of one another). Is it any surprise that the curve fits well? The other models have at least two parameters, for a data set that effectively has only two or three unique values; those near the peak and those in the 2000's. Again, it is not at all surprising that these can be calibrated to fit the data.

Reply: This criticism is also invalid because it takes no account of the wealth of evidence for this particular compound LPM (i.e. the DDM) presented in Stewart and Thomas (2008). This evidence is referred to in the sentence (P10-L22): "The DDM was used because δ18O and Cl measurements showed that there were two separate water systems contributing to the Main Spring (a shallow system and a deep system)". We also reject the description of the tritium dataset as "effectively … only two or three unique values". There are six fully independent measurements covering 40 years (including most of the rise and fall of the bomb tritium peak).

Evidence in Stewart & Thomas includes determination of recharge and discharge for the catchment and Main Spring based on extensive flow measurements and the 18O balance. The two flow systems are identified by the 18O values and most clearly by their chloride concentrations, since chloride is a powerful tracer here because the deep system contains a small proportion of sea water while the shallow one does not. The proportions of the flows (shallow fraction, b) for the Main Spring comes from the relationships between the flows, 18O and chloride and concentrations. The dating was based on tritium, CFC-11 and 18O evidence (CFC-12 gave anomalous results as it frequently has done in other groundwater studies).

Ref #3: 3. Figure 10c is presented as evidence that "the mean residence times were sharply constrained close to 8 years". This is at best unproven and at worst misleading.

Reply: See replies to sections 3 and 3a – 3d following 3e and 3f.

Ref #3: 3.a. Consider, for example, the red curve for the DDM. In the DDM, the mean residence time (MRT) is a function of three parameters (b, tau_s, and tau_d), and the tritium curve, and thus the fit to the data (SD) is determined by these three parameters, plus two others (P_Ds and P_Dd). It is mathematically impossible for the relationship between MRT (which depends on three independent parameters) and SD (which depends on five independent parameters) to be described by a single

curve. There will be multiple combinations of b, tau_s, and tau_d that give the same MRT but different values of SD, and the range of SD will be inflated further by variations in P_Ds and P_Dd.

3.b. The same problem arises, in simpler form, for the EPM and DM. The DM, for example, depends on a residence time and a dispersion parameter P_D; for any individual value of the dispersion parameter, one can draw a curve relating the residence time to the misfit parameter SD. But to describe the relationship between SD and the residence time, one needs a full family of curves, to represent the range of possible values of the dispersion parameter.

3.c. It is impossible to know for sure (since the methods are unacceptably vague on this point), but it seems likely that Figure 10c was generated by choosing fixed values for all-but-one parameter in each model, and then varying just one parameter and tracing out the resulting relationship between MRT and SD.

3.d. From a parameter estimation standpoint, this is a fundamentally flawed procedure, because (1) it ignores the extra degrees of freedom from the other parameters that are arbitrarily held constant, and (2) it therefore underestimates the uncertainty in the MRT, possibly by large factors. This is true even if the parameters were fixed by "expert judgment" rather than algorithms, as long as the experts were free to revise their "judgment" based on whether the tritium curves made sense.

3.e. Methods for multi-variable parameter estimation and uncertainty analysis are widely available. There is no valid excuse for not using them. The revised manuscript must eliminate all claims (explicit or implied) about MRT's estimated from tritium measurements using multi-parameter models, unless and until proper parameter estimation and uncertainty analysis are done.

Reply: This demand would have eliminated all tritium papers in the past up to that recently published by Gallart et al. (2016)! As far as we know, their paper is the only attempt to apply multi-variable parameter estimation methods to tritium measurements, and it was by no means a trivial exercise. Presumably the Referee would have cited any relevant earlier work if any had been available. Gallart et al. produced complex parameter diagrams with multiple solutions, from which likelihood-weighted cumulative density functions were determined. However, their work was concerned with a Northern Hemisphere location with a relatively short record (1996 to 2013), so multiple solutions were to be expected (Stewart and Morgenstern, 2016). The very different Southern Hemisphere tritium input function and the sample record from 1966 would give very different and very much more straight-forward parameter diagrams for the Waikoropupu Springs (Stewart and Morgenstern, 2016). It will be interesting to apply the methods to Southern Hemisphere locations and we plan to carry out such calculations, but we think it is beyond the scope of this paper.

Ref #3: 3.f. There is likewise no valid reason for ignoring the uncertainties in the tritium measurements themselves, and their consequences for parameter uncertainties. Looking at the error bars in Figure 10a, for example, one can estimate that the pooled standard deviation (due to the measurement uncertainties themselves) is about 1-2 TU. Therefore, Figure 10c implies that the MRT is only constrained within about plus or minus two years (for a standard deviation of 1 TU) or about plus or minus four years (for a standard deviation of 2 TU), which is quite a contrast to the paper's assertions that the MRT is "sharply constrained". And this estimate does not even begin to account for the additional uncertainty introduced by

the other four parameters. Again, there are standard methods for propagating these uncertainties in parameter estimation, and there is no valid excuse for not using them.

Reply: We have estimated the effects of the tritium measurement errors and considered the effects of adjusting the (previously unadjusted) model parameters below. Tritium data from the Waikoropupu Spring are given in Table 1 (data from Table 7 in Stewart and Thomas, 2008). The fractional tritium measurement error decreased between 1966/76 and 1998/2006 because of methodological improvements (Morgenstern and Taylor, 2009).

Table 1: Tritium concentrations in the Main Spring of the Waikoropupu Springs. Errors are one standard deviation.

Date	Tritium (TU)	Error (TU)	Fractional error
27-05-66	14.0	0.9	0.064
29-07-72	15.2	1.9	0.125
20-03-76	11.0	1.2	0.109
26-02-98	2.25	0.06	0.027
16-03-99	2.08	0.08	0.038
21-03-06	1.53	0.05	0.033

Table 2 gives the mean transit times determined taking account of the measurement errors. We have considered a worst case scenario by fitting LPMs to "Low", "Mid" and "High" cases, in which all of the errors are subtracted from the tritium measurements in the "Low" case, and all are added on in the "High" case. This will give a larger range of MTTs than a Monte Carlo sampling technique would have done, because the measurements have uncorrelated errors and there is very low likelihood that all of the measurements would have been low or high together. The fits were optimised using different numbers of the model parameters (as shown in column 2). The parameter b (fraction of the shallow system) was set at 0.26 ± 0.10 based on flow and 18O balance measurements (Table 4, Stewart and Thomas, 2008). (This produced a relatively minor change in MTT of ±0.2 years about the value with b=0.26.) The range of variation around the mid MTT (1.4 to 2.2 years, with one outlier) gives a good indication of the uncertainty of the mean transit times. We consider that this qualifies as "sharply constrained".

Table 2: Mean transit times (τ_m) determined by fitting the LPMs to the tritium concentrations by adjusting the parameters shown. "Low", "Mid" and "High" columns show the τ_m obtained for tritium value minus error, tritium value and tritium value plus error. The plus/minus column shows the average variation around the mid age. Parameters not adjusted in a particular LPM (i.e. not listed in column 2) were set at their optimised mid values. Parameter b for the DDM was set at 0.26 based on other measurements.

LPM	Parameters	Low τ_m (yr)	Mid τ_m (yr)	High τ_m (yr)	$\pm$ (yr)
EPM	τ_m	9.5	7.9	6.7	1.4
EPM	τ_m, f	9.5	8.7	6.6	1.5
DM	τ_m	10.4	8.2	6.5	1.9
DM	τ_m, P_D	13.0	8.7	5.7	3.8
DDM	τ_d	10.0	7.9	6.3	1.9
DDM	τ_s, τ_d,	9.3	7.8	6.2	1.6
DDM	τ_s, τ_d, P_{Ds}, P_{Dd}	10.2	7.7	5.8	2.2

Ref #3: 4. As was also pointed out by another reviewer, the claims that compound LPM's have less aggregation bias are not supported by clear lines of reasoning. For example:

Reply: We think that there <u>are</u> clear lines of reasoning for this claim, which perhaps we have not explained clearly enough. Simple LPMs assume homogeneous systems, compound ones are binary or more complicated systems. In terms of our virtual experiments, the simple LPMs yield the "apparent" MTTs while the binary LPMs yield the "true" MTTs; in these experiments the "true" MTTs actually are true because we have built the systems by adding two subsystems together (in the proportions of 1:1 for convenience) making binary systems. When the two subsystems have the same MTTs, the simple and compound LPMs yield the same MTTs (and therefore plot on the 1:1 line in Fig. 4). As the subsystem MTTs become more and more different, so the simple (i.e. apparent) and compound (i.e. true) MTTs become more different and the aggregation error increases. This is quite clear and so is the reason for it, i.e. that the young water component outweighs the old water component (or as shown in Fig. 3, the relationship between MTT and tritium concentration is non-linear).

When it comes to applying simple and compound LPMs to real systems (the four case studies), the simple LPMs yield apparent MTTs, but the compound LPMs yield true MTTs <u>if and only if</u> they capture the separation between young and old water subsystems in the overall system. If the compound model does not capture this separation, then it may not help in terms of aggregation error. In the case of the virtual experiments, the compound (i.e. binary) model would of course be perfect, because that is how we have set the experiments up.

To summarise, application of a compound LPM to real systems will greatly reduce the possibility of aggregation error, if it captures approximately the separation between young and old water subsystems. The question then becomes: how well do the compound models capture this separation in the four case studies described in Section 3.4? The examples were chosen to cover the tritium range and because there was evidence that simple LPMs had been inadequate in their cases (the evidence included both hydrological/chemical/geological and tritium evidence). We think that the compound models are well-chosen (in the sense defined above) in these four cases. But we agree that explanation of this background and further analysis of the compound models is required (as pointed out by Refs #2 and #3).

Ref #3: 4.a. In 3.4.1, the manuscript says that there is little aggregation bias because the simple and compound LPM's have similar mean residence times. But why does this imply an absence of aggregation bias, rather than a similar aggregation bias across all three LPM's? The manuscript also argues that we should expect little aggregation bias because the two model components have MRT's that are similar to, or shorter than, the half-life of tritium. This is only a valid argument if we have independent evidence about the ages of the system components. What evidence do we have that the deep aquifer really contributes 74% of the flow and has a MRT of 10.2 years, instead of (say) 35% of the flow with a MRT of 100 years? If such independent information exists, the reader should be made aware of it. Alternatively, the manuscript needs to demonstrate that the MRT's of the individual system components can be reliably constrained through parameter estimation (which will not be easy).

Reply: We have explained why there is little aggregation bias if the simple and well-chosen compound LPMs give the same MTTs (just above). All three of the optimised LPMs for the Waikoropupu Main Spring match the tritium concentrations very well, and they all give very similar TTDs.

There would have been no need to invoke a compound LPM, except that there is overwhelming evidence that the springs are fed by two flow subsystems (Stewart and Thomas, 2008). There is no possibility that (for example) the deep aquifer contributes 35% of the flow with an MTT of 100 years, as perusal of Stewart and Thomas (2008) would make clear. 1. The recharge/discharge model for the springs and overall system would need to be completely different. 2. The 18O concentrations would not balance, and its variations in time would need to be different. 3. The chloride concentrations would need to be different. 4. The relationships between the flows, chloride concentrations and 18O in the springs would need to be different. 5. The tritium concentrations in the Springs would need to be lower. 6. The CFC-11 concentrations in the springs would need to be lower. We will include this explanation in the revised paper as requested by the Referee.

Ref #3: 4.b. In 3.4.2 and 3.4.3, the claim seems to be that the simple LPM's are subject to aggregation bias because they disagree with each other or with the compound LPM, which fits the data better. But again, the compound LPM has at least twice as many parameters as the simple LPM's, so one would need to somehow show that the better fit does not simply arise from this rather obvious explanation. And of course the simple LPM's will disagree with each other; they have different shapes, so it is unsurprising that they may have different MRT's when fitted to data.

Reply: We applied one simple and one compound LPM in each of these two cases. Our claim is that there is aggregation bias in each case because there is disagreement between the simple and the compound LPMs. This of course requires that the compound LPMs give good representations of their catchments in regard to separation of young and old water subsystems, i.e. that the compound LPMs be "well-chosen" for their catchments. Compound LPMs have the possibility of correctly combining different parts of the catchment with different characteristics when they are optimised to the data, whereas simple LPMs do not.

We agree that we need to establish more conclusively that these compound LPMs are in fact well chosen for these catchments. For the Kuratau River (Sect. 3.4.2), geological evidence strongly supports two subsystems within the catchment. The

area within the catchment with the very impermeable Whakamaru Group ignimbrites and andesitic and basaltic lavas produces 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 also substantiated these flows and their contrasting ages (Gusyev et al., 2013; 2014).

For Hangarua Spring and Hamurana Stream (Sect. 3.4.3), 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).

Ref #3: 5. One needs to recognize that the abstract's claim that "The choice of a suitable lumped parameter model can be assisted by matching simulations to time series of tritium measurements (underlining the value of long series of tritium measurements)" is mostly a statement about the past, and is misleading as a generalization about the future.

Reply: We thank the Referee for this comment, and acknowledge that there is some truth in this. Note that we really meant "(underlining the value of long series of past tritium measurements)". There is no doubt that identifying LPMs from tritium data will become problematical in the future. However, Northern Hemisphere hydrological systems still contain some bomb tritium and although this can cause problems with ambiguous ages (see Gusyev et al. (2016)), it can also assist identification of LPMs (although identifying suitable *compound* LPMs may be a step too far). Gallart et al. (2016) used Monte Carlo sampling to account for measurement error in tritium and parameter estimation errors to demonstrate that tritium measurements taken now combined with future measurements will enable effective identification of MTTs provided high quality tritium measurements are used (their Fig. 13). They used the EPM model.

Southern Hemisphere systems contain much less bomb tritium and identification of LPMs (i.e. mixing models) is becoming more difficult unless past tritium data is available. High quality measurements are essential, especially in the Southern Hemisphere, because of the low levels of cosmogenic tritium. There are small seasonal variations in cosmogenic tritium (see data for the last 25 years for Kaitoke NZ in Fig. 1) that may become more useful in the future. This situation has already been reported for some areas in Japan (Gusyev et al. 2016) and will occur in other Northern Hemisphere areas.

Ref #3: 5.a. In the (few) springs and aquifers where tritium analyses were performed decades ago, during and after the bomb peak, those analyses have turned out to be quite useful for comparison with the more recent measurements. Indeed, as Figure 12 shows, it is these early samples that allow one to distinguish between the differently shaped LPM's, and the more recent samples have almost no power to discriminate between those same LPM's.

5.b. And that is precisely the problem: going forward into the future, long time series will be much less useful, for the simple reason that the bomb pulse tritium is largely
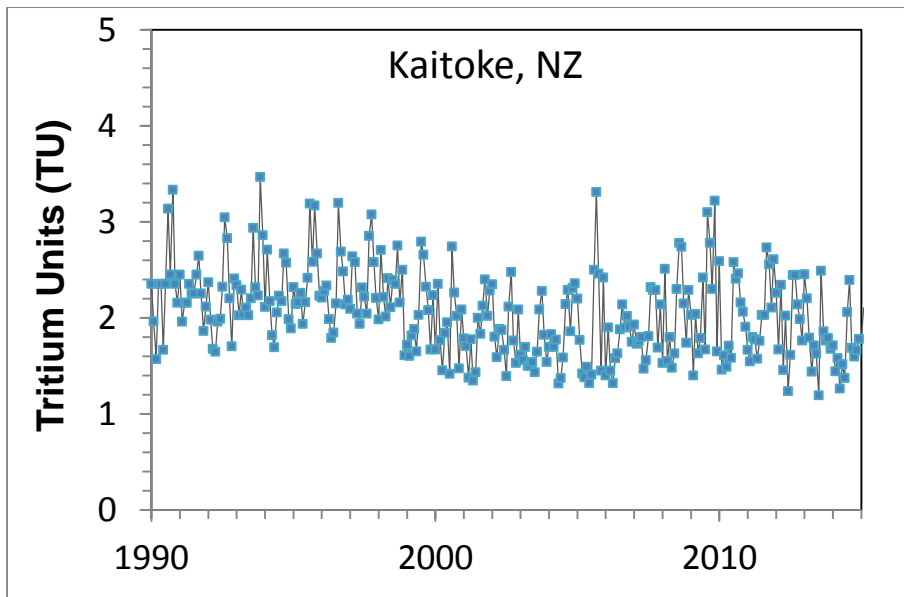
Figure 1. Tritium concentrations in monthly precipitation at Kaitoke, New Zealand. The sizes of the points show the measurement errors (±0.03 TU).

gone and we are approaching an equilibrium between tritium production and decay.

Thus, going forward, long time series will not help, because tritium concentrations are becoming less and less dynamic over time. As the bomb pulse tritium vanishes, we will just be measuring the same value over and over.

Reply: We are not at this stage yet in the Northern Hemisphere, but when we are in 10 to 20 years we will not be able to use tritium for identifying the type of mixing model. To determine MTTs will require assuming a mixing model based on other criteria.

Actually we are unlikely to be in the situation of measuring the same value over and over because systems (especially streams) are inherently unsteady even during baseflow and as the MTT changes so will the tritium concentration. (But again we will need time-series with high-quality measurements to exploit this.)

Ref #3: 5.c. I am sure the authors know this, and it is disingenuous not to make it clear to the reader, particularly because they celebrate the one clear benefit of the fading of the bomb pulse (the end of double solutions for many tritium models).

Reply: We have made the situation clear in other papers, e.g. we urged Northern Hemisphere researchers to start sampling now before all of the bomb tritium is gone (Stewart et al., 2012; Stewart and Morgenstern, 2016). However, we will re-emphasize the point in the revised paper.

Ref #3: 5.d. The fading of the bomb pulse will make the parameter estimation problem outlined above even more impossible than it is already. Consider the red curves in Figure 10 as an example. As mentioned above, these are five-parameter fits to six data points.

In the future, anywhere that we do not already have measurements of bomb pulse tritium, we will instead have a five-parameter fit to what is effectively just ONE data point (because in equilibrium, all future measurements are redundant).

Reply: There will certainly be complications in the future use of tritium, e.g. those connected with the fading of the bomb pulse, those with variations in flow, those due to aggregation bias (unfortunately aggregation bias still applies with cosmogenic tritium input). However, as pointed out in the paper, not much aggregation bias is expected in systems (such as many catchments) that have MTTs in the range of one to two decades. One cannot completely predict how things will turn out, but we can foresee that future measurements of tritium will not be redundant.

Ref #3: 5.e. There will still be value in sampling across a range of discharges in order to quantify how modeled tritium ages vary with different wetness conditions, as previous work from the New Zealand group has very nicely demonstrated.

Reply: Agreed.

Ref #3: Specific comments
1. As other reviewers have pointed out, the organization and clarity of the presentation must be improved. Many necessary details have also been left unmentioned.

Reply: We will reorganise the paper as noted in the replies to this and the other referees.

Ref #3: 2. Needless confusion is created by the alphabet soup of acronyms. Saying "dispersion model", "exponential model", "lumped parameter model", and so on is preferable to forcing your readers to learn a dozen acronyms just so they can get through your paper.

Reply: Different people have different preferences in terms of jargon, but we agree with Ref #3 and will reduce the jargon to improve the readability of the paper.

Ref #3: 3. Inconsistencies abound. The double exponential piston flow model is called both DEPM and (apparently) BMM. Using inconsistent terminology like this is bad enough, but what's even worse is that readers are never told, and are left to figure this out for themselves. Most of the text uses MTT but some of the figures and captions use MRT, and again readers are never told whether these are the same things or different things.

These are just a few examples of a general problem, and it should not be a reviewer's job to flag all these issues.

Reply: We thank the Referee for spotting these inconsistencies, which we will remove.

Ref #3: 4. The wiggles in the black curves in Figures 3b and 3c are obvious numerical artifacts, since the real theoretical curves should be smooth. It is troubling that such visually obvious numerical errors have not been noticed and corrected. One naturally wonders whether there are other technical issues that are less visually obvious, and also have not been caught.

Reply: We will correct these in the revised paper.

Ref #3: 5. at line 8 on page 3, Bethke and Johnson (2008) should be cited; otherwise it looks like the authors are taking credit for this observation.

Reply: We will cite this reference.

References

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

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, 2014. doi:10.5194/hess-18-3109-2014

Gusyev M.A., Toews M., Morgenstern U., Stewart M., White P., Daughney C. and Hadfield J. Calibration of a transient transport model to tritium data in streams and simulation of ground water ages in the western Lake Taupo catchment, New Zealand, Hydrol. Earth Syst. Sci. 17, 1217-1227, 2013, doi:10.5194/hess-17-1217-2013,

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. Hydrol. Earth Syst. Sci. 20, 3043-3058, 2016. doi:10.5194/hess-20-3043-2016

Gusyev, M.A., Toews, M., Morgenstern, U., Stewart, M., White, P., Daughney, C., 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. Hydrology and Earth System Sciences 17(3), 1217-1227, 2013.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. doi:10.5194/hess-19-803-2015

Morgenstern U, Taylor CB. Ultra low-level tritium measurement using electrolytic enrichment and LSC. Isotopes in Environmental and Health Studies 45: 96–117, 2009. DOI:10.1080/10256010902931194

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.

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

Stewart, M.K., Morgenstern, U., McDonnell, J.J., Pfister, L. The 'hidden streamflow' challenge in catchment hydrology: A call to action for streamwater transit time analysis. Hydrol. Process. 26(13), 2061-2066, 2012. doi: 10.1002/hyp.9262.

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 (18O, Cl, 3H and CFC) evidence, Hydrol. Earth Syst. Sci. 12, 1-19, 2008. http://www.hydrol-earth-syst-sci.net/12/1/2008/hess-12-1-2008.html