Combined response to reviewers

We thank the two referees for their helpful and insightful comments that will help improve this paper. We have addressed the various comments below (responses in blue) and have indicated how we will incorporate the suggestions in the revised paper. There are some issues that come down to editorial preferences or house style and we have indicated these in green. In addition, we understand from previous papers that a data availability section should be included after the Conclusions. We propose it to state: "All geochemistry data used in this study are contained in Table 1. Streamflow data are publicly available from the Victorian State Government Department of Environment, Land, Water, and Planning (http://data.water.vic.gov.au/monitoring.htm)."

Reviewer 1 (Edison Timbe)

The manuscript HESSD "Contrasting transit times of water from peatlands and eucalypt forests in the Australian Alps determined by tritium: implications for vulnerability and the source of water in upland catchments" by Cartwright and Morgenstern, 2016, assesses, in terms of mean transit times, the hydrological differences between wet-lands (peatlands) and eucalypt forest ecosystems. Although this study is mainly based on lumped parameter models for which uncertainties are commonly large, the inclusion in the analysis of diverse datasets like major chemical elements, stable and radioactive isotope data, allows the authors to crosscheck findings from different perspectives. The authors perform an appropriate analysis of the contrasting mean transit times found between both analyzed ecosystems, allowing to hypothesize about the hydrological functioning of the related aquifers and flow paths. Considering the scarce studies dealing with wetlands and more specifically, peatlands (e.g., as compared to mountain forest head water catchments), this study is very timely and therefore I recommend it for publication after some minor revisions which I detail below.

We thank the reviewer for these positive comments and while we recognise that the lumped parameter model approach is subject to uncertainty, it is probably the most practical approach to use in these situations where catchments are not intensively instrumented. In the southern hemisphere, relative mean transit times derived from ^{3}H generally hold regardless of the uncertainties in the models (i.e. low ^{3}H waters will be older than high ^{3}H waters). This is important in this study when considering the gross differences between mean transit times in the forests and peatland. We did note this in the paper (Page 5, lines 4-9) but we can emphasise the point later as well (which in part will address comments by Reviewer 2).

Specific questions/issues

Page 2, line 8. Define the acronym TU when first used, e.g., Tritium Units (TU). Besides, consider using an acronym for 3H activity/activities, this term is widely used along the manuscript (around 100 times).

We will define "TU" on first usage both in the abstract and the main text. I tend to avoid acronyms in papers as too many make the text difficult to read, but in the case of activities, we could easily use a ^{3}H without too much confusion.

Please avoid beginning a sentence or a paragraph with an acronym or an abbreviation, this basic grammar rule is circumvented throughout the manuscript. Just to mention some few paragraphs starting with acronyms: pag. 8, line 17; pag. 11, lines 16 and 25, pag. 12, lines 10 and 19.

We agree that this is poor style and will change it throughout.

Furthermore, three from four paragraphs of the Section 4.3 begin with “3H activities of…”

We will try to vary the style of sentences a little more to improve readability.

Page 4, lines 20 to 22. There are very few studies dealing with the appropriate tracer data resolutions to obtain reliable transit time estimations through lumped parameter models. Please consider mentioning the study by Timbe et al., 2015, who investigated this topic using stable isotopes of water.

We thank the reviewer for pointing out this paper. We will incorporate it into our general discussion of lumped parameter models (Page 4, lines 19-22) for which we do not otherwise have any detail on sampling frequencies or a reference (as noted by Reviewer 2).

Page 5, lines 17 to 19. Consider adding another citation, there is a more recent study for a similar ecosystem (peatlands), located in the tropics by Mosquera et al., 2016, in which mean transit times of less than one year have been also found (it uses Oxigen-18 and Deuterium as tracers).

We became aware of the Mosquera et al. (2016) paper following the submission of our paper. As well as being one of a few studies that have addressed the mean transit times in peatlands there are results in that study (specifically the correlation of mean transit times with catchment attributes) that are also worth discussing in the context of our study and we will certainly refer to it.

Page 9, lines 12 to 15. Is it really possible to outline small catchment, like 0.5 km2, from the mentioned coarse google resolution?

The Google Earth Pro images are high resolution and when used in conjunction with the topographic maps, the catchment areas are reasonably clearly defined. The exact catchment areas are not particularly important as there is no correlation of ^{3}H activities with area (this is a recurring theme in our studies of Australian catchments). We can add a sentence noting that the catchment areas are approximate and review our error bars on Fig. 7, but this does not change any of our conclusions. We also have annotated Google images of the sampling sites that we can include as supplementary material to provide context

Technical Corrections

Page 2, line 6. Please write out the full name when the acronyms are first mentioned. E.g., define acronym 3H: “This study uses Tritium (3H) to estimate…”

Agreed (as above)

Page 2, line 13. Should say: “are higher in the eucalypt forest stream than in the peatland…”

Agreed, we will rephrase this sentence

Page 6, line 2. Rephrase: “This study is based in the upland areas of the Victorian Alps, southeast Australia and was designed to...” to something like: “This study, located in the upland areas of the Victorian Alps (southeast Australia), was designed to…”

This is an awkward sentence that we will rephrase

Page 7, line 17. Please check this “the soils alci include…”?

This sentence does not convey what it is supposed to and needs rewriting (the colluvium and alluvium are nor related to the soils so need splitting out).

Page 8, line 15. Please check, sentence is currently written in present tense: “Observations indicate that this is at or close…”

We will change this sentence so that the tense is consistent.

Page 8, line 16. Correct: “At least three bore volumes of water WERE extracted prior to sampling”

Agreed that it should be “were” not “was”.

Page 8, lines 20 to 22. Rephrase the complete sentence written in these lines.

We agree that this is not clearly written and will rephrase it.

Page 9, line 3. This device is more commonly known as Thermo-Finningan Delta plus.

The device is as named on the manual etc.

Page 10, line 9. Delete “transit times” once, it is repeated.

We will correct this as suggested.

Page 11, lines 16. Please rephrase the description of the stable isotope values, it is a bit confusing in its current state. I would first describe the range for O-18 and then for Deuterium (e.g., from -8.3 to -5.0 per mil for O-18 and from -43 to -23 per mil for Deuterium).

Yes, the ranges should be for $\delta^{18}O$ and $\delta^{2}H$. We will correct this as suggested.

Page 12, line 1. Rephrase “Three 2 week to 3 months aggregated samples…”

Probably “Three aggregate samples from periods of 2 weeks to 3 months…” would be clearer and more grammatically correct.

Page 12, line 4. Correct “activities in rainfall FOR this region”.

We will correct this as suggested.

Page 12, line 14. Delete the word “and” at the end of the sentence.

We will correct this as suggested.

Page 15, line 14. Define acronym “Eqs” when first used.

In other HESS papers “Eqs” is generally used without definition (i.e., it is one of the standard acronyms), but we can correct it if it fits with the house style.

Page 15, line 19. “Given that the water WAS sampled” or “Given that waters WERE sampled”?

Should be “waters were”.

Page 16, lines 3 to 5. Rephrase the sentence contained in these lines.

“For a water with a ^{3}H activity of 3 TU, propagating the analytical uncertainty of… results in…” would be clearer.

Page 16, line 8. “yields mean transit times of up to 3.9 years” Do you mean 4.0 years (Table 2, last column)?.

Yes this should be 4 years. We will change this and ensure that the other values are the same as those in the table.

Page 17, line 1. Insert a comma after the word “increase”.

We will correct this as suggested.

Page 20, line 15. Delete the word “at” after “mean transit times”.

We will correct this as suggested.

Page 21, lines 25 – 26. Rephrase: “…comparable studies such as this will become possible there which will allow...”

This would be better as “As the remaining bomb-pulse ^{3}H declines in the northern hemisphere, the utility of ^{3}H in determining mean transit times will be improved, which will allow a fuller assessment…”

Figure 1. Units are missing in the legend for elevation.

We will add the units to the legend (they are m)

Figures 4 and 6a: Correct the axes labelling

For some reason hidden layers (from Adobe CS) became visible in these figures following uploading (they were not visible in the files that were uploaded or .pdf’s that I generated from these on my computer). We will remove the hidden layers in the final version of the figures.

Reviewer 2

The paper presents a comparison between mean transit time, MTT, between peatlands and eucalyptus forest in the Australian Alps. The authors used tritium as a tracer to model MTT using a lumped parameter approach. The authors also integrate geochemistry data and water stable isotopes in the analysis to yield important interpretations related to water storage and availability in these ecosystems. The paper is relevant to the scientific community because it provides information about underrepresented ecosystems with scare hydrologic data. The paper could eventually make an important contribution to the scientific literature. However, I believe there are issues with the paper structure and writing style. Thus I recommend major revisions before it can be considered for publication.

We also thank this reviewer for the comments that recognise the importance of this study and the dearth of information currently available from peatlands. One of the reviewer's major comments relates to the structure of the paper. HESS does not have a fixed house style and looking through recent issues, the structure that we have used is one that is common but not ubiquitous. We discuss this in more detail below, but perhaps the editor can comment on what structure they think would best fit with the journal. Another group of comments relates to the application of the lumped parameter models and there are some important differences between the use of single ^{3}H measurements in the southern hemisphere and other tracers that are applied as time series. Although we note this in section 1.2 (and it is discussed in detail elsewhere, notably Morgenstern et al., 2010), it would be worth adding more details here. While not wishing to turn this into a review, a few more details in the introduction and throughout the paper would be helpful. This would also help address the general comments of Reviewer 1.

I provide below some specific comments and suggestions:

Abstract: Lines 1-2: the first sentence sounds redundant (use of word "that"). Please rewrite.

The second "that" should be a "which"

Line 8: define the acronym before using it (TU).

Agreed we will define it on first use both in the abstract and main text (also raised by Reviewer 1)

1. Introduction:

I believe the introduction should provide more information about the use of geochemical data the context of mean transit time modelling.

While not trying to turn what is an applied study into a review article, we agree that a few more key details would be useful. We have considered the issue of aggregation and time-invariance and will add more detail on these topics in the introduction and where appropriate later in the paper (e.g., Section 5.2) (see below).

Page 3, line 21-23: This sentence should be written.

We will clarify this sentence.

Page 4, section 1.1: I think the authors should include a paragraph about the recent advances and challenges in the determination of transit times and the use of lumped parameter models. For instance, the authors should clarify that the calculated times are most likely representative of base flow conditions and spell out the underlying assumptions in the use of these models. There is a vast new literature dealing with time variant modelling of transit times.

As indicated above we are happy to do this without turning the paper into a review. The recent work by Kirchner (2016. Hydrol. Earth Syst. Sci., 20, 299–328) provides a good summary of the problem of transience for mean transit time calculations and will be used to frame this discussion. The assumption regarding time-invariance is important where mean transit times are calculated from time series measurements (which is the methodology used for stable isotope, major ion tracers, or tritium in the northern hemisphere), which was explored in depth in the Kirchner paper.

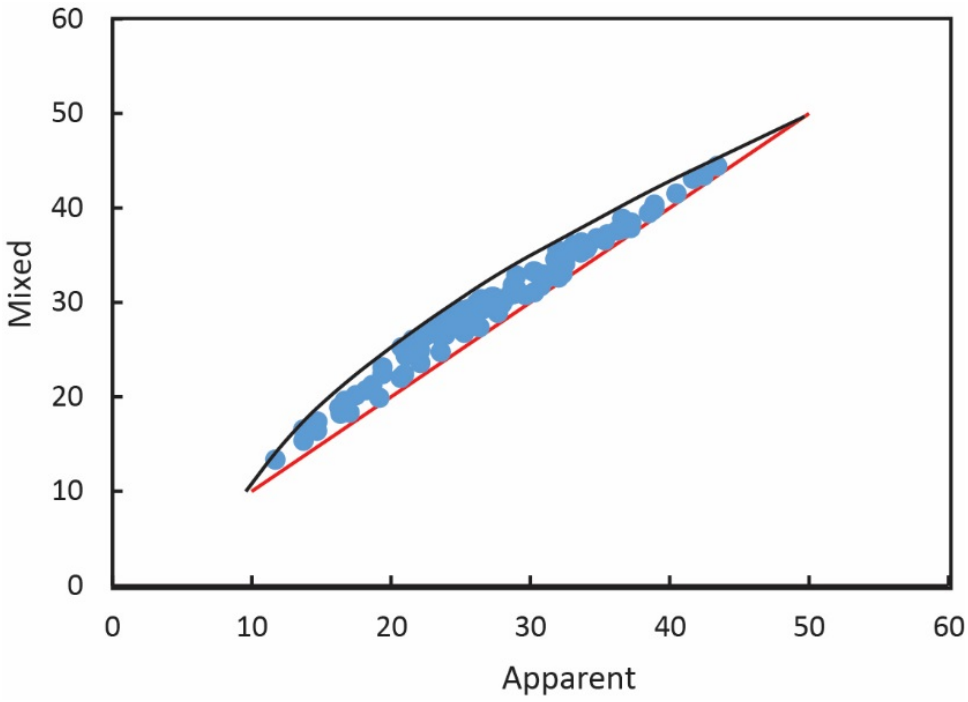
Our approach was to estimate the mean transit times from individual ^{3}H measurements rather than to use a time series approach, which due to the attenuation of the bomb-pulse tritium is not now practicable for ^{3}H in the southern hemisphere. The attenuation of the bomb-pulse tritium, however, now allows for robust interpretation of individual ^{3}H measurements. These ^{3}H activities reflect the residence time in the catchment and do not require an assumption of time invariance. What is required is that the flow path geometry remains relatively constant and able to be approximated by an LPM. The disadvantage of the individual ^{3}H measurement approach is that one must assume a LPM.

By contrast, the time series approach using ^{3}H, which is still possible in the northern hemisphere, does allow one to assess how appropriate the assumed LPM is. The disadvantages are that there is an assumption of time invariance between the various sampling times, which is unlikely in reality. The time series approach also only yields a single mean transit time estimate per site which would not apply to different times if the flow system changed, and therefore it is not suited for interrogating whether the mean transit times vary with flow conditions or in different years.

It was noted in this section that in the southern hemisphere it is possible to use ^{3}H to provide estimates of mean transit times from single measurements (Page 5, lines 6-7) rather than having to use time series data. We will explain more fully the differences in these two approaches. This issue came up again in Section 3.2 and we will reiterate the points there also. Also, again due to the attenuation of the bomb-pulse tritium, the relative ages derived from ^{3}H activities in the southern hemisphere are not model dependant (Page 5, lines 7-9) and it is worth emphasising this throughout the paper.

Aggregation (as in the mixing of waters from different flow systems with different mean transit times) is a problem for all methods that determine mean transit times (e.g., Kirchner 2016, Hydrol. Earth Syst. Sci., 20, 279–297), and we will also add a comment on that to this part of the text (and discuss the potential impacts later). A recent paper by Stewart et al. (2016, Hydrol. Earth Syst. Sci. Discussions) directly addresses aggregation for mean transit times from single ^{3}H activities. That study demonstrates that for single ^{3}H measurement, aggregation is most problematic when waters with very different mean transit times are mixed. The figure below shows the results of aggregating up to 5 samples with mean transit times of between 10 and 50 years in random amounts (150 simulations). The mixed mean transit time is calculated from the proportion of the end members (i.e., a 1:1 mix of waters with mean transit times of 10 and 50 years has a mixed MTT of 30 years) and the apparent MTT is calculated from the ^{3}H activity of the mixed sample. All the mean transit times were calculated from the exponential-piston flow model with f = 0.75. The red line is agreement between the mixed and apparent mean transit times.

Aggregation in this case results in an underestimation of the mean transit times (which is in agreement with the Kirchner and Stewart studies). The maximum error is when there are only two end-members being mixed and there are approximately equal amounts of each end-member (the black line shows the apparent ages from binary mixes of waters with mean transit times of 10 and 50 years). Repeated aggregation reduces the error. For example, aggregating two waters with mean transit times of 10 and 50 years in a 1:1 ratio results in a difference between apparent and mixed mean transit times of ~13%; by contrast, if nine waters with MTT's of 10, 15, 20, 25, 30, 40, 45 & 50 years were aggregated in equal amounts the difference is 3.6%.



The percentage differences are similar for other LPM's and for end-members with different mean transit times.

Given the recent discussion of aggregation in the literature, we propose to include a statement on it in the introduction to determining mean transit times and a paragraph in section 5.2 that addresses the impacts along the lines of the above discussion (albeit shorter). Overall, the uncertainty resulting from aggregation is similar to that of the other uncertainties and does not change the overall conclusions.

The reviewer is correct that we are characterising baseflow conditions and we will note this where we outline the aims of the study (Section 1.3). This is also important for the assumption of a single water store in our calculations (Section 5.3) and again we will reiterate it there.

Page 4, line 19-22: there is a missing "the" before "use". In addition are there any relevant references to this statement?

Reviewer 1 noted that the Mosquera et al. (2016) reference has information on sampling frequency, and we can incorporate the details from that paper here.

Page 4 line 19: consider using "to determine" instead of "to determining"

Will change as suggested.

Page 6: I suggest you eliminate section 1.3 and have the objectives of the study be the last paragraph of the introduction.

We are happy to do this as it makes little difference to the paper's structure. However, previously we have been requested to provide a specific "Objectives" section in papers. Perhaps the Editor could comment on what fits better with HESS's house style.

Page 6 line 11-17: this information should appear before the paragraph with the objectives. Could be part of the last paragraph in page 5.

We agree that this doesn't fit with the objectives (whether they are in a separate section or not) and we will move it as suggested.

I suggest this section be part of Methods.

We are happy to do this although it makes little difference to the paper's overall structure. Again HESS doesn't appear to have a consistent house style. Perhaps the Editor could comment

Page 6 line 19-22. I suggest you rewrite this sentence.

Would probably be better as "Water from peatlands and eucalypt forest streams…"

Page 7 line 1-2, line 13-14: Please use the correct notation for the scientific name of species.

We will italicise the species names

3. Methods

Line 9: Explain "aggregated" over what time frame? In addition, how many, when, how frequent were the samples collected?

"Aggregated" refers to the total rainfall collected over the time period. The duration and the frequency of rainfall samples is in Table 1 but that information can be added here so that it is clear what timespans and when the rainfall is collected.

Line 9-10: That just means grab samples, right?

We can use the term "Grab samples" if it is clearer.

Page 9 section 3.2. More information is required about the modeling procedure, how were the best parameters identified, what objective function was used, how many possible parameter combinations were implemented, can you include dotty plots? How did you chose among the 3 different transit time functions (exponential, piston flow, and dispersion). Why did you chose these 3 and not others?

As discussed above, to calculate mean transit times from single ^{3}H measurements one has to assume a LPM model. Our choice of models spans those commonly used in the literature. For example of the ~120 studies listed by McGuire & McDonnell (Journal of Hydrology, 330, 543– 563) some 100 utilise the three LPMs that we discuss. The exponential-piston flow model corresponds to the most likely form of the flow system (near vertical recharge through the unsaturated zone and exponential flow in the unconfined shallow aquifers and soils). Where ^{3}H time series data are available, this model with f values between 1.0 and 0.4 has reproduced the ^{3}H activities of river water and groundwater (e.g. Maloszewski et al. 1983, J. Hydrol., 66, 319–330; Stewart et al. 2007, Hydrol. Process., 21, 3340–3356).

There are differences between the mean transit times from the different models and we have used this (recognising that we do not know which is the most appropriate LPM) to place some uncertainties on the mean transit times. Given that the mean transit times are so different between

the peatland and eucalyptus forest waters, the main conclusion of the paper (that water retention times in the peat are much shorter than those in the forests) remains unaffected by the choice of model (we note this point on Page 20, lines 12-14).

We will add more details here as to the rationale behind the modelling and our choice of models, which will echo the additional material that will appear in the introduction (discussed above). Additionally, we will re-emphasise these points in Section 5.2 where we discuss the results of the modelling.

4. Results

Here is where the major structural issues arise: The results section does not present any of the MTT related findings. This is odd considering that this is precisely the main topic of the paper. Residence time results are mentioned in the abstract, discussion, and conclusions. The manuscript must include a residence time results section in which the modelling findings are presented.

The paper is structured such that it presents the data first (Section 4) and then discusses it (Section 5). While this is not a strict requirement in HESS, it is a common structure used by many journals and one that is followed by numerous papers in HESS and elsewhere (and from experience it is a structure that some reviewers are insistent on). While it is possible to write papers that mix results and interpretation into a single section, the drawback of doing this is that the reader may be unclear as to what is data and what is interpretation or that the paper begins to present interpretations ahead of describing the data that was used to make those interpretations. Much of what precedes section 5.2 is required to produce the interpretations in that section.

We take the point that section 5.2 is a fair way into the paper, but we would prefer to keep the structure as is. What we can do is to provide a brief explanation as to what is in each section at the start of Sections 4 and 5 (i.e. that this is the section that presents the data that is interpreted in Section 5) which will make it clearer what the reader can expect.

Page 11 lines 2-5: This sentence is awkward. Please rewrite.

We can rephrase this sentence to ensure that it is clear.

Page 11, line 8: Avoid starting the sentence with a delta symbol, instead say" Water stable isotopes (18O and 2H).

A similar point was raised by Reviewer 1. We will ensure that sentences do not start with symbols or acronyms, which we agree is poor grammar.

Page 11: section 4.2. Figure 3 should be cited sooner.

We will cite this on first usage (top of Section 4.2).

In addition it is not clear when or how many samples were collected.

The samples are listed in Table 1 together with their sampling dates. We will ensure that that is clear in the figure caption / legend and in the methods.

Page 11 line 11-12: is this slope different from the GMWL. The MMWL in Fig. 3 is very step is that correct. In addition, the text says that the samples line to the left of this line which is not true.

The MMWL has a slope of ~7.5 (vs. 8 for the GMWL), which we will note in the text. The line through the samples is a best fit to the data with a slope of 5 (it is not the MMWL), we will ensure that that is clear in the figure caption and the text.

Page 11 line 13: please provide the range if deuterium execs.

We will add these to the text.

Page 11 line 22-25: This sentence is too long.

Agreed that it would be clearer if we split it.

Page 12 line 5: the word "and" makes no sense.

This is also a long sentence that we could make clearer by splitting.

Page 12 Lines 3-10: Did you test the data for normality to make sure a parametric test was appropriate?

Although there is no indication that the data are not normally distributed, the datasets are probably too small for the statistics to be rigorous. Rather than presenting the p values, we can just point out that the ^{3}H activities between the sites overlap (the p values were not used for any other purpose).

Page 12 line 11-14: this sentence is too long.

We can split the sentence to make it clearer.

Discussion:

Page 13 line 19-21: Please avoid single sentence paragraphs.

The lines listed are not a single sentence paragraph. The first paragraph to section 5 is a single sentence, but it introduces the section. Given that we indicated that we would provide a bit more detail of what Sections 4 and 5 contain, that section will be a little longer.

Page 14 line 22: reword sentence.

We will clarify this sentence

Page 15: The MTT results should be move to the result section. Also the selection of the best model should be justified both physically and statistically. What objective function was use to qualify the goodness of fit?

We discussed the location of this material and the way we approached the modelling above. Our preference is to retain this material here as it constitutes interpretation (discussion) rather than data presentation (results). As noted above we will provide more detail as to the choice of models and the assumptions that we have made, and some of those details will go into the discussion of the results in this section. We start off discussing the exponential-piston flow model as this accounts for

vertical flow in the unsaturated zone and exponential flow in the saturated zone (which is how the flow system is most likely to operate). However, we do recognise that this is only one possibility, which is why we utilised the other LPM's for comparison. However, to reiterate that adopting different models does not alter the primary conclusion that the mean transit times in the peatlands must be much shorter than those in the forests.

Tables: in general the captions need to be more comprehensive. For instance, the caption for Table 1 should indicate if the isotopic values averages? If so what is the time period over which they were calculated, how many samples are included, and are there metrics of uncertainty.

These data are from single samples. The error bars on the figures reflect analytical errors not sample variability (as is stated in the figure captions). We can specify this in the Table caption (it would make it clearer for the rainfall samples that were collected over a period of time).

Figures

Figure 1: Please use different markets to indicate the location of peatlands and eucalyptus forest.

It would be difficult to show the outcrops of peat at this scale. As noted above, we have annotated high-resolution Google Earth images that show the distribution of peatlands and eucalypt forest more clearly that we could include as a supplement.

Figure 2b. Please use probability scale in the x-axis.

Although %time exceeded is commonly used, probability is more accurate (the graph remains the same).

Fig. 3 not clear what the dash (–) line is.

The dashed line is a linear best fit to the rainfall samples. We will add that to the figure caption.

The error bars for the grab samples should be representative of the accuracy (from the analysis of duplicate samples).

This is the case (as in the figure caption).

If the precipitation signature corresponds to weighted means, then the error should be weighed errors.

They are individual samples.

Please add the number of samples (n=xx) associated to each (peat, Eucalypts, rain)

We agree that this would be useful and will add it to the figure.

Figure 4: Not sure what the equation and R2= 0.89 mean versus the R2=0.69. The legend has a "series 3" and "linear (series 3)", automatically generated from excel, that are not identified.

As explained above, some hidden layers appeared in the final versions of these figures when they were uploaded. We will remove them.