**In the following text please find the corrections and comments to the referee's response (for better understanding, comments from the referees were copied are black and our comments in blue).**

**Replies to Referee #2 (Anonymous):**

**General comments:** Timbe et al. show, by using a set of 7 lumped parameter models to determine water mean transit times in different compartments of a tropical catchment, that the choice of an appropriate transit time distribution function (TTD) is crucial and may be ambiguous. Besides the goodness of fit also the uncertainties of the model results are analyzed and compared. One of the most valuable messages is that models that yield the best fit may provide highly uncertain results yet. The concept of the study is coherent and the results are relevant for scientists working with lumped parameter models. The paper focusses predominantly on technical aspects which is fine since an artificially extended input data series was used, making it critical to interpret long transit times regarding processes. In this respect, the title should not imply that process analysis is a main goal of the study. It should therefore be modified to more emphasize the technical component of model testing and comparison (specific comment below). The language is clear in the first half of the manuscript. Particularly in the discussion section sentences are often long and difficult to follow. The authors may try to shorten and clarify the longest sentences (see also comments below).

General reply: We really appreciate and thank for the comments from Referee #2. Specific suggestion referred to change the title of the paper has been implemented and the title now reads as follows:

*"Understanding uncertainties when inferring mean transit times trough tracer based lumped parameter models in Andean tropical montane cloud forest catchments"*

Regarding to the changes suggested for the 'Discussion' section, it has been re-written, and also analyses and discussions related to TTDs have been added (according suggestion from Referee#1). Figures related to the analysis of TTD for the retained models are now shown in Figs. 8 and 9 for soil waters and Figs. 14 and 15 for stream, creek, and spring waters (see figures below). Besides, 'Results' and 'Discussion' sections have been condensed (regarding to the selection of the best performing model). As the changes of these sections are extensive, please check the complete modified sections below.

Please notice that the inclusion of new figures (8, 9, 14 and 15) changes the numbering of figures of the previous version of the manuscript (Fig. 7 is now 10; Fig. 8 is now 7; Figs. 9, 10 and 11 are now 11, 12 and 13 respectively).

In the new version of Section 4, the following references have been added:

*"Botter, G., Bertuzzo, E. and Rinaldo, A.: Catchment residence and travel time distributions: The master equation, Geophys. Res. Lett., 38, L11403, doi:10.1029/2011GL047666, 2011."*

*"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."*

*“Roa-Garcia, M. C. and Weiler, M.: Integrated response and transit time distributions of watersheds by combining hydrograph separation and long-term transit time modeling, Hydrol. Earth Syst. Sci., 14, 1537–1549, doi:10.5194/hess-14-1537-2010, 2010.”*

*“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, doi:10.1002/hyp.7576, 2010.”*

Modified sections:

# 3 Results

## 3.1 Soil water

*“Of all predictions the best matches of the models, with respect to the NSE objective function, ranged between 0.64 and 0.91 (Fig. 5a). When only the best goodness of fit was considered, the GM and the EPM models performed best in most of the sampled sites (13 from 18), followed by the DM, LM and LPM models (Fig. 5b). Only these models were considered for further mutual comparison. Even when the derived MTT values were similar among the models that best fitted the objective function (Fig. 6a, Table 4 and Annex 1), the LPM model performed best taking into consideration additional selection criteria, as shown in Figs. 6b and 6c. Fig. 7 depicts, for the LPM model applied to site C2, the uncertainty and the range of behavioral solutions for the two model parameters.”*

*“Considering results from the LPM model (Table 4), differences between observed and predicted values described by the RMSE are up to 1.72‰ and the larger absolute bias accounts for 0.181‰ (Table 4). Bearing in mind the ranges of behavioral solution, MTT results were between 2.3 to 6.3 weeks for pastures soils and between 3.7 to 9.2 weeks for forested soils, while parameterizations for $\eta$ (ratio of the total volume to the volume in which linear flow applies) ranged from 0.84 to 2.23 and from 0.76 to 1.61 respectively.”*

*“Regarding to the shapes of the distribution functions, Fig. 8 shows the best matching results for two representative and comparable sampling sites (C2 for pastures and E2 for forest) for each lumped model (results for LM model are not included since best matching results for LPM were achieved with $\eta \approx 1$, see Table 4). These probability (PDF) and cumulative density functions (CDF) depict how water is routed through the system. In this sense, pasture sites generally show a faster and higher response of the tracer peak when compared to forest sites. The CDF (Figs. 8b and d) of all models are quite similar for the major part of the flows, even including the linear function LPM that averages the shape of the peaks described by the other models. Models based on exponential functions (EPM, DM, or GM in Figs. 8b and d) predict a small portion of the flow with an exponentially delayed tail, which is larger for forested sites than for pastures. Best distribution function results (based on highest NSEs) for all sampled sites, according to the type of land cover, are shown in Figs. 9a and b for the LPM and GM models applied to pasture sites, and in Figs. 9c and d for forest sites. Considering the range of possible or behavioral solutions (e.g., shaded area represents range of solutions for C2 site in Figs. 9a and b, and for E2 in Figs. 9c and d), distributions functions for each type of model and land cover are very similar between each sampled site.”*

### 3.2 River and tributaries

*"Considering all sites and models the criteria NSE > 0.45 was exceeded in 41 of the 63 predictions (9 sites per 7 models, Fig. 5a). Among the analyzed sites the TPLR model yielded the best matches for PL, SF, FH, QZ, QN, QM and QC, while the EPM model for the QR and QP sites (Fig. 5b). The GM model reached closest efficiencies when compared to the best match for every site. Consequently only the TPLR, EPM and GM models were further considered. Differences between MTT predictions for all sites are depicted in Fig. 10a and results from retained models in Table 5 and Annex 2. Although MTT results according to the best NSEs were reached using the TPLR model, compared to the GM or the EPM, these predictions also showed the largest uncertainties (Fig. 10b) and at the same time depicted the lowest number of observations inside the predicted range of behavioral solutions (Fig. 10c). Considering these additional selection criteria, EPM performed better. For stream water at the main outlet, Figs. 11-13 show the parameter uncertainties and behavioral solutions for the TPLR, GM and EPM models, respectively."*

*"Considering results from the EPM model (Table 5, Fig. 10a), the fitting efficiencies reached a maximum NSE of 0.56 for the main stream, and NSEs between 0.48 and 0.58 for the main tributaries (Fig. 5a). The predicted MTT at catchment outlet was 2.0 yr with a $\eta$ parameter of 1.84 (a similar value was estimated for the main river at the SF sampling site, MTT = 2.0 yr and $\eta$ = 1.85) and varied from 2.0 (QM, $\eta$ = 1.85) to 3.9 yr (QC, $\eta$ = 1.97) for the main tributaries. Uncertainties of MTT predictions between sites were similar with a maximum range between 14.1% and 20.4% of the predicted MTT, as derived for the FH and QM sites (Table 5). Similarly, $\eta$ ranged from 1.61 (QZ) to 2.21 (QP), the average value of $\eta$ = 1.85 implies a 54% of volume portion of exponential flow and a 46% volume of piston flow; the uncertainty for the $\eta$ parameter was 25% on average."*

*"Figures 14a and 14b show the shape of the TTD for the main river outlet (PL), corresponding to the highest NSEs for EPM, GM and TPLR models. The curve for EPM shows a delayed peak that is not accounted in the GM or TPLR models (Fig. 14a), which in turn are very similar between them (at least after a short initial time since GM tends to infinity for times closes to cero). Besides, the latter models show a more delayed flow tail when compared to EPM, which show in general a faster transit time (Fig. 14b). Differences between stream water TTDs from the main sub-catchments considering EPM and GM models are shown in Figs. 15a and b. For comparison of the degree of similarities between sites, these plots include the range of behavioral solutions for the main outlet (PL), thereby being clear that apart from QC or QP, the remaining sites have similar (EPM or GM) transit time distribution functions."*

### 3.3 Springs and creeks

*"Of 35 predictions (7 models for 5 sites) the criterion NSE > 0.45 was fulfilled in 20 cases. Sites with reduced isotope signal (small $\sigma$) yielded lower efficiencies (Fig. 5a, Table 5 and Annex 2). Apart from TP and QRS, in the remaining sites the criterion NSE > 0.45 was reached at least by 5 models. TP, PLS and SFS sites were best described by using a TPLR model (Fig. 5b). In this regard, GM and EPM were the second and third best models. Figure 10a shows the MTT results predicted by the three models, while detailed information is given in Table 5 and Annex 2. As for stream waters, the EPM model performed best when looking at the uncertainties and the number of observed data inside the range of behavioral solutions (Figs. 10b and c)."*

*"Considering EPM, MTTs of 4.5 yr (NSE = 0.49, η = 1.74) for TP and 2.1 yr (NSE = 0.65, η = 1.84) for Q3 were estimated; while for springs, 2.0 yr (NSE = 0.69, η = 1.85) for PLS and 3.3 yr (NSE = 0.47, η = 1.42) for SFS. Results for the QRS site showed poor reliability due to the reduced amplitude of $\delta^{18}O$ in the observed data (Table 5), the lowest among the observed sites (σ = 0.17). Estimations of MTTs for this site was larger than 5 yr, and therefore beyond the level of applicability of the method for natural isotopic tracers."*

*"Figures 14c and d show the TTD results of EPM, GM and TPLR models, for a representative site with long MTT (creek TP). This site show a distinctive more delayed time to the peak (for EPM model) and longer duration of flow tails compared to stream water (Figs. 14a and b). In Figs. 15c and d, the TTDs for all spring and creek sampled sites are shown for the EPM and GM models. In these figures, it is noticeable that the sites Q3 and PLS show the same patterns described previously for most of the stream waters (Figs. 14a and b), while some differences related to more delayed flow responses can be accounted for SFS, TP or QRS sites (Figs. 15c and d), which are more similar to QP and QC stream waters."*

## 4 Discussion

*"For each soil water site, similar MTT results of a few weeks to months were obtained regardless of the lumped parameter model used (Fig. 6a, Table 4 and Annex 1). Although the LPM model did not yield predictions with the highest efficiencies (Fig. 5a), provided smaller ranges of uncertainty (Fig. 6b) and a larger number of observations inside them (Fig. 6c), advantages that could not be inferred by using only the best matches to NSE, for which GM and EPM models performed better than others (Fig. 5b). Using a LPM model, suitable to describe a partially confined aquifer with increasing thickness (Maloszewski and Zuber, 1982), we found MTTs varying from 2.3 to 6.3 weeks for pastures sites and from 3.7 to 9.2 weeks for forested soils. If we consider that only the top soil horizon was sampled (maximum sampled depth was 0.4 meters), these results are comparable to values between 7.5 and 31 weeks found in 2.0 meter soil columns of typical Bavarian soil using the DM model (Maloszewski et al., 2006). When analyzing the distribution function for soil waters, similarities between model results are evident (Figs. 8 and 9). Considering the range of possible solutions of each site (shaded areas in Figs. 9a-d), it is noticeable that the major part of the flow's transit can be described similarly by all models, even using the simpler function (LPM). For these sites, when considering exponential models (EPM, GM or DP), a small portion of the flow is depicted as having a delayed tail; however, compared to the magnitude of the total volume, an LPM distribution could still be considered as a reliable method to estimate MTTs."*

*"Considering the LPM results for MTTs of soil water from pastures (4.3 weeks on average) and forest sites (5.9 weeks on average) as independent data sets, a two tailed p-value of 0.0075 for a Student's t-test was calculated, meaning that the difference between the two groups was statistically significant, although physical characteristics, like length, slope and altitude and meteorological conditions of the respective hill slopes were more or less similar. Land use effects, affecting soil hydraulic properties controlling the infiltration and flow of water, were detected in previous studies within the research area (Huwe et al., 2008). Confirming findings in other tropical catchments were published by Zimmermann et al. (2006) and by Roa-Garcia and Weiler (2010), who stated that under grazing the hydraulic conductivity decreased, overland and near surface flows increased, the storage capacity of the soil matrix declined, with feedbacks on the MTT of soil water. Similar insights were found*

*by Tetzlaff et al. (2007) comparing two small catchments in Central Scotland Highlands of different land use."*

*"For larger MTTs (> 1 yr), as derived for sampled surface waters and shallow springs, there were differences when predicted results among models were compared (Fig. 10a, Table 5 and Annex 2), especially for sites with strong damped signals of measured $\delta^{18}O$ (e.g. QRS and TP sites). When considering uncertainties, the EPM model performed significantly better when compared to the TPLR or GM models (Figs. 10b and c), although the latter two performed best for most of the sampled surface waters according to the NSE objective function (Figs. 5a and b)."*

*"When analyzing results from different models, dotty plots of model parameter uncertainty are very useful to display not only the magnitude of uncertainty but also its tendency. Similarly, the uncertainty bands of behavioral solutions can help to account for the sensitivity of the parameter uncertainty on $\delta^{18}O$ modeled results. For example, when predicted results for the PL site are compared, larger parameter uncertainty and skewness are notorious for TPLR than for EPM or GM models (Figs. 11a-c for TPLR; 12a-c for GM; 13a and b for EPM). At the same time EPM shows the highest sensitivity in modeled results (Figs. 11d, 12d, 13c). In order to contrast the signature of the effluent with younger waters such as rainfall, Figs. 11e, 12e, or 13d show the damped observed (and predicted) $\delta^{18}O$ signatures at the main outlet; a characteristic present in all analyzed surface waters. Considering the efficiencies reached by the predictions, we should keep in mind that ranges of behavioral solutions derived from a fixed 5% of the top NSE are generally smaller than a predefined lower limit for all waters, e.g., a predefined lower efficiency limit of 0.30 and 0.45 were used by Speed et al. (2010) and Capell et al., (2012), respectively."*

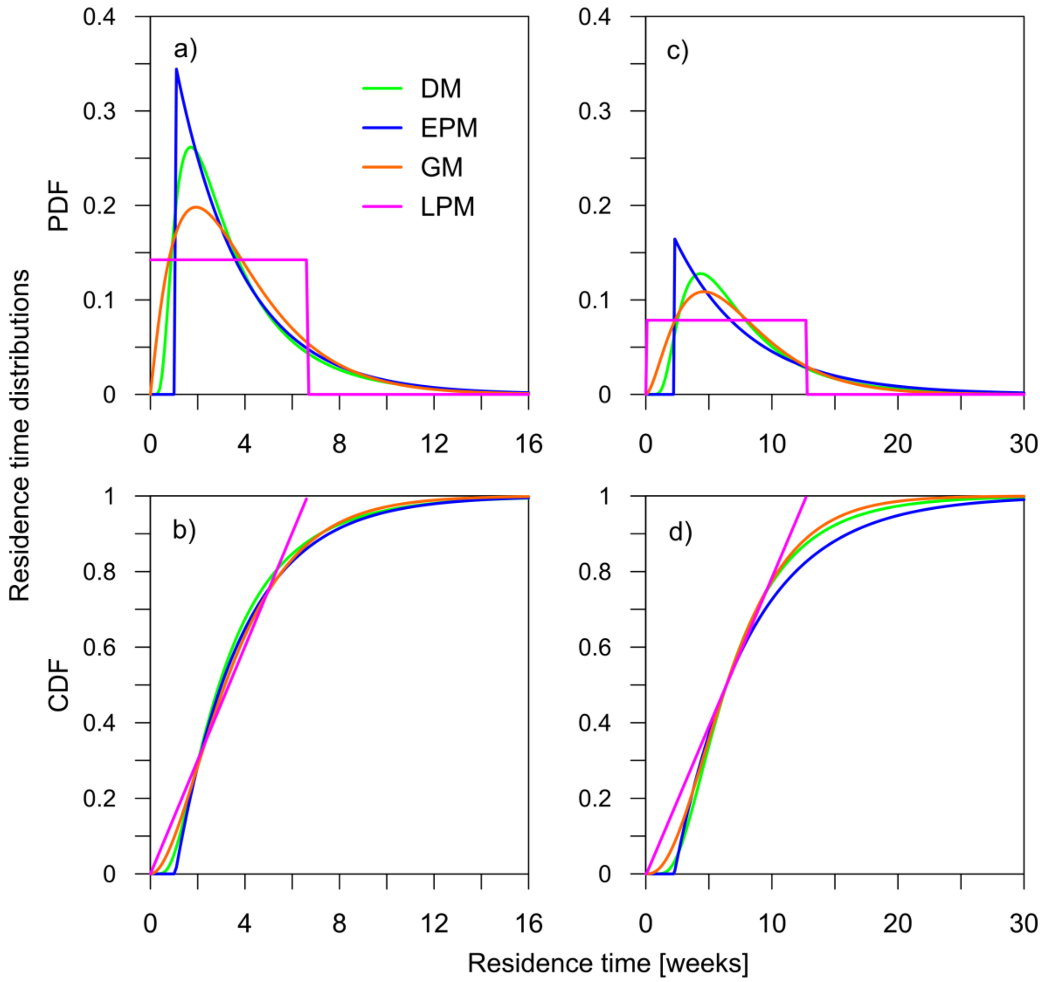
*"For stream waters, as for springs and creeks, the main differences between EPM and GM (or TPLR) results consisted first in a delayed response of the tracer signal in the outlet, modeled by a parameter $\eta > 1$ (Table 5), while for GM or TPLR the response of the flow occurred instantaneously after the spread of the tracer along the catchment (Figs. 14 and 15, Annex 2); and secondly by a comparatively smaller exponential flow tails, which also means that in general the flow transport is faster considering EPM than GM or TPLR models. For these cases, regardless of the degree of efficiencies or uncertainties, the decision on which TTD is more reliable would depend on the conceptual knowledge of the functioning of the catchment. For the San Francisco catchment this can be gained through additional field experiments in selected sites or sub-catchment using either higher resolution samples from the effluents in order to analyze non steady conditions (Botter et al., 2011) or considering different mixing assumptions (Hrachowitz et al., 2013). Another approach could be to analyze longer time series of stable isotopes, or even to include radioactive isotopes as tritium, which would help to crosscheck results, as it has been claimed that, in some cases, the inferences of the processes using solely stables isotopes, underestimate the delayed part of the flow (Stewart et al., 2010)."*

*"Regardless of the used model, efficiencies of MTT for stream waters were lower than for soil waters. This was somehow expected, since the dampening effect on a catchment to sub-catchment scale generates a smoother signal filtering/averaging the heterogeneity observed at a single point along a precise transect. Since for most of the cases MTTs for soil waters showed an increasing trend according to increasing soil depth, longer MTTs corresponding to deeper soil layers are to be expected. Soil water below 0.4 m was not monitored within this study, given the shallow soil depth and the increasing fraction of rock material with depth, preventing the use of wick samplers."*

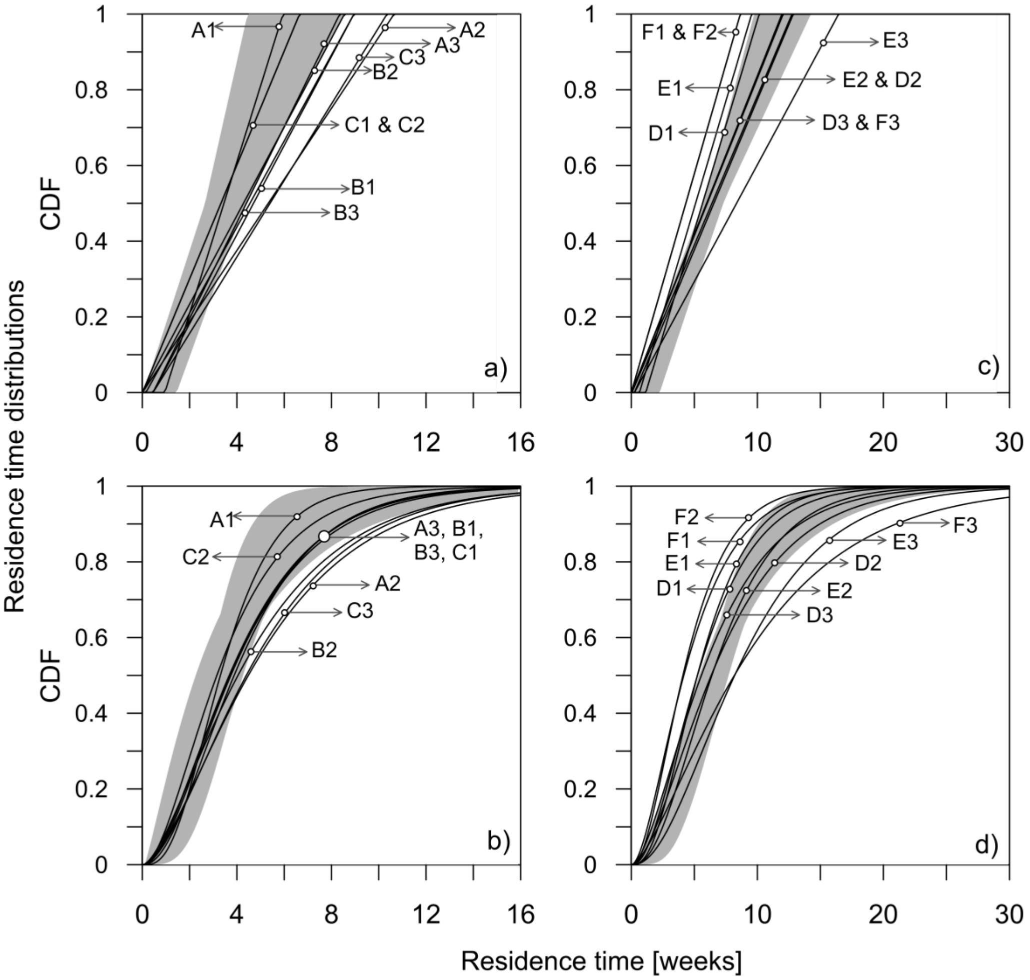
*"The similarities and differences between models for sites with MTTs > 1 yr, as for stream and spring waters, gave insights about the importance to account for a proper TTD, defined according to the conceptual knowledge of the catchment's functioning, before calculating MTT. In this regard, the use of a multi-model approach and uncertainty analysis is believed essential as to be able of defining which functions describes in a better way the parameter identifiability and bounds of behavioral solutions. By considering best matches to NSE for stream waters, best predictions were obtained with the TPLR, EPM and GM models; being more flexible versions of a pure exponential distribution function (i.e. EM model), which help to account for non-linearities of the system. The same distribution functions were identified as good predictors of observed data in a related study by Weiler et al. (2003). When comparing the TPLR to EPM or GM models, the latter two take the non-linearity of the flow without splitting it in two reservoirs with different exponential behaviors into account, therefore yielding more identifiable results. However, findings by Weiler et al. (2003) suggest that the TPLR distribution function could achieve better predictions for runoff events generated by mixed fast and slow flows. In related studies using multiple models, the EPM model yielded the best predictions for surface and spring waters (Viville et al., 2006). Considering this model, in the San Francisco catchment, the average $\eta$ = 1.85 value for surface waters (similar values were found for creeks: $\eta$ = 1.79 and springs: $\eta$ = 1.64) implies that a significant portion of old water (46%) is released previous to the new one (54%). The $\eta$ value in this study is larger than the $\eta$ value found in studies for stream water in temperate small headwaters catchments ($\eta$ = 1.09, Kabeya et al., 2006; $\eta$ = 1.28, McGuire et al., 2002; $\eta$ = 1.37, Asano et al., 2002), and close to results published by Katsuyama et al. (2009) for two riparian groundwater systems ($\eta$ = 1.6 and 1.7)."*

*"Regarding to the Gamma model, it was also identified as an applicable distribution function in headwater montane catchments with dominant baseflow in temperate climate (Hrachowitz et al., 2009a, 2010; Dunn et al., 2010). For our study area, a characteristic shape parameter $\alpha$ < 1 (e.g. Fig. 12b and Annex 2) was found in all stream and spring sites meaning that an initial peak or a significant part of the flow was quickly transported to the river. Similar results were found recently for mountain catchments of comparable size in Scotland by Kirchner et al. (2010), who also stated the importance for accounting the best distribution shape, which is usually assumed as purely exponential ($\alpha$ = 1). MTTs derived without the use of observed data, using a purely exponential model, frequently led to an overestimation of $\alpha$ and consequently an underestimation of MTTs. The higher flexibility of the GM model permits to account for the non-linearity in the behavior of a catchment system (Hrachowitz et al., 2010)."*

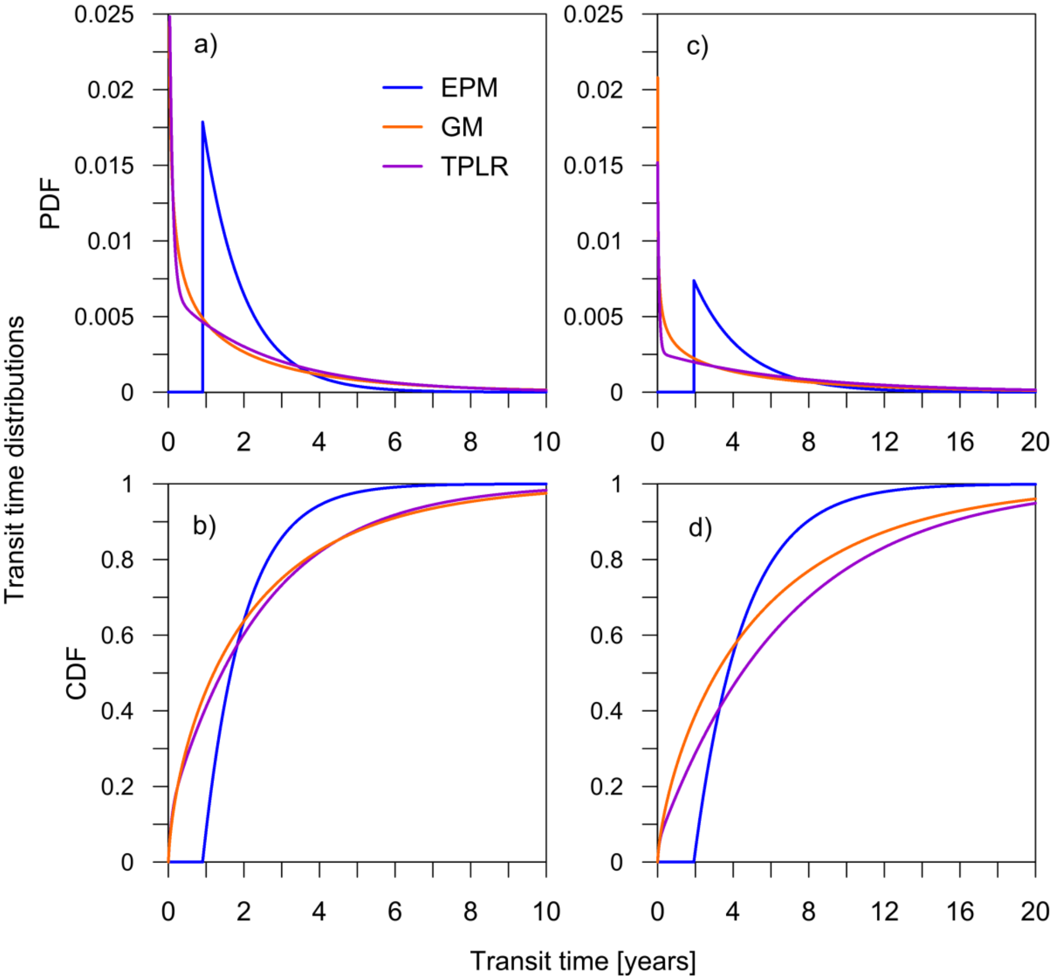
New figures:



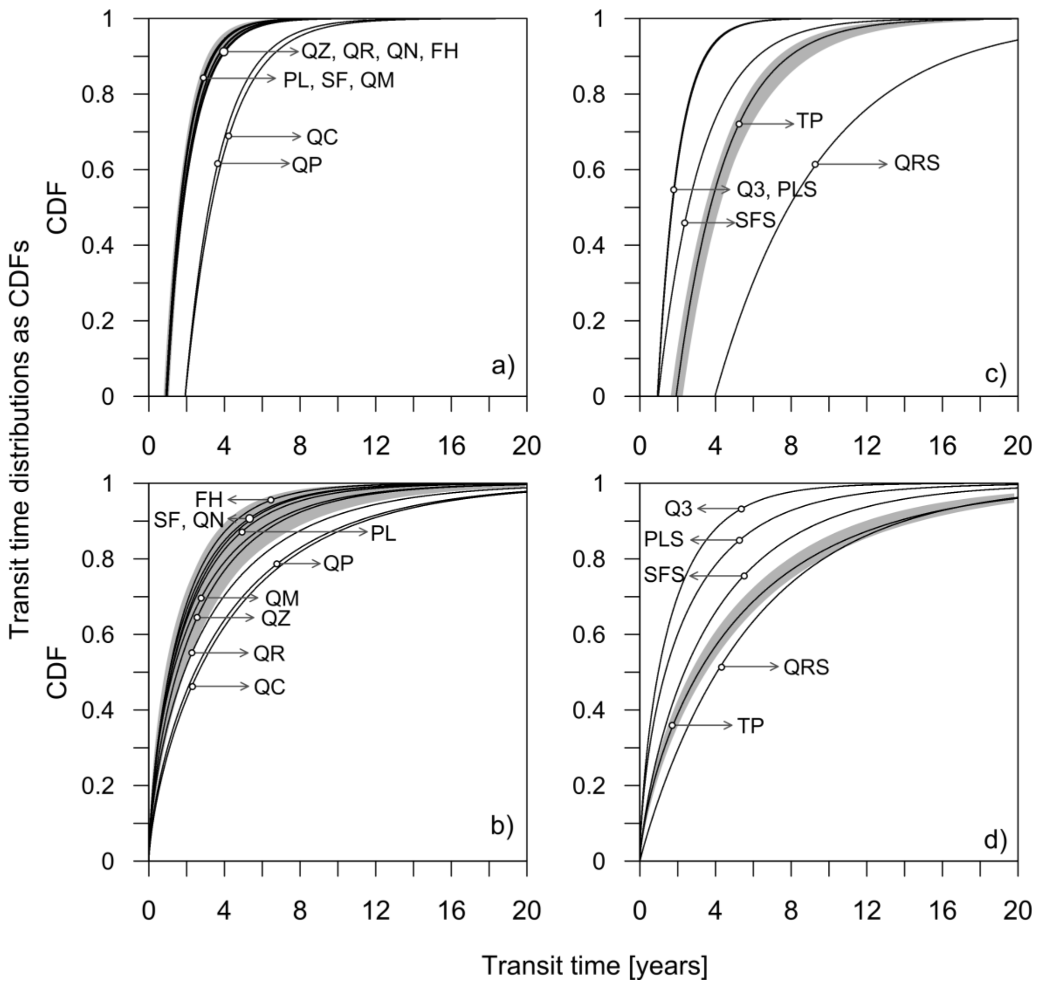

*“**Fig. 8.** Comparative characteristic shapes of residence time distribution functions corresponding to the best NSE using four lumped parameter models (DM, EPM, GM and LPM): **(a)** and **(b)** for the soil site C2 located in a pastures land cover; **(c)** and **(d)** for the soil site E2 located in a forest land cover.”*




*"**Fig. 9.** Comparative results between LPM and GM models of soil water residence time distributions functions corresponding to the best NSE for every sampling site: **(a)** pastures sites using LPM; **(b)** pastures sites using GM; **(c)** forest sites using LPM; **(d)** forest sites using GM. Gray shaded area in each plot corresponds to the range of possible shapes of the distribution function for one of the sampling sites: C2 in sub-plots **(a)** and **(b)**, and E2 in sub-plots **(c)** and **(d)**."*




*"Fig. 14. Comparative characteristic shapes of the transit time distribution functions corresponding to the best NSE using three lumped parameter models (EPM, GM and TPLR): (a) and (b) for the stream water sampled at the main outlet PL; (c) and (d) for the small creek TP."*

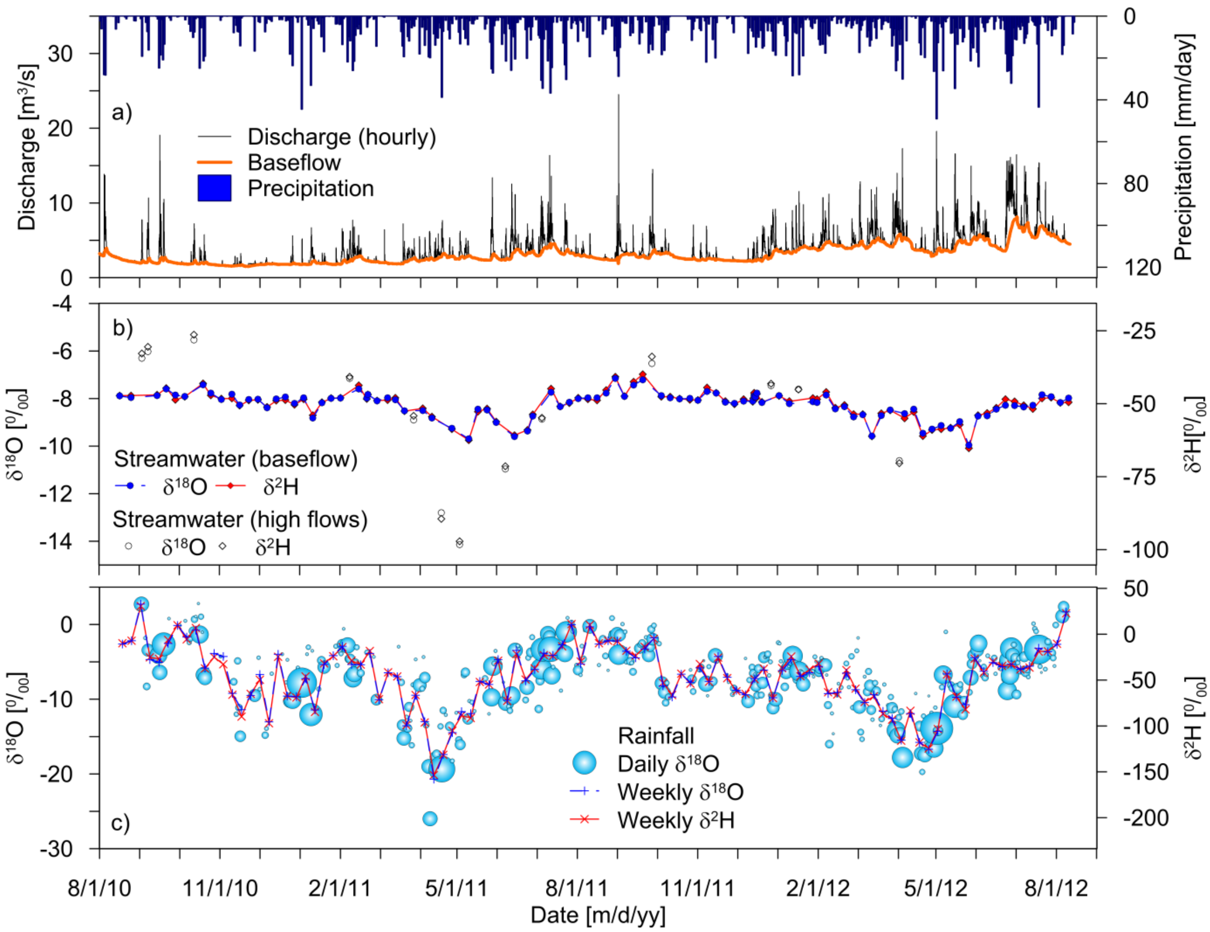



*“**Fig. 15.** Comparative results between EPM and GM models of soil water transit time distributions functions corresponding to the best NSE for every sampling site: **(a)** stream water of main outlet and sub-catchments using EPM, and **(b)** using GM; **(c)** spring waters and creeks using LPM, and **(d)** using GM. Gray shaded area in each plot corresponds to the range of possible shapes of the distribution function for one of the sampling sites: the main outlet (PL) in sub-plots **(a)** and **(b)** and TP creek in sub-plots **(c)** and **(d)**.”*

**Specific comments:** Reorder Tables and Figures: They should appear in the same order as they are referred to in the text. (e.g. Table 2 is referred to on p. 15876, table 1 not before p. 15878)

Numbering and sequence of Figures and tables have been checked for their correct sequential order (as referred in the text). In this sense, figures 2a and 2c are now 2c and 2a respectively, while Table 1 is now 2 and former Table 2 is now 1. Additional changes have been also performed due to the inclusion of new figures in the Results and Discussion sections (please check reply to "General Comment").

New version of figure 2:




*"**Fig. 2.** **(a)** Time series of rainfall for ECSF meteorological station, hourly discharge and baseflows at the catchment outlet (PL); **(b)** weekly $\delta^{18}O$ and $\delta^{2}H$ of streamwater at PL for baseflow and high flow conditions; and **(c)** weekly $\delta^{18}O$ and $\delta^{2}H$ at the ECSF rainfall sampling collector; light blue bubbles indicate daily $\delta^{18}O$ and relative volume of daily rainfall."*

**p. 15877, l. 26:** How is surface water velocity transferred into the mean velocity?

The referred sentence has been deleted, accordingly to suggestion of Referee #1 (in order to condense this specific paragraph). For your interest: The records from the radar instrument (RQ24) were used (for our case) only to crosscheck water level records (from pressure transducers) at the main outlet. The radar uses an empirical factor that relates the measured velocity in the surface and the mean velocity through the cross section. This factor is calibrated to discharges measured manually.

**p. 15878, ll. 1-2:** The Manning equation is based on the wetted perimeter and the cross-sectional area and the result is the stream velocity.

We thank for the technical correction. However this part was deleted in the revised version of the paper since we also considered the suggestion#11 from Referee #1 in terms of shortening the paragraph.

**Eq. (2) on p. 15880:** Now there are three time variables. If you substitute g(t-t') by g(τ) then you have also to change C(t') by C(t-τ) and the integration variable is τ, integrated from 0 to ∞.

We thank for the technical correction and changed Equation 2 accordingly. It now reads:

$$C_{out}(t) = \int_0^{\infty} C_{in}(t-\tau)g(\tau)d\tau$$

**Chapter 2.6:** I understand that the input time series was too short and, therefore, had to be repeated. This proceeding is acceptable as long as the focus of the study is the comparison of different models regarding their uncertainties and applicability for different compartments. But interpreting these results and the absolute mean transit times in terms of site characterization might be risky, in particular if the MTTs are >2 yr. If this is done, the additional uncertainty arising should be taken into account. This is not an obstacle for this manuscript since, as it is said in the Conclusions, the analysis of the catchment's functioning was beyond its scope. But then you should more clearly stick to the technical aspects throughout and also modify the title: "Understanding mean residence times…" implies that process understanding is a central part of your work.

As mentioned previously, the title was changed to:

*"Understanding uncertainties when inferring mean transit times through tracer based lumped parameter models in Andean tropical montane cloud forest catchments"*

Besides, we added the following sentence in the first paragraph of Section 2.6:

*"It is common practice to extend the time series artificially by duplicating it (Hrachowitz et al., 2010 and 2011). This does not change the results; it rather gives the model more room to find stable results."*

The following references are related to the later sentence:

*"Hrachowitz, M., Soulsby, C., Tetzlaff, D., Malcolm, I. A. and Schoups, G.: Gamma distribution models for transit time estimation in catchments: physical interpretation of parameters and implications for time-variant transit time assessment, Water Resour. Res., 46, W10536, doi:10.1029/2010WR009148, 2010."*

*"Hrachowitz, M., Soulsby, C., Tetzlaff, D. and Malcolm, I. A.: Sensitivity of mean transit time estimates to model conditioning and data availabilityle, Hydrol. Process., 25, 980–990, doi:10.1002/hyp.7922, 2011."*

**p. 15884 l. 28 – p. 15885 l. 3:** Is it reasonable that, at baseflow conditions, MTTs in the tributaries are larger than in the mainstream river? This question leads more to the involved processes, however, a discussion might help to understand uncertainties of model results. Could this maybe be attributed to the synthetic input data series?

In general, it is correct that the water of the main river cannot be older than sum of its tributaries or vice versa. In this sense, taking into account the local cascade of the tributaries (and discharge of each tributary) the reported results are feasible and within the given range of uncertainties. Considering the mixing ratio of the monitored subcatchments, representing 60.1 $km^2$ of the 76.9$km^2$, the MTT for PL would be 2.2 yrs which is concordant with predictions from lumped models for PL 2.0 yrs (range 1.8 - 2.2 yrs).

In order to improve the understanding of results and to have insights on the processes and their uncertainties, now the results and discussion section includes the analysis of the distribution functions (please check reply to "General Comment").

**Table 2:** Runoff from the catchments is mostly >2500 mm/yr while precipitation was only 2000 – 2500 mm/yr in the study period. How does that fit?

Please note: Table 2 now is numbered as Table 1.

When calculating the average rainfall amount in the catchment, three potential sources of uncertainties, could explain the observed discrepancies:

- Scarce number of rainfall/meteorological stations in the catchment (considering the area and the high variability of rainfall amount between stations).
- uncertainties due to the interpolation method,
- horizontal rainfall accounting for additional precipitation input.

**Technical corrections:**

**Response: All the suggested technical corrections will be performed.**

p. 15876 l. 15: Delete the power after 100 m

Power number has been deleted

p. 15876 l. 27: total runoff volume (or delete "volume")

"Volume" has been deleted

p. 15880 l. 8: as function of time

Now it reads *"as function of time"*

p. 15880 l. 16: delete "the" in…the Eq. (1)…

Word *"the"* has been deleted

p. 15883 l. 13: expressed as average values

This expression belongs to a sentence that has been deleted in the revised version. Please check modified version of Sections 3 and 4 (reply to General Comment).

p. 15883 l. 22: clearer…a decreasing trend with increasing sampling depth…

This expression belongs to a sentence that has been deleted in the revised version. Please check modified version of Sections 3 and 4 (reply to General Comment).

p. 15883 l. 25: with increasing soil depth

This expression belongs to a sentence that has been deleted in the revised version. Please check modified version of Sections 3 and 4 (reply to General Comment).

p. 15886 l. 15: regardless of

Now it reads: *"regardless of…"*

p. 15888 ll. 7-9: Can you rearrange this sentence – it is difficult to understand

Sentence now reads as follows:

*"Regardless of the used model, efficiencies of MTT for stream waters were lower than for soil waters."*

p. 15888 ll. 12-14: This sentence is also not clear – please reword.

Sentence now reads as follows:

*"Since for most of the cases MTTs for soil waters showed an increasing trend according to increasing soil depth, longer MTTs corresponding to deeper soil layers are to be expected."*

Table 1: "m a.s.l." and "(weeks)" have to be shifted one column to the right

This mistake has been corrected.

Table 4: Is the superscript "a" for N, $\tau$, NSE and RMSE of relevance?

We deleted the superscript "a" in Table 4.

Table 5: $\tau$ is given in years, not in weeks. For a better clearness of the table it could help to separate the observed and the modelled data by a vertical line.

We corrected the units from weeks to years. Besides we plotted a vertical line in Table 4 in order to differentiate between modeled and observed data.