<u>Editor Comments</u>

<u>Comment:</u>

the reviewers clearly indicated a considerable improvement of the paper, however, one reviewer still has several points that should be addressed and looking at these points, I can only support the assessment of the reviewer. Hence, I would propose that the authors should careful examine the points and they can hopefully improve the paper regarding these observations.

<u>Reply:</u>

We appreciate the Editor's interest in our work and his additional efforts to help strengthening our work. We have addressed all remaining comments in detail below and revised the manuscript accordingly.

<u>Comment:</u>

Personally, I am also not happy, that the authors did not consider to use additional data collected in the Wüstebach catchment - certainly, there are only point data, but there is a wealth of point data and there has been many studies showing how to use such a wealth of data to improve the models. Uncertainty should not be an issue as discharge and isotope data as well as the forcing data are also uncertain which is not considered explicitly in the model either.

<u>Reply:</u>

We thank the Editor for insisting on this point. Although very challenging, mostly for spatial commensurability issues, we believe that we have found an effective way of incorporating soil moisture and throughfall observations as additional constraints on the model. More specifically, we have included them as constraints following a simplified limits-of-acceptability approach. As such we believe that we managed to strike a reasonable balance to limit both Type I and Type II errors. Overall, the additional constraints proved to be very valuable to identify and reject unsuitable parameter sets. The model solutions retained as feasible do now not only simultaneously reproduce multiple system signatures (as quantified by 14 individual performance metrics) but also reasonably well mimic the observed time series of soil moisture and the average throughfall ratio P_E/P. Please note that this resulted in some minor changes in the results but did not affect the overall interpretation of the results.

(in the track-changed revised manuscript: p.5, l.160ff; p.13, l.396-406; p.16, l.482-486; p.18, l.542-543; p.22, l.693-694; Figure 7)

<u>Comment:</u>

Interception - many studies show that interception in forest can make up easily 30% of actual ET, so why not considering this process explicitly and limiting I_Max after removing of the forest. I do not understand how Si(t) would be different after the forest removal when I_max is constant.

Reply:

We completely agree with the Editor that interception evaporation can easily account for 30% of actual ET. We therefore decided already in the original analysis to explicitly represent this process in our model. That is why we are rather confused by the notion of the Editor that we kept the model parameter for interception capacity I_{max} constant after deforestation. We of course explicitly accounted for a reduction of that storage capacity I_{max} after deforestation, as would be expected. This was described, illustrated and discussed at length in all previous versions and also here, in the latest revised version of this manuscript. More specifically, the reduction of I_{max} is shown in Figure 8 and Table 2, where it can be seen that, in particular in the riparian zone that was fully deforested, $I_{max,R}$ is reduced from 2.5 mm (1.9 – 4.8 mm) to 1.1 mm (0.1 – 1.3 mm). These effects are also described and discussed in some detail in the main text (in the track-changed revised manuscript: p.18, l.572-574; p.23, l.715-717; p.23, l.737-741, p.24, l.742-744; p.24, l.749-752).

Comment:

Model performance with respect to isotope data: Does this mean that the value of the isotope data is rather limited after all? (assuming that if the fit is poor for the best parameterization, this criterion might not be that helpful in distinguishing between good and not so good parameter sets.

Reply:

Although the Nash-Sutcliffe efficiency of the stream tracer signal is rather modest (E_{NS}<0.4) this does not entail that the model does a poor job in reproducing the tracer response: the modelled tracer composition remains much better than the mean of the observations. As there is very little fluctuation around this mean, with a variance of ~ 0.1‰, it is realistically seen close to impossible for a model to achieve much higher values of E_{NS}, given the limitations in the available data, as described in more detail in the direct reply to the reviewer below. In addition, the model does a rather good job in reproducing the second and here probably even more important aspect of the tracer response: the high level of damping between the precipitation and stream water tracer signals. To explicitly acknowledge and emphasize this, we have now added one more calibration objective function: the relative error between the observed and modelled damping ratios (see below for more detail; in the track-changed revised manuscript: p.13, l.393; p.14, l.413; p.16, l.487-489; p.17, l.527-529; p.18, l.551-554; Figure 6; Table 3).

Comment:

Monte Carlo runs

Reply:

The flux tracking module based on SAS-functions used to model the tracer response requires considerable computational power and is associated with very long run times. We have now added a further $2*10^6$ model realizations to a total of $3*10^6$ (in the track-changed revised manuscript: p.13, l.391). In spite of some minor improvements in terms of performance metrics, the additional parameter sets did not change the overall results nor the major conclusions of our analysis. We believe that this, together with *the relatively well constrained posterior distributions of most parameters as shown in Figure 8 and Table 2 of the revised manuscript provides a robust description of the feasible parameter space. For a more detailed response please see below the direct reply to the reviewer.*

<u>Comment:</u>

Due to the lengthy combined result and discussion section, I would also support that the results and discussion should be separated.

<u>Reply:</u>

We have now separated results (in the track-changed revised manuscript: section 5, p.14-21) and discussion (section 6, p.21-27)

<u>Comment:</u>

I would also like to ask the authors to provide a manuscript with track changes, so it is easier for me and the reviewer to follow the changes between the manuscript.

<u>Reply:</u>

Such as in the previous round of revisions we have in detailed followed the HESS guidelines and again attached the track changed version of the revised manuscript at the end of this authors' response document and we have, in addition, now also uploaded an individual track-changed version.

<u>Comments Reviewer #1</u>

<u>Comment:</u>

The authors have done a good job in implementing my suggestions and revising the paper. My only remaining concern is data accessibility. I have not found links or indications on how to access the catchment data and model data. I think the authors should provide such data.

<u>Reply:</u>

We highly appreciate the reviewers' positive assessment. We have added this information in the revised manuscript.

<u>Comments Reviewer #2</u>

<u>Comment:</u>

The authors took great care in revising the manuscript and replying to reviewers' comments. The manuscript goals are clearer and results are easier to follow. I believe the paper is now ready for publication on HESS.

Reply:

We highly appreciate the reviewer's positive assessment of our manuscript.

Comments Reviewer #3

Comment:

Overall the authors did a good job in revising the manuscript. The rebuttal reads a bit strange as it often says that something will be changed, but I assume this is just the text from the discussion phase and things actually have been changed by now. The manuscript has clearly improved by the revision, however, I have still a few points that should be considered.

Reply:

We highly appreciate the overall positive assessment of the reviewer and thank him/her for the additional time invested for providing further valuable comments to strengthen the manuscript.

Comment:

I appreciate the clarifications regarding the representation of interception. The authors argue that forest removal affects transpiration more than interception. While this obviously is case-specific I would, in general, maintain that the changed interception is the more obvious and often also more important change. While decreased transpiration will be partly compensated for by increased soil evaporation, interception is more of a 'completely lost process'.

Reply:

We fully agree with the reviewer that after deforestation interception becomes almost a "lost process". This is illustrated in our results by the significant reduction of the model parameter representing the interception capacity. In particular, in the fully deforested riparian zone, the interception capacity is significantly reduced from of $I_{max,R}$ = 1.9 – 4.8 mm before deforestation to a value close to zero ($I_{max,R}$ = 0.1 – 1.3 mm) after deforestation, which indicates that this process is indeed almost "lost", as remarked by the reviewer (Figure 8, Table 2). In contrast, no discernible change in the interception capacity $I_{max,H}$ was found for the hillslopes of which only ~ 10% were deforested (Figure 8, Table 2).

We have further emphasized this in the revised version of the manuscript (in the track-changed revised manuscript: p.18, l.572-574; p.23, l.715-717; p.23, l.737-741, p.24, l.742-744; p.24, l.749-752).

Comment:

I am not fully satisfied with the responses regarding passive water storage. Yes, water below the wilting point cannot be extracted by the plants, but for most of that water, it can be assumed that some mixing/exchange occurs with more mobile water. I am still wondering what it means if this is not represented in the modelling.

Reply:

We fully agree with the reviewer, that although the water volumes below wilting point are not accessible to plants, they will experience some exchange with the more mobile water. Water stored below the wilting point will thus act as an additional passive mixing volume.

The overall effect of different mixing volumes on young water fractions however, critically depends on the magnitude of those individual volumes. For example, the passive mixing volume $S_{s,p}$ in the model (Figures 4, 8; p.12, l.363-370), which defines the hydrologically and hydraulically passive mixing volume of the groundwater and which is constant over time (i.e. $dS_{s,p}/dt = 0$), was here and in many tracer studies found to be several orders of magnitude larger than the hydrologically and hydraulically active volume $S_{s,a}$, which is variable over time (i.e. $dS_{s,a}/dt \neq 0$) – here: $S_{s,p}$ > 7600 mm, Table 2, Figure 8; $S_{s,a}$ ~ 10 – 200 mm (see also e.g. Birkel et al., 2010,2012,2014; Fenicia et al., 2010; McMillan et al., 2012; Hrachowitz et al., 2013,2015; Harman, 2015; Benettin et al., 2016,2017). This passive mixing volume $S_{s,p}$ therefore has a significant and discernible effect on the overall water volume $S_{s,tot} = S_{s,a} + S_{s,p}$ from which water outfluxes are "sampled" and thus on the age composition and consequently on the young water fractions of the outflows.

In contrast, the volume of water stored in the passive mixing volume below the wilting point is typically much lower than the volume of the more mobile water and thus also of the total mixing volume. More specifically, for silty clay loam soils such as in the Wuestebach catchment the water content at the wilting point is typically found at ~ 0.1 m^3/m^3, while the water content at field capacity is up to 4 times higher at ~ 0.35 – 0.45 m^3/m^3 (e.g. Romano and Santini, 2002). The hydrologically and hydraulically active mixing volume combines the volumes between wilting point and field capacity as well as the transient water volume above field capacity that can eventually not be held against gravity. Depending on the wetness conditions, this active mixing volume can in the extreme case be up to 10 times (under fully saturated conditions) higher than the passive mixing volume and still ~ 4 times higher under average conditions when field capacity is reached. Water ages and in particular the young water fractions in the unsaturated zone are therefore largely controlled by the larger, hydrologically active mixing volume. This is further exacerbated by the fact that under drier conditions, i.e. when the soil water content approaches the water content at the wilting point, by definition the fraction of young water is very low in any case, as the water released from the system cannot have entered the system very recently under such dry conditions. In addition, the rather tightly bound water below the wilting point will, due to the strong adhesive forces close to the surface of the soil particles, only experience rather low diffusive exchange rates with mobile water. By extension, this implies that water volumes stored below the wilting point are, on average, characterized by rather old ages, which will only have very limited effect on the young water fractions (i.e. younger than 3 months) analysed here.

Altogether this implies that while omitting this exchange in the model will cause a slight under-/over-estimation of older ages, the absolute effects of this omission will be very minor: under very dry situations when a proportionally larger effect of exchange with water stored below the wilting point may be expected (as the water volume in the passive store is then larger than the volume in the active store), the fractions of young water, which are analysed here in this manuscript, are close to zero (Figure 9c,d,j; Figure 10) and the absolute effect of the passive storage volume below wilting point cannot be meaningfully discerned anymore.

In Figure R1 hereafter we provide a simplified illustration of the effects of including the water content below the wilting point as an additional passive mixing volume on the estimates of young water fractions F_{yw}, i.e. water younger than 3 months, under (a) very wet and under (b) dry conditions. As can be seen, the differences are rather limited.

We have clarified this and also briefly discussed the potential effects of this in the revised version of the manuscript (p.12, l.369-372).

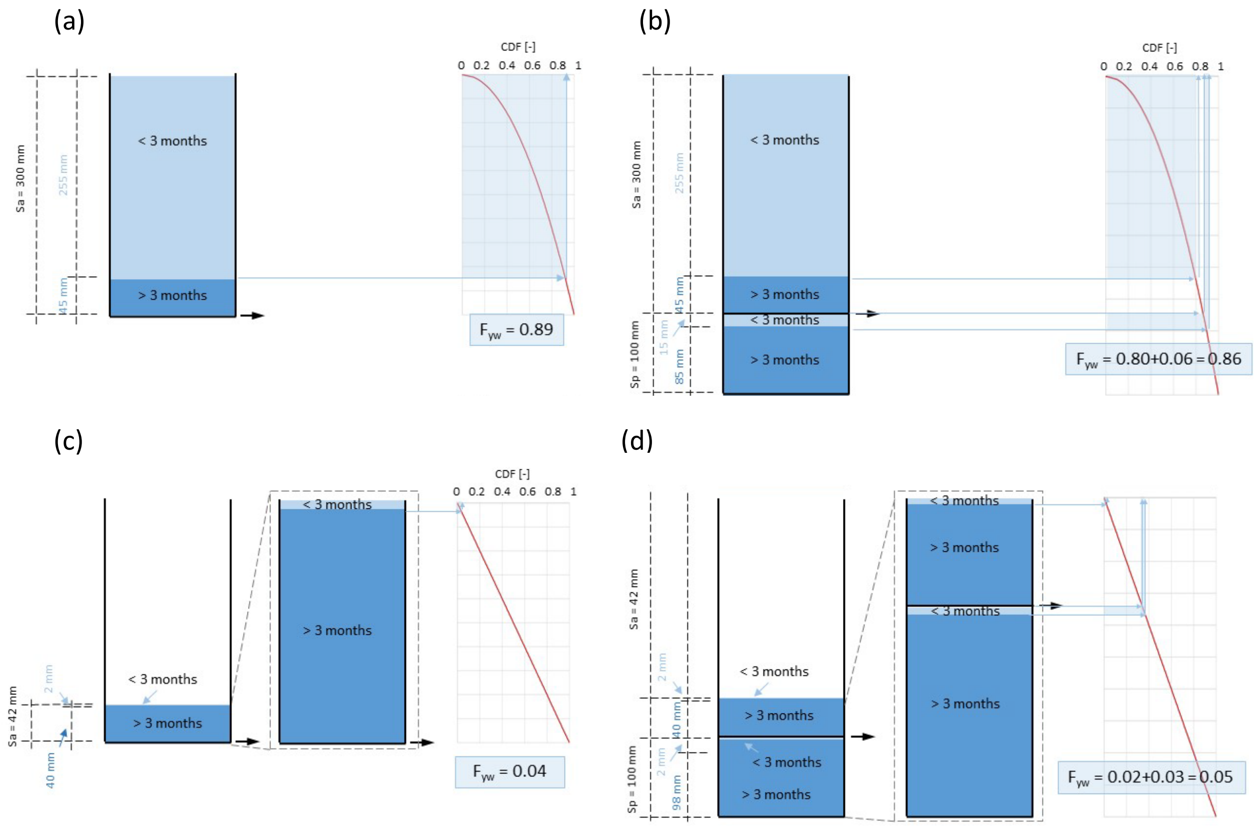


Figure R1: *Fractions of young water F_{yw} released from (a) the active mixing volume of the soil storage S_a under wet conditions after a few weeks of rainfall, i.e. soil storage filled to field capacity, with most of the water being younger than 3 months due to recent rainfall. Outflow released from that example is sampled with a SAS-function with strong preference for young water following a Beta-distribution (α=0.75, β=1; Eq.40 in the manuscript). (b) under wet conditions accounting for the additional passive mixing volume S_p below the wilting point, assuming a volume ratio of 4:1 between the total water content S_a+S_p and the water content below wilting point S_p, thereby reflecting the soil hydraulic properties of silty clay loams. Due to the rather low diffusive exchange rates between S_a and S_p, the fraction of water < 3 months is limited in S_p. The same SAS-function to sample water of different ages as in (a) is used. However, in this case the outflows are sampled from the total volume S_a+S_p. (c) shows the situation under dry conditions without considering the passive mixing volume S_p below the wilting point. After a prolonged dry period, only a small fraction of water in the soil is younger < 3months, as it did not rain for several weeks. Under such dry conditions the preference for sampling young water is reduced following Eq.40 in the manuscript and the SAS-function therefore converges towards a uniform distribution, i.e. a Beta-distribution with parameters α=1 and β=1. (d) shows dry conditions, sampling water according to the same SAS-function as in (c), but explicitly considering the passive mixing volume below the wilting point equivalent to (b). Note, that in this illustrative example F_{yw} is higher under wet conditions than in the results reported in the manuscript. The reason for this is that here, in contrast to the results in the manuscript, F_{yw} describe exclusively the time until release from the soil storage and before having travelled through the groundwater reservoir.*

Comment:

Looking closer at the model performance with respect to the isotope data (Fig 3), I am wondering whether the model really can be said to represent these data. The authors also discuss this issue and report some rather low model performance values. Modelling isotopes is hard, so I don't want to blame the authors here, but as the novelty of this study is the inclusion of isotope data in the modelling, I am a bit puzzled by these poor model fits. Does this mean that the value of the isotope data is rather limited after all? (assuming that if the fit is poor for the best parameterization, this criterion might not be that helpful in distinguishing between good and not so good parameter sets.

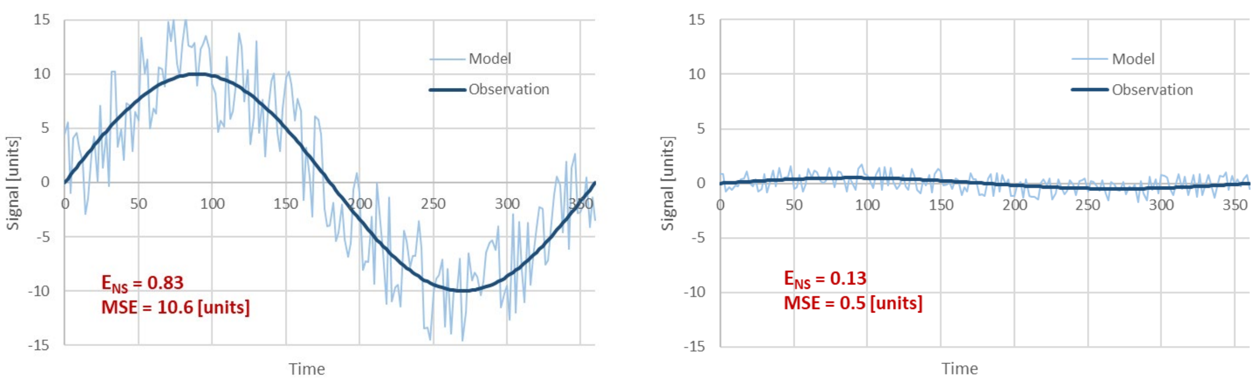
Reply:

We highly appreciate this comment and there are several complementary points in response to this comment that we would like to clarify.

It is true that the model provides only a rather modest representation of the high-frequency dynamics of the stream tracer concentration, as illustrated by the values of Nash-Sutcliffe efficiencies $E_{NS} < 0.4$. The limited ability of models to reproduce high-frequency dynamics of tracer compositions in stream water, in particular in strongly dampened systems, is widely reported in a wide range of environments, and reflected by Nash-Sutcliffe efficiencies of modelled stream tracer concentration that, similar to our study, very rarely exceed values of $E_{NS} \sim 0.4$ (e.g. Page et al., 2007, Shaw et al., 2008; Birkel et al., 2010,2011a,b; Fenicia et al., 2010; McMillan et al., 2012; Hrachowitz et al., 2013; Benettin et al., 2015, 2016; Birkel and Soulsby, 2016; van Huijgevoort et al., 2016; Ala-aho et al., 2017; etc.).

However, the second important feature describing the tracer storage and release properties of this catchment is very well captured in this study: the degree to which the seasonal amplitudes in the precipitation tracer are dampened in the stream. This system property reflects the buffer function of the catchment. The model's ability to closely reproduce this buffer function is evidence that the model meaningfully represents the low pass filter characteristics, and thus the major characteristics of solute transport in the study catchment. This is also the reason why we decided to show the precipitation and stream water tracer signals on the same y-axis scale in Figure 3a, as this allows to visually appreciate the strong degree of dampening between the precipitation and stream signals in the study catchment and the model's ability to reproduce that. The damping ratio R_D, here expressed as the ratio of the standard deviation of the stream water signal to the standard deviation of the precipitation signal, describes this damping effect. The modelled dampening ratio matches the observed dampening ratio for most solutions very well, as indicated by the error metric $E_{RD,\delta O18} = 1 - R_{D,mod}/R_{D,obs}$, which remains > 0.95 with a value of 1 indicating a perfect fit. This further underlines the model's ability to reproduce this critical characteristic of the study catchment. To better and explicitly emphasize the relevance of this catchment signature, we have added the relative error between the observed and the modelled damping ratio in the revised manuscript as additional, 14th objective function to the analysis (Table 3; Figure 6)

To further clarify the interpretation of the Nash-Sutcliffe efficiency E_{NS}, it is critical to see that this error metric essentially evaluates how much better a modelled variable is than the <u>mean of the observed values</u>. This implies that the lower the variance around that mean in the observed data is, the more difficult it is for a model to provide a better fit than the mean. This is because the mean provides already a rather strong representation of the signal. In contrast, for signals with much more pronounced amplitudes (and thus higher variance), it is easier for models to generate high values of E_{NS}. In the *following we illustrate this with a illustrative hypothetical example of two signals with considerably distinct amplitudes (Figure R2 here below). It can be seen that in the example with the high amplitudes, the E_{NS} of the modelled signal reaches E_{NS} = 0.83 while the mean squared error reaches MSE = 10.6. However, in the second example, where the variance in the observed signal is much lower, such that the signal itself plots very close around the mean, E_{NS} = 0.13 and thus <u>much lower</u> than before, while according to MSE = 0.5 the model provides a <u>much better</u> fit. In other words, already very small errors can cause dramatic reductions of E_{NS} in low-amplitude (and thus low variance) signals, while very high values of E_{NS} can be sustained with much larger errors in high-amplitude signals, as these higher errors still allow improvements as compared to the mean. This further implies that the magnitude of E_{NS} <u>always</u> needs to be interpreted together with the magnitude of the variance in the signal. In the specific case of the stream tracer signal in the analysis in our manuscript, we observe a very high degree of damping. The stream tracer compositions, both observed and modelled, therefore show only very limited variance around the mean, indicating that the absolute errors in the model are very low, but also that it is very difficult for a model to perform much better than the mean. Thus, while it is indeed questionable to use E_{NS} as metric to compare signals from different systems, E_{NS} is nevertheless useful to compare different models and to identify feasible parameter sets in one single system.*



<u>Figure R2:</u> the dark line indicates an observed high-amplitude signal (amplitude = 10; left) and on observed low-amplitude signal (amplitude = 0.5; right). The light blue lines indicate models for these observations with high absolute errors (and thus also the squared errors MSE) for the high-amplitude signal (MSE=10.6; left) and much lower absolute errors for the low-amplitude signal (MSE=0.5). However, in spite of this, E_{NS} remains much higher for the high-amplitude (E_{NS}=0.83) than for the low amplitude signal (E_{NS}=0.13)

The rather small variability around the mean in the observed stream tracer signal introduces an additional challenge, related to the information content of the available data: the fluctuations around the mean are very close to the observational uncertainty as the analytical precision of the spectrometer used for the isotope analysis is ~0.1‰ while the variance in the observed (and modelled) signals is equally ~ 0.1‰. Much of the observed fluctuations may therefore merely be errors/uncertainties in the observations, which the model cannot and should not reproduce. Another issue with respect to the information content in the available data is the fact that here, such as in the vast majority of tracer studies, precipitation isotope data were available as weekly bulk samples, while stream water samples were taken as instantaneous grab samples. This entails that the weekly bulk samples average out the *effects of potential extreme events, while the variability in the stream water composition from grab samples may reflect these events. As the model is forced with bulk samples taken at weekly resolutions, it can represent fluctuations at any higher frequencies only to a very limited degree.*

Besides the effect of data on E_{NS}, please also note that we used an extended multi-objective model calibration and evaluation strategy based on 14 performance metrics. As such, the modelled output of any parameter combination had to simultaneously exceed thresholds of all 14 objective functions and, in addition, satisfy two newly added limits-of-acceptability constraints to ensure a meaningful representation of throughfall and soil moisture (for more detail please see replies to the Editor). Parameter combinations that failed to do so were discarded. If we had used a common calibration strategy based on only one or 2 calibration objectives, such as $E_{NS,Q}$ and $E_{NS,\delta O18}$, as is used in the vast majority of research papers, most of the discarded solutions would have been wrongly kept as feasible (although they cannot reproduce any of the other objectives), many of which will potentially have misleadingly resulted in $E_{NS,Q}$ and $E_{NS,\delta O18}$ much higher than the ones obtained here – we simply would have obtained the right results for the wrong reasons (cf. Kirchner, 2006). This phenomenon of pareto optimal solutions is well known and exhaustively described in literature (e.g Gupta et al., 1998; Yapo et al., 1998; Vrugt et al., 2003; Efstratiadis and Koutsoyiannis, 2010; Hrachowitz et al., 2014).

Overall, we have exposed our model to much more rigorous model constraints (i.e. 14 objective functions and 2 limits-of-acceptability constraints) than is done in most research papers. This resulted in the rejection of deceptively high-performance models, in favour of somewhat lower performance with respect to individual performance metrics, but which ensure a much higher internal consistency of the model and thus provide more reliable representations of real world dynamics.

Following the reviewer's observation and in a further step to improve the robustness of our model and to emphasize the relevance of the degree of damping in the isotope signal we have now added with the relative error of the damping ratio one more model objective function. This allowed us to move the emphasis away from considering only the isotope dynamics as evaluated by $E_{NS,\delta O18}$ to also explicitly consider the buffer effect of the catchment in a separate evaluation metric. As can be seen, this metric, the relative error in the damping ratio remains very well reproduced in all model cases with $E_{RD,\delta O18}$ = 0.95 – 0.99 (in the track-changed revised manuscript: p.13, l.393; p.14, l.413; p.16, l.487-489; p.17, l.527-529; p.18, l.551-554; Figure 6; Table 3)

<u>Comment:</u>

Term effective precipitation: the authors use this term not in its usual way. Usually, effective precipitation is the part of precipitation that makes it to groundwater (=recharge), e.g., http://www.fao.org […]. Here it is the precipitation minus interception (but not minus other evaporative fluxes!). I would recommend avoiding this term totally to prevent confusion.

<u>Reply:</u>

Agreed, we have replaced the term "effective precipitation" by "throughfall" in the revised manuscript.

Comment:

Monte Carlo runs: while 10e6 sounds like a lot, with 14 parameters this means just about 2.7 values along each (parameter)dimensions, i.e., the parameter spaces is sampled rather scarcely. This is an issue of many MC studies, but these days, much larger numbers of runs should be possible, why only one million? This was the standard years ago. Today more runs should be doable, any reason for using one million?

Reply:

Indeed, under-sampling can become a limiting issue in inverse modelling approaches.

The challenge we were facing here was that process-based hydrological models which are coupled with tracer modules based on the SAS-function approach are computationally very demanding and require very long run-times, which is also one of their major limitations (cf. personal communication Paolo Benettin). These can be traced back to two major reasons: (1) for each time step SAS-functions need to be generated from programming language internal functions of the associated probability distributions, which is a comparably time-consuming process, and even more important (2) to avoid numerical instabilities related to problems in maintaining water and tracer mass balances, which occur when SAS-functions dictate to sample a higher proportion of a specific age than is actually present in a storage component. In such a case, the water ages and the relates water and tracer masses need to be accordingly redistributed and resampled in an iterative process, which can lead to long running loops in the code and which often are very time-consuming. In our specific case, 10^6 model realizations require a model run time of approximately 1.5 months(!).

The initially 10^6 model realizations led to most parameters being reasonably well identified between rather limited ranges as illustrated in Figure 8 and Table 2. It is thus not implausible to assume that these limited ranges largely delimit the region of the parameter space that contains the optimal parameter combination. Thus even if the actual optimal solution is not found, there will be only limited deviation from the solutions found and kept as feasible in this analysis.

To strengthen the model we have followed the suggestion of the reviewer and have for the revised version of the manuscript increased the number of model realizations by 200% to a total of $3*10^6$ (in the track-changed revised manuscript: p.13, l.391). Given the computational resources required by the model we unfortunately do not have the capacity to further extend this analysis. In any case and notwithstanding some minor improvements in terms of performance metrics, the additional parameter sets did not change the overall results nor the major conclusions of our analysis. However, we believe that the significantly increased number of realizations together with the rather well identified parameters allow a meaningful, albeit quite clearly not complete, exploration of the parameter hyperspace.

Comment:

Personally, I would prefer if results and discussion could be separated, this could help to make the manuscript more readable

Reply:

We have now separated results (in the track-changed revised manuscript: section 5, p.14-21) and discussion (section 6, p.21-27)

Comment:

Eq 42 looks fancy, but I find this a bit confusing, there is only one measure related to O-18, or?

Reply:

Yes, indeed, Eq.42 initially only considered one metric for O-18. We have now added a second one: the relative error of the damping ratio (see also reply above; in the track-changed revised manuscript: p.14, l.413; Table 3; Figure 6).

References:

Ala-Aho, P., Tetzlaff, D., McNamara, J. P., Laudon, H., & Soulsby, C. (2017). Using isotopes to constrain water flux and age estimates in snow-influenced catchments using the STARR (Spatially distributed Tracer-Aided Rainfall–Runoff) model. Hydrology and Earth System Sciences, 21(10), 5089-5110.

Benettin, P., Bailey, S. W., Campbell, J. L., Green, M. B., Rinaldo, A., Likens, G. E., ... & Botter, G. (2015). Linking water age and solute dynamics in streamflow at the H ubbard B rook E xperimental F orest, NH, USA. Water Resources Research, 51(11), 9256-9272.

Benettin, P., Soulsby, C., Birkel, C., Tetzlaff, D., Botter, G., & Rinaldo, A. (2017). Using SAS functions and high-resolution isotope data to unravel travel time distributions in headwater catchments. Water Resources Research, 53(3), 1864-1878.

Birkel, C., & Soulsby, C. (2016). Linking tracers, water age and conceptual models to identify dominant runoff processes in a sparsely monitored humid tropical catchment. Hydrological Processes, 30(24), 4477-4493.

Birkel, C., Dunn, S. M., Tetzlaff, D., & Soulsby, C. (2010). Assessing the value of high-resolution isotope tracer data in the stepwise development of a lumped conceptual rainfall–runoff model. Hydrological Processes, 24(16), 2335-2348.

Birkel, C., Tetzlaff, D., Dunn, S. M., & Soulsby, C. (2011a). Using time domain and geographic source tracers to conceptualize streamflow generation processes in lumped rainfall-runoff models. Water Resources Research, 47(2).

Birkel, C., Tetzlaff, D., Dunn, S. M., & Soulsby, C. (2011b). Using lumped conceptual rainfall–runoff models to simulate daily isotope variability with fractionation in a nested mesoscale catchment. Advances in Water Resources, 34(3), 383-394.

Efstratiadis, A., & Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrological Sciences Journal–Journal Des Sciences Hydrologiques, 55(1), 58-78.

Fenicia, F., Wrede, S., Kavetski, D., Pfister, L., Hoffmann, L., Savenije, H. H., & McDonnell, J. J. (2010). Assessing the impact of mixing assumptions on the estimation of streamwater mean residence time. Hydrological Processes, 24(12), 1730-1741.

Harman, C. J. (2015). Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed. Water Resources Research, 51(1), 1-30.

Hrachowitz, M., Savenije, H., Bogaard, T. A., Tetzlaff, D., & Soulsby, C. (2013). What can flux tracking teach us about water age distribution patterns and their temporal dynamics?. Hydrology and Earth System Sciences, 17(2), 533-564.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., ... & Gascuel-Odoux, C. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. Water resources research, 50(9), 7445-7469.

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3).

McMillan, H., Tetzlaff, D., Clark, M., & Soulsby, C. (2012). Do time-variable tracers aid the evaluation of hydrological model structure? A multimodel approach. Water Resources Research, 48(5).

Page, T., Beven, K. J., Freer, J., & Neal, C. (2007). Modelling the chloride signal at Plynlimon, Wales, using a modified dynamic TOPMODEL incorporating conservative chemical mixing (with uncertainty). Hydrological Processes: An International Journal, 21(3), 292-307.

Romano, N., & Palladino, M. (2002). Prediction of soil water retention using soil physical data and terrain attributes. Journal of Hydrology, 265(1-4), 56-75.

Shaw, S. B., Harpold, A. A., Taylor, J. C., & Walter, M. T. (2008). Investigating a high resolution, stream chloride time series from the Biscuit Brook catchment, Catskills, NY. Journal of Hydrology, 348(3-4), 245-256.

Van Huijgevoort, M. H. J., Tetzlaff, D., Sutanudjaja, E. H., & Soulsby, C. (2016). Using high resolution tracer data to constrain water storage, flux and age estimates in a spatially distributed rainfall-runoff model. Hydrological Processes, 30(25), 4761-4778.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., & Sorooshian, S. (2003). Effective and efficient algorithm for multiobjective optimization of hydrologic models. Water resources research, 39(8).

Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. Journal of hydrology, 204(1-4), 83-97.

Reduction of vegetation-accessible water storage capacity after deforestation affects catchment travel time distributions and increases young water fractions in a headwater catchment

Markus Hrachowitz[1], Michael Stockinger[2,3], Miriam Coenders-Gerrits[1], Ruud van der Ent[1], Heye Bogena[2], Andreas Lücke[2], Christine Stumpp[3]

[1]Department of Watermanagement, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628CN Delft, Netherlands
[2]Institute of Bio- and Geosciences, Agrosphere Institute (IBG-3), Forschungszentrum Jülich, Wilhelm-Johnen-Straβe, 52425 Jülich, Germany
[3]Institute for Soil Physics and Rural Water Management, University of Natural Resources and Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria

Correspondence to: Markus Hrachowitz (m.hrachowitz@tudelft.nl)

Abstract. Deforestation can considerably affect transpiration dynamics and magnitudes at the catchment-scale and thereby alter the partitioning between drainage and evaporative water fluxes released from terrestrial hydrological systems. However, it has so far remained problematic to directly link reductions in transpiration to changes in the physical properties of the system and to quantify these changes of system properties at the catchment-scale. As a consequence, it is difficult to quantify the effect of deforestation on parameters of catchment-scale hydrological models. This in turn leads to substantial uncertainties in predictions of the hydrological response after deforestation but also to a poor understanding of how deforestation affects principal descriptors of catchment-scale transport, such as travel time distributions and young water fractions. The objectives of this study in the Wüstebach experimental catchment are therefore to provide a mechanistic explanation of *why* changes in the partitioning of water fluxes can be observed after deforestation and how this further affects the storage and release dynamics of water. More specifically, we test the hypotheses that (1) post-deforestation changes in water storage dynamics and partitioning of water fluxes are largely a direct consequence of a reduction of the catchment-scale effective vegetation-accessible water storage capacity in the unsaturated root-zone ($S_{U,max}$) after deforestation and that (2) the deforestation-induced reduction of $S_{U,max}$ affects the shape of travel time distributions and results in shifts towards higher fractions of young water in the stream. Simultaneously modelling stream flow and stable water isotope dynamics using meaningfully adjusted model parameters both for the pre- and post-deforestation periods, respectively, a hydrological model with integrated tracer routine based on the concept of storage age selection functions is used to track fluxes through the system and to estimate the effects of deforestation on catchment travel time distributions and young water fractions F_{yw}.

It was found that deforestation led to a significant increase of stream flow, accompanied by corresponding reductions of evaporative fluxes. This is reflected by an increase of the runoff ratio from C_R = 0.55 to 0.68 in the post-deforestation period despite similar climatic conditions. This reduction of evaporative fluxes could be linked to a reduction of the catchment-scale water storage volume in the unsaturated soil ($S_{U,max}$) that is within the reach of active roots and thus accessible for vegetation transpiration from ~258 mm in the pre-deforestation period to ~ 101 mm in the post-deforestations period. The hydrological model, reflecting the changes in the parameter $S_{U,max}$ indicated that in the post-deforestation period stream water was characterized by slightly, yet statistically not significantly higher mean fractions of young water (F_{yw} ~ 0.13) than in the pre-deforestation period (F_{yw} ~ 0.~~11~~12). In spite of these limited effects on the overall F_{yw}, changes were found for wet periods, during which post-deforestation fractions of young water increased to values F_{yw} ~ 0.~~40~~ 37 for individual storms. Deforestation also caused a significantly increased sensitivity of young water fractions to discharge under wet conditions from dF_{yw}/dQ = 0.25 to 0.~~43~~36.

Overall, this study provides quantitative evidence that deforestation resulted in changes of vegetation-accessible storage volumes $S_{U,max}$ and that these changes are not only responsible for changes in the partitioning between drainage and evaporation and thus the fundamental hydrological response characteristics of the Wüstebach catchment, but also for changes in catchment-scale tracer circulation dynamics. In particular for wet conditions, deforestation caused higher proportions of younger water to reach the stream, implying faster routing of stable isotopes and plausibly also solutes through the subsurface.

1 Introduction

Plant transpiration is, globally, the largest continental water flux (Jasechko, 2018). Notwithstanding considerable uncertainties (Coenders-Gerrits, 2014), its magnitude depends on the interplay between canopy water demand and subsurface water supply (Eagleson, 1982; Milly and Dunne, 1994; Donohue et al., 2007; Yang et al., 2016; Jaramillo et al., 2018; Mianabadi et al., 2019). The latter is regulated by water volumes that are within the reach of roots and can be taken up by plants. Many plant species across humid climate zones develop only rather shallow root systems (Schenk, 2005) that do not directly tap the groundwater (Fan et al., 2017). In regions that are dominated by such shallow-rooting vegetation, the pore volume between field capacity and permanent wilting point that is *within the reach of active roots* becomes a core property of many terrestrial hydrological systems (Rodriguez-Iturbe et al., 2007). This maximum vegetation-accessible water storage volume in the unsaturated root-zone of soils, hereafter referred to as vegetation-accessible water storage capacity $S_{U,max}$ [mm], constitutes a major partitioning point of water fluxes. It regulates the temporally varying ratio between drainage, such as groundwater recharge or shallow lateral flow, on the one hand and transpiration fluxes on the other hand (Savenije and Hrachowitz, 2017), which can in turn generate considerable feedback effects on downwind precipitation and drought generation (e.g. Seneviratne et al., 2013; Ellison et al., 2017; Teuling, 2018; Wang-Erlandsson et al., 2018; Wehrli et al., 2019).

Traditionally, $S_{U,max}$ is determined as the product of root-depths or root-distributions and pore water content between field capacity and permanent wilting point. Although correct in principle, this method has several weaknesses for applications at the catchment-scale as much of the required data are typically not available at sufficient levels of detail. While soil maps and the associated soil water retention curves have become globally available at resolutions < 1km (Arrouays et al., 2017; Hengl et al., 2017), they are characterized by considerable uncertainties. Similarly, direct and detailed observations of root-systems are very scarce. They are, globally, limited to a few thousand individual plants only (e.g. Schenk and Jackson, 2002; Fan et al., 2017) and many of the observations are based on biomass extrapolations after excavating only the first meter of soil or less (Schenk and Jackson, 2003). Consequently, soil and root data largely remain inaccurate snapshots in space. As such, they are likely to be inadequate reflections of the spatial heterogeneity of soils and roots. In addition, these available data are also mostly snapshots in time and therefore disregard the adaptive behaviour of plant communities, whose compositions, and thus characteristics, at ecosystem level continuously evolve over multiple scales in space and time in response to changes in ambient conditions (e.g. Laio et al., 2006; Brunner et al., 2015; Tron et al., 2015).

There is increasing evidence that vegetation does not only actively adapt to its (changing) environment, but that it does so in an way that allows the most efficient use of available energy and resources (e.g. Guswa, 2008; Schymanski et al., 2008). The vegetation, i.e. a collective of individual different plants within an area of interest that is present at any given moment at any given location has survived past conditions. This in itself is a manifestation of the successful adaption of individual plants to their environment in the past. They have optimally allocated resources to balance sub- and above-surface growth to simultaneously meet water, nutrient and light requirements. This implies that these plants developed root-systems that, amongst other factors, ensure continuous access to *sufficient* water – but not more – to bridge dry periods. An individual plant that is not adapted to meet its water and nutrient requirements through its root-system as well as its light requirements through its foliage system in competition with other plants will disappear and be replaced by a better adapted plant. The root-system of vegetation at ecosystem level, and the associated vegetation-accessible water storage capacity $S_{U,max}$, is therefore at a dynamic equilibrium with and responding to the ever changing conditions of its environment. Similarly, any type of direct human interference with vegetation, such as deforestation, has an impact on transpiration water demand, the extent and structure of active root-systems and consequently on $S_{U,max}$ (Nijzink et al., 2016a).

For a meaningful quantification of $S_{U,max}$ at larger scales, such as the catchment-scale, it is therefore necessary to adopt a Darwinian perspective (Harman and Troch, 2014) and to estimate effective values of $S_{U,max}$, reflecting the collective and adaptive behaviour of all individual plants within a catchment. Results from many previous studies suggest, broadly speaking, three methods to do so. The first is the use of inverse approaches that treat $S_{U,max}$ as model calibration parameter (Fenicia et al., 2008; Speich et al., 2018; Bouaziz et al., 2020; Knighton et al., 2020). Alternatively, the second type of methods is based on optimality principles that maximize variables such as net primary production or carbon gain (Kleidon, 2004; Guswa, 2008; Hwang et al., 2009; Yang, et al., 2016; Speich et al., 2018, 2020), nitrogen uptake (McMurtrie et al., 2012) or transpiration rates (Collins and Bras, 2007; Sivandran and Bras, 2012). Lastly, $S_{U,max}$ and its evolution over time can be directly estimated through magnitudes of annual water deficits as determined from observed water balance data (Gentine et al., 2012; Donohue et al., 2012; Gao et al., 2014; DeBoer-Euser et al., 2016; van Oorschot et al., 2021).

For transpiration, shallow-rooting plants extract pore water of unsaturated soils that is held against gravity, i.e. between field capacity and permanent wilting point, and within the reach of roots. Significant vertical or lateral drainage only occurs at water contents above field capacity. By extracting soil water below that, transpiration therefore generates a root-zone water storage reservoir between field capacity and permanent wilting point that is characterized by a storage capacity $S_{U,max}$, i.e. a *maximum* vegetation-accessible storage volume, and that is at any given moment filled with a specific water volume S_U(t), depending on the past sequence of water inflow and release.

Storage reservoirs as $S_{U,max}$, or others such as groundwater bodies, are key for hydrological functioning (Sprenger et al., 2019b) as they provide a buffer against hydrological extremes, such as floods and droughts. With larger storage reservoirs, the hydrological memory of a system can increase as more water can be stored and held over longer periods of time (e.g. Hrachowitz et al., 2015; Sprenger et al., 2019b). This also implies that while increased actual volumes of water stored in and thus the degree of filling of storage reservoirs, e.g. $S_U(t)$, can reduce water ages (Harman, 2015), increased sizes of storage reservoirs, e.g. $S_{U,max}$, can increase water ages, thereby both controlling catchment travel time distributions (TTD; Soulsby et al., 2010). As fundamental descriptors of hydrological functioning TTDs describe the age structure of water held in and released from catchments (Birkel et al., 2015; Rinaldo et al., 2015), which is critical for regulating solute transport and thus nutrient and contaminant dynamics (Hrachowitz et al., 2016).

However, neither the effects of land cover change (Blöschl et al., 2019) nor the individual roles of different storage compartments in terrestrial hydrological systems are well understood (McDonnell et al., 2010; Penna et al., 2018, 2020). This is mostly a consequence of the lack of suitable observational technology to directly observe their respective volumes at larger scales. It remains therefore also unclear how deforestation affects $S_{U,max}$ (e.g. due to a less developed and complex rooting system for subsequent younger vegetation) and how changes in $S_{U,max}$ may propagate to affect both, the partitioning of water fluxes as well as the age structure of water stored in and released from catchments as described by residence and travel time distributions.

For the study site of this paper, the Wüstebach experimental catchment (Germany), a previous study quantified the effects of deforestation on the partitioning of water fluxes (Wiekenkamp et al., 2016). It was found that forest removal significantly reduced evaporative fluxes. This led to more persistent higher soil moisture levels and eventually to increases in stream flow. Similarly, in the same catchment, Wiekenkamp et al. (2020) found evidence for increased post-deforestation occurrence of preferential flows while Stockinger et al. (2019) reported minor post-deforestation reductions in travel times.

To establish a quantitative mechanistic link between these studies we here aim to trace back and attribute the above reported post-deforestation changes in the hydrological response of the Wüstebach to deforestation-induced changes in (subsurface) system properties. The overall objective of this study is thus to analyse whether changes in these (subsurface) properties can explain *why* deforestation affects water flux partitioning and reduces travel times in the Wüstebach in an attempt to improve our quantitative understanding of critical zone processes (Brooks et al., 2015). Specifically we test the hypotheses that (1) post-deforestation changes in water storage dynamics and partitioning of water fluxes are largely a direct consequence of a reduction of the catchment-scale effective vegetation-accessible water storage capacity in the unsaturated root-zone ($S_{U,max}$) after deforestation and that (2) the deforestation-induced reduction of $S_{U,max}$ affects the shape of travel time distributions and results in shifts towards higher fractions of young water in the stream.

2 Study site

The experimental Wüstebach headwater catchment (0.39 km^2; Fig. 1a) is part of the Lower Rhine/Eifel Observatory of the Terrestrial Environmental Observatories network (TERENO; Bogena et al., 2018) located in the Eifel National Park in Germany (50°30'16"N, 06°20'00"E). The catchment is characterized by a humid, temperate climate with warm summers, mild winters and a mean annual temperature of around 7° C (Zacharias et al., 2011). Mean annual precipitation is about 1200 mm yr^{-1} and mean annual runoff about 700 mm yr^{-1} (Fig. 2). Although most of the precipitation occurs in the winter months, the fraction that falls as snow is typically less than 10 % of the annual precipitation and snow cover is present for no more than 3-4 weeks per year.

The catchment is drained by a perennial 2nd-order stream and extends from 595 to 630 m asl. The landscape is characterized by the gentle slopes of the surrounding hills and a flatter riparian area close to the stream, covering approximately 10 % of the catchment (Fig. 1a). The underlying bedrock is largely Devonian shales with sandstone inclusions (Richter, 2008) covered by periglacial layers (Borchardt, 2012). While cambisols dominate the hillslopes, gleysols and histosols characterize much of the riparian area (Bogena et al., 2015). The average soil depth in the catchment reaches about 1.6 m with a maximum of 2 m (Graf et al., 2014). In 1946, after the Second World War, the catchment was homogeneously and completely afforested (Fig. 1) with Sitka spruce (*Picea sitchensis*) and Norway spruce (*Picea abies*; Etmann, 2009). The maximum observed rooting depth of these spruce trees in the catchment is 50 cm and no roots were observed below this depth. In the course of the development of the area into a national park approximately 21 % of the catchment, including the entire riparian zone, were deforested in September 2013 and kept largely vegetation free since (Wiekenkamp et al., 2016; Fig. 1).

3 Data

3.1 Hydro-meteorological data

Daily hydro-meteorological data were available for the period 01/10/2009 – 30/09/2016 (Fig. 2). Precipitation P [mm d^{-1}] and mean daily temperature T [°C] were available from the Monschau-Kalterherberg meteorological station operated by the German Weather Service (Deutscher Wetterdienst DWD station 3339), located 9 km northwest of the Wüstebach catchment. The precipitation data were corrected for evaporation and wind drift losses according to Richter (1995) and as described in detail by Graf et al. (2014). Stream discharge Q [mm d^{-1}] at the outlet of the Wüstebach was observed with a V-notch weir for low flow measurements and a Parshall flume for medium to high flows (Bogena et al., 2015). Daily potential evaporation E_P [mm d^{-1}] was estimated using the Penman-Monteith equation. Daily depth-weighted average daily soil water content for the study period was estimated from a network of soil moisture sensors placed at 5, 20 and 50 cm depths at >100 locations across the study catchment as described by Graf et al. (2014) and Bogena et al. (2015). In addition, throughfall rates P_E [mm d^{-1}] were measured at one continuously forested location in the study catchment (Fig. 1) with an array of samplers as described in detail by Stockinger et al. (2015) over irregular intervals over the period 01/10/2012 – 30/09/2016.

3.2 Stable isotope data

Regular weekly $\delta^{18}O$ data from bulk precipitation samples collected in a cooled wet deposition gauge at the meteorological station Schleiden-Schöneseiffen (Meteomedia station) 3 km northeast of the catchment, were available for the period 01/10/2010 – 24/09/2012. After that, precipitation was sampled at half-daily intervals until 30/09/2016 using an automatic, cooled sampler (Eigenbrodt GmbH, Germany). The half-daily samples were precipitation volume-weighed to daily sampling intervals (Stockinger et al., 2016, 2017). Weekly stream water grab samples for stable water isotope analysis were taken at the outlet of the Wüstebach catchment in the 01/10/2010 – 30/09/2016 period (Fig. 3a; Bogena et al., 2020).

Isotope analysis was carried out using laser-based cavity ringdown spectrometers (L2120-i/L2130-i, Picarro Inc.). Internal standards calibrated against VSMOW, Greenland Ice Sheet Precipitation (GISP) and Standard Light Antarctic Precipitation (SLAP2) were used for calibration and to ensure long-term stability of analyses (Brand et al., 2014). The long-term precision of the analytical system was ≤ 0.1 ‰ for $\delta^{18}O$.

4 Methods

To quantify effects of deforestation on $S_{U,max}$ and, due to the role of $S_{U,max}$ as a mixing volume also on the age structure of water as described by TTDs and the associated young water fractions F_{yw}, the following stepwise experiment was designed: (1) quantify changes in the partitioning of annual water fluxes between the pre- and the post-deforestation periods based on observed water balance data; (2) estimate the effect of these changes on the magnitudes of pre- and post-deforestation $S_{U,max}$, respectively, using the same data; (3) calibrate a hydrological model to simultaneously reproduce stream flow and stream $\delta^{18}O$ dynamics for the pre-deforestation period; (4) use the calibrated parameter sets to run the model in the post-deforestation period and evaluate the model's post-deforestation performance without further calibration; (5) re-calibrate the model for the post-deforestation period and evaluate if changes in calibrated $S_{U,max}$ (and other parameters) are plausible and reflect changes in $S_{U,max}$ directly estimated from water balance data in step (2); and finally (6) use the calibrated pre- and post-deforestation parameter sets, respectively, to track modelled water fluxes through the system and quantify changes in TTDs and F_{yw} between the pre- and the post-deforestation periods.

4.1 Water balance-based estimation of $S_{U,max}$

To survive, plants need continuous access to water to satisfy canopy water demand. The root-systems of vegetation are therefore adapted to provide access to water volumes that correspond to annual water deficits that result from the combination of (1) the phase lag between and (2) the difference in the respective magnitudes of seasonal precipitation and solar radiation signals (Donohue et al., 2012; Gentine et al., 2012; Gao et al., 2014). On a daily basis, these water deficits $S_{D,j}(t)$ can be estimated as the cumulative sum of daily ~~effective precipitation~~throughfall P_E [mm d^{-1}] minus transpiration E_T [mm d^{-1}]. The maximum deficit $S_{D,j}$ for a specific year j is then equivalent to the soil water volume that was accessible to and actually accessed by vegetation through its root system for transpiration during the dry season over that period when E_T exceeded P_E (deBoer-Euser et al., 2016; Nijzink et al., 2016a):

$$S_{D,j}(t) = \begin{cases} \int_{t_0}^{t} \left(P_E(t) - E_T(t)\right) dt, & if\ S_{D,j}(t) \leq 0 \\ 0, & if\ S_{D,j}(t) > 0 \end{cases}$$

(Eq. 1)

$$S_{D,j} = max\left(\left|S_{D,j}(t)\right|\right)$$

(Eq. 2)

where t is the time step [d], and t_0 is the last preceding time step for which the storage deficit $S_{D,j}(t) = 0$. As an approximation, Equation 1 implies that if $S_{D,j}(t) = 0$, the water content in the root-accessible pore space at day t is at field capacity and cannot hold additional water. If water supply then exceeds canopy water demand on that day, i.e. $P_E(t) - E_T(t) > 0$, this water surplus is drained from the root zone, e.g. to recharge groundwater or directly to the stream, and cannot be used for transpiration.
Daily ~~effective precipitation~~throughfall P_E, i.e. precipitation that actually reaches the soil, was estimated on basis of the water balance of a canopy interception storage (Nijzink et al., 2016a):

$$\frac{dS_I(t)}{dt} = P(t) - E_I(t) - P_E(t)$$

(Eq. 3)

Where E_I [mm d^{-1}] is daily interception evaporation and S_I [mm] the canopy interception storage. For each time step, E_I can then be computed as:

$$E_I(t) = \begin{cases} E_P(t), & if\ E_p(t)dt < S_I(t) \\ \frac{S_I(t)}{dt}, & if\ E_P(t)dt \geq S_I(t) \end{cases}$$

(Eq. 4)

This then further allows to estimate P_E according to:

$$P_E(t) = \begin{cases} 0, & if\ S_I(t) < I_{max} \\ \frac{S_I(t) - I_{max}}{dt}, & if\ S_I(t) \geq I_{max} \end{cases}$$

(Eq. 5)

where I_{max} [mm] is the canopy interception capacity. In the absence of more detailed information P_E was estimated with a range of different interception capacities, i.e. I_{max} = 0, 1, 2, 3, and 4 mm, in a sensitivity analysis approach.

Note that the catchment average P_E after deforestation was estimated as the areal weighted mean of P_E in the deforested area (21% of catchment area) computed with an assumed I_{max} = 0 mm and P_E from the remaining area computed based on the above range of I_{max} between 0 and 4 mm. In a next step, assuming negligible groundwater imports or exports (cf. Bouaziz et al., 2018), data errors and storage changes, long-term mean transpiration $\overline{E_T}$ was estimated according to the water balance:

$$\overline{E_T} = \overline{P_E} - \overline{Q} \tag{Eq. 6}$$

Where $\overline{P_E}$ [mm d^{-1}] is the long-term mean ~~effective precipitation~~throughfall and $\overline{Q}$ [mm d^{-1}] is the long-term mean observed stream discharge. Daily transpiration E_T [mm d^{-1}] for use in Eq. (1) is then estimated by scaling the long-term mean transpiration to the signal of daily potential evaporation to approximate the seasonal fluctuation of energy input (Bouaziz et al., 2020):

$$E_T(t) = \left(E_P(t) - E_I(t)\right)\frac{\overline{E_T}}{\overline{E_P} - \overline{E_I}} \tag{Eq. 7}$$

A range of previous studies provided evidence that mature forests develop root-systems that allow access to sufficiently large pore water storage volumes $S_{U,max}$ to bridge droughts with return periods T_R ~ 40 years (Gao et al., 2014; deBoer-Euser et al., 2016; Nijzink et al., 2016a; Wang-Erlandsson et al., 2016). The maximum annual water deficits $S_{D,j}$ (Eq. 2) for all j years in the pre-deforestation study period were therefore used to fit a Gumbel extreme value distribution (Gumbel, 1941). This subsequently allowed the estimation of a water deficit with a 40-year return period, which is for this study defined as vegetation-accessible water storage $S_{U,max}$ so that $S_{U,max} = S_{D,40yr}$.

Note that due to the limited length of the data series the $S_{U,max}$ estimates are rather uncertain and need to be understood as merely indicative approximations. This is in particular true for the post-deforestation period, where attempts to explicitly link $S_{U,max}$ to a specific return period are subject to additional uncertainty: as the catchment was not reforested and natural recovery of vegetation is negligible (see aerial images in Figure 1), it is not implausible to assume that the development of the root-system after the disturbance is far from equilibrium and likely to be actively evolving over time. Also note that although E_T is, for brevity, referred to as transpiration throughout this manuscript, it also contains soil evaporation. However, no explicit and quantitative distinction could be made between these two fluxes with the available data. A further critical assumption of the above method required that roots do not tap the groundwater and that water for transpiration is exclusively extracted from the unsaturated soil. In contrast to other landscapes (Fan et al., 2017; Roebroek et al., 2020), it is likely that this assumption largely holds in the Wüstebach as throughout the catchment the groundwater levels, also in the riparian zone, remains largely below a depth of 50 cm during the relatively dry growing season (Bogena et al., 2015) when storage deficits S_D typically accumulate (~ May to October) and no roots have so far been observed for the dominant *picea* species below that depth in the Wüstebach catchment. This is also broadly consistent with the results of Evaristo and McDonnell (2017), who show rather limited groundwater use by *picea* species.

4.2 Model architecture

A semi-distributed, process-based catchment model, iteratively customized and tested within the previously developed DYNAMITE modular modelling framework (Hrachowitz et al., 2014; Fovet et al., 2015), was adapted with additional, hydrologically passive storage volumes to allow for simultaneous representation of water fluxes and tracer transport (Hrachowitz et al., 2013) based on the general concept of storage-age selection functions (SAS; Rinaldo et al., 2015). This model type was chosen over simpler, more data-based methods (e.g. McGuire and McDonnell, 2006; Kirchner, 2016) as it did not only allow a simultaneous representation of water and tracer fluxes but also allowed to attribute observed pattern to specific process hypotheses and the associated model parameters that represent (subsurface) system properties, thereby providing potential quantitative mechanistic explanations of why deforestation affects the hydrology in the Wüstebach. As an intermediate model type between purely data-driven (e.g. Kirchner, 2016) and spatially explicit physically-based models (e.g. Maxwell et al., 2016), it requires assumptions on underlying processes and effective parameters and does not allow a detailed spatial analysis. Yet this model type provides the possibility to test these process hypotheses at the scale of the semi-distributed model units thereby integrating and accounting for the natural heterogeneity of system properties across the model domain (Hrachowitz and Clark, 2017).

4.2.1 Hydrological model

The model domain of the Wüstebach catchment was spatially discretized into two functionally distinct response units, i.e. hillslopes and riparian areas. These are represented in the model as two parallel suites of storage components, linked by a common groundwater body as shown in Figure 4 (e.g. Euser et al., 2015; Nijzink et al., 2016b). According to elevation data and distribution of soil types (Fig.1), 90% of the catchment area was classified as hillslope and the remaining 10% as riparian area. Below a threshold temperature T_T [°C] precipitation P [mm d^{-1}] accumulates as snow P_S [mm d^{-1}] in S_{Snow} [mm]. Above that temperature precipitation is falling as rain P_R [mm d^{-1}] and snow melt P_M [mm d^{-1}] is released from S_{Snow} according to a melt factor F_M [mm d^{-1} °C^{-1}] using a simple degree-day method (e.g. Arsenault et al., 2015; Ala-aho et al., 2017; Gao et al., 2017). The total liquid water input $P_R + P_M$ [mm d^{-1}] entering the hillslope is routed through the canopy interception storage $S_{I,H}$ [mm]. Water that is not evaporated as $E_{I,H}$ [mm d^{-1}] enters the unsaturated root-zone $S_{U,H}$ [mm], whose storage capacity is defined by the calibration parameter $S_{U,max,H}$ [mm]. Water can be released from $S_{U,H}$ as combined root-zone transpiration and soil evaporation flux $E_{T,H}$ [mm d^{-1}] or eventually recharge the groundwater $S_{S,a}$ [mm] over a fast, preferential recharge pathway as $R_{F,H}$ [mm d^{-1}] and a slower percolation flux $R_{S,H}$ [mm d^{-1}]. Similarly, water entering the riparian zone, i.e. $P_R + P_M$ [mm d-

[1]], is routed through $S_{I,R}$ [mm]. Excess water $P_{E,R}$ [mm d^{-1}] that is not evaporated infiltrates into the unsaturated root-zone $S_{U,R}$ [mm], defined by calibration parameter $S_{U,max,R}$ [mm]. In addition, a fraction of the upwelling groundwater $R_{S,R}$ [mm d^{-1}] replenishes $S_{U,R}$ and thus, in addition to precipitation, sustains soil moisture levels in the riparian zone (e.g. Hulsman et al., 2021a), while the remainder Q_S [mm d^{-1}] drains directly into the stream. While water stored in $S_{U,R}$ is available for transpiration (and soil evaporation) $E_{T,R}$ [mm d^{-1}], water that cannot be held is released as $R_{F,R}$ [mm d^{-1}] to a fast responding reservoir $S_{F,R}$ [mm] from where it reaches the stream as Q_R [mm d^{-1}]. The relevant model equations can be found in Table 1.

4.2.2 Tracer transport model

The $\delta^{18}O$ composition of water fluxes and storages was tracked through the model using the storage age selection approach (SAS; Rinaldo et al., 2015), which allows a catchment-scale description of conservative transport based on time-variant travel time distributions. The method builds on the fact that a water volume S [mm] stored in any storage component can, at any moment t [d], consist of parcels of water of different age T [d]. The composition of ages in the stored volume at t depends on the history of water inflows and outflows. Consequently, it evolves over time as new inputs enter into and outflows are released from the storage component, whereby each inflow I [mm d^{-1}] and outflow volume O [mm d^{-1}] can have a different age composition. A convenient way to implement the SAS approach is the use of age-ranked storage $S_T(T,t)$ [mm], which represents, “at any time t the cumulative volumes of water in a storage component as ranked by their age T” (Benettin et al., 2017). Similarly, decomposing each inflow and outflow of a storage component into their respective cumulative, age-ranked volumes $I_T(T,t)$ *and* $O_T(T,t)$ [mm d^{-1}], respectively, then allows to update the age-ranked storage $S_T(T,t)$ at each time step according to the general water age balance (Botter et al., 2011; van der Velde et al., 2012; Benettin et al., 2015a, 2017; Harman, 2015):

$$\frac{\partial S_{T,j}(T,t)}{\partial t} + \frac{\partial S_{T,j}(T,t)}{\partial T} = \sum_{n=1}^{N} I_{T,n,j}(T,t) - \sum_{m=1}^{M} O_{T,m,j}(T,t) \tag{Eq.36}$$

where the term $\partial S_T/\partial T$ represents the aging of water in storage. Reflecting the slightly more abstract approach by Rodriguez and Klaus (2019) and similar to previous studies based on the functionally equivalent mixing coefficient approach (e.g. Fenicia et al., 2010; McMillan et al., 2012; Birkel and Soulsby, 2016; Hrachowitz et al., 2015), the water age balance is here individually formulated for each storage reservoir j (e.g. $S_{I,H}$, $S_{U,H}$, etc.), which each can have varying numbers N and M of inflows I (e.g. P_R, P_M, $R_{S,H}$, etc.) and outflows O (e.g. P_M, $R_{S,H}$, Q_S, etc.), respectively (see Figure 4). It is assumed that the entire volume of a precipitation signal P(t) entering the system at t has an age T of zero so that the associated $I_{T,P,j}(T,t) = P_T(T,t) = P(t)$ for all T. As all other inflows to any following storage component in the system are outflows of storage components prior in the sequence (see Figure 4), the corresponding $I_{T,n,j}(T,t)$ entering a storage component are identical to the $O_{T,m,j}(T,t)$ released from the storage component above.

Each age-ranked outflow $O_{T,m,j}(T,t)$ of a specific storage component j depends on the outflow volume $O_{m,j}$(t) along this outflow pathway and the cumulative age distribution $P_{o,m,j}(T,t)$ of that outflow:

$$O_{T,m,j}(T,t) = O_{m,j}(t)P_{O,m,j}(T,t)$$

(Eq.37)

The outflow volume $O_{m,j}$(t) is estimated via the hydrological model (see Section 4.2.1; Figure 4) and thus assumed to be known. In contrast, the cumulative age-distribution $P_{o,m,j}(T,t)$ can in general not be directly parametrized, as it depends on the temporally varying age distribution of water in the storage component j represented by $S_{T,j}(T,t)$ and thus on the history of past inflows and outflows (Botter et al., 2011; Harman, 2015). Instead, it is possible to define a SAS function $\omega_{o,m,j}$ (or $\Omega_{o,m,j}$ in its cumulative form) for each outflow m from each storage component j that describes how outflow is sampled (or selected) from the temporally varying water volumes of different age present in the age-ranked storage $S_{T,j}(T,t)$ at any time t:

$$P_{O,m,j}(T,t) = \Omega_{O,m,j}\left(S_{T,j}(T,t),t\right)$$

(Eq.38)

From the cumulative age-distribution $P_{o,m,j}(T,t)$ the associated probability density function, which represents the outflow age distribution $p_{o,m,j}(T,t)$, frequently also referred to as backward travel time distribution of that outflow (TTD; e.g. Benettin et al., 2015a; Wilusz et al., 2017), can be obtained according to:

$$p_{O,m,j}(T,t) = \varpi_{O,m,j}\left(S_{T,j}(T,t),t\right)\frac{\partial S_{T,j}}{\partial T}$$

(Eq.39)

Note that conservation of mass requires that any SAS function $\omega_{O,m,j}$ integrates to the total storage volume $S_j(t)$ present in j at any time t. To avoid the resulting need for rescaling $\omega_{O,m,j}$ at each time step, it is helpful to normalize the age-ranked storage to $S_{T,norm,j}(T,t) = S_{T,j}(T,t)/S_j(t)$ so that it remains bounded to the interval [0,1] and defines a residence time distribution (RTD). For this study beta distributions, which are conveniently bound between the limits [0,1] and defined by two shape parameters α and β, were used as SAS functions $\omega_{o,m,j}$ to sample water of different age for outflows from storage components. The parameters β were fixed at a value of 1 for all SAS functions $\omega_{o,m,j}$ used here. However, there is substantial evidence for preferential flow through macropores in the shallow subsurface (e.g. Weiler and Naef, 2003; Zehe et al., 2006, 2007; Weiler and McDonnell, 2007; Beven, 2010; Beven and Germann, 2013; Klaus et al., 2013; Angermann et al., 2017; Loritz et al., 2017). Such preferential flow can, with increasing wetness, increasingly bypass water volumes stored in small pores with little exchange (Sprenger et al., 2016, 2018, 2019a; Cain et al., 2019; Evaristo et al., 2019; Knighton et al., 2019). This then leads to an increasing preferential release of younger water as the system becomes wetter (Brooks et al., 2010). To mimic this, the shape parameters α of the preferential fluxes $R_{F,H}$ and $R_{F,R}$ released from the two unsaturated root-zone storage components S_j

$= S_{U,H}$ and $S_{U,R}$ (Figure 4), were allowed to vary as a function of the water volumes stored in $S_{U,H}$ and $S_{U,R}$, respectively (Hrachowitz et al., 2013; van der Velde et al., 2015):

$$\alpha_{m,j}(t) = 1 - \left(\frac{S_j(t)}{S_{U,max,j}}(1-\alpha_0)\right)$$

(Eq.40)

Where α_0 is a calibration parameter representing a lower bound so that $\alpha_{m,j}(t)$ can vary between α_0 and 1. A value of $\alpha_{m,j} = 1$ indicates complete mixing in dry conditions. Any value below that entails incomplete mixing and thus increases the preference towards releasing younger water in wet conditions (Benettin et al., 2017). Although there is evidence for the presence of preferential flow in other components of the system, such as in the groundwater (e.g. Berkowitz and Zehe, 2020), initial model testing suggested that the inclusion of the additional calibration parameters is not warranted by the available data. For simplicity and following the principle of model parsimony we assumed complete mixing for all other outflows from all other storage components (Figure 4; cf. Fenicia et al., 2010; Kuppel et al., 2018a; Rodriguez et al., 2018). Parameter α was therefore fixed to value of 1 for these SAS functions.

The $\delta^{18}O$ precipitation input signals are damped to the level of fluctuation observed in the stream by subsurface storage volumes that remain to some extent hydrologically passive (e.g. Birkel et al., 2011b). While the hydrologically active storage volumes are represented by the individual storage components of the model (Figure 4; Equations 8-14), an additional hydrologically passive storage volume $S_{S,p}$ [mm] was added as a calibration parameter to the active groundwater storage $S_{S,a}$. (Zuber, 1986; Hrachowitz et al., 2015, 2016), so that $S_{S,tot} = S_{S,a} + S_{S,p}$ (Figure 4). While $dS_{S,p}/dt = 0$, the age-ranked groundwater storage was computed as $S_{T,Ss,tot}$ and the outflows from the groundwater component consequently thus sampled from the entire storage volume $S_{S,tot}$, thereby representing the combined contributions from $S_{S,a}$ and $S_{S,p}$ to the age structure of the outflow Q_S according to Eq. 39. Note that the effects of the hydrologically passive water volume stored in the unsaturated soil below the wilting point are assumed to be negligible due to the small size of that storage volume and the low diffusive exchange rates with the hydrologically active storage volume in the unsaturated zone.

Each individual volume with different age in $I_{T,n,j}(T,t)$ and, as a consequence, also in $S_{T,j}(T,t)$ is also characterized by a different tracer concentration $C_{I,n,j}(I_{T,n,j}(T,t),t)$ and $C_{S,j}(S_{T,j}(T,t),t)$, respectively. For a conservative tracer such as $\delta^{18}O$ that is not significantly affected by decay, evapoconcentration, retention or any other biogeochemical transformation (e.g. Bertuzzo et al., 2013; Benettin et al., 2015b; Hrachowitz et al., 2015) the concentration $C_{O,m,j}(t)$ in any outflow at any time t can then be obtained from:

$$C_{O,m,j}(t) = \int_0^{S_j} C_{S,j}\big(S_{T,j}(T,t),t\big)\varpi_{O,m,j}\big(S_{T,j}(T,t),t\big)dS_T$$

(Eq.41)

Due to data availability, age tracking was here limited to 4 years in the pre- and 3 years in the post-deforestation period. For age beyond that it can only be said that water is older than these 4 and 3 years, respectively. The TTDs reported hereafter are thus truncated at these ages. The model generates TTDs for all fluxes and storage components (Figure 4) for each time step.

As a summary metric, we will here use the fraction of young water F_{yw} as robust descriptor of the left tail of TTDs. Following the definition of Kirchner (2016), F_{yw} is here the fraction of water that is younger than 3 months, which can be extracted directly from any TTD generated by the model. Note, we here only analyse water ages in stream flow as these are the only ones that are directly constrained by available data, while for all other model components, such as transpiration E_T, such direct data support was not available, and the resulting age estimates may thus be characterized by considerable additional uncertainty.

4.3 Model calibration and post-calibration evaluation

The model was run with a daily time step and has a total of 14 free calibration parameters, which were calibrated for the model to simultaneously reproduce flow and $\delta^{18}O$ dynamics in the stream. The uniform prior parameter distributions (Table 2) were sampled using a Monte Carlo approach with $3*10^6$ realizations. To limit equifinality (Beven, 2006) and to ensure robust posterior parameter distributions for a meaningful process representation (e.g. Kuppel et al., 2018b), an extensive multi-objective calibration strategy was applied. Briefly, this was done using a total of 14 performance metrics that describe the model's skill to reproduce different signatures associated to streamflow (E_Q) and $\delta^{18}O$ dynamics ($E_{\delta^{18}O}$) as shown in Table 3. To be accepted as feasible, solutions had to exceed a threshold value of 0.5 for all performance metrics, with the exception of $E_{NS,\delta18O}$ for which a threshold of 0.2 was used. To further constrain the model, we only accepted solutions that could reproduce the dynamics in observed soil moisture as well as the average observed magnitudes of canopy throughfall. To do so we used a simplified limits-of-acceptability approach (e.g. Coxon et al., 2014) with a rectangular step function so that all solutions that fall within the limits of the step function receive a weight of one while all others are assigned a weight of zero and thus rejected (Bouaziz et al., 2021). More specifically, we rejected solutions whose modelled normalized relative soil moisture fell outside the acceptable limits, here defined as ± 0.15 of the observed relative soil moisture, in more than 75% of the time steps in the calibration periods. Similarly, we rejected solutions for which the modelled mean ratio P_E/P in the continuously forested part of the catchment was outside ± 0.15 of the observed mean ratio $P_E/P = 0.71$. This strategy was chosen instead of directly calibrating the time-series of associated model variables S_U (Eq.20) and P_E (Eq.18) to explicitly account for commensurability errors between the point-scale and the scale of the model application (Bouaziz et al., 2021). Subsequently, the 14 metrics of the solutions retained as feasible were combined into two equally weighted classes, describing stream flow (Q) and tracer ($\delta^{18}O$) dynamics, respectively. This then allowed to obtain solutions with balanced overall model performances ~~were then obtained~~ using the mean Euclidean Distance D_E [-] from the "perfect" model (i.e. $D_E = 1$; Hrachowitz et al., 2014; Hulsman et al., 2020):

$$D_E = 1 - \sqrt{\frac{1}{2}\left(\frac{\sum_{n=1}^{N}\left(1 - E_{Q,n}\right)^2}{N} + \frac{\sum_{m=1}^{M}\left(1 - E_{\delta^{18}O,m}\right)^2}{M}\right)} \quad \text{(Eq.42)}$$

Where N = 12 is the number of different performance metrics describing streamflow and M = 2 the number of different performance metrics for δ^{18}O. To construct the posterior parameter distributions and the corresponding model uncertainty intervals, the retained parameter sets where then weighted according to a likelihood measure $L = D_E^p$ (cf. Freer et al., 1996), where the exponent p was set to a value of 10 to emphasize models with good overall calibration performance.

In a first step, the model was calibrated for the pre-deforestation period 01/10/2009 – 31/08/2013. Note that due to a lack of regular and weekly δ^{18}O precipitation data before 01/10/2010, the performance metric $E_{\delta^{18}O}$ describing the δ^{18}O dynamics was computed from that date onwards only. The feasible parameter sets were then used to test the model without further calibration in the post-deforestation period. In a second step, the model was re-calibrated for the 01/09/2013 – 30/09/2016 post-deforestation period and the changes in the resulting model performance and posterior distributions compared to those from the pre-deforestation calibration. The estimation of the effects of deforestation on TTDs is based on model parameter sets obtained from calibration in the pre-deforestation and post-deforestation periods, respectively.

5 Results ~~and Discussion~~

5.1 Observed ~~D~~deforestation effects on the hydrological system

Initial analysis of water balance data suggests that the hydro-meteorological conditions as expressed by the aridity index $I_A = \overline{E_P}/\overline{P}$, do not show significant differences between the pre-deforestation (I_A = 0.50 ± 0.02) and the post-deforestation periods (I_A = 0.51 ± 0.03), respectively (Figure 5a). However, and in spite of these comparable climatic conditions, the results show a shift in the partitioning of water fluxes between runoff Q and actual evaporation E_A (note that $E_A = E_I + E_T$). While the fraction of precipitation that was released into the atmosphere as vapour was reduced ($\overline{E_A}/\overline{P}$; Figure 5a), the mean runoff ratio ($C_R = 1 - \overline{E_A}/\overline{P}$) increased correspondingly from C_R = 0.55 ± 0.04 to C_R = 0.68 ± 0.03 after deforestation of 21 % of the catchment with p = 0.049 based on a Wilcoxon rank sum test. ~~These results correspond well with the findings of an earlier study in the Wüstebach, based on a shorter study period (2011 – 2015; Wiekenkamp et al., 2016), which estimated an increase of C_R from ~ 0.58 to ~ 0.66 during that period using eddy-covariance measurements.~~ In absolute terms this entails that, notwithstanding rather stable mean annual precipitation P = 1269 ± 24 mm yr^{-1} and potential evaporation E_P = 632 ± 9 mm yr^{-1} over the entire study period, the annual actual evaporation E_A decreased from 576 ± 11 mm yr^{-1} to 401 ± 6 mm yr^{-1} whereas annual runoff Q increased by ~ 25 % from 694 ± 47 mm yr^{-1} to 870 ± 63 mm yr^{-1}.

~~The overall pattern found here also broadly reflect the effects of land cover/use change in many different environments (Creed et al., 2014; Jaramillo and Destouni, 2014; Renner et al., 2014, van der Velde et al., 2014; Moran-Tejada et al., 2015; Nijzink et al., 2016; Zhang et al., 2017; Jaramillo et al., 2018). The vast majority of these studies suggest that forest removal leads to an increase in the runoff ratio C_R at the cost of reduced evaporation E_A, although the magnitudes of these changes do substantially vary between individual catchments and studies, which is consistent with our physical understanding of the~~

~~importance of forest for transpiration in hydrological systems. Under the assumption that reduction of E_A is largely a direct consequence of forest removal in the Wüstebach, a plausible hypothesis to directly attribute this shift in water partitioning from E_A to Q to a physical process can be formulated as follows: the roots of harvested trees stopped extracting water for transpiration from the subsurface. In addition, the decrease of turbulent exchange of vapour with depth effectively limits soil evaporation to the first few centimetres of the soil (e.g. Brutsaert, 2014). Thus, the felling of trees led to a situation where under comparable atmospheric water demand E_P, water volumes held at depths below that and previously within the reach of active roots became largely unavailable for transpiration and evaporation after deforestation. This implies that the water volumes accessible to satisfy atmospheric water demand, i.e. $S_{U,max}$ and I_{max}, are drastically reduced.~~

~~In our study, this becomes evident when comparing the catchment-scale maximum annual storage deficits $S_{D,j}$ (Eq. 2) of the pre- and post-deforestation periods, respectively, which are indicative of differences in soil depths affected by E_A in the two periods.~~ In spite of similar climatic conditions, the above is reflected in a significantly higher (p = 0.047)~~e~~ mean annual maximum storage deficit in the pre-deforestation ~~period is significantly higher (p = 0.047)~~ than in the post-deforestation period. In the pre-deforestation period values between 105 ± 23 mm for I_{max} = 0 mm and 95 ± 21 mm for I_{max} = 4 mm, respectively, were found (Figure 5b). Whereas in the post-deforestation period the mean storage deficit only reached between 49 ± 10 mm and 33 ± 7 mm for the same values of I_{max} (Figure 5b). Note that in both periods, $S_{D,j}$ is relatively insensitive to the magnitude of I_{max} (cf. Gerrits et al., 2009). From the above maximum annual storage deficits $S_{D,j}$, the corresponding catchment-scale vegetation-accessible water storage capacity, assuming vegetation adaptation to dry conditions with 40-year return periods (see Section 4.1), was estimated at values of $S_{U,max}$ = 258 ± 125 mm for the pre-deforestation (R^2 = 0.91, p = 0.04; Figure 5c) and $S_{U,max}$ = 101 ± 149 mm for the post-deforestation period (R^2 = 0.83, p = 0.27; not shown). ~~Directly reflecting reductions of E_A, these estimated reductions in storage deficits are consistent with observed post-deforestation increases in soil moisture (Wiekenkamp et al., 2016). Note, however, that in particular the estimates for the post-deforestation period are characterized by considerable uncertainty and therefore need to be understood as merely indicative as they are inferred from only 3 years of data, and a system that is likely to be far from equilibrium, because the deforested part cannot have adapted yet (e.g. Nijzink et al., 2016; Teuling and Hoek van Dijke, 2020). These considerable uncertainties are also reflected in the surprisingly low post-deforestation $S_{U,max}$. Notwithstanding these limitations, the above results illustrate that here the reduction of transpiration due to deforestation is likely to go hand in hand with a considerably reduction of $S_{U,max}$ and thus the catchment-scale sub-surface pore volume between field capacity and permanent wilting point that is actively accessed by vegetation to satisfy the evaporative demand.~~

5.2 Modelled ~~D~~deforestation effects on the hydrological system

5.2.1 Model calibration for pre-deforestation period

The model parameter sets retained as feasible after calibration in the 2009-2013 pre-deforestation period reproduce the general features of the hydrograph in that period rather well (Figures 2c,d), similar to a previous modelling study (Cornelissen et al., 2014). This is true for both, the timing and magnitudes of high flows, with an associated Nash-Sutcliffe Efficiency $E_{NS,Q}$ = 0.~~79~~ 83 for the best performing model in terms of D_E (Figure 6a) but also for low flows ($E_{NS,log(Q)}$ = 0.~~79~~70), with the exception of some overestimation in summer 2011. The modelled runoff ratio comes with C_R = 0.54 (5/95th IQR: 0.52 – 0.58) very close to the observed runoff ration of C_R = 0.55 ($E_{R,CR}$ = 0.98). In addition, the model could also simultaneously mimic most other observed flow signatures reasonably well (Figure 6a), in particular the flow duration curve ($E_{NS,FDC}$ = 0.79; Figure 6b), the peak distribution ($E_{NS,PD}$ = 0.~~91~~85; Figure 6d)~~,~~ and the auto correlation function ($E_{NS,AC}$ = 0.~~90~~98; Figure 6f) ~~and the runoff ratio ($E_{R,CR}$ = 0.97)~~. The limits-of-acceptability constraints for P_E/P allowed the identification and removal of a few additional parameter sets (~ 5%) that likely overestimate throughfall P_E (Figure 7a). The soil moisture constraint was more effective as it allowed to reject a considerable additional proportion of solutions (> 90%) that did not sufficiently well match the observed soil moisture dynamics according to the pre-defined limits-of-acceptability (Figure 7c). With the parameter sets eventually retained as feasible the modelled temporal dynamics of relative soil moisture broadly reflect the observed ones (Figure 7e) Similarly, the model captures the substantial attenuation of the precipitation $\delta^{18}O$ variability ($E_{R,RD}$ = 0.98; Figure 6a), while at the same time largely preserving the limited but visible low-frequency temporal fluctuations in the stream $\delta^{18}O$ composition (Figures 3a,b). In comparison to the flow performance metrics the Nash-Sutcliffe Efficiency of the $\delta^{18}O$ composition for the best model is somewhat lower ($E_{NS,\delta18O}$ = 0.~~38~~37; Figure 6a), which mostly results from the low variability of such a damped signal, where even very small absolute errors (MAE = 0.11 ‰) and a few scattered outliers can lead to very low Nash-Sutcliffe Efficiencies (cf. Hrachowitz et al., 2009).

The posterior distributions (Table 2, Figure ~~7~~8) show that most model parameters are reasonably well identified. Individually calibrated for their respective landscape class, i.e. hillslope and riparian zone, $S_{U,max,H}$ = ~~246~~ 242 mm (5/95th IQR: ~~233~~ 213 – 3~~09~~11 mm) and $S_{U,max,R}$ = ~~234~~ 213 mm (~~194~~ 186 – 28~~7~~0 mm) showed similar optimal values and distributions (Figures 7a,b), reflecting the catchment-wide relatively homogenous forest cover in the pre-deforestation period (Figure 1). Remarkably, these calibrated values also come close to ~~the water balance-derived~~ catchment-scale estimates of $S_{U,max}$ = 258 ± 125 mm that were directly derived from water balance data without any calibration, as described in 5.1 (Figure 5c).

5.2.2 Application of pre-deforestation model to post-deforestation period

In a next step, the parameter sets obtained from the above calibration in the pre-deforestation period were used to run the model without further re-calibration in the post-deforestation period. This entails the implicit and clearly wrong assumption that the physical characteristics of the system remained unaffected by deforestation. The consequence of that can be seen in Figures 2c and 2d (red line). While the low flows remain well reproduced, the post-deforestation application of the model substantially and systematically underestimates high flows, partly by 50% or more, such as in November 2013 or August 2014. The inability of the model to reproduce several aspects of post-deforestation high-flow dynamics of the system is also evident in the lower model performance metrics associated with high flows (Figure 6a). Besides the time series of flow ($E_{NS,Q}$ = 0.~~63~~65), notably the model's skill to capture the rising limb density ($E_{NS,PD}$ = 0.~~81~~78), the autocorrelation function ($E_{NS,AC}$ = 0.~~71~~58; Figure 6g) and the runoff ratio ($E_{R,CR}$ = 0.~~73~~81) were negatively affected. In contrast to the pre-deforestation period, the modelled runoff ratio C_R = 0.~~52~~ 55 (0.~~48~~ 54 – 0.~~67~~58) in the post-deforestation period considerably underestimates the observed C_R = 0.68 ± 0.03 (Figure 5a). ~~The above implies that the model also overestimates post-deforestation evaporative fluxes E_A. Therefore, it can, without re-calibration, not deal with the observed changes in the partitioning between drainage and evaporative fluxes (Figure 5a). A likely explanation for the pattern produced by the model is that, in contrast to the real world, no reduction in E_A due to the reduced forest cover is achieved because the model still relies on the catchment-scale vegetation-accessible storage volume $S_{U,max}$ that characterizes the extent of the catchment-scale active root system before deforestation. This $S_{U,max}$ falsely provides sufficient water supply to sustain E_A at high levels comparatively close to E_P throughout the year (see red line in Figure 2b), although, in the parts of the catchment where trees were removed, water stored at depths below a few centimetres is not available for significant evaporation anymore. Such an overestimation of $S_{U,max}$ implies also that in the model a more pronounced water storage deficit can and does develop throughout dry periods. The model therefore assumes that soils dry out to deeper depths. Consequently, to establish connectivity and to eventually generate flow during and after rainstorms, more water needs to be stored in the model than in the real world system to overcome this deficit. This water is then in the model held against gravity and thus only available for evaporation but not for drainage. Although it is reasonable to assume that groundwater recharge is affected in a similar way, the model can better reproduce low flows. The reason for this is that the draining groundwater body, which sustains summer low flows, is, due to limited recharge during these drier periods, largely disconnected from and thus largely unaffected by subsurface – vegetation interaction in shallower parts of the subsurface. In the parts of the catchment where trees were removed a similar reasoning also holds for the interception capacity I_{max} and the associated likely overestimation of interception evaporation E_I, yet, due to the smaller magnitude of I_{max}, to a lesser extent than for $S_{U,max}$.~~ The ~~above described~~ problems to describe ~~for~~ the high flow periods are accompanied by the model's reduced ~~in~~ability to describe the post-deforestation $\delta^{18}O$ dynamics in stream water ($E_{NS,\delta 18O}$ = 0.11), although the observed general degree of damping of the $\delta^{18}O$ signal ($E_{R,RD}$ = 0.98) remains well reproduced as shown in Figures 3 and 6a. While the low $E_{NS,\delta 18O}$ values are partly an effect of the above explained low signal-to-noise ratio of such a damped signal and thus of the chosen performance metric, the model also struggles to adequately reproduce the lower-frequency fluctuations, such as between February and July 2014, when the model indicated rather stable $\delta^{18}O$ values while the observed values show a slight yet clear increasing trend over the same period (Figure 3b). Together with the lower overall model performance metric D_E (Figure 6a), these results illustrate that the pre-deforestation model parameter sets provide an unsuitable characterization of the system characteristics in the post-deforestation period.

5.2.3 Recalibrate model for post-deforestation period

To estimate the effect of forest removal on the characteristics of the hydrological system and thus on the model parameters, the model was in a next step recalibrated for the post-deforestation period. This led to a slight improvement of the overall model performance from D_E = 0.~~22~~ 77 to 0.~~58~~ 80 (Figure 6a). Most notably, it can be observed that the recalibrated model can much better reproduce the increased high flows in that period (Figures 2c,d), as reflected by improvements in the performance metrics associated with high flows (Figure 6a), but most notably $E_{NS,Q}$ = 0.~~69~~70, $E_{NS,FDC}$ = 0.~~93~~ 95 (Figure 6c) or $E_{NS,AC}$ = 0.~~87~~

92 (Figure 6g). Similarly, the limits-of-acceptability constraints ensured a choice of solutions that broadly reflect the observed throughfall ratios P_E/P (Figure 7b) as well as the observed soil moisture dynamics (Figures 7d, f). In addition and perhaps most importantly, the runoff ratio also increased and was with a modelled value of C_R = 0.~~58~~ 62 (0.56 – 0.~~61~~63) closer to the observed C_R = 0.68 ($E_{R,CR}$ = 0.~~84~~91). This further implies that, in contrast to the initial model, the recalibrated model also features expected reductions of evaporative fluxes E_A by about 10%, which can be seen in Figure 2b. ~~In addition, analysis of the modelled fluxes indicates that a higher proportion of flows, mostly during wet-up periods, is rapidly released from the root-zones as fluxes $R_{F,H}$ and $R_{F,R}$ (Figure 4; Table 1), representing preferential flows. Such a post-deforestation increase in preferential flow occurrence is supported by observations recently reported by Wiekenkamp et al. (2020).~~ Mirroring the improvements in the reproduction of flows, recalibration also allowed the model to better capture the stream water $\delta^{18}O$ dynamics ($E_{NS,\delta18O}$ = 0.24; MAE = 0.10 ‰; Figure 6a). While there is little change in the model's ability to mimic the general level of damping of the $\delta^{18}O$ signal ($E_{R,RD}$ = 0.99) and its low-frequency fluctuations, the more pronounced, albeit in absolute terms still small, high-frequency fluctuations, as short-term response to individual storms are better described (Figures 3a,b).

~~It is of course unsurprising that recalibration leads to an improved model performance in the post-deforestation period. Without further analysis, such a mere model fitting exercise allows in the presence of model equifinality only little insight into the underlying processes (Beven, 2006; Kirchner, 2006). To gain more confidence that the improvements in the recalibrated model are at least partly due to the right reasons (Kirchner, 2006), the changes in the posterior parameter distributions resulting from the two calibration runs were thus analysed. It was hypothesized above that reductions in evaporative fluxes are directly linked to reduced water volumes accessible and available for evaporation and transpiration at the catchment-scale. In the theoretical ideal case, the representations of the associated storage capacities in the model, i.e. the parameters $S_{U,max}$ and I_{max}, should thus be the only ones to significantly change after deforestation. However, note that this is unlikely for two reasons. First, while it is plausible to assume that these storage capacities are significantly affected by forest removal, it is not unlikely that other system characteristics and their mutual interactions, thus far unknown and not considered, are similarly influenced, potentially causing considerable ontological uncertainty. Second, model parameter interactions that arise as artefacts to compensate overly simplistic process representations and/or data uncertainty are also likely to affect parameters seemingly unrelated to deforestation.~~

Inspection of the posterior parameter distributions reveals that the catchment-scale $S_{U,max}$ experienced considerable reductions after recalibration. While in the hillslope parts of the catchment, which were less affected by deforestation (~ 10% of the hillslope area; Figure 1) an average decrease by ~ ~~75~~ 50 mm to $S_{U,max,H}$ = ~~137~~ 212 mm (~~118~~ 137 – ~~249~~ 270 mm) can be seen (Figure ~~7a~~8a), the completely deforested riparian area exhibits an average decrease by ~ ~~130~~ 100 mm to $S_{U,max,R}$ = ~~67~~ 93 mm (~~53~~ 92 – 1~~26~~90 mm; Figure ~~7b~~8b). As an indicative value, the area-weighted catchment-average $S_{U,max}$ = 120 mm of the best performing parameter set falls into the plausible range of $S_{U,max}$ = 101 ± 149 mm as described in Section 5.1. While there is little evidence for reductions of I_{max} on the less deforested hillslopes (Figure ~~7d~~8d), a clear decrease of interception capacities by on average ~ 2mm to $I_{max,R}$ = 1.1 mm (0.1-1.3 mm; Figure ~~7e~~8e) can be observed in the fully deforested riparian zone. ~~Comparing to the posterior distributions of other parameters, the results illustrate that the storage parameters $S_{U,max}$ and I_{max} of~~

~~the completely deforested riparian zone, and to a lesser extent of the hillslope, were subject to the most pronounced changes. In contrast, for most other parameters, the pre- and post-deforestation posterior distributions exhibit a closer resemblance (Figure 7). Yet, it can also be observed that the individual parameter values associated with the best model solutions in the pre- and post-deforestation periods, respectively, do vary to a stronger degree for most parameters. Notwithstanding the distinct overall effects of forest removal on the individual posterior distributions, this clearly highlights the influence of parameter compensation effects and related uncertainties. This is also illustrated by a few parameters, such as~~ ~~$R_{S,max}$~~ ~~(Figure 7c, Equation 22), that remain poorly constrained. Note that in spite of uncertainties introduced by the associated compensation effects, in particular~~ ~~$S_{U,max}$~~ ~~remains rather well constrained. However, after preliminary unsuccessful testing, no further attempts were made to re-calibrate only the above discussed four storage parameters, i.e.~~ ~~$S_{U,max,H}$, $S_{U,max,R}$, $I_{max,H}$~~ ~~and~~ ~~$I_{max,R}$,~~ ~~acknowledging the limitations introduced by parameter compensation effects.~~

~~Overall the results suggest that the model formulation together with the multi-objective calibration strategy ensured the identification of solutions that provide a robust description of the system and allow a simultaneous representation of flow and isotope dynamics in the stream. There are indications that at least some processes and parameters can be directly linked to real world quantities. In particular, the results provide supporting evidence that the parameters~~ ~~$S_{U,max,H}$~~ ~~and~~ ~~$S_{U,max,R}$~~ ~~are not merely abstract quantities, but that it is not implausible to assume that they, taken together, provide a catchment-scale representation of vegetation-accessible and -accessed water volumes as defined by Equation 2.~~

5.3 Deforestation effects on travel time distributions, SAS-functions and young water fractions

While the volume weighted mean $\delta^{18}O$ compositions of observed precipitation with -7.9 ‰ and stream water with -8.2 ‰ are comparable, a substantial difference in their fluctuations, with standard deviations of 3.6 ‰ and 0.2 ‰, respectively, is evident (Figures 3a,b). This difference suggests a remarkably elevated degree of damping rarely found elsewhere (e.g. Speed et al., 2010), indicative of the importance of old water contributions to the stream in the study catchment. No significant difference in damping ratios was observed between the pre- and post-deforestation period, which further corroborates the prevalence of old water.

Tracking the $\delta^{18}O$ signals through the model then allowed to estimate travel time distributions (TTD). Note that any results reported hereafter are necessarily conditional on the assumptions made in and the uncertainties arising from the modelling process.

In general and consistent with the observed high degree of damping, it was found that pre-deforestation the system was characterized by rather old water. The range of truncated TTDs of stream water exhibits considerable variability in response to changing wetness conditions with on average about ~~24~~ 27 % of the discharge younger than 3 years (Figure ~~8b~~9b,c). In spite of the low mean F_{yw} ~ 0.12 (Figure 10a), ~~S~~stream water can contain up to ~~30~~ 34 % water younger than 3 months (i.e. F_{yw} ~ 0.~~30~~34) for individual storm events in the wet period, while frequently dropping to < 1 % during elongated summer dry periods (Figures 8c, 9a), similar to what has been reported elsewhere (e.g. Gallart et al., 2020b). It can also be observed that the age composition of stream water (Figure ~~8c~~9c) and the associated F_{yw} (Figure ~~9a~~10a) do considerably vary throughout wet periods.

Dry periods are characterized by considerably less variability and more stable stream water TTDs.~~ This is a consequence of increased bypass flow that has little interaction with resident water as the system gets wetter and which may reach the stream over preferential flow paths and increased contributions from the riparian zone with its shorter flow paths. In other words, in a wet system where little additional water can be stored, the precipitation volumes of individual storm events control the shape of TTDs (Heidbüchel et al, 2020). In the summer dry season, however, precipitation is to a higher degree buffered in the root-zone and used for transpiration (Stockinger et al., 2014). Conversely, stream flow is then mostly sustained by groundwater which is characterized by large volumes of older water. This effectively attenuates fluctuations by the proportionally much lower volumes of younger precipitation water that cannot be stored and is thus quickly released to the stream.~~ This is further corroborated by the significantly higher sensitivity of F_{yw} to changes in stream flow in wet-up and wet periods (dF_{yw}/dQ_n ~ 0.35 and 0.25, respectively) as compared to dry periods (dF_{yw}/dQ_n ~ 0.05; Figure ~~9c~~10c). In spite of the low mean F_{yw} ~ 0.~~11~~ 12 (Figure ~~9a~~10a), the above also entails that very fast switches towards higher young water fractions can be observed when the system is wetting up after dry periods as well as for storm events throughout the wet season. In general, the above observations are also encapsulated in the catchment-overall storage age selection functions ω, that represent the ratio of stream water TTD over the combined RTD of all model storage elements (Benettin et al., 2015a). While for dry periods under-sampling of young water ages with relatively little variability is evident, it can also be seen that in particular during wet-up and wet periods a considerable, yet highly variable preference for very young water can be seen (Figure ~~10a~~11a), similar to what has been reported previously in other environments (e.g. Benettin et al, 2015a; Remondi et al., 2018).

The overall picture did not change in the post-deforestation period.~~,~~ Similar to the pre-deforestation period, the TTDs can exhibit considerable variability. However, in contrast to the pre-deforestation period and depending on the wetness conditions, considerable shifts towards younger water can be observed for the TTDs (Figure ~~8d~~9d-g). ~~While individual summer storms led to increases of almost exclusively very young water <10 – 20 days in the stream~~ There are little discernible changes in F_{yw} during the dry summer months (Figure ~~8d~~9d). However, storms in wet-up periods, mostly during autumn, led to considerable increases in the fractions of water younger than 10 – 20 days (Figure 9e). During wet periods clear shifts towards younger water can be observed throughout the entire spectrum of tracked ages ~~during wet conditions~~ (Figure ~~8f~~9f). During the wet period ~ 36 % of the stream water are on average younger than the tracked three years (Figure ~~8i~~9i). The mean F_{yw} only slightly increased to 0.13 (Figure ~~9b~~10b), compared to 0.~~11~~ 12 in the pre-deforestation period (Figure ~~9a~~10a), which corroborates earlier results by Stockinger et al. (2019) that suggested only minor fluctuations in mean F_{yw} over multiple moving time windows. For individual winter storm events, ~~however,~~ F_{yw} slightly increased to up to ~ 0.~~40~~ 37 (Figures ~~8j~~9j, ~~9b~~10b) compared to F_{yw} of up to ~ 0.~~30~~ 34 in the pre-deforestation period (Figures ~~8c~~9c, ~~9a~~10a). Besides the generally higher F_{yw} during wet periods, the F_{yw} became more sensitive to flow during ~~wet-up and~~ wet conditions, with dF_{yw}/dQ_n ~ 0.~~27~~ 36~~and 0.43, respectively~~ (Figure ~~9d~~10d), similar to what has been previously reported by von Freyberg et al. (2018) and Gallart et al. (2020a). ~~t the end of dry periods and the beginning of the wet period, elsewhere also referred to as "autumn flush" (e.g. Dawson et al., 2011), the switches towards younger water at given flow levels occur considerably faster in the post-deforestation period than in the pre-deforestation period. Therefore, where, at the same discharge, previously relatively little young water reached~~

~~the stream, a much higher fraction of young water can now be observed in the stream. Underlining the role of transpiration (e.g. Douinot et al., 2019; Kuppel et al., 2020), this is a direct effect of the reduced evaporative removal of relatively young near-surface water (Maxwell et al., 2019), which in turn is intimately linked to the reduced water supply for evaporative fluxes, i.e. smaller storage volumes $S_{U,max}$ and I_{max}. This modelled relatively young, surface-near water, not taken up by vegetation anymore is thus to a higher degree flushed from the system mostly via preferential flow paths to the stream (i.e. $R_{F,H}$, $R_{F,R}$) and thus bypassing older resident water with little exchange, which is consistent with recent observations of more frequent activation of preferential flow paths (Wiekenkamp et al., 2020). Once connectivity and the associated higher degree of bypass flow are established in the wet period, the peak sensitivity of dF_{yw}/dQ to flow increased to ~ 0.43, as under these conditions when little additional water can be stored in the shallow subsurface, F_{yw} is largely controlled by magnitude of the individual precipitation signals and to a lesser extent by the footprint of the pre-storm history of evaporative fluxes in the shallow subsurface storage. In contrast, no significant changes could be observed for the sensitivity of F_{yw} to discharge during dry periods, as during that period, the composition of water ages is controlled by large volumes of old water.~~ The above described post-deforestation changes are also manifest in the corresponding storage age selection function ω (Figure ~~10b~~11b) for that period. While the degree of under-sampling of young water during dry periods significantly decreased, a substantially higher preference for young water during wet-up and wet periods can be observed than during the pre-deforestation period, with a clear overall shift towards younger water for all wetness conditions.

6 Discussion

6.1 Observed deforestation effects on the hydrological system

The observed post-deforestation changes to the hydrological response, in particular the increase of C_R from ~ 0.55 to ~ 0.68 correspond well with the findings of an earlier study in the Wüstebach, based on a shorter study period (2011 – 2015; Wiekenkamp et al., 2016), which estimated an increase of C_R from ~ 0.58 to ~ 0.66 during that period using eddy-covariance measurements. The overall pattern found here also broadly reflect the effects of land cover/use change in many different environments (Creed et al., 2014; Jaramillo and Destouni, 2014; Renner et al., 2014, van der Velde et al., 2014; Moran-Tejada et al., 2015; Nijzink et al., 2016; Zhang et al., 2017; Jaramillo et al., 2018). The vast majority of these studies suggest that forest removal leads to an increase in the runoff ratio C_R at the cost of reduced evaporation E_A, although the magnitudes of these changes do substantially vary between individual catchments and studies, which is consistent with our physical understanding of the importance of forest for transpiration in hydrological systems.

Under the assumption that reduction of E_A is largely a direct consequence of forest removal in the Wüstebach, a plausible hypothesis to directly attribute this shift in water partitioning from E_A to Q to a physical process can be formulated as follows: the roots of harvested trees stopped extracting water for transpiration from the subsurface. In addition, the limited turbulent exchange of vapour at depth effectively limits soil evaporation to the first few centimetres of the soil (e.g. Brutsaert, 2014).

Thus, the felling of trees led to a situation where under comparable atmospheric water demand E_P, water volumes held at depths below that and previously within the reach of active roots became largely unavailable for transpiration and evaporation after deforestation. This implies that the water volumes *accessible* to satisfy atmospheric water demand, i.e. $S_{U,max}$ and I_{max}, are drastically reduced. Most notably, the available water balance data suggest that catchment-scale $S_{U,max}$ decreased from pre-deforestation $S_{U,max}$ = 258 ± 125 mm to post-deforestation $S_{U,max}$ = 101 ± 149 mm.

Note, however, that in particular the estimates for the post-deforestation period are characterized by considerable uncertainty and therefore need to be understood as merely indicative as they are inferred from only 3 years of data, and a system that is likely to be far from equilibrium, because the deforested part cannot have adapted yet (e.g. Nijzink et al., 2016; Teuling and Hoek van Dijke, 2020). These considerable uncertainties are also reflected in the surprisingly low post-deforestation $S_{U,max}$.

Notwithstanding these limitations, the above results illustrate that here the reduction of transpiration due to deforestation is likely a direct consequence of the considerable reduction of $S_{U,max}$ and thus the catchment-scale sub-surface pore volume between field capacity and permanent wilting point that can be actively accessed by vegetation to satisfy the evaporative demand. These post-deforestation decreases in transpiration due to reductions in accessible water volumes $S_{U,max}$, further lead to reduced soil water storage deficits $S_{D,j}$ (Eq.2) in dry seasons, which is consistent with observed post-deforestation increases in soil moisture (Wiekenkamp et al., 2016).

6.2 Modelled deforestation effects on the hydrological system

The model application provided further evidence for the central role of $S_{U,max}$ as dominant control on the hydrological response as well as for the direct effects of deforestation on $S_{U,max}$. The model calibration in pre-deforestation period resulted in a set of solutions that could simultaneously reproduce multiple signatures, as expressed by 14 individual performance metrics, while also satisfying two additional limits-of-acceptability constraints. Overall this suggests a rather robust representation of the system.

In a next step, the parameter sets obtained from the calibration in the pre-deforestation period were used to run the model without further re-calibration in the post-deforestation period. This entails the implicit and clearly wrong assumption that the physical characteristics of the system remained unaffected by deforestation. As a consequence, that model exhibited a considerably reduced ability to reproduce the hydrological response in the post-deforestation period, in particular high flows as well as the runoff ratio C_R. The latter implies that the model also overestimates post-deforestation evaporative fluxes E_A. Therefore, it can, without re-calibration, not deal with the observed changes in the partitioning between drainage and evaporative fluxes (Figure 5a). A likely explanation for the pattern produced by the model is that, in contrast to the real world, no reduction in E_A due to the reduced forest cover is achieved because the model still relies on the catchment-scale vegetation-accessible storage volume $S_{U,max}$ that characterizes the extent of the catchment-scale active root-system before deforestation. This $S_{U,max}$ *falsely* provides sufficient water supply to sustain E_A at high levels comparatively close to E_P throughout the year (see red line in Figure 2b), although, in the parts of the catchment where trees were removed, water stored at depths below a few centimetres is not available for significant evaporation anymore in reality. Such an overestimation of $S_{U,max}$ implies also that in the model a more pronounced water storage deficit can and does develop throughout dry periods. The model therefore assumes that soils dry out to deeper depths. Consequently, to establish connectivity and to eventually generate flow during and after rainstorms, more water needs to be stored in the model than in the real world system to overcome this deficit. This water is then in the model held against gravity and thus only available for evaporation but *not* for drainage. Although it is reasonable to assume that groundwater recharge is affected in a similar way, the model can better reproduce low flows. The reason for this is that the draining groundwater body, which sustains summer low flows, is, due to limited recharge during these drier periods, largely disconnected from and thus largely unaffected by subsurface – vegetation interaction in shallower parts of the subsurface. In the parts of the catchment where trees were removed a similar reasoning also holds for the interception capacity I_{max} and the associated likely overestimation of interception evaporation E_I, yet, due to the smaller magnitude of I_{max}, to a lesser extent than for $S_{U,max}$.

Recalibration of the model in the post-deforestation period led to a considerably improved representation of the hydrological response and in particular of the high-flows as well as the runoff ratio C_R. The latter implies that the modelled partitioning of water fluxes and in particular E_A (see orange line in Figure 2b) is more consistent with the observed post-deforestation reductions in E_A. In addition, analysis of the modelled fluxes indicates that a higher proportion of flows, mostly during wet-up periods, is rapidly released from the root-zones as fluxes $R_{F,H}$ and $R_{F,R}$ (Figure 4; Table 1), representing preferential flows. Such a post-deforestation increase in preferential flow occurrence is supported by observations recently reported by Wiekenkamp et al. (2020).

It is of course unsurprising that recalibration leads to an improved model performance in the post-deforestation period. Without further analysis, such a mere model fitting exercise allows in the presence of model equifinality only little insight into the underlying processes (Beven, 2006; Kirchner, 2006). To gain more confidence that the improvements in the recalibrated model are at least partly due to the right reasons (Kirchner, 2006), the changes in the posterior parameter distributions resulting from the two calibration runs were thus analysed. In the pre-deforestation period, the range of the posterior distributions of $S_{U,max,H}$ and $S_{U,max,R}$ (Figure 8a,b) as well as the modelled catchment-average $S_{U,max}$ = 240 mm, estimated as area-weighted average of $S_{U,max,H}$ and $S_{U,max,R}$, come close to the catchment-scale estimate of $S_{U,max}$ = 258 ± 125 mm that was directly derived from water balance data without any calibration (Figure 5c). The modelled post-deforestation reductions of $S_{U,max,H}$ and $S_{U,max,R}$ are evident in the shifts of their respective posterior distributions (Figure 8a,b) and the lower catchment-average $S_{U,max}$ = 199 mm of the best performing parameter set, falling into the plausible range of $S_{U,max}$ = 101 ± 149 mm as estimated from water balance data. In addition and quite remarkably, the re-calibrated model is able to broadly represent the differences in forest removal on the hillslopes and in the riparian zone. While in the fully deforested riparian area $S_{U,max,R}$ decreased by ~ 100 mm (Figure 8b), $S_{U,max,H}$ on the only partly deforested hillslopes decreased by merely ~ 50mm (Figure 8a). Similarly, there is little evidence for reductions of $I_{max,H}$ on the less deforested hillslopes (Figure 8d). Yet, a clear decrease of interception capacities by on average ~ 2mm to $I_{max,R}$ = 1.1 mm (0.1 – 1.3 mm; Figure 8e) can be observed in the riparian zone. Comparing to the posterior distributions of other parameters, the results illustrate that the storage parameters $S_{U,max}$ and I_{max} of the completely deforested riparian zone, and to a lesser extent of the hillslope, were subject to the most pronounced changes. For most other parameters, the pre- and post-deforestation posterior distributions exhibit much less pronounced differences (Figure 8). Together, these results suggest that deforestation mostly affects $S_{U,max}$ and I_{max}, while there is less evidence for systematic changes in other parameters. However, it can also be observed that the individual parameter values associated with the best model solutions in the pre- and post-deforestation periods, respectively, do vary to a stronger degree for most parameters. Notwithstanding the distinct overall effects of forest removal on the individual posterior distributions, this clearly highlights the influence of parameter compensation effects and related uncertainties. This is also illustrated by a few parameters, such as $R_{S,max}$ (Figure 8c, Equation 22), that remain poorly constrained.

It was hypothesized above that reductions in evaporative fluxes are directly and exclusively linked to reduced water volumes $S_{U,max}$ and I_{max}, respectively, which are accessible and available for evaporation and transpiration at the catchment-scale. In the theoretical ideal case, the representations of the associated storage capacities in the model, i.e. the parameters $S_{U,max}$ and I_{max}, should thus be the only ones to significantly change after deforestation. However, note that this is unlikely for two reasons. First, while it is plausible to assume that these storage capacities are significantly affected by forest removal, it is not unlikely that other system characteristics and their mutual interactions, so far unknown and not considered, are similarly influenced, potentially causing considerable ontological uncertainty. Second, model parameter interactions that arise as artefacts to compensate overly simplistic process representations and/or data uncertainty are also likely to affect parameters seemingly unrelated to deforestation. Note that in spite of these uncertainties and the associated compensation effects, in particular $S_{U,max}$ remains rather well constrained. However, after preliminary unsuccessful testing, no further attempts were made to re-calibrate only the above discussed four storage parameters, i.e. $S_{U,max,H}$, $S_{U,max,R}$, $I_{max,H}$ and $I_{max,R}$, acknowledging the limitations introduced by parameter compensation effects.

Overall these results suggest that the model formulation together with the multi-objective calibration strategy ensured the identification of solutions that provide a robust description of the system and allow a simultaneous representation of flow and isotope dynamics in the stream. There are indications that at least some processes and parameters can be directly linked to real world quantities. In particular, the results provide strong evidence that the parameters $S_{U,max,H}$ and $S_{U,max,R}$ are not merely abstract quantities, but that it is plausible to assume that they, taken together, provide a catchment-scale representation of vegetation-accessible and -accessed water volumes, which can be estimated based on water balance data without calibration as defined by Equation 2, thereby providing an alternative to small-scale in-situ observations. As such, the parameter $S_{U,max}$ is also a means to directly and independently estimate the catchment-scale effects of deforestation, and plausibly other types of land cover disturbances, on sub-surface system properties, which underlie and control the changes in the post-disturbance partitioning of water fluxes into drainage and evaporative fluxes.

6.3 Deforestation effects on travel time distributions, SAS-functions and young water fractions

Tracking water fluxes through the system it was observed that wet periods are characterized by substantially more variability and more stable stream water TTDs than dry periods. This is largely a consequence of increased bypass flow that has little interaction with resident water as the system gets wetter and which may reach the stream over preferential flow paths and increased contributions from the riparian zone with its shorter flow paths. In other words, in a wet system where little additional water can be stored, the precipitation volumes of individual storm events control the shape of TTDs, resulting in considerable variability (Heidbüchel et al, 2020). In the summer dry season, however, precipitation is to a higher degree buffered in the root-zone and used for transpiration (Stockinger et al., 2014). Conversely, stream flow is then mostly sustained by groundwater which is characterized by large volumes of older water. This effectively attenuates fluctuations by the proportionally much lower volumes of younger precipitation water that cannot be stored and is thus quickly released to the stream.

In particular, at the beginning of the wet period, elsewhere also referred to as "autumn flush" (e.g. Dawson et al., 2011), the switches towards younger water at given flow levels occur considerably faster in the post-deforestation period than in the pre-deforestation period. Therefore, where, at the same discharge, previously relatively little young water reached the stream, a much higher fraction of young water can now be observed in the stream. Underlining the role of transpiration (e.g. Douinot et al., 2019; Kuppel et al., 2020), this is a direct effect of the reduced evaporative removal of relatively young near-surface water (Maxwell et al., 2019) in the post-deforestation period, which in turn is intimately linked to the reduced water supply for evaporative fluxes, i.e. smaller storage volumes $S_{U,max}$ and I_{max}. This modelled relatively young, surface-near water, not taken up by vegetation anymore is thus to a higher degree flushed from the system mostly via preferential flow paths to the stream (i.e. $R_{F,H}$, $R_{F,R}$) and thus bypassing older resident water with little exchange, which is consistent with recent observations of more frequent activation of preferential flow paths (Wiekenkamp et al., 2020). Once connectivity and the associated higher degree of bypass flow are established in the wet period, the post-deforestation peak sensitivity of dF_{yw}/dQ to flow increased to ~ 0.36, as under these conditions when little additional water can be stored in the shallow subsurface, F_{yw} is largely controlled by magnitude of the individual precipitation signals and to a lesser extent by the footprint of the pre-storm history of evaporative fluxes in the shallow subsurface storage. In contrast, no significant post-deforestation changes could be observed for the sensitivity of F_{yw} to discharge during dry periods, as during that period, the composition of water ages is controlled by large volumes of old water.

Altogether these results suggest that even in systems dominated by old water, such as the Wüstebach, the removal of forest has the potential to increase the importance of bypass flow through fast flow paths and thus increase the risk of fast, often underestimated propagation of contaminant pulses into ground- and stream water (e.g. Hartmann et al., 2021).

~~5~~6.~~2~~4 Uncertainties, unresolved questions and limitations

As emphasized above, all results are conditional on the assumptions taken throughout the modelling process. These assumptions, present in model structure, parameterization and parameters, can lead to uncertainties. Yet, notwithstanding these potential uncertainties, extensive preliminary model testing together with the use of multiple model calibration and evaluation criteria suggest that there is relatively strong evidence to support the main results in this study: the post-deforestation reduction of evaporative fluxes can, at least partially, be linked to a relatively clear reduction in the catchment-scale storage capacities $S_{U,max}$ and I_{max}, which in turn triggered a shift towards younger water ages in the stream, particularly during wet-up and wet conditions.

This is further corroborated when comparing the estimates of $S_{U,max}$ to estimates of physically plausible upper limits of $S_{U,max}$. By definition, $S_{U,max}$ is physically bound by the depth of the groundwater table. Although fluctuating, the groundwater table in the Wüstebach remains at depths below 1 m for much of the year even in the riparian zone (Bogena et al., 2015) and can be expected to be considerably deeper on the hillslopes. Thus assuming a conservative upper bound of catchment-average depth of the groundwater table at ~ 5 m, assuming that the lowest groundwater table at each point in the catchment is at the elevation of the nearest stream, a porosity of the silty clay loam soil of 0.4 (Bogena et al., 2018) and field capacity at a relative pore water content of 0.5 suggests an upper limit of $S_{U,max,GW}$ ~ 1000 mm. However, actual roots are very often shallower than these 5 m of the groundwater table. Although sufficient detailed data on root depths are not available in the study catchment, there is no evidence for systematic and wide-spread roots extending to below 2 m. This is broadly consistent with direct experimental evidence that roots of temperate forests in general (Schenk and Jackson, 2002) and *Picea* species in particular mostly remain rather shallow (< 1 m; e.g. Schmid and Kazda, 2001) and with indirect evidence that *Picea* species rarely tap groundwater and are thus comparatively shallow (e.g. Evaristo and McDonnell, 2017). As a conservative back-of-the-envelope calculation, assuming thus a maximum plausible catchment-average root depth of 2 m, which comes close to the average observed soil depth reported in Graf et al. (2014), rather suggests a physically plausible upper limit of $S_{U,max,RD}$ ~ 400 mm, which is not exceeded by the water balance inferred catchment-scale estimates of $S_{U,max}$ = 258 ± 125 mm.

Note that the above also suggests the presence of an unsaturated transition zone between the root-zone and the groundwater table, i.e. $S_{U,max,TZ} = S_{U,max,GW} - S_{U,max,RD}$ ~ 600 mm. In the absence of root water uptake and likely negligible soil evaporation in that zone the water content will remain close to field capacity for much of the year, except for days when a wetting front infiltrates towards the groundwater. This transition zone can therefore be considered as hydrologically largely passive so that at time scales of more than a few days $dS/dt \sim 0$. However, this zone also provides a mixing volume that affects tracer circulation and thus water ages (Hrachowitz et al., 2015). Given its hydrologically passive nature and following the idea of a parsimonious model to limit uncertainty, we here, in a simplification, implicitly added the mixing volume $S_{U,max,TZ}$ to the passive groundwater mixing volume $S_{S,p}$.

For a meaningful interpretation, two specific observations resulting from our analysis warrant special scrutiny. First, model calibration-based estimations of hillslope $S_{U,max,H}$ (Figures 7a) suggest post-deforestation median $S_{U,max,H}$ reductions of ~ 25 % as a consequence of clear cutting only ~10 % of the hillslope part of the catchment (Figure 1). While this may be surprising at the first, it can be plausibly explained by considerable further thinning of the remaining forest on the hillslopes in 2015, two years after deforestation and thus by reduced catchment-scale transpiration demand. Yet, no detailed and systematic data on the degree of forest thinning is available to meaningfully test this hypothesis.

Second, our results suggest that a passive mixing volume $S_{S,p}$ of at least ~ 8.000 mm is necessary for the model to attenuate the amplitudes of the precipitation $\delta^{18}O$ signals to those in the stream water. Although, $S_{S,p}$ is rather well constrained (Figure ~~7h~~8h), there has in the past been no hydrogeological evidence for the presence of such a surprisingly large groundwater volume nor for its hydrological relevance in the study catchment. Indeed, the authors are not aware of any catchment-scale study that reported similarly high values for $S_{S,p}$ or functionally equivalent parameters (e.g. Birkel et al., 2011a,b; Hrachowitz et al., 2013,2015; Benettin et al., 2013,2015a; Harman, 2015; van der Velde et al., 2015). Yet, to achieve the degree of damping observed in the stream water, such a volume is necessary, if the current understanding of conservative tracer dynamics holds e.g. Maloszewski and Zuber, 1982; McGuire and McDonnell, 2006). Reflecting our insufficient knowledge to which depths exchange with surface water occurs (e.g. Condon et al., 2020), a potential explanation for this observation is that the frequently layered and fractured structure of the Devonian shale bedrock may provide relatively high-permeability pathways for the circulation of and exchange with water at depth. Another, yet, given the current understanding of the Wüstebach (e.g. Graf et al., 2014), less likely hypothesis is the presence of significant lateral groundwater exchange (e.g. Bouaziz et al., 2018; Hulsman et al., 2021b). In other words the possibility that the subsurface catchment does not match with the surface catchment (Figure 1) and that older groundwater is imported from "outside" the surface catchment, while an equivalent volume of younger groundwater is exported, maintaining the mass balance. These are hypotheses to be tested in future studies, as the currently available data do not allow a conclusive answer to this question.

~~6~~ 7 Conclusions

The small Wüstebach catchment experienced significant deforestation in 2013. Analyzing the effects of this deforestation on the hydrology and stable isotope circulation dynamics in the study catchment our main findings are:

(1) Water balance data suggest that deforestation led to a significant increase of stream flow, accompanied by corresponding reductions of evaporative fluxes. This is reflected by an increase of the runoff ratio from C_R = 0.55 to 0.68 in the post-deforestation period despite similar climatic conditions, supporting previous results based on eddy covariance measurements (Wiekenkamp et al., 2016).

(2) Based on water balance data, this reduction of evaporative fluxes, as a consequence of reduced vegetation water uptake, could at least partly be linked to a reduction of the catchment-scale water storage volume in the unsaturated soil ($S_{U,max}$) that is within the reach of active roots and thus accessible for vegetation transpiration from ~258 mm in the pre-deforestation period to ~ 101 mm in the post-deforestations period.

(3) Estimating $S_{U,max}$ as calibration parameter of a process-based hydrological model led to similar conclusions. The catchment-average calibrated model parameters representing $S_{U,max}$ for both, the pre- and deforestation periods, respectively, correspond with ~ 240 mm and ~ ~~120~~ 199 mm broadly with $S_{U,max}$ directly estimated from water balance data. Other model parameters, assumed to have a less direct link to vegetation, exhibited much lower levels of systematic change following deforestation.

(4) Using the model to track the age composition of stream water suggested that, in general, water reaching the stream in the pre-deforestation period was rather old with a mean young water fraction F_{yw} ~ 0.~~11~~12. In spite of the overall low F_{yw}, clear shifts in the shape of travel time distributions towards younger water can be seen under wet conditions with young water fractions increasing up to F_{yw} ~ 0.~~30~~34.

(5) Deforestation and the associated reduction of $S_{U,max}$ led to shifts in travel time distributions towards younger water. Under wet conditions, this resulted in increases of young water fractions to up to F_{yw} ~ 0.~~40~~ 37 for individual storms. In contrast, dry period travel time distributions exhibited only minor changes. Overall the mean fraction of young water in the stream increased to F_{yw} ~ 0.13.

(6) Deforestation resulted in a considerable increase of the sensitivity of young water fractions to discharge under wet conditions from dF_{yw}/dQ = 0.25 to 0.~~43~~36. This implies faster switches towards younger water and thus faster routing of solutes during and shortly after storm events and thus faster routing of solutes with increasing wetness.

The above results suggest that deforestation has not only the potential to affect the partitioning between drainage and evaporation, and thus the fundamental hydrological response characteristics of catchments, but also catchment-scale tracer circulation dynamics. In particular for wet and wet-up conditions, sometimes also referred to as "autumn flush", deforestation in the Wüstebach caused higher proportions of younger water to reach the stream, implying faster routing of water and plausibly also solutes through the subsurface, thereby also increasing the risk for faster propagation of contaminants into stream- and groundwater.

Overall, this study demonstrates that post-deforestation changes in both, the hydrological response and travel times, can to a large extent be traced back and attributed to changes in $S_{U,max}$, a readily quantifiable catchment-scale subsurface property (and model parameter) representing the maximum water volume that can be stored within the reach of roots. As such, $S_{U,max}$ and changes therein provide a quantitative, mechanistic hypothesis that can explain *why* deforestation in the Wüstebach decreased evaporative fluxes, increased stream flow – particularly generated by preferential flows – and reduced travel times. The catchment-scale quantification of $S_{U,max}$ based on water balance data therefore provides a potentially valuable way towards meaningful and data-based catchment-scale representation of vegetation-accessible water where soil and root observations are not available at sufficient spatial and temporal detail to meaningfully represent their respective natural heterogeneities. In addition and perhaps more importantly, the method may also hold considerable potential for the formulation of temporally adaptive root-zone parameterizations in catchment-scale hydrological models for more reliable predictions in a changing environment.

Data availability. The meteorological and hydrological data of the Wüstebach TERENO site used in this study can be made available by the co-author HB upon request. The model results, including states, fluxes, hydrological signatures, parameter sets, and performance metrics underlying this paper are available online in the 4TU data repository at https://doi.org/10.4121/14626050.v1.

Code availability. The model code used can be made available by the first author upon request. The equations used in the model are described in the paper.

Author contributions. MH and MS designed the experiment. MH did the analysis and wrote the first draft. All authors discussed the design, results and the first draft and contributed to writing the final manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgments. We would like to thank the editor and four anonymous reviewers for providing a list of critical and very valuable comments that helped to considerably improve the manuscript.

References

Ala-Aho, P., Tetzlaff, D., McNamara, J. P., Laudon, H., & Soulsby, C. (2017). Using isotopes to constrain water flux and age estimates in snow-influenced catchments using the STARR (Spatially distributed Tracer-Aided Rainfall-Runoff) model. Hydrology and Earth System Sciences 21, 5089-5110.

Angermann, L., Jackisch, C., Allroggen, N., Sprenger, M., Zehe, E., Tronicke, J., ... & Blume, T. (2017). Form and function in hillslope hydrology: Characterization of subsurface ow based on response observations. Hydrology and Earth System Sciences 21, 3727-3748.

Arrouays, D., Lagacherie, P., & Hartemink, A.E. (2017). Digital soil mapping across the globe. Geoderma Regional 9

Arsenault, R., Poissant, D., & Brissette, F. (2015). Parameter dimensionality reduction of a conceptual model for streamflow prediction in Canadian, snowmelt dominated ungauged basins. Advances in water resources, 85, 27-44.

Benettin, P., Kirchner, J. W., Rinaldo, A., & Botter, G. (2015a). Modeling chloride transport using travel time distributions at Plynlimon, Wales. Water Resources Research, 51(5), 3259-3276.

Benettin, P., Rinaldo, A., & Botter, G. (2015b). Tracking residence times in hydrological systems: Forward and backward formulations. Hydrological Processes, 29(25), 5203-5213.

Benettin, P., Soulsby, C., Birkel, C., Tetzlaff, D., Botter, G., & Rinaldo, A. (2017). Using SAS functions and high-resolution isotope data to unravel travel time distributions in headwater catchments. Water Resources Research, 53(3), 1864-1878.

Berkowitz, B., & Zehe, E. (2020). Surface water and groundwater: Unifying conceptualization and quantification of the two water worlds. Hydrology and Earth System Sciences, 24.

Bertuzzo, E., Thomet, M., Botter, G., & Rinaldo, A. (2013). Catchment-scale herbicides transport: Theory and application. Advances in water resources, 52, 232-242.

Beven, K. (2006). A manifesto for the equifinality thesis. Journal of hydrology, 320(1-2), 18-36.

Beven, K. J. (2010). Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models. Hydrological Processes, 24(12), 1537-1547.

Beven, K., & Germann, P. (2013). Macropores and water flow in soils revisited. Water resources research, 49(6), 3071-3092.

Birkel, C., Tetzlaff, D., Dunn, S. M., & Soulsby, C. (2011a). Using lumped conceptual rainfall–runoff models to simulate daily isotope variability with fractionation in a nested mesoscale catchment. Advances in Water Resources, 34(3), 383-394.

Birkel, C., Soulsby, C., & Tetzlaff, D. (2011b). Modelling catchment-scale water storage dynamics: Reconciling dynamic storage with tracer-inferred passive storage. Hydrological Processes, 25(25), 3924-3936.

Birkel, C., Soulsby, C., & Tetzlaff, D. (2014). Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables. Water Resources Research, 50(4), 3481-3501.

Birkel, C., Soulsby, C., & Tetzlaff, D. (2015). Conceptual modelling to assess how the interplay of hydrological connectivity, catchment storage and tracer dynamics controls nonstationary water age estimates. Hydrological Processes, 29(13), 2956-2969.

Birkel, C., & Soulsby, C. (2016). Linking tracers, water age and conceptual models to identify dominant runoff processes in a sparsely monitored humid tropical catchment. Hydrological Processes, 30(24), 4477-4493.

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., ... & Stumpp, C. (2019). Twenty-three unsolved problems in hydrology (UPH)–a community perspective. Hydrological Sciences Journal, 64(10), 1141-1158.

Bogena, H. R., Bol, R., Borchard, N., Brüggemann, N., Diekkrüger, B., Drüe, C., ... & Missong, A. (2015). A terrestrial observatory approach to the integrated investigation of the effects of deforestation on water, energy, and matter fluxes. Science China Earth Sciences, 58(1), 61-75.

Bogena, H.R., Montzka, C., Huisman, J.A., Graf, A., Schmidt, M., Stockinger, M., ... & Lücke, A. (2018). The TERENO-Rur Hydrological Observatory: A multiscale multi-compartment research platform for the advancement of hydrological science. Vadose Zone Journal, 17(1).

Bogena, H.R., Stockinger, M. & Luecke, A. (2020). Long-term stable water isotope data for the investigation of deforestation effects on the hydrological system of the Wüstebach catchment, Germany. Hydrological Processes. e14006.

Borchardt, H. (2012). Einfluss periglazialer Deckschichten auf Abflusssteuerung am Beispiel des anthropogen überprägten Wüstebaches (Nationalpark Eifel), PhD diss., Lehrstuhl für Physische Geogr. und Geo€okologie, Fak. für Georessourcen und Materialtechnik, Rheinisch-Westfälische Tech. Hochsch. Aachen, Aaachen, Germany.

Botter, G., Bertuzzo, E., & Rinaldo, A. (2011). Catchment residence and travel time distributions: The master equation. Geophysical Research Letters, 38(11).

Bouaziz, L., Weerts, A., Schellekens, J., Sprokkereef, E., Stam, J., Savenije, H., & Hrachowitz, M. (2018). Redressing the balance: quantifying net intercatchment groundwater flows. Hydrology and Earth System Sciences, 22(12), 6415-6434.

Bouaziz, L., Steele-Dunne, S.C., Schellekens, J., Weerts, A., Stam, J., Sprokkereef, E., Winsemius, H.C., Savenije, H.H.G., & Hrachowitz, M. (2020). Improved understanding of the link between catchment-scale vegetation accessible storage and satellite-derived Soil Water Index. Water Resources Research, 56, e2019WR026365.

Bouaziz, L. J., Fenicia, F., Thirel, G., de Boer-Euser, T., Buitink, J., Brauer, C. C., ... & Hrachowitz, M. (2021). Behind the scenes of streamflow model performance. Hydrology and Earth System Sciences, 25(2), 1069-1095.

Brand, W.A., Coplen, T.B., Vogl, J., Rosner, M., & Prohaska, T. (2014). Assessment of international reference materials for isotope-ratio analysis (IUPAC Technical Report). Pure and Applied Chemistry, 86(3), 425-467.

Brooks, J. R., Barnard, H. R., Coulombe, R., & McDonnell, J. J. (2010). Ecohydrologic separation of water between trees and streams in a Mediterranean climate. Nature Geoscience, 3(2), 100-104.

Brooks, P. D., Chorover, J., Fan, Y., Godsey, S. E., Maxwell, R. M., McNamara, J. P., & Tague, C. (2015). Hydrological partitioning in the critical zone: Recent advances and opportunities for developing transferable understanding of water cycle dynamics. Water Resources Research, 51(9), 6973-6987.

Brunner, I., Herzog, C., Dawes, M.A., Arend, M., & Sperisen, C. (2015). How tree roots respond to drought. Frontiers in plant science, 6, 547.

Brutsaert, W. (2014). Daily evaporation from drying soil: Universal parameterization with similarity. Water Resources Research, 50(4), 3206-3215.

Cain, M. R., Ward, A. S., & Hrachowitz, M. (2019). Ecohydrologic separation alters interpreted hydrologic stores and fluxes in a headwater mountain catchment. Hydrological Processes, 33(20), 2658-2675.

Coenders-Gerrits, A.M.J., Van der Ent, R.J., Bogaard, T.A., Wang-Erlandsson, L., Hrachowitz, M., & Savenije, H.H.G. (2014). Uncertainties in transpiration estimates. Nature, 506(7487), E1-E2.

Collins, D.B.G., & Bras, R.L. (2007). Plant rooting strategies in water-limited ecosystems. Water Resources Research, 43(6).

Condon, L. E., Markovich, K. H., Kelleher, C. A., McDonnell, J. J., Ferguson, G., & McIntosh, J. C. (2020). Where is the bottom of a watershed?. Water Resources Research, 56(3), e2019WR026010.

Cornelissen, T., Diekkrüger, B., & Bogena, H. R. (2014). Significance of scale and lower boundary condition in the 3D simulation of hydrological processes and soil moisture variability in a forested headwater catchment. Journal of hydrology, 516, 140-153.

Coxon, G., Freer, J., Wagener, T., Odoni, N. A., & Clark, M. (2014). Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments. Hydrological Processes, 28(25), 6135-6150.

Creed, I. F., Spargo, A. T., Jones, J. A., Buttle, J. M., Adams, M. B., Beall, F. D., ... & Green, M. B. (2014). Changing forest water yields in response to climate warming: Results from long-term experimental watershed sites across North America. Global change biology, 20(10), 3191-3208.

Criss, R. E., and W. E. Winston (2008), Do Nash values have a value? Discussion and alternate proposals, Hydrol. Proc., 22, 2723–2725.

deBoer-Euser, T., McMillan, H.K., Hrachowitz, M., Winsemius, H.C., & Savenije, H.H.G. (2016). Influence of soil and climate on root zone storage capacity. Water Resources Research, 52(3), 2009-2024.

Donohue, R., Roderick, M., & McVicar, T.R. (2007). On the importance of including vegetation dynamics in Budyko's hydrological model.

Donohue, R.J., Roderick, M.L., & McVicar, T.R. (2012). Roots, storms and soil pores: Incorporating key ecohydrological processes into Budyko's hydrological model. Journal of Hydrology, 436, 35-50.

Douinot, A., Tetzlaff, D., Maneta, M., Kuppel, S., Schulte-Bisping, H., & Soulsby, C. (2019). Ecohydrological modelling with EcH2O-iso to quantify forest and grassland effects on water partitioning and flux ages. Hydrological Processes, 33(16), 2174-2191.

Eagleson, P.S. (Ed.) (1982). Land surface processes in atmospheric general circulation models. Cambridge University Press.

Ellison, D., Morris, C. E., Locatelli, B., Sheil, D., Cohen, J., Murdiyarso, D., ... & Gaveau, D. (2017). Trees, forests and water: Cool insights for a hot world. Global Environmental Change, 43, 51-61.

Etmann, M. (2009). Dendrologische Aufnahmen im Wassereinzugsgebiet Oberer Wüstebach anhand verschiedener Mess- und Schätzverfahren, MSc thesis, Institut für Landschaftsökologie, Univ. of Münster, Münster, Germany.

Euser, T., H. C. Winsemius, M. Hrachowitz, F. Fenicia, S. Uhlenbrook, and H. H. G. Savenije (2013), A framework to assess the realism of model structures using hydrological signatures, Hydrol. Earth Syst. Sci., 17, 1893–1912.

Euser, T., Hrachowitz, M., Winsemius, H.C., & Savenije, H.H. (2015). The effect of forcing and landscape distribution on performance and consistency of model structures. Hydrological Processes, 29(17), 3727-3743.

Evaristo, J., & McDonnell, J. J. (2017). Prevalence and magnitude of groundwater use by vegetation: a global stable isotope meta-analysis. Scientific reports, 7(1), 1-12.

Evaristo, J., Kim, M., van Haren, J., Pangle, L. A., Harman, C. J., Troch, P. A., & McDonnell, J. J. (2019). Characterizing the fluxes and age distribution of soil water, plant water, and deep percolation in a model tropical ecosystem. Water Resources Research, 55(4), 3307-3327.

Fan, Y., Miguez-Macho, G., Jobbágy, E.G., Jackson, R.B., & Otero-Casal, C. (2017). Hydrologic regulation of plant rooting depth. Proceedings of the National Academy of Sciences, 114(40), 10572-10577.

Fenicia, F., Savenije, H.H.G., Matgen, P., & Pfister, L. (2008). Understanding catchment behavior through stepwise model concept improvement. Water Resources Research, 44(1).

Fenicia, F., Wrede, S., Kavetski, D., Pfister, L., Hoffmann, L., Savenije, H. H., & McDonnell, J. J. (2010). Assessing the impact of mixing assumptions on the estimation of streamwater mean residence time. Hydrological Processes, 24(12), 1730-1741.

Fovet, O., Ruiz, L., Hrachowitz, M., Faucheux, M., & Gascuel-Odoux, C. (2015). Hydrological hysteresis and its value for assessing process consistency in catchment conceptual models. Hydrology and Earth System Sciences, 19(1), 105-123.

Freer, J., Beven, K., & Ambroise, B. (1996). Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resources Research, 32(7), 2161-2173.

Gallart, F., von Freyberg, J., Valiente, M., Kirchner, J. W., Llorens, P., & Latron, J. (2020a). An improved discharge sensitivity metric for young water fractions. Hydrology and Earth System Sciences, 24(3), 1101-1107.

Gallart, F., Valiente, M., Llorens, P., Cayuela, C., Sprenger, M., & Latron, J. (2020b). Investigating young water fractions in a small Mediterranean mountain catchment: both precipitation forcing and sampling frequency matter. Hydrological Processes.

Gao, H., Hrachowitz, M., Schymanski, S.J., Fenicia, F., Sriwongsitanon, N., & Savenije, H.H.G. (2014). Climate controls how ecosystems size the root zone storage capacity at catchment scale. Geophysical Research Letters, 41(22), 7916-7923.

Gao, H., Ding, Y., Zhao, Q., Hrachowitz, M., & Savenije, H. H. (2017). The importance of aspect for modelling the hydrological response in a glacier catchment in Central Asia. Hydrological Processes, 31(16), 2842-2859.

Gentine, P., D'Odorico, P., Lintner, B.R., Sivandran, G., & Salvucci, G. (2012). Interdependence of climate, soil, and vegetation as constrained by the Budyko curve. Geophysical Research Letters, 39(19).

Gerrits, A. M. J., Savenije, H. H. G., Veling, E. J. M., & Pfister, L. (2009). Analytical derivation of the Budyko curve based on rainfall characteristics and a simple evaporation model. Water Resources Research, 45(4).

Graf, A., Bogena, H.R., Drüe, C., Hardelauf, H., Pütz, T., Heinemann, G., & Vereecken, H. (2014). Spatiotemporal relations between water budget components and soil water content in a forested tributary catchment. Water resources research, 50(6), 4837-4857.

Gumbel, E. J. (1941). The return period of flood flows. The annals of mathematical statistics, 12(2), 163-190.

Guswa, A.J. (2008). The influence of climate on root depth: A carbon cost-benefit analysis. Water Resources Research, 44(2).

Harman, C., & Troch, P.A. (2014). What makes Darwinian hydrology" Darwinian"? Asking a different kind of question about landscapes. Hydrology and Earth System Sciences, 18(2), 417.

Harman, C. J. (2015). Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed. Water Resources Research, 51(1), 1-30.

Hartmann, A., Jasechko, S., Gleeson, T., Wada, Y., Andreo, B., Barberá, J. A., ... & Wagener, T. (2021). Risk of groundwater contamination widely underestimated because of fast flow into aquifers. Proceedings of the National Academy of Sciences, 118(20).

Heidbüchel, I., Yang, J., Musolff, A., Troch, P., Ferré, T., & Fleckenstein, J. H. (2020). On the shape of forward transit time distributions in low-order catchments. Hydrology and Earth System Sciences 24, 2895–2920.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., ... & Guevara, M. A. (2017). SoilGrids250m: Global gridded soil information based on machine learning. PLoS one, 12(2).

Hrachowitz, M., & Clark, M. P. (2017). HESS Opinions: The complementary merits of competing modelling philosophies in hydrology. Hydrol. Earth Syst. Sci, 21(8), 3953-3973.

Hrachowitz, M., Soulsby, C., Tetzlaff, D., Dawson, J. J. C., Dunn, S. M., & Malcolm, I. A. (2009). Using long-term data sets to understand transit times in contrasting headwater catchments. Journal of Hydrology, 367(3-4), 237-248.

Hrachowitz, M., Savenije, H., Bogaard, T.A., Tetzlaff, D., & Soulsby, C. (2013). What can flux tracking teach us about water age distribution patterns and their temporal dynamics?. Hydrology and Earth System Sciences, 17, 533–564.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., ... & Gascuel-Odoux, C. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. Water resources research, 50(9), 7445-7469.

Hrachowitz, M., Fovet, O., Ruiz, L., & Savenije, H.H.G. (2015). Transit time distributions, legacy contamination and variability in biogeochemical $1/f^{\alpha}$ scaling: how are hydrological response dynamics linked to water quality at the catchment scale?. Hydrological Processes, 29(25), 5241-5256.

Hrachowitz, M., Benettin, P., Van Breukelen, B.M., Fovet, O., Howden, N.J., Ruiz, L., ... & Wade, A.J. (2016). Transit times—The link between hydrology and water quality at the catchment scale. Wiley Interdisciplinary Reviews: Water, 3(5), 629-657.

Hulsman, P., Winsemius, H.C., Michailovsky, C., Savenije, H.H.G. & Hrachowitz, M. (2020). Using altimetry observations combined with GRACE to select parameter sets of a hydrological model in data scarce regions. Hydrology and Earth System Sciences, 24, 3331-3359.

Hulsman, P., Savenije, H. H., & Hrachowitz, M. (2021a). Learning from satellite observations: increased understanding of catchment processes through stepwise model improvement. Hydrology and Earth System Sciences, 25(2), 957-982.

Hulsman, P., Hrachowitz, M., & Savenije, H. H. (2021b). Improving the representation of long-term storage variations with conceptual hydrological models in data-scarce regions. Water Resources Research 57, e2020WR028837.

Hwang, T., Band, L., & Hales, T.C. (2009). Ecosystem processes at the watershed scale: Extending optimality theory from plot to catchment. Water Resources Research, 45(11).

Jaramillo, F., & Destouni, G. (2014). Developing water change spectra and distinguishing change drivers worldwide. Geophysical Research Letters, 41(23), 8377-8386.

Jaramillo, F., Cory, N., Arheimer, B., Laudon, H., Van Der Velde, Y., Hasper, T. ., ... & Uddling, J. (2018). Dominant effect of increasing forest biomass on evapotranspiration: interpretations of movement in Budyko space. Hydrology and Earth System Sciences, 22(1), 567-580.

Jasechko, S. (2018). Plants turn on the tap. Nature Climate Change, 8(7), 562-563.

Jothityangkoon, C., M. Sivapalan, and D. L. Farmer (2001), Process controls of water balance variability in a large semi-arid catchment: Downward approach to hydrological model development, J. Hydrol., 254(1), 174–198.

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3).

Klaus, J., Zehe, E., Elsner, M., Külls, C., & McDonnell, J. J. (2013). Macropore flow of old water revisited: experimental insights from a tile-drained hillslope. Hydrology and Earth System Sciences, 17(1), 103.

Kleidon, A. (2004). Global datasets of rooting zone depth inferred from inverse methods. Journal of Climate, 17(13), 2714-2722.

Knighton, J., Souter-Kline, V., Volkman, T., Troch, P. A., Kim, M., Harman, C., ... & Walter, M. T. (2019). Seasonal and topographic variations in ecohydrological separation within a small, temperate, snow-influenced catchment. Water Resources Research, 55(8), 6417-6435.

Knighton, J., Singh, K., & Evaristo, J. (2020). Understanding Catchment-Scale Forest Root Water Uptake Strategies Across the Continental United States Through Inverse Ecohydrological Modeling. Geophysical Research Letters, 47(1), e2019GL085937.

Kuppel, S., Tetzlaff, D., Maneta, M., & Soulsby, C. (2018a). EcH2O-iso 1.0: water isotopes and age tracking in a process-based, distributed ecohydrological model. Geoscientific Model Development.

Kuppel, S., Tetzlaff, D., Maneta, M. P., & Soulsby, C. (2018b). What can we learn from multi-data calibration of a process-based ecohydrological model?. Environmental Modelling & Software, 101, 301-316.

Kuppel, S., Tetzlaff, D., Maneta, M. P., & Soulsby, C. (2020). Critical zone storage controls on the water ages of ecohydrological outputs. Geophysical Research Letters, 47(16), e2020GL088897.

Laio, F., D'Odorico, P., & Ridolfi, L. (2006). An analytical model to relate the vertical root distribution to climate and soil properties. Geophysical Research Letters, 33(18).

Loritz, R., Hassler, S. K., Jackisch, C., Allroggen, N., van Schaik, L., Wienhöfer, J., & Zehe, E. (2017). Picturing and modeling catchments by representative hillslopes. Hydrology and Earth System Sciences, 21, 1225-1249.

Małoszewski, P., & Zuber, A. (1982). Determining the turnover time of groundwater systems with the aid of environmental tracers: 1. Models and their applicability. Journal of hydrology, 57(3-4), 207-231.

Maxwell, R. M., Condon, L. E., Kollet, S. J., Maher, K., Haggerty, R., & Forrester, M. M. (2016). The imprint of climate and geology on the residence times of groundwater. Geophysical Research Letters, 43(2), 701-708.

Maxwell, R. M., Condon, L. E., Danesh-Yazdi, M., & Bearup, L. A. (2019). Exploring source water mixing and transient residence time distributions of outflow and evapotranspiration with an integrated hydrologic model and Lagrangian particle tracking approach. Ecohydrology, 12(1), e2042.

Mezentsev, V. (1955). Back to the computation of total evaporation. Meteorologia i Gidologia, 5. 24-26.

McDonnell, J.J., McGuire, K., Aggarwal, P., Beven, K.J., Biondi, D., Destouni, G., ... & Lyon, S. (2010). How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis. Hydrological Processes, 24(12), 1745-1754.

McGuire, K. J., & McDonnell, J. J. (2006). A review and evaluation of catchment transit time modeling. Journal of Hydrology, 330(3-4), 543-563.

McMillan, H., Tetzlaff, D., Clark, M., & Soulsby, C. (2012). Do time-variable tracers aid the evaluation of hydrological model structure? A multimodel approach. Water Resources Research, 48(5).

McMurtrie, R.E., Iversen, C.M., Dewar, R.C., Medlyn, B.E., Näsholm, T., Pepper, D.A., & Norby, R.J. (2012). Plant root distributions and nitrogen uptake predicted by a hypothesis of optimal root foraging. Ecology and evolution, 2(6), 1235-1250.

Milly, P.C., & Dunne, K.A. (1994). Sensitivity of the global water cycle to the water-holding capacity of land. Journal of Climate, 7(4), 506-526.

Mianabadi, A., Coenders-Gerrits, M., Shirazi, P., Ghahraman, B., & Alizadeh, A. (2019). A global Budyko model to partition evaporation into interception and transpiration. Hydrology and Earth System Sciences, 23(12), 4983-5000.

Montanari, A., and E. Toth (2007), Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, Water Resour. Res., 43, W05434, doi:10.1029/2006WR005184.

Morán-Tejeda, E., Zabalza, J., Rahman, K., Gago-Silva, A., López-Moreno, J. I., Vicente-Serrano, S., ... & Beniston, M. (2015). Hydrological impacts of climate and land-use changes in a mountain watershed: Uncertainty estimation based on model comparison. Ecohydrology, 8(8), 1396-1416.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models: part 1—A discussion of principles, J. Hydrol., 10, 282–290.

Nijzink, R., Hutton, C., Pechlivanidis, I., Capell, R., Arheimer, B., Freer, J., ... & Hrachowitz, M. (2016a). The evolution of root-zone moisture capacities after deforestation: a step towards hydrological predictions under change?. Hydrology and Earth System Sciences, 20(12), 4775-4799.

Nijzink, R.C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., ... & Hrachowitz, M. (2016b). The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models. Hydrology & Earth System Sciences, 20(3), 1151–1176

Penna, D., Hopp, L., Scandellari, F., Allen, S. T., Benettin, P., Beyer, M., ... & Volkmann, T. H. (2018). Ideas and perspectives: Tracing terrestrial ecosystem water fluxes using hydrogen and oxygen stable isotopes-challenges and opportunities from an interdisciplinary perspective. Biogeosciences, 15(21), 6399-6415.

Penna, D., Geris, J., Hopp, L., & Scandellari, F. (2020). Water sources for root water uptake: Using stable isotopes of hydrogen and oxygen as a research tool in agricultural and agroforestry systems. Agriculture, Ecosystems & Environment, 291, 106790.

Renner, M., Brust, K., Schwärzel, K., Volk, M., & Bernhofer, C. (2014). Separating the effects of changes in land cover and climate: a hydro-meteorological analysis of the past 60 yr in Saxony, Germany. Hydrology and Earth System Sciences, 18, 389-405.

Richter, F. (2008), Bodenkarte zur Standorterkundung. Verfahren Quellgebiet Wüstebachtal (Forst), Geologischer Dienst Nordrhein-Westfalen, Krefeld, Germany.

Richter, D. (1995), Ergebnisse methodischer Untersuchungen zur Korrektur des systematische Meßfehlers des Hellmann-Niederschlagmessers, in Berichte des Deutschen Wetterdienstes, vol. 194, 93 pp, LRG (Lysimeter Research Group), Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main, Germany.

Rinaldo, A., Benettin, P., Harman, C. J., Hrachowitz, M., McGuire, K.J., Van Der Velde, Y., ... & Botter, G. (2015). Storage selection functions: A coherent framework for quantifying how catchments store and release water and solutes. Water Resources Research, 51(6), 4840-4847.

Rodriguez, N. B., McGuire, K. J., & Klaus, J. (2018). Time-varying storage–water age relationships in a catchment with a mediterranean climate. Water Resources Research, 54(6), 3988-4008.

Rodriguez, N. B., & Klaus, J. (2019). Catchment travel times from composite StorAge Selection functions representing the superposition of streamflow generation processes. Water Resources Research.

Rodriguez-Iturbe, I., D'Odorico, P., Laio, F., Ridolfi, L., & Tamea, S. (2007). Challenges in humid land ecohydrology: Interactions of water table and unsaturated zone with climate, soil, and vegetation. Water Resources Research, 43(9).

Roebroek, C. T., Melsen, L. A., Hoek van Dijke, A. J., Fan, Y., & Teuling, A. J. (2020). Global distribution of hydrologic controls on forest growth. Hydrology and Earth System Sciences, 24(9), 4625-4639.

Savenije, H.H.G., & Hrachowitz, M. (2017). HESS Opinions "Catchments as meta-organisms–a new blueprint for hydrological modelling". Hydrol. Earth Syst. Sci, 21(2), 1107-1116.

Sawicz, K., T. Wagener, M. Sivapalan, P. A. Troch, and G. Carrillo (2011), Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15, 2895–2911.

Seneviratne, S.I., Wilhelm, M., Stanelle, T., van den Hurk, B., Hagemann, S., Berg, A., ... & Claussen, M. (2013). Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment. Geophysical Research Letters, 40(19), 5212-5217.

Schenk, H.J. (2005). Vertical vegetation structure below ground: scaling from root to globe. In Progress in botany (pp. 341-373). Springer, Berlin, Heidelberg.

Schenk, H.J., & Jackson, R.B. (2002). Rooting depths, lateral root spreads and below-ground/above-ground allometries of plants in water-limited ecosystems. Journal of Ecology, 90(3), 480-494.

Schenk, H. J., & Jackson, R. B. (2003). Global distribution of root profiles in terrestrial ecosystems. ORNL DAAC.

Schymanski, S.J., Sivapalan, M., Roderick, M.L., Beringer, J., & Hutley, L B. (2008). An optimality-based model of the coupled soil moisture and root dynamics. Hydrology and earth system sciences, 12(3), 913-932.

Shamir, E., B. Imam, H. V. Gupta, and S. Sorooshian (2005), Application of temporal streamflow descriptors in hydrologic model parameter estimation, Water Resour. Res., 41, W06021, doi:10.1029/2004WR003409.

Sivandran, G., & Bras, R.L. (2012). Identifying the optimal spatially and temporally invariant root distribution for a semiarid environment. Water Resources Research, 48(12).

Speich, M.J., Lischke, H., & Zappa, M. (2018). Testing an optimality-based model of rooting zone water storage capacity in temperate forests. Hydrology and Earth System Sciences, 22(7), 4097-4124.

Speich, M. J., Zappa, M., Scherstjanoi, M., & Lischke, H. (2020). FORests and HYdrology under Climate Change in Switzerland v1. 0: a spatially distributed model combining hydrology and forest dynamics. Geoscientific Model Development, 13(2), 537-564.

Soulsby, C., Tetzlaff, D., & Hrachowitz, M. (2010). Are transit times useful process-based tools for flow prediction and classification in ungauged basins in montane regions?. Hydrological Processes, 24(12), 1685-1696.

Soulsby, C., Braun, H., Sprenger, M., Weiler, M., & Tetzlaff, D. (2017). Influence of forest and shrub canopies on precipitation partitioning and isotopic signatures. Hydrological Processes, 31(24), 4282-4296.

Speed, M., Tetzlaff, D., Soulsby, C., Hrachowitz, M., & Waldron, S. (2010). Isotopic and geochemical tracers reveal similarities in transit times in contrasting mesoscale catchments. Hydrological Processes: An International Journal, 24(9), 1211-1224.

Sprenger, M., Leistert, H., Gimbel, K., & Weiler, M. (2016). Illuminating hydrological processes at the soil-vegetation-atmosphere interface with water stable isotopes. Reviews of Geophysics, 54(3), 674-704.

Sprenger, M., Tetzlaff, D., Buttle, J., Laudon, H., & Soulsby, C. (2018). Water ages in the critical zone of long-term experimental sites in northern latitudes. Hydrology and Earth System Sciences.

Sprenger, M., Llorens, P., Cayuela, C., Gallart, F., & Latron, J. (2019a). Mechanisms of consistently disjunct soil water pools over (pore) space and time.

Sprenger, M., Stumpp, C., Weiler, M., Aeschbach, W., Allen, S.T., Benettin, P., ... & McDonnell, J.J. (2019b). The demographics of water: a review of water ages in the critical zone. Reviews of Geophysics, 57(3), 800-834.

Stockinger, M. P., Bogena, H. R., Lücke, A., Diekkrüger, B., Weiler, M., & Vereecken, H. (2014). Seasonal soil moisture patterns: Controlling transit time distributions in a forested headwater catchment. Water resources research, 50(6), 5270-5289.

Stockinger, M. P., Lücke, A., McDonnell, J. J., Diekkrüger, B., Vereecken, H., & Bogena, H. R. (2015). Interception effects on stable isotope driven streamwater transit time estimates. Geophysical research letters, 42(13), 5299-5308.

Stockinger, M. P., Bogena, H. R., Lücke, A., Diekkrüger, B., Cornelissen, T., & Vereecken, H. (2016). Tracer sampling frequency influences estimates of young water fraction and streamwater transit time distribution. Journal of hydrology, 541, 952-964.

Stockinger, M. P., Lücke, A., Vereecken, H., & Bogena, H. R. (2017). Accounting for seasonal isotopic patterns of forest canopy intercepted precipitation in streamflow modeling. Journal of Hydrology, 555, 31-40.

Teuling, A. J. (2018). A hot future for European droughts. Nature Climate Change, 8(5), 364-365.

Teuling, A. J., & van Dijke, A. J. H. (2020). Forest age and water yield. Nature, 578(7794), E16-E18.

Tron, S., Perona, P., Gorla, L., Schwarz, M., Laio, F., & Ridolfi, L. (2015). The signature of randomness in riparian plant root distributions. Geophysical Research Letters, 42(17), 7098-7106.

Turc, L. (1954). Le bilan d'eau des sols. Relations entre les precipitations, l'evaporation et l'ecoulement. Ann. Agron., 5, 491-596

Van Der Velde, Y., Torfs, P. J. J. F., Van Der Zee, S. E. A. T. M., & Uijlenhoet, R. (2012). Quantifying catchment-scale mixing and its effect on time-varying travel time distributions. Water Resources Research, 48(6).

Van der Velde, Y., Vercauteren, N., Jaramillo, F., Dekker, S. C., Destouni, G., & Lyon, S. W. (2014). Exploring hydroclimatic change disparity via the Budyko framework. Hydrological Processes, 28(13), 4110-4118.

van der Velde, Y., Heidbüchel, I., Lyon, S. W., Nyberg, L., Rodhe, A., Bishop, K., & Troch, P. A. (2015). Consequences of mixing assumptions for time-variable travel time distributions. Hydrological Processes, 29(16), 3460-3474.

von Freyberg, J., Allen, S. T., Seeger, S., Weiler, M., & Kirchner, J. W. (2018). Sensitivity of young water fractions to hydro-climatic forcing and landscape properties across 22 Swiss catchments. Hydrology and Earth System Sciences, 22(7), 3841-3861.

van Oorschot, F., van der Ent, R. J., Hrachowitz, M., & Alessandri, A. (2021). Climate controlled root zone parameters show potential to improve water flux simulations by land surface models. Earth System Dynamics Discussions, 1-26.

Wang-Erlandsson, L., Bastiaanssen, W.G., Gao, H., Jägerrmeyr, J., Senay, G.B., Van Dijk, A., ... & Savenije, H.H. (2016). Global root zone storage capacity from satellite-based evaporation.

Wang-Erlandsson, L., Fetzer, I., Keys, P. W., Van Der Ent, R. J., Savenije, H. H., & Gordon, L. J. (2018). Remote land use impacts on river flows through atmospheric teleconnections. Hydrology and Earth System Sciences, 22(8), 4311-4328.

Wehrli, K., Guillod, B.P., Hauser, M., Leclair, M., & Seneviratne, S.I. (2019). Identifying Key Driving Processes of Major Recent Heat Waves. Journal of Geophysical Research: Atmospheres, 124(22), 11746-11765.

Weiler, M., & Naef, F. (2003). An experimental tracer study of the role of macropores in infiltration in grassland soils. Hydrological Processes, 17(2), 477-493.

Weiler, M., & McDonnell, J. J. (2007). Conceptualizing lateral preferential flow and flow networks and simulating the effects on gauged and ungauged hillslopes. Water resources research, 43(3).

Wiekenkamp, I., Huisman, J.A., Bogena, H.R., Graf, A., Lin, H.S., Drüe, C., & Vereecken, H. (2016). Changes in measured spatiotemporal patterns of hydrological response after partial deforestation in a headwater catchment. Journal of hydrology, 542, 648-661.

Wilusz, D. C., Harman, C. J., & Ball, W. P. (2017). Sensitivity of catchment transit times to rainfall variability under present and future climates. Water Resources Research, 53(12), 10231-10256.

Yadav, M., T. Wagener, and H. V. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Adv. Water Resour., 30, 1756–1774.

Yang, Y., Donohue, R.J., & McVicar, T.R. (2016). Global estimation of effective plant rooting depth: Implications for hydrological modeling. Water Resources Research, 52(10), 8260-8276.

Yilmaz, K. K., H. V. Gupta, and T. Wagener (2008), A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resour. Res., 44, W09417, doi:10.1029/2007WR006716.

Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., ... & Bens, O. (2011). A network of terrestrial environmental observatories in Germany. Vadose Zone Journal, 10(3), 955-973.

Zehe, E., Lee, H., & Sivapalan, M. (2006). Dynamical process upscaling for deriving catchment scale state variables and constitutive relations for meso-scale process models.

Zehe, E., Elsenbeer, H., Lindenmaier, F., Schulz, K., & Blöschl, G. (2007). Patterns of predictability in hydrological threshold systems. Water Resources Research, 43(7).

Zhang, M., Liu, N., Harper, R., Li, Q., Liu, K., Wei, X., ... & Liu, S. (2017). A global review on hydrological responses to forest change across multiple spatial scales: Importance of scale, climate, forest type and hydrological regime. Journal of Hydrology, 546, 44-59.

Zuber, A. (1986). On the interpretation of tracer data in variable flow systems. Journal of Hydrology, 86(1-2), 45-57.

Table 1: Water balance, state and flux equations used in the hydrological model. Symbols shown in bold are model parameters. Subscripts H and R indicate hillslope and riparian zone, respectively. Model variables: P is total precipitation [mm d^{-1}], P_S is solid precipitation (snow)]mm d^{-1}], P_M is snow melt [mm d^{-1}], P_R is rain [mm d^{-1}], P_E is ~~effective precipitation~~throughfall [mm d^{-1}], E_P is potential evaporation [mm d^{-1}], E_I is interception evaporation [mm d^{-1}], R_F is preferential recharge [mm d^{-1}], R_S is slow recharge [mm d^{-1}], E_T is transpiration [mm d^{-1}], Q_S is flow from slow responding reservoir [mm d^{-1}], Q_R is flow from the fast responding riparian reservoir [mm d^{-1}], Q is the total flow [mm d^{-1}] and E_A is the total actual evaporation [mm d^{-1}]. Model parameters: T_T is the threshold temperature [^{o}C], F_M is a melt factor [mm d^{-1} ^{o}C^{-1}], I_{max} is the interception capacity [mm], $S_{U,max}$ is the root-zone storage capacity [mm], γ is a shape factor [-], $R_{S,max}$ is the maximum percolation rate [mm d^{-1}], L_p is a transpiration water stress factor [-], f_{QS} is a factor determining the fraction of groundwater flow that is upwelling into the riparian zone [-], k_S is the storage coefficient of the slow responding reservoir [d^{-1}], k_R is the storage coefficient for the fast responding riparian reservoir [d^{-1}] and f is the areal fraction of the riparian zone [-].

Landscape unit	Storage component	Water balance	Eq.	Constitutive equations	Eq.
	Snow storage	$dS_{snow}/dt = P_S - P_M$	(8)	$P_S = \begin{cases} P, & T < \boldsymbol{T_T} \\ 0, & T \geq \boldsymbol{T_T} \end{cases}$	(15)
				$P_M = \begin{cases} 0, & T < \boldsymbol{T_T} \\ min\left(\boldsymbol{F_M}(T - \boldsymbol{T_T}), \frac{S_{snow}}{dt}\right), & T \geq \boldsymbol{T_T} \end{cases}$	(16)
Hillslope	Interception storage	$dS_{I,H}/dt = P_R + P_M - P_{E,H} - E_{I,H}$	(9)	$P_R = \begin{cases} 0, & T < \boldsymbol{T_T} \\ P, & T \geq \boldsymbol{T_T} \end{cases}$	(17)
				$P_{E,H} = max\left(0, \frac{S_{I,H} - \boldsymbol{I_{max,H}}}{dt}\right)$	(18)
				$E_{I,H} = min\left(E_P, \frac{S_{I,H} - P_{E,H}}{dt}\right)$	(19)
	Unsaturated root-zone storage	$dS_{U,H}/dt = P_{E,H} - R_{F,H} - R_{S,H} - E_{T,H}$	(10)	$S'_{U,H} = (1 + \boldsymbol{\gamma})\boldsymbol{S_{U,max,H}}\left(1 - \left(1 - \frac{S_{U,H}}{\boldsymbol{S_{U,max,H}}}\right)^{\left(\frac{1}{1+\gamma}\right)}\right)$	(20)
				$R_{F,H} = P_{E,H} - \left(\boldsymbol{S_{U,max,H}} + S_{U,H} + \boldsymbol{S_{U,max,H}}\left(1 - \frac{P_{E,H}dt + S_{U,H}'}{(1 + \boldsymbol{\gamma})\boldsymbol{S_{U,max,H}}}\right)^{(1+\gamma)}\right)\boldsymbol{dt^{-1}}$	(21)
				$R_{S,H} = min\left(\boldsymbol{R_{S,max}}\frac{S_{U,H}}{\boldsymbol{S_{U,max,H}}}, \frac{S_{U,H}}{dt}\right)$	(22)
				$E_{T,H} = min\left((E_P - E_{I,H})min\left(\frac{S_{U,H}}{\boldsymbol{S_{U,max,H}}}\frac{1}{\boldsymbol{L_P}}, 1\right), \frac{S_{U,H}}{dt}\right)$	(23)
	Slow responding storage	$dS_{S,a}/dt = (1 - \boldsymbol{f})(R_{F,H} + R_{S,H}) - R_{S,R} - Q_S$	(11)	$R_{S,R} = \boldsymbol{f_{QS}}S_{S,a}(1 - e^{-\boldsymbol{k_S}t})dt^{-1}$	(24)
				$Q_S = (1 - \boldsymbol{f_{QS}})S_{S,a}(1 - e^{-\boldsymbol{k_S}t})dt^{-1}$	(25)
Riparian zone	Interception storage	$dS_{I,R}/dt = P_R + P_M - P_{E,R} - E_{I,R}$	(12)	$P_{E,R} = max\left(0, \frac{S_{I,R} - \boldsymbol{I_{max,R}}}{dt}\right)$	(26)
				$E_{I,R} = min\left(E_P, \frac{S_{I,R} - P_{E,R}}{dt}\right)$	(27)
	Unsaturated root-zone storage	$dS_{U,R}/dt = P_{E,R} + R_{S,R}/\boldsymbol{f} - R_{F,R} - E_{T,R}$	(13)	$S'_{U,R} = (1 + \boldsymbol{\gamma})\boldsymbol{S_{U,max,R}}\left(1 - \left(1 - \frac{S_{U,R}}{\boldsymbol{S_{U,max,R}}}\right)^{\left(\frac{1}{1+\gamma}\right)}\right)$	(28)
				$R_{F,R} = P_{E,R} + \frac{R_{S,R}}{f} - \left(\boldsymbol{S_{U,max,R}} + S_{U,R} + \boldsymbol{S_{U,max,R}}\left(1 - \frac{P_{E,R}dt + S_{U,R}'}{(1 + \boldsymbol{\gamma})\boldsymbol{S_{U,max,R}}}\right)^{(1+\gamma)}\right)\boldsymbol{dt^{-1}}$	(29)
				$E_{T,R} = min\left((E_P - E_{I,R})min\left(\frac{S_{U,R}}{\boldsymbol{S_{U,max,R}}}\frac{1}{\boldsymbol{L_P}}, 1\right), \frac{S_{U,R}}{dt}\right)$	(30)
	Fast responding storage	$dS_{F,R}/dt = R_{F,R} - Q_R$	(14)	$Q_R = S_{F,R}(1 - e^{-\boldsymbol{k_R}t})dt^{-1}$	(31)
				$Q = Q_S + \boldsymbol{f}Q_R$	(32)
				$E_I = (1 - f)E_{I,H} + fE_{I,R}$	(33)
				$E_T = (1 - f)E_{T,H} + fE_{T,R}$	(34)
				$E_A = E_I + E_T$	(35)

Table 2: Parameter prior distributions and 5/95th percentiles of the posterior distributions. Note that *) parameter *f*, characterizing the areal proportion of the riparian zone was fixed according to soil and elevation data and **) the interception capacity I_{max} was assumed to be identical on the hillslopes and the riparian zone in the pre-deforestation period.

Model	Parameter	Prior distribution	Posterior distribution	
			Pre-deforestation	Post-deforestation
Hydrological model	f [-]*	0.1	0.1	0.1
	F_M [mm d^{-1} ^{o}C^{-1}]	1.0 – 5.0	2.0 – 4.8	1.4 – 4.7
	f_{QS} [-]	0.00 – 0.20	0.02 – 0.11	0.01 – 0.11
	$I_{max,H}$ [mm]	0.0 – 6.0	1.9 – 4.8	2.5 – 4.1
	$I_{max,R}$ [mm]**	0.0 – 6.0	1.9 – 4.8	0.1 – 1.3
	k_R [d^{-1}]	0.01 – 2.00	0.26 – 1.28	0.29 – 1.01
	k_S [d^{-1}]	0.01 – 0.20	0.02 – 0.15	0.03 – 0.17
	L_p [-]	0.0 – 1.0	0.2 – 0.8	0.1 – 0.3
	$R_{S,max}$ [mm d^{-1}]	0.0 – 4.0	0.5 – 2.8	0.9 – 3.1
	$S_{U,max,H}$ [mm]	0 – 400	213 – 311	137 – 270
	$S_{U,max,R}$ [mm]	0 – 400	186 – 280	92 – 190
	T_T [^{o}C]	-1.5 – 1.5	-0.6 – 1.1	-0.2 – 1.1
	γ [-]	0.0 – 5.0	0.2 – 1.0	0.5 – 4.3
Tracer model	α_0 [-]	0.00 – 1.00	0.80 – 0.99	0.61 – 0.96
	$S_{S,p}$ [mm]	1000 – 30 000	7999 – 16228	7612 - 13920

Table 3: Signatures of flow and $\delta^{18}O$ and the associated performance metrics used for model calibration and evaluation. The performance metrics used include the Nash-Sutcliffe efficiency (E_{NS}), the volume error (E_V) and the relative error (E_R).

Variable/Signature	Symbol	Performance Metric	Reference
Time series of flow	Q	$E_{NS,Q}$	Nash and Sutcliffe (1970)
	log(Q)	$E_{NS,log(Q)}$	
	Q	$E_{V,Q}$	Criss and Winston (2008)
Flow duration curve	FDC	$E_{NS,FDC}$	Jothityangkoon et al. (2001)
Flow duration curve high flow period	FDC,h	$E_{NS,FDCh}$	Yilmaz et al. (2008)
Peak distribution	PD	$E_{NS,PD}$	Euser et al. (2013)
Rising limb density	RLD	$E_{R,RLD}$	Shamir et al. (2005)
Declining limb density	DLD	$E_{R,DLD}$	Sawicz et al. (2011)
Autocorrelation function of flow	AC	$E_{NS,AC}$	Montanari and Toth (2007)
Lag-1 autocorrelation	AC1	$E_{R,AC1}$	Hrachowitz et al. (2014)
Lag-1 autocorrelation low flow period	AC1,l	$E_{R,AC1,l}$	Fovet et al. (2015)
Runoff ratio	CR	$E_{R,CR}$	Yadav et al. (2007)
Time series of $\delta^{18}O$ in stream water	$\delta^{18}O$	$E_{NS,\delta18O}$	Birkel et al. (2011a)
Damping ratio of $\delta^{18}O^{*)}$	RD	$E_{R,RD}$	

$^{*)}RD = \frac{stdev_Q(\delta^{18}O)}{stdev_P(\delta^{18}O)}$

Figure 1: Map of the Wüstebach study catchment showing the spatial distribution of soil types. The riparian zone is defined by the parts of the catchment covered by Gleysols, Planosols and Halfbogs. The red line indicates the outline of the deforested part of the catchment, as can also be seen on the aerial images (Google Earth, Maxar Technologies 2020) from 2013 and 2016.

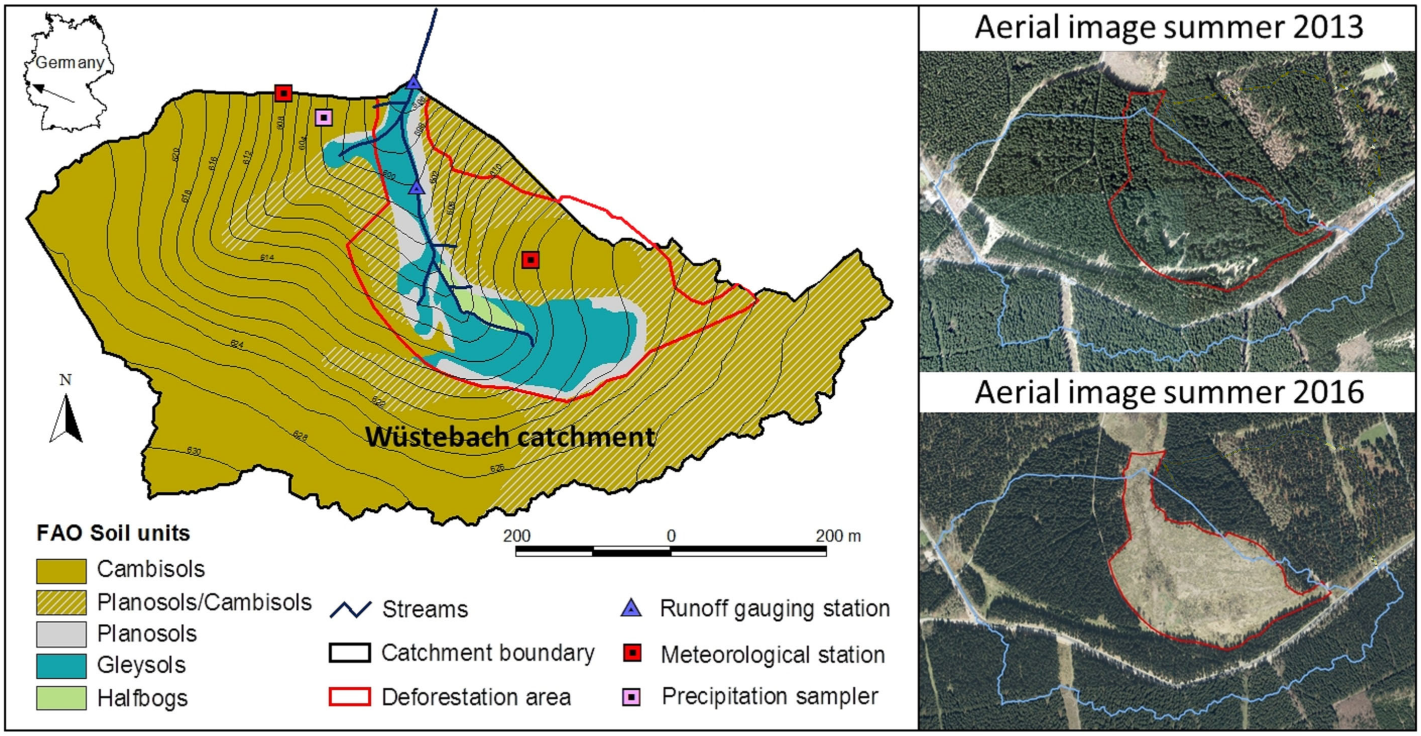


Figure 2: (a) Time series of observed weekly precipitation *P*; (b) daily cumulative evaporative fluxes for the pre- and post-deforestation period, where the dark brown line indicates potential evaporation E_P and the orange lines and the yellow shaded areas show the actual evaporation E_A modelled using the best fit parameter sets and the associated 5/95th percentiles of all feasible solutions of the pre- and post-deforestation periods, respectively. The dashed red line indicates the modelled E_A in the post-deforestation period using the best fit pre-deforestation parameter set; (c) observed (dark blue line) and modelled daily stream flow *Q*; light blue line indicates best fit model and the shaded area the 5/95th percentile of all feasible solutions for the pre- and post-deforestation periods, respectively. The dashed red line indicates the modelled stream in the post-deforestation period using the best fit pre-deforestation parameter set; (d) zoom-in to the observed and modelled stream flow for the 10/2012 – 10/2014 period. The grey shaded area indicates the deforestation period.

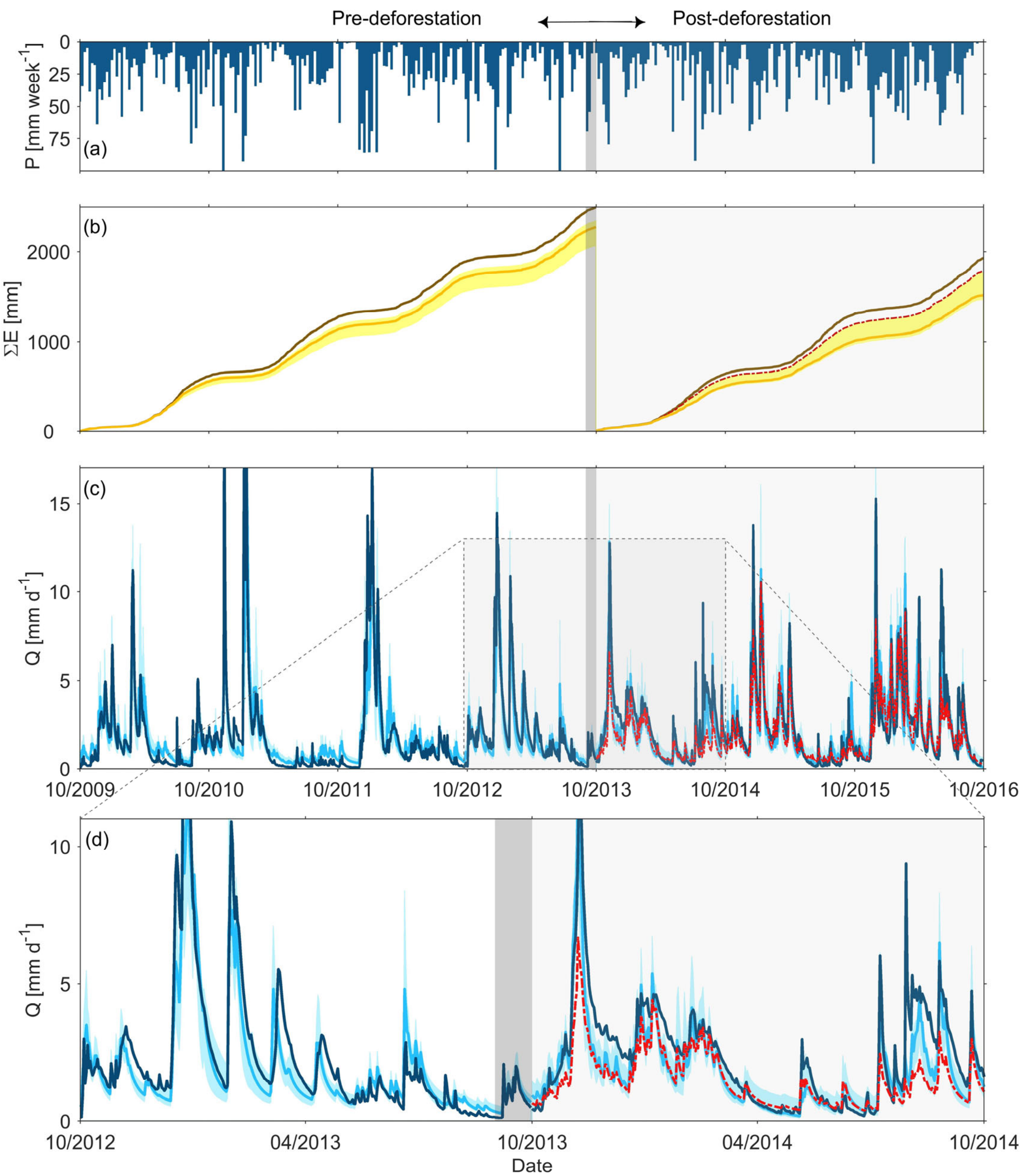


Figure 3: (a) Observed volume weighted monthly δ^{18}O signals in precipitation (grey dots; size of dots indicates the precipitation volume) and stream flow (green dots) as well as the best fit modelled δ^{18}O signal in the stream (green line) and the 5/95th percentile of all feasible solutions from pre- and post-deforestation calibration (green shaded area); (b) zoom-in of observed and modelled δ^{18}O signal in the stream for the 10/2012 – 10/2014 period.

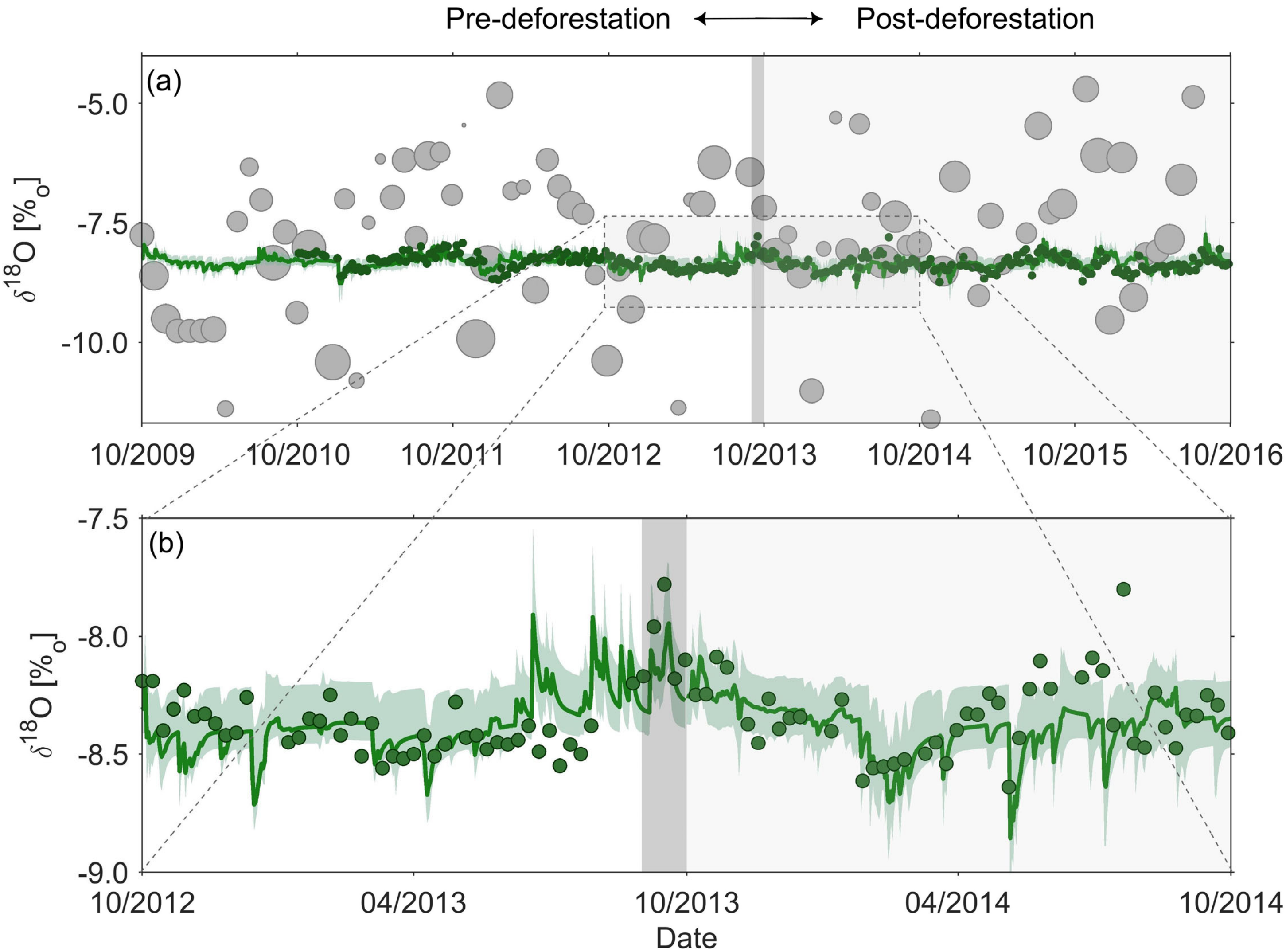


Figure 4: Model structure used in this study. The light blue boxes indicate the hydrologically active individual storage volumes in the hillslope and riparian zones, respectively. The darker blue box $S_{S,p}$ indicates a hydrologically passive, i.e. d$S_{S,p}$/dt = 0, mixing volume. The blue lines indicate liquid water fluxes, the green lines indicate vapour fluxes. Model parameters are shown in red, adjacent to the model component they are associated with. All symbols are defined in Table 1.

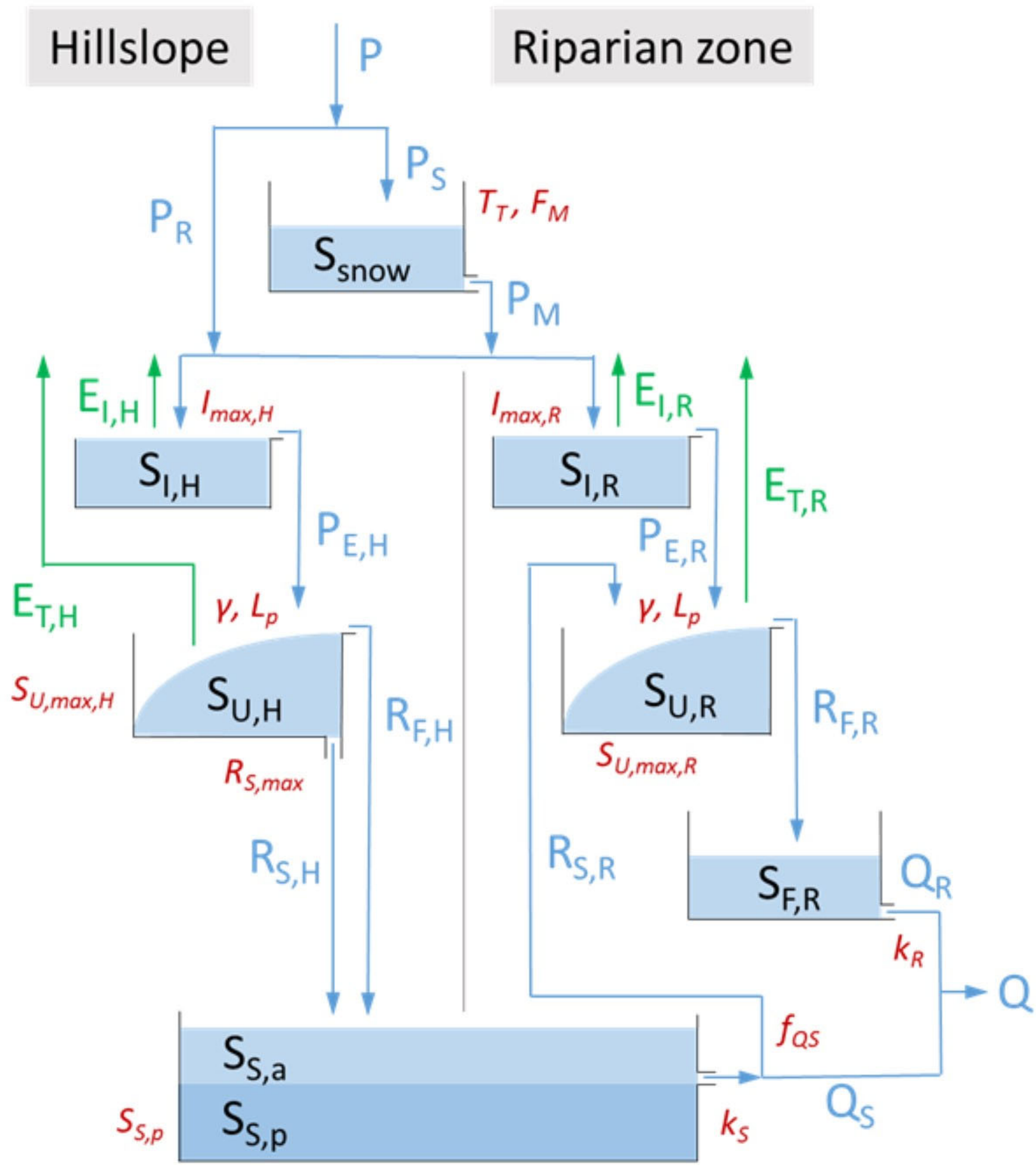


Figure 5: (a) Positions of the individual years of the study period in the Budyko framework. The x-axis shows the aridity index $I_A = E_P/P$, the y-axis indicates the evaporative ration E_A/P and the runoff ratio $C_R = 1 - E_A/P$. Pre-deforestation years are shown with blueish shades, post-deforestation years with greenish shades. The bold black lines indicate the energy and water limits, respectively. The dashed grey line is the theoretical-analytical Turc-Mezentsev relationship (Turc, 1954; Mezentsev, 1955). (b) The range of time series of storage deficits as computed according to equation 2, using values of I_{max} from 0 to 4 mm. The maximum annual storage deficits $S_{D,j}$ are indicated by the arrows. The grey shaded area indicates the deforestation period. (c) Estimation of $S_{U,max}$ as the storage deficit associated with a 40-year return period $S_{D,40yr}$ using the Gumbel extreme value distribution for the pre-deforestation period. The blueish dots indicate the range of maximum annual storage deficits $S_{D,j}$ for each year in the four year pre-deforestation period. The dark grey shaded area indicates the envelop of least-square fits for the individual values of I_{max}. The light grey shaded area indicates the envelope of the 5/95th confidence intervals. The red line shows the plausible range for $S_{U,max}$.

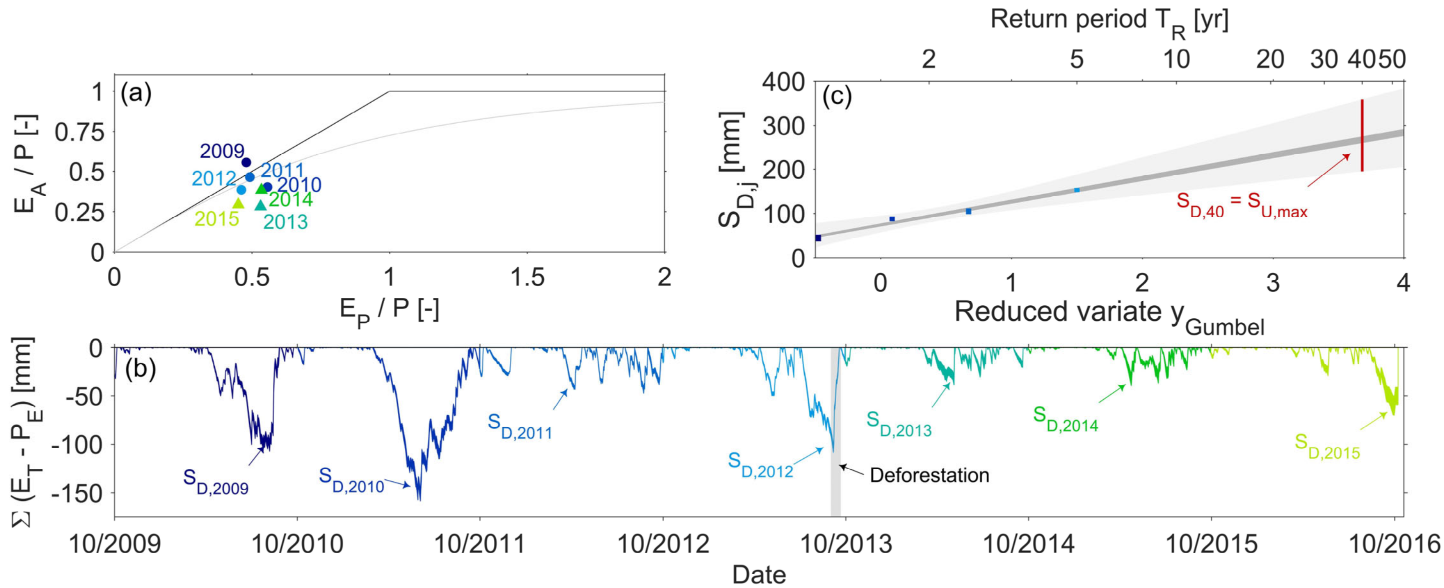


Figure 6: (a) Model performance metrics for all variables and signatures. D_E is the Euclidean distance to the perfect model. It combines all other performance metrics (Table 3) into one number (Eq.42). All performance metrics are formulated in a way that a value of 1 indicates a perfect fit. The boxplots summarize the performances of all parameter sets retained as feasible. The circle symbols indicate the performance of the best performing model in terms of D_E. The dark red shades indicate pre-deforestation model performance based on calibration in the pre-deforestation period. Orange shades indicate post-deforestation performance using the pre-deforestation parameter sets without further re-calibration. Yellow shades show the post-deforestation performance after model re-calibration in the post-deforestation period. (b)-(c) show flow duration curves, (d)-(e) show the peak distributions and (f)-(g) the autocorrelation functions for the pre- (red) and the post deforestation periods (orange and yellow), respectively. The black lines indicate the observed values, the dashed lines indicate the best fits and the shaded areas the 5/95th uncertainty interval of all solutions retained as feasible. The dark red shades indicate pre-deforestation model results based on calibration in the pre-deforestation period. Orange shades indicate post-deforestation model results using the pre-deforestation parameter sets without further re-calibration. Yellow shades show the post-deforestation model results after model re-calibration in the post-deforestation period.

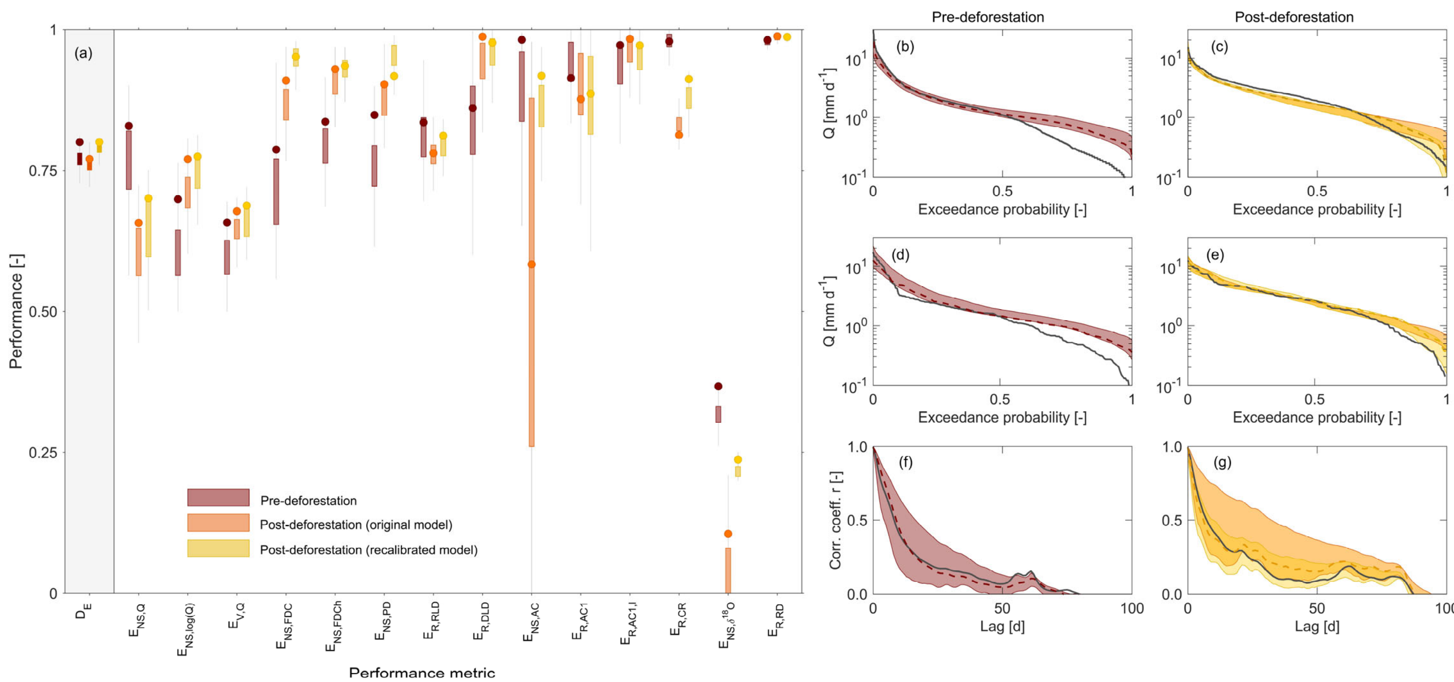


Figure 7: Observed mean P_E/P (dashed line), the range around observed mean P_E/P defined as acceptable (grey shaded area), the distribution of modelled mean P_E/P from all solutions that satisfy the behavioural thresholds for all performance metrics (Table 3) as well as the mean P_E/of the best solution in terms of D_E (orange symbol) for (a) the 2009-2013 pre-deforestation period and (b) the 2013-2016 post-deforestation period. Note, only modelled solutions (yellow) that fall into the acceptable observed range (grey shaded) are kept as feasible. The fractions of time steps in the pre-deforestation (c) and the post-deforestation periods (d) in which the modelled relative soil moisture $S_{U,rel}$ falls within the pre-defined acceptable range around the observed relative catchment-average soil moisture. The blue symbols indicates the best solution in terms of D_E, the distributions indicate the set of solutions that satisfy the behavioural thresholds for all performance metrics (Table 3). The grey shaded areas indicate the region of acceptable solutions, i.e. solutions that fall at least 75% of the time steps into the acceptable interval. the fractions time steps. Note that only modelled solutions (light blue) that fall into the acceptable observed range (grey shaded) are kept as feasible. Pre-deforestation (e) and post-deforestation (f) time series of the acceptable range around the observed normalized, relative soil moisture (light grey shade), range of modelled normalized relative soil moisture for all solutions that satisfy all performance metrics ("unconstrained"; light blue) and for the set of feasible solutions that satisfy both, P_E/P and soil moisture constraints as shown in (a)-(d) ("constrained"; blue). The dark blue line indicates the modelled normalized, relative soil moisture of the best solution in terms of D_E.

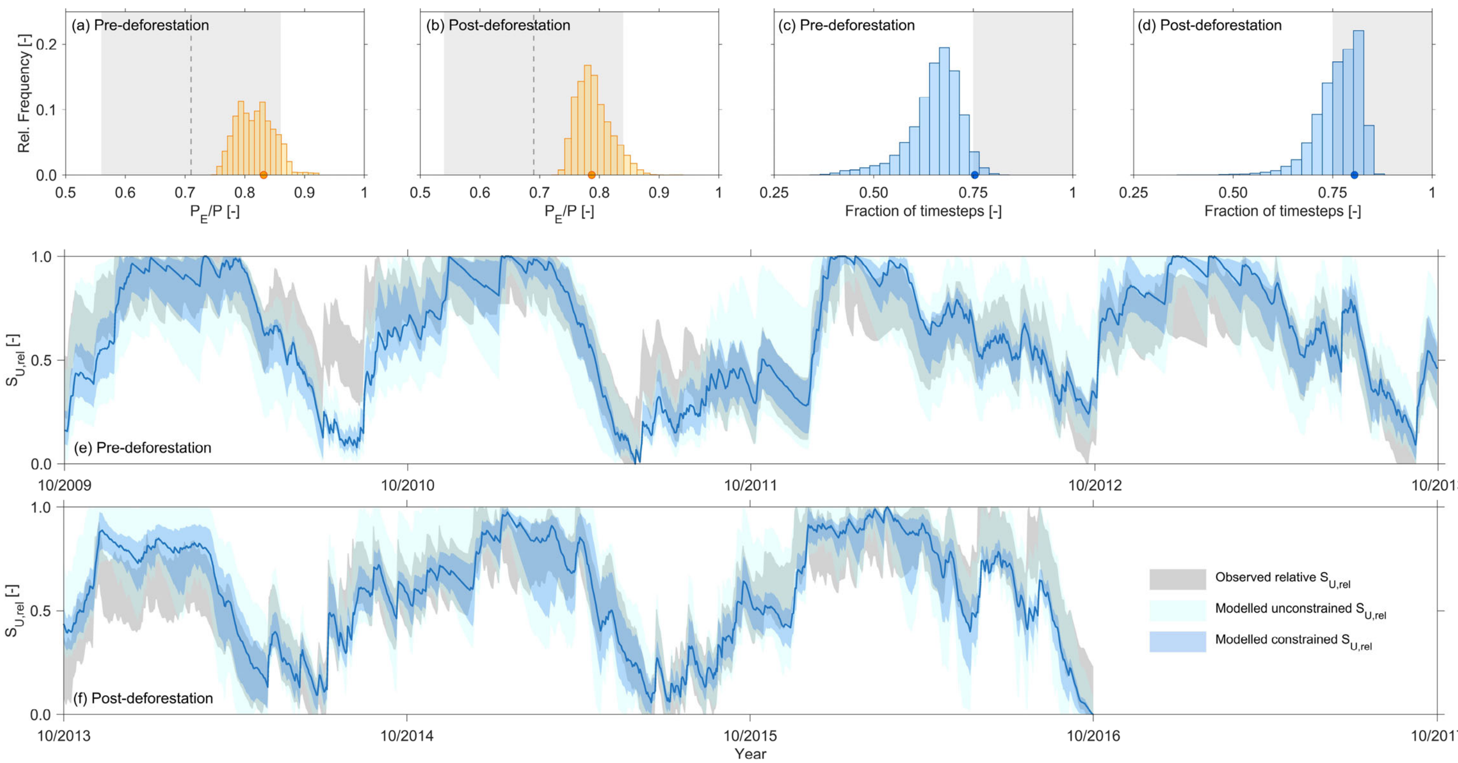


Figure ~~7~~8: Posterior distributions of selected parameters shown as empirical cumulative distribution function (lines) and the associated relative frequency distributions (bars). Red shades indicate calibration in the pre-deforestation period, Yellow shades indicate post-deforestation calibration. The dots indicate the parameter values associated with respective best fit models.

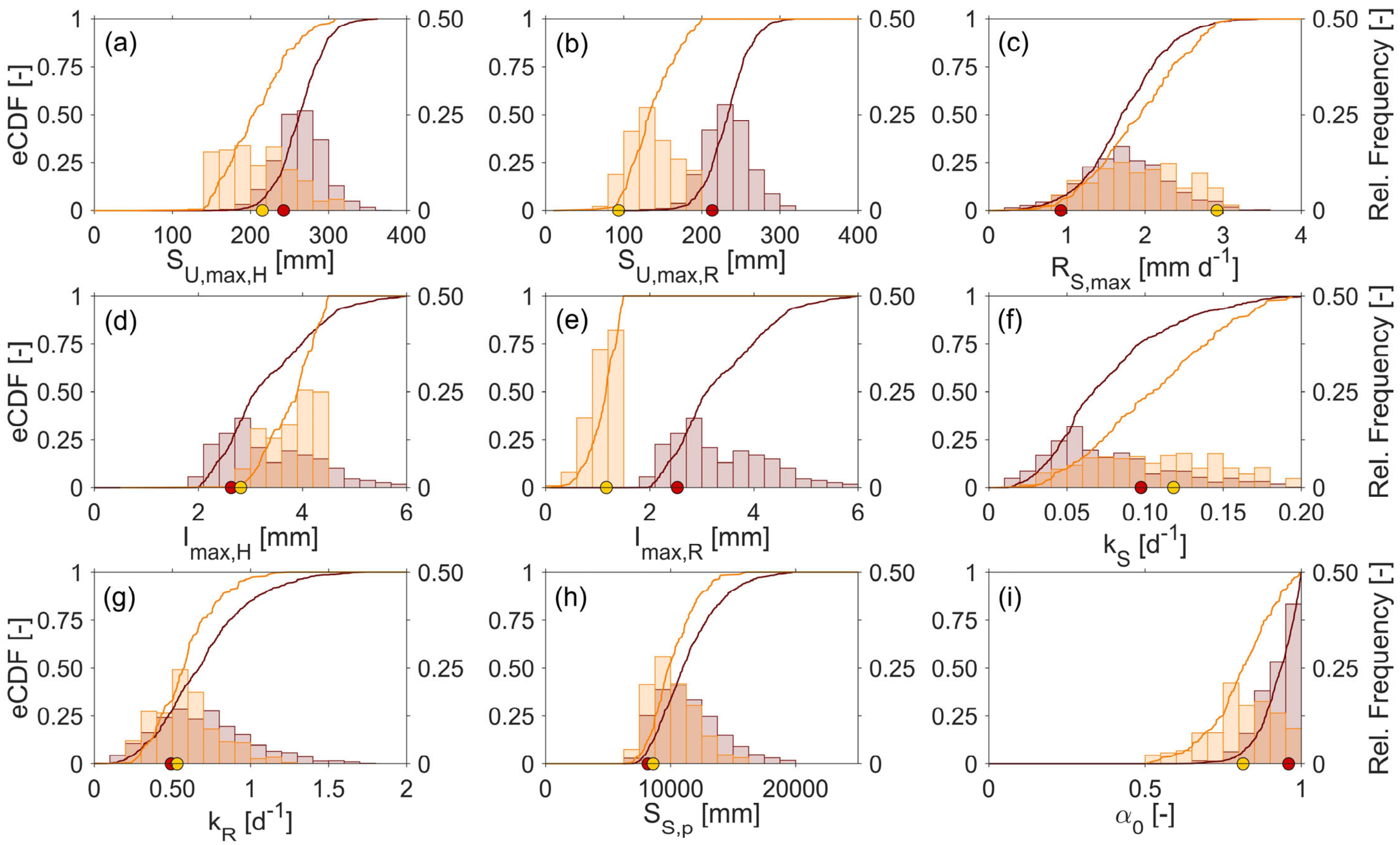


Figure ~~8~~9: Panels in the left column show pre-deforestation (a) discharge, the coloured dots indicate to which period (dry, wet-up, wet, drying) the individual selected time steps belong; (b) the 5/95th percentiles of the empirical cumulative TTDs for wet (blue) and dry (red) periods, respectively; (c) the ensemble of the individual TTDs at the time steps indicated in (a). Panels in the middle column (d-g) compare the 5/95th percentiles of empirical cumulative TTDs between pre-deforestation (dark shades) and post-deforestation (light shades) periods for dry, wet-up, wet and drying conditions, respectively. Panels in the right column show post-deforestation (h) discharge, the coloured dots indicate to which period (dry, wet-up, wet, drying) the individual selected time steps belong; (i) the 5/95th percentiles of the empirical cumulative TTDs for wet (blue) and dry (red) periods, respectively; (j) the ensemble of the individual TTDs at the time steps indicated in (h). All distributions shown are truncated at 3 (post-deforestation) for 4 years (pre-deforestation), which coincides with the tracked period. For the remaining fractions, i.e. the difference to 1, it can only be said that they are older than 3 years but nothing more than that. The grey shaded areas indicate regions with ages > 3 months, thereby exceeding F_{yw}.

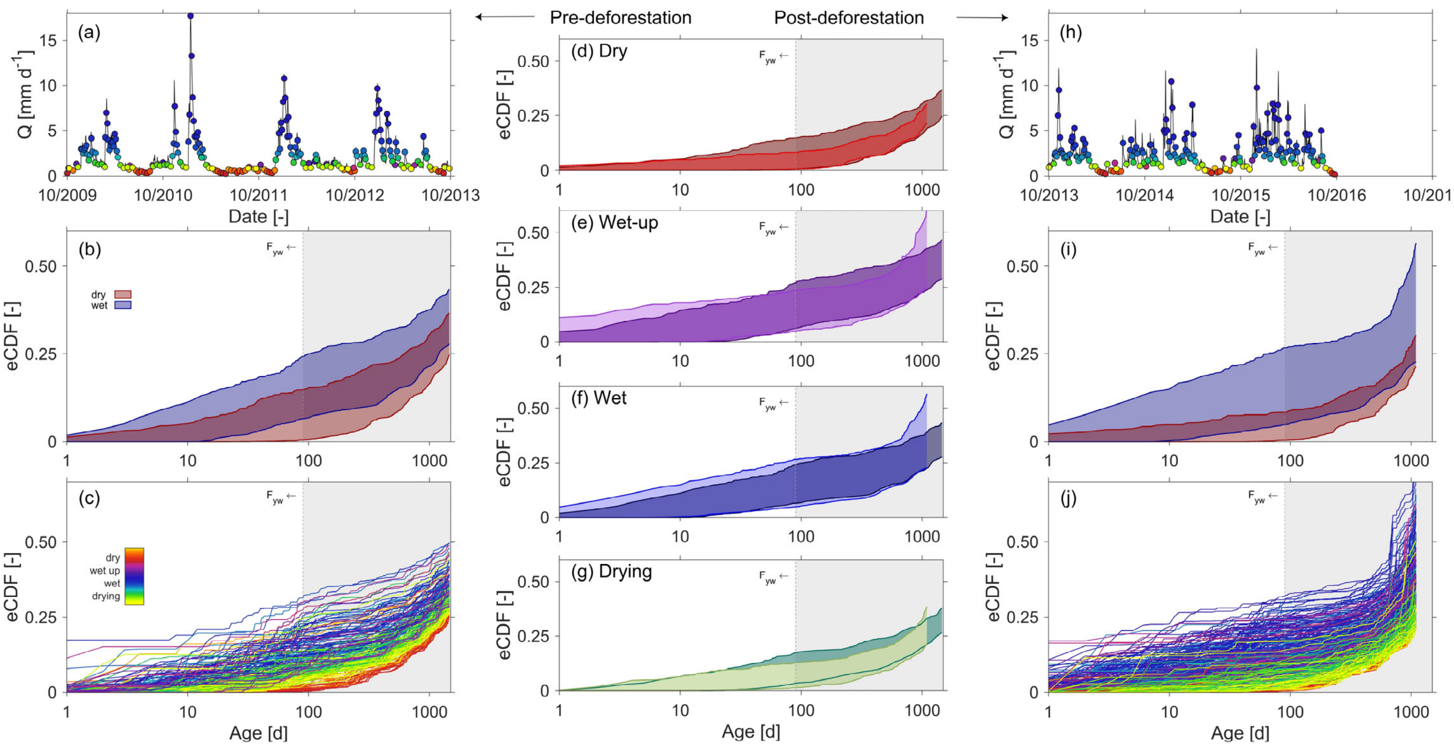


Figure ~~9~~10: (a)-(b) Pre- and post-deforestation time series of young water fractions F_{yw} in discharge. The colour code indicates the transition between dry, wetting-up, wet and drying conditions. The bold black line shows the mean F_{yw} of the best model fit, the grey shaded area shows the 5/95th percentile of F_{yw} for all feasible model solutions; (c)-(d) pre- and post-deforestation sensitivity of F_{yw} to discharge, using the same colour code as above to indicate dry, wetting-up, wet and drying conditions. The arrows in (d) indicate if there are statistically significant (↑; $p < 0.05$) changes or not (↔) in the sensitivities between the post-deforestation period and the pre-deforestation period.

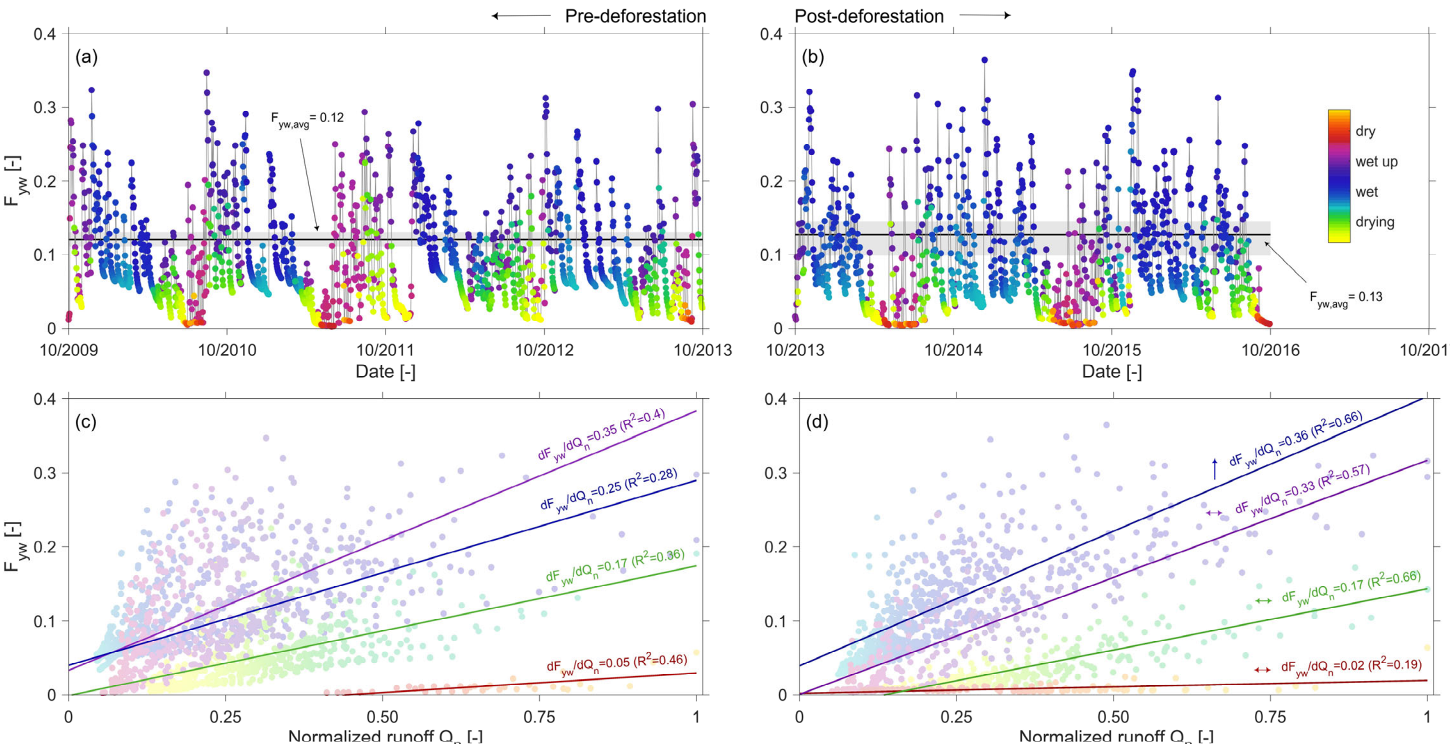


Figure ~~10~~11: Individual catchment overall SAS ω-functions for individual time steps under different wetness conditions in the (a) pre-deforestation period and (b) post-deforestation period. The insets show the relative water content in $S_{U,rel,mod} = S_U/S_{U,max}$ at the individual time steps.