Comment:

This paper analyses the results of a deforestation experiment on the Wüstebach experimental catchment using conceptual flow and transport models. Overall the paper uses interesting approaches, but needs better structuring, and a more appropriate title.

Reply:

We thank the reviewer for the overall positive assessment of our study as well as for her/his thoughtful and detailed comments.

Comment:

This paper essentially focuses on the Wüstebach catchment, a 0.39 km2 catchment in Germany. Clearly this catchment is a small and specific area. Whatever is found on this catchment cannot have general relevance, considering the place and scale dependence of hydrological processes. The title, however, is very trenchant and general, which contrasts with the specificity of the case study. I suggest a more specific title, more reflective of the individual headwater catchment that has been used in the analyses.

Reply:

On re-reading the title we agree with the reviewer and realize that the title is more general than intended. To better reflect the location specificity but also to place more emphasis on the actual objective and innovative aspect of this study we will reformulate the title into: "Reduction of vegetation-accessible water storage capacity after deforestation affects travel time distributions and increases young water fractions in a headwater catchment."

Comment:

The introduction is very general, and projects a status quo that is much broader of what is needed to introduce this specific study. The readers unavoidably end up asking themselves: what is already known about this specific catchment? The Wüstebach catchment has been the object of a countless number of studies, which analysed the results of the deforestation experiment, and modelled its behaviour using many modelling approaches (https://experimental-hydrology.net/wiki/index.php?title=W%C3%BCstebach_long-term_experimental_catchment). In particular, Wiekenkamp et al. (2016) already analysed issues related to water balance, potential evaporation, and water storage associated to deforestation. I can see that the authors here use different methods. However, that in this catchment "Deforestation reduces the vegetation-accessible water storage in the unsaturated soil" (part of the paper title) is already clear from Wiekenkamp et al. (2016), and other studies (e.g. works by Stockinger) have analysed the isotope data and evaluated MRT. These references are cited in the current manuscript. But they are not discussed to provide a clear motivation for the current paper, and to justify the novelty of the results.

Reply:

The reviewer is completely correct in saying that the Wüstebach study catchment has over the past years been subject to numerous studies. We highly appreciate the notion of the reviewer that for the reader the added value and innovative aspects of this study may not be completely clear yet. We agree that it will therefore be helpful for the reader to more explicitly describe how this study fits into the context of previous work in the Wüstebach and which knowledge it specifically adds.

*Briefly, the Wiekenkamp et al. (2016) paper quantified in an elegant study the effects of deforestation on the partitioning of water fluxes, based on discharge and eddy covariance observations. Overall, they found that deforestation decreased actual evaporation/transpiration, which led to higher levels of soil moisture and eventually to increases in discharge. In contrast, the aim of our study was to establish a quantitative mechanistic link between these observed decreases in transpiration and changes to sub-surfaces properties of the system (and thus model parameters) to provide an explanation of **why** deforestation reduces evaporative fluxes and increased soil moisture and discharge – after all the atmospheric water demand (here approximated by potential evaporation) remained stable pre- and post-deforestation.*

*Similarly, detailed previous work of some of our co-authors (Stockinger at al., 2019) quantified travel times and changes therein over time while Wiekenkamp et al. (2020) found evidence for increased post-deforestation occurrence of preferential flows. As an extension of that work, we here quantify the effect of deforestation on subsurface system properties and found that changes in these subsurface properties can to some extent explain **why** deforestation affects travel times and in particular young water ages in the Wüstebach.*

*Overall our results provide evidence that post-deforestation changes in the water balance and the hydrological response dynamics can, to a large part, be traced back to changes in the soil pore space that can be ***accessed*** by evaporative fluxes (i.e. evaporation and/or transpiration) to satisfy atmospheric water demand. Vegetation, through its root system, can access pores and efficiently extract water from them for transpiration in, relatively spoken, deep parts of the soil. Soil evaporation, in contrast, can only extract water from comparably much shallower parts of the soil, due to increasingly limited turbulent exchange with depth. In other words, after the top soil is dry, soil evaporation is largely "blocked" while root-water uptake from deeper soil layers for transpiration can continue. In case of deforestation, plants are removed and the associated transpiration stops. The pore space that was accessed by transpiration before, can after deforestation then not be accessed by evaporative fluxes (i.e. soil evaporation) anymore and soil moisture levels in these pores remains longer close to field capacity. Informally spoken, this pore volume is thus lost as "reservoir" for evaporative fluxes.*

*In contrast to previous studies, where this pore volume is either estimated as model calibration parameter or based on very scarce data of root depths and soil porosities, our results here suggest (1) that this root-accessible pore volume ($S_{U,max}$) and its change due to deforestation can be meaningfully estimated from water balance data at the catchment-scale, (2) that it is therefore also a meaningful and directly observable catchment-scale effective parameter for use in process-based models and, most importantly (3) that changes in that hydrologically relevant subsurface property/model parameter can explain much of **why** the hydrological response as well as travel times and young water fractions changed after deforestation.*

*We will provide a more detailed context of previous studies together with a more explicit definition of the objectives of this study and how they fit into the context and actually **link** the results of previous studies.*

Comment:

The motivation for the choice of methods is unclear. E.g. why conceptual models, and SAS are chosen for the problem at hand? Wouldn't the result be obtainable with much simpler methods? It seems to me that if I look at the abstract or conclusions, a simple water balance calculation, and simple hydrograph separation techniques could have been sufficient to end up with the same outcomes. I understand that the authors are proposing a more elaborate approach. But why is that necessary?

Reply:

The motivation for the choice of the modelling strategy, i.e. a process-based hydrological model with an integrated SAS-function based tracer module, is closely linked to our overall objective of this study, i.e. to develop a better understanding of ***why*** *the water balance, the hydrological response dynamics as well as the travel times in the Wüstebach catchment were subject to change after deforestation even though climatic conditions did largely not change. We wanted to understand (1) which (subsurface) system property was most affected by deforestation, (2) if this real-world property (or when used in models: parameter) and its change after deforestation can be meaningfully quantified at the catchment-scale and (3) if changes in this property can explain the observed post-deforestation changes in the water balance, response dynamics and travel times.*

The simpler methods referred to by the reviewer are powerful tools to quantify overall hydrological system characteristics. However, due to their simplicity they mostly need to remain process-agnostic and can therefore provide us only with limited information on the underlying processes and mechanistic reasons. We will clarify this in the revised manuscript.

Comment:

The authors mention that they use an extensive multi-objective strategy, but in the end they use a single objective function. True that the objective function aggregates multiple objectives, but this cannot be defined multi-objective optimization, which would require determining the Pareto-front between the various objectives.

Reply:

We respectfully but strongly disagree with this comment. Only because multiple performance metrics f (i.e. objectives and criteria) were compressed into one summary statistic does not mean that the parameter selection strategy was not based on multiple objectives.

*It is true that the solution space of a multi-objective optimization problem typically spans a front of pareto-optimal solutions. However, this pareto front only quantifies uncertainties arising from the choice and weight of and the trade-offs between the individual objectives. It does not account for data uncertainty and the idea of behavioural models, like in the GLUE framework (Beven and Binley, 1992). To embrace that we here did not treat *****only****** the actual pareto optimal solutions as feasible but also potential solutions behind the pareto front, i.e. we applied the GLUE method in a multi-dimensional case. This is schematically illustrated in Figure R1 below for a theoretical multi-objective evaluation based on two performance metrics f_1 and f_2 (note that the analysis of this manuscript was based on n=12 performance metrics instead, but based on the same reasoning).*

We first defined a performance threshold for each individual performance metric f. Only solutions that fell within that threshold for each performance metric were kept as behavioural. The resulting solution space then implicitly contains the pareto front.

Based on that we then further evaluated each solution that was kept as behavioural based on its multi-dimensional distance D_E from the "perfect model". The "best" solution is the one with the shortest distance D_E, thus inherently also being a point on the pareto-front. All other solutions accepted as feasible are further away from the "perfect model" and can be located on the pareto front or behind it.

If less (or in the extreme case only one) performance metrics than used in our analysis (n=12) had been used, the subset of retained behavioural solutions would be larger and their associated performances in terms of D_E different, which clearly demonstrates the effect of multiple performance metrics even when compressed into one summary statistic D_E. The above outlined strategy not only includes all pareto optimal solutions but also allows for additional uncertainty, thereby not only being defined as but actually extending and going beyond classic multi-objective parameter selection strategies as also similarly described and formalized by various recent papers (e.g. Efstradiadis and Koutsoyiannis, 2010; Gharari et al., 2013).

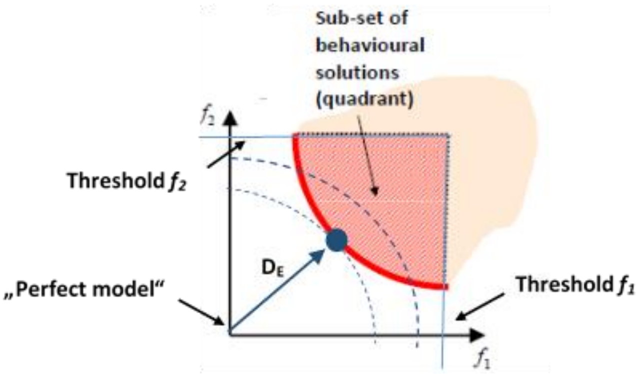

Figure R1: Schematic sketch of two-dimensional model performance space as defined by two performance metrics f_1 and f_2 (i.e. multi-objective). The theoretical "perfect model" would plot in the origin. The yellow shade indicates all model solutions. The light blue lines indicate the respective behavioural thresholds for f_1 and f_2 below which solutions are kept as behavioural and thus feasible. The red shaded area indicates the two-dimensional space of solutions accepted as behavioural. The bold red line indicates the set of pareto optimal solutions and the blue circle the single "best" solution with the lowest distance D_E to the "perfect model". The dashed blue circles-arcs indicate iso-performance levels in terms of D_E, i.e. solutions that have equal distances from the "perfect model" (after Gharari et al., 2013)

Comment:

I would have preferred to see separate results and discussion sections, to see separate the outcome of this work from the outcomes of other works. Currently the blend adds to the confusion of not being able to appreciate the value of the current work compared to earlier work on the specific catchment.

Reply:

We acknowledge and appreciate this comment. However, given the stepwise nature of the analysis we would strongly prefer to have the results of the individual steps closely associated to the interpretation thereof. However, we will strengthen and clarify the link to previous studies in the revised manuscript.

Comment:

Not clear to me why the Su,max of the model would be reflective of available water storage. For example, the model could have an Su,max of 200, but the variable storage between the reservoir can vary between e.g. 30 and 40 over one year, or between 10 and 150. So, Su,max sets an upper bound, but the real variability of the storage can be much smaller. The observation that 10.000 mm of water is necessary to attenuate the isotope signal suggests that indeed Su,max can be much larger than the dynamic range experienced by the catchment.

Reply:

The reviewer is absolute right in stating that $S_{U,max}$ is an upper bound. To further reply to this interesting comment it will be helpful to upfront remind us again of how $S_{U,max}$ is actually defined. In the original *manuscript, $S_{U,max}$ is described as the "pore volume between field capacity and permanent wilting point that is within the reach of active roots" and in the text referred to as "vegetation-accessible water storage **capacity**" and thus the **maximum** vegetation-accessible water storage volume in the unsaturated soil.*

As such, the upper limits of $S_{U,max}$ are by definition physically bound by

1. *Depth of groundwater table. Although fluctuating, the groundwater table in the study catchment 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), porosity of the silty clay loam (Bogena et al., 2018) soil of 0.4 and field capacity at a relative pore water content of 0.5 suggests an upper limit of $S_{U,max}$ at ~ 1000 mm.*
2. *Root-depths. However, actual roots are often shallower than these 5m 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 0.5 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 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), porosity of the silty clay loam (Bogena et al., 2018) soil of 0.4 and field capacity at a relative pore water content of 0.5 suggests an upper limit of $S_{U,max}$ at ~ 400 mm.*

*Thus given these estimates, we think it is very unlikely that $S_{U,max}$ is much larger than that. In addition, there is increasing and robust evidence that the magnitude of $S_{U,max}$ is actually based on optimality principles (e.g. Kleidon, 2004; Schymanski et al., 2008; Guswa et al., 2008; Yang et al., 2016; Speich et al., 2018 and many more): individual plants which have survived and are thus present in an ecosystem were well adapted to past conditions. This entails that they have root systems **large enough** to provide them with access to enough water, **but not more than that**. Unnecessary allocation of resources to subsurface growth (and maintenance) would have led to a lack of resources for above-surface growth. These plants would have eventually lost out in competition for light. By extension, the question of how much water was necessary to be stored (i.e. held below field capacity and thus against gravity) and accessible for vegetation (i.e. $S_{U,max}$) can then be inferred from what vegetation *actually* transpired under extreme conditions (here: dry spells with 20 year return periods), as done in this manuscript.*

*However, please also note that as mentioned above, here and in many other landscapes world-wide (e.g. Fan et al., 2017) most roots do not extend to the depth of the groundwater table. This leads to the presence of an unsaturated zone **below** the root zone. This zone is hydrologically largely "passive" as it is outside the influence of roots and given its relatively deep position essentially also outside the influence of soil evaporation: evaporative fluxes can therefore not extract water from pores in this part of the subsurface and water entering this zone only undergoes some delay before it is released from that zone again as groundwater recharge or direct drainage to a stream (e.g. through preferential flow paths). As no significant volumes of water can be drained below field capacity the unsaturated soil below the root zone remains close to field capacity for much of the year (except for the moments when a wetting front passes through). This is illustrated by the much lower variability in soil moisture content in these deeper soil layers (see e.g. Graf et al., 2014). Albeit being hydrologically largely passive (at least at time-scales of more than a few days $dS/dt \sim 0$), this transitional zone between the root zone and the groundwater provides a mixing volume for tracers. This was implemented in one of our recent* *studies (Hrachowitz et al., 2015). However, here we finally decided not to explicitly account for this effect as it requires 2 additional model calibration parameters, i.e. the size of the mixing volume and the time lag for releasing water. After intensive model testing it was found that these two parameters cannot be well constrained at all. In particular the size of the mixing volume showed considerable trade-offs with the parameter of the passive mixing volume* $S_{S,p}$ *here. Instead of explicitly representing this volume as individual model component, we therefore implicitly accounted for this additional mixing volume in* $S_{S,p}$. *It can thus be assumed that, in reality, part of the 10.000 mm of* $S_{S,p}$ *is actually a mixing volume in the transitional zone. It plausible to assume that here the magnitude of this volume does realistically not exceed ~500-600mm (i.e. the difference between the storage capacity above the groundwater and the root zone – see (1) and (2) above).*

We will clarify this in the revised manuscript.

Comment:

In terms of isotopes, it seems from the figure that there is an increase in the variability of the inputs. This leads to the question of how different are the inputs in the two periods, and whether the increase in young water fraction can be partly attributed to non-stationary inputs.

Reply:

The variability in the observed precipitation isotope inputs indeed slightly increased after deforestation (mostly due to higher sampling frequency) from a standard deviation of ~3.2 ‰ to ~3.8 ‰, while the stream water isotope variability remained stable at a standard deviation of ~0.2 ‰. This entails that the damping ratio between precipitation slightly increased and the input signal was thus proportionally more strongly damped. Following our current understanding of tracer dynamics, a higher damping ratio is associated with older water (e.g. McGuire and McDonnell, 2006). We therefore believe that, if this slight change in variability had an effect, it would rather ***decrease*** *than increase young water fractions.*

Comment:

From the uncertainty analysis, it appears that some parameters, e.g. Rs,max, are poorly constrained. But I am guessing that this parameter can strongly affect the behaviour of the Su reservoir, which is the key storage analysed in this paper. Any comment on this?

Reply:

As correctly observed by the reviewer, some model parameters are less well constrained than others, which remains a problem in essentially all inverse model implementations in environmental sciences. It is also true that the specifically mentioned parameter $R_{S,max}$ *has some influence on* $S_{U,max}$. *However, the results suggest that the absolute magnitudes of the associated flux* $R_{S,H}$ *are too low (i.e. < 10% of the water balance) to significantly influence* $S_{U,max}$. *This is further supported by the fact that* $S_{U,max}$ *is, for conceptual model standards, very well constrained: no matter the value of* $R_{S,max}$, $S_{U,max}$ *remains within the same narrow range. We will add this to the discussion in the revised manuscript.*

References:

Beven, K., & Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. Hydrological processes, 6(3), 279-298.

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).

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.

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.

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.

Gharari, S., Hrachowitz, M., Fenicia, F., & Savenije, H. H. G. (2013). An approach to identify time consistent model parameters: sub-period calibration. Hydrology and Earth System Sciences, 17(1), 149.

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.

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

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

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

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

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.

Schmid, I., & Kazda, M. (2001). Vertical distribution and radial growth of coarse roots in pure and mixed stands of Fagus sylvatica and Picea abies. Canadian Journal of Forest Research, 31(3), 539-548.

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.

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.

Stockinger, M. P., Bogena, H. R., Lücke, A., Stumpp, C., & Vereecken, H. (2019). Time variability and uncertainty in the fraction of young water in a small headwater catchment. Hydrology & Earth System Sciences, 23(10).

Wiekenkamp, I., Huisman, J. A., Bogena, H. R., Lin, H. S., & Vereecken, H. (2016). Spatial and temporal occurrence of preferential flow in a forested headwater catchment. Journal of hydrology, 534, 139-149.

Wiekenkamp, I., J.A. Huisman, H.R. Bogena, and H. Vereecken (2020): Spatiotemporal Changes in Sequential and Preferential Flow Occurrence after Partial Deforestation. Water 12(1), 35.

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.