Reply (in blue) to « RC1: 'Comment on hess-2023-152', Anonymous Referee #1, 24 Jul 2023 (https://doi.org/10.5194/hess-2023-152-RC1) » (*in italic*)

This works provides an analysis of tritium observations to estimate groundwater mean transit times in the Luxembourg Sandstone aquifer. The analysis benefits from the interpretation of a large tritium observation dataset that is distributed in both space and time. The authors apply an uncertainty analysis framework to quantify some of the salient sources of uncertainty in estimating mean groundwater transit times from the tritium observations. A powerful component of this work is the assimilation of spring discharge response times and major ion solute chemistry to regularize the tritium analysis. The primary conclusion from the analysis is that groundwater mean transit times can successfully be interpreted at this field study site, despite numerous sources of uncertainty and short observation timeseries. While the manuscript is easy to follow, it often lacks clarity and work is needed to update the language.

We thank Referee #1 for reviewing our manuscript and providing several constructive suggestions. Despite overall positive feedback, Referee #1 did also express some concerns about our work. We believe that her/his main concerns are partly based on misunderstandings. In this document, we provide our detailed responses to Referee #1's comments and also mention how we plan to address them in a revised version of this manuscript.

My main concern is that this work does not provide novel nor significant insights into the interpretation of tritium observations nor hydrologic processes at the investigated site. Rather, it is generally an application of established methods to estimate mean groundwater transit times. There is little interpretation into what the inferred mean transit times add to our understanding of this aquifer.

We are sorry that Referee #1 did not perceive the significant/novel aspects associated with our work.

As the reviewer rightly said, and our article does not contradict this, our work applies existing methods for estimating mean groundwater transit times using tritium. However, our study focuses on investigating new sites, whereas recent studies using tritium for transit time assessments in the Northern Hemisphere have essentially focused on sites that already have historical tritium records and thus benefit from longer time series (see lines 77-82 of our manuscript). This is one of the novel aspects of our work. We show how to face current limitations of tritium dating in the northern hemisphere for characterising new groundwater sites. This is in our opinion valuable for scientists that may like to assess transit times of young groundwater bodies which lack historical tritium data in this part of the world.

While Referee #1 claims the opposite, we also believe that our study provides key learnings on hydrological processes in the Luxembourg Sandstone aquifer. Our assessment of the mean transit times of the springs of Luxembourg city and their relationship with both the vadose zone and the saturated zone of the Luxembourg Sandstone aquifer, as significantly supported by several other types of data, are in our opinion an undeniable asset for better monitoring, regulating, and managing the transport of contaminants in this key drinking water resource.

Furthermore, it is not clear whether the uncertainty analysis applied in this work results in a substantial advantage over the already established tritium interpretation techniques, many of which also propagate uncertainties.

As no reference is given, we were unfortunately unable to target the "already established tritium interpretation techniques propagating uncertainties" that Referee #1 is referring to here. To our knowledge, tritium dating studies using lumped parameter models typically follow a basic best-fit calibration approach that does not incorporate tritium analytical uncertainties (which is not appropriate in the context of our study, as argued in the manuscript section 3.1.4, lines 233-243).

Major Points:

While this work propagates the analytic tritium errors in the uncertainty analysis, I feel that it misses the key point that mean groundwater transit times can be non-unique because tritium is insensitive to the long residence time groundwater fractions (the tails of the transit time distribution). This non-uniqueness poses a major challenge in estimating mean transit times with tritium alone and is potentially the larger source of uncertainty compared to analytical errors. I think that the introduction can be revised to better account for the multiple sources of uncertainty in tritium mean transit times and clearly state in what circumstances having a tritium timeseries is most likely to reduce this uncertainty and in what situations it cannot.

Referee #1 is right; tritium is not sensitive to long residence time groundwater fractions. This information is implicit in our manuscript when we indicate in the introduction the range of mean transit times to which tritium provides access (i.e., the sentence lines 50-52), but this is obviously not clear enough in this version of our manuscript. We propose to rephrase it in the revised version of our article. In this age range (i.e., up to 100-200 years maximum), we nonetheless agree that there is currently a problem of non-unicity of the tritium dating results. But this information is also already mentioned in our introduction (lines 71-80) as well as where/why this ambiguity happens (i.e., in the Northern Hemisphere because the remnant bomb tritium activities have not yet decayed to below those of current rainfall) and how having a tritium time series can help facing this issue (i.e., the longer the time series, the better).

Estimating groundwater mean transit times with tritium and the convolution integral is not new. Thus, I do not entirely agree with the claim that this work is pioneering interpretation of sparse tritium measurements in the northern hemisphere.

We assume that Referee #1 is referring here to the final sentence of the introduction (i.e., "By testing the potential of rather short but high-accuracy recent tritium data time series, our work has pioneering character for the dating of young groundwater bodies in Central Europe."; lines 95-96). Whether Referee #1 comment highlights the inappropriate use of the term "pioneering" or a lack of precision (e.g., not specifying that our study concerns only new sites characterisation), note that we will rephrase this sentence in the revised version of the manuscript.

I suggest revising the manuscript to focus more on the new hydrologic processes we learn from the tritium analysis, rather than inferring mean transit times from tritium. Alternatively, if the more accurate tritium measurements are proving to be essential to estimate mean travel times compared to previous studies, I think this needs to be explicitly shown in this work. While the mean travel time estimates in the Luxembourg aquifer can have value, I feel any innovation in this work is not properly highlighted.

We do not agree with Referee #1's suggestions. We indeed believe that explaining our rationale to assess mean groundwater transit times of new 'undated' sites using tritium and lumped parameter models is just as important as deciphering transit times in terms of hydrological processes. Besides, these two aspects are inextricable in the context of our study. This is why our discussion section deliberately develops the two aspects in a mixed and balanced way. Moreover, as indicated in our introduction (lines 86-88), the current advantage of using high-precision tritium analyses for obtaining more accurate water dating results has already been demonstrated in previous studies and we thus do not think it is necessary to specifically document it in our study. Additional references on this subject could nevertheless be added to the revised version of the manuscript, for instance Stewart et al. (2010).

I have concerns on the authors approach to constrain the parameters during the Monte Carlo uncertainty analysis. For instance, the mean transit time parameter is forced to be less than 35 years during the Monte Carlo sampling. Yet, the evidence and rationale to make this assumption is not strong and needs further discussion. Falsifying old transit times due to the high storage volumes is one line of evidence; however, I feel the highly simplified interpretation of storage volumes in this work is not a particularly strong argument and further analysis or comparison to alternative environmental tracers is needed. Furthermore, if it is evident *that mean transit times cannot be older than 35 years, it is unclear why the first Monte Carlo was performed. Generally, our prior bounds should encompass our best estimate of the uncertainty ranges for the given parameter. If the reader is to trust the uncertainty analysis, I suggest that the implications of these strong assumptions on the plausible mean transit times need to be further explored and presented.*

As done in several other tritium dating studies (e.g., Morgenstern et al., 2010; Stewart et al., 2010, 2012; Gallart, et al., 2016; Gusyev et al., 2016; Stewart and Morgenstern, 2016), we considered important to explore in our study the full range of mean residence times to which tritium provides access (i.e., 0-200 years). Indeed, we believe that this first broad exploration of the parameter space is key to illustrate comprehensively the current problem of non-uniqueness of tritium dating results in the Northern Hemisphere.

Concerning the justification for restricting the range of residence times in a second stage, we maintain that in our case the rejection of the oldest subpopulation of solutions (i.e., whose mean value varies between 64 and 98 years depending on the springs) due to the excessively high storage volumes that this would imply is a sufficiently solid decision criterion. We agree that Referee #1's concerns would have been justified if a connection with a deeper and more extended aquifer layer had been probable, but this is not the case here. In the context of our study, the Luxembourg Sandstone is isolated from such a connection (because cut by several valleys and perched on an impermeable underlying layer; cf. section 2.2 and Fig. 2). Note that this explanation will complement our revised manuscript.

Referee #1 also states that further analysis or comparison to alternative environmental tracers is needed, but we think this is exactly what we already propose in our work as we use the hydrogeological information at hand, the spring water physico-chemistry, as well as the spring discharge reaction time to effective precipitation inputs to corroborate our final modelling results.

It is unclear whether the mass balance is closed using the spring discharge measurements alone, or if river discharge or regional groundwater flow are additional fluxes that need to be considered.

Referee #1 refers here to the water balance calculations mentioned in lines 146-148 of the manuscript, which have been carried out for the different groups of springs used by the City of Luxembourg. As a reminder, these calculations have not been done in this study but were carried out by consultancy firms for the Luxembourg government as part of the official delimitation of the catchment areas of groundwater drinking water resources. To the best of our knowledge, only mean spring flows have been considered for these calculations. It is worth reminding that we also highlighted a potential problem with the recharge area of Group B springs, which seems too small considering our tritium dating results (see lines 544-551).

In general, there needs to be more explanation of the soil model that calculates the ET fluxes and how the uncertainties in the effective recharge rates impact the mean transit times. This uncertainty can potentially be more important than the tritium analytical error. While it can be argued this point is beyond the scope of this manuscript, I think this work needs to better acknowledge the many sources of uncertainty when quantifying mean transit times with tritium observations.

See our response to Referee #1 specific comment regarding the soil model parameterization (i.e., L218).

My opinion is that major revisions are required to address the comments above and to highlight the novelty in this work.

We hope that the answers we provide in this document will help to dispel Referee #1 concerns.

Specific Comments:

L19: I suggest that the abstract needs to explicitly state what the identified mean transit times are.

We will modify the abstract according to Referee #1 comment.

29: Duplicated references.

This is not duplicated references, this is two different ones.

L34: The use of 'by far' and 'obviously' are too strong of descriptions and are not needed. These are examples of the many places where the clarity of the manuscript can be improved.

In the revised version of the manuscript, 'by far' and 'obviously' will be removed.

L67: I think at this point the Introduction needs to mention the power of 3H/3He dating, which has largely super-seeded using 3H alone. Using 3H alone represents a major limitation of this study in my opinion.

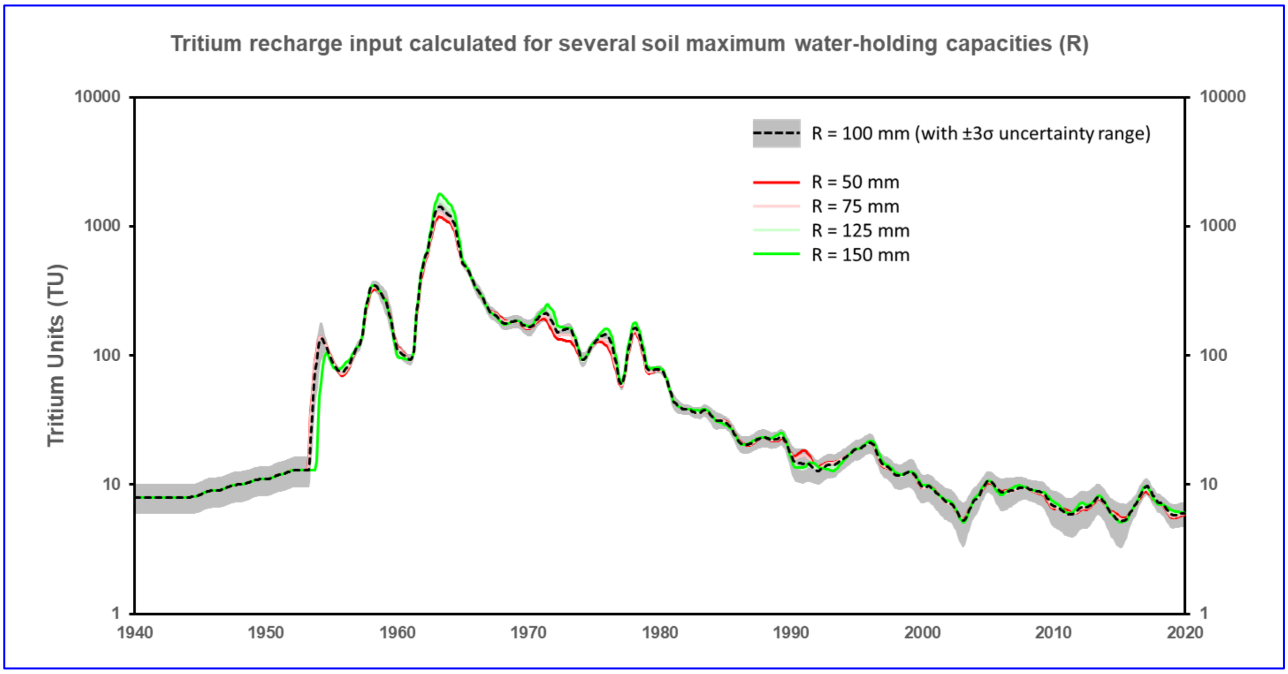
We do not agree with Referee #1. As explained a few lines before in the introduction (lines 53-59), the use of the $^{3}H/^{3}He$ dating technique faces issues specific to gaseous tracers which we wanted to avoid in our study. First, water sampling in spring locations is particularly thorny due to the risk of contamination with the surrounding air. But above all, as ^{3}He generated in the vadose zone is effectively lost from the water phase, the $^{3}H/^{3}He$ clock does not start until water is below the water table and completely isolated from a gas phase. In our study, we didn't want the vadose zone to be neglected.

L208: The information "initiated in 1958 by the IAEA (International Atomic Energy Agency) and the WMO (World Meteorological Organization)" is not needed.

This information will be removed in the revised manuscript.

L218: Why was a water-holding capacity of 100 mm assumed? In general, there needs to much more information on the soil model.

This value is an overall mean value derived from laboratory measurements carried out on samples taken from several soil pits and is usually used for soils overlying the Luxembourg Sandstones (Hissler et al., 2015; Hissler and Gourdol, 2015). This information will complement the revised version of the manuscript. In addition, as shown in the figure hereafter, it should be noted that using a soil maximum water-holding capacity varying between 50 and 150 mm does not have a strong impact on the tritium recharge input.



L225: Equation 3 for the tritium input weighting does not seem correct. Will Peff just not cancel out because it is in the numerator and denominator unchanged?

Referee #1 misunderstood Equation 3.

If the equation had been:

$C_{in\,i} = \frac{\sum_{j=i-11}^{i} C_j \times \sum_{j=i-11}^{i} C_j Peff_j}{\sum_{j=i-11}^{i} Peff_j}$, then yes there would indeed have been cancellation.

But the equation is:

$C_{in\,i} = \frac{\sum_{j=i-11}^{i} C_j Peff_j}{\sum_{j=i-11}^{i} Peff_j}$, which is correct.

For instance, for i = 12:

$$C_{in\,12} = \frac{\sum_{j=1}^{12} C_j\, Peff_j}{\sum_{j=1}^{12} Peff_j} = \frac{C_1 Peff_1 + C_2 Peff_2 + C_3 Peff_3 + \cdots + C_{10} Peff_{10} + C_{11} Peff_{11} + C_{12} Peff_{12}}{Peff_1 + Peff_2 + Peff_3 + \cdots + Peff_{10} + Peff_{11} + Peff_{12}}$$

L241: It states that the input signal uncertainties are propagated by randomly sampling from a normal distribution; yet there is no information on what the distribution parameters are.

Means and standard errors values of the recharge input signal used to generate the replicated time-series are shown in Fig. 5b,d. These values will also be supplied in a dataset and archived in an online repository, whose DOI will be included in the final published version of the paper.

L255: It is not explained why the comparison to hydraulic response times and major ions is done in first place. Adding some rationale for why this analysis was performed into the introduction seems necessary.

We thank Referee #1 for this advice. We have chosen to include this information in the "Material and methods" section so as to not overload the introduction. But if required, the reason for using the spring chemistry, as well their hydraulic response time to the precipitation inputs, will be additionally mentioned in the introduction section of the revised version of the manuscript.

L345: It is not clear how the measured tritium suggests the presence of bomb-peak water. For instance, it seems that the measured values could be obtained through recent recharge and radioactive decay.

We agree with Referee #1, it is theoretically possible for some springs considering the modelling results presented in the Supplementary file. Note that in the revised version of the manuscript, the last part of the sentence (i.e, "emphasizing the presence of bomb tritium in the Luxembourg Sandstone aquifer"; line 346) will be removed.

L357: I think there needs to be much more discussion on how these samples with old mean transit times can be confidently rejected. I am not convinced of this assumption given the terse description provided.

See our response to Referee #1 general comment.

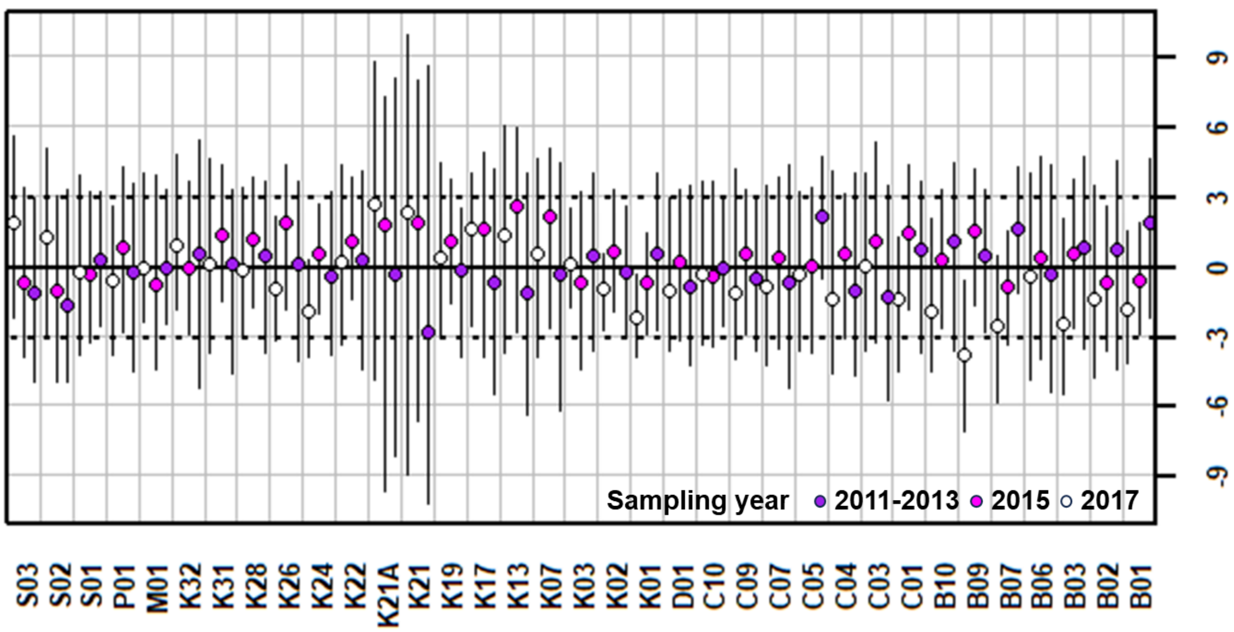
L366: Why were only a third of the samples retained? The seemingly subjective 'throwing away' of samples concerns me and seems to deteriorate the uncertainty quantification significantly. The assumptions and rationale need to be explained better. L369: Same comment as above. Is keeping only the dominant population actually providing a robust measure of the uncertainty?

Referee #1 is concerned here about our approach to use an additional 2D kernel NSE weighted density filter to obtain a more coherent ensemble of solutions from the raw modelling results obtained from the second EPM parameter space screening and asks for clarification.

As shown in the Supplementary material (Fig. S1-S35) and mentioned in the manuscript (lines 363-366), the 5000 behavioural solutions resulting solely from the NSE threshold of 0.5 were not distributed around an unambiguous parameter distribution, but spread into several subpopulations - most of the time overlapping, and thus leading to an ensemble with a complex multimodal character. We wanted to isolate the most probable unimodal distribution of parameters to tend toward, what we called in line 366, “a more coherent ensemble of solutions”. It is to face this issue that we opted for a 2D kernel NSE weighted density filter. In practice, its use allowed the multimodal distribution to retain the (almost) unimodal set of solutions with respect to both EPM parameters with the highest probability (referred to as the “most likely solution ensemble” in the manuscript), and it provided acceptable results for all springs. The decision to keep only a third of the samples was taken, after some trials, as a good compromise on the one hand to reject the least possible solutions and to limit cases requiring rejection of a secondary weaker subpopulation on the other hand. This will be explained in the revised version of the manuscript.

It is true that the use of such an additional filter has a considerable impact on the dating results provided and their associated uncertainties (compared to if we had used all 5000 solutions), but, as mentioned in lines 502-503 and supported with Fig. 14, the simulated spring tritium concentrations resulting from theses ensembles are overall consistent with the tritium measurements of spring samples, their downward trends, as well as with their analytical uncertainties. To better document this statement, we propose to add a second graphical panel to Fig.14 that will show the simulated concentrations in tritium standardized with the tritium observations (subtracting the mean and dividing by standard deviation) as in the figure hereafter. Moreover, as mentioned in the discussion section of the manuscript, the distribution of the most probable parameters is supported by several independent data sources (lines 514-521) and we thus believe that our decision is further reinforced.

EPM simulated concentrations in tritium standardized with the tritium observations (subtracting the mean and dividing by standard deviation)



The colored circles and the vertical thin lines stand for the average and the min-max range simulated concentrations, respectively.

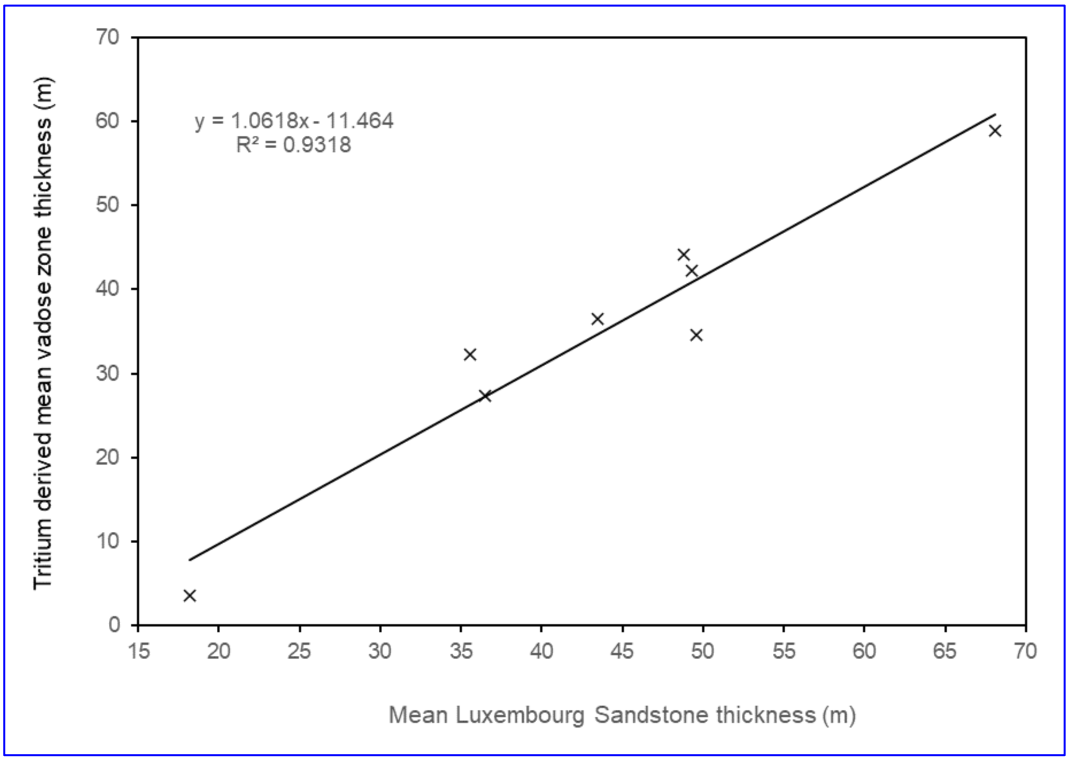
L472: This sentence is awkward and is not needed.

We will follow Referee #1's advice and remove this sentence in the revised version of the manuscript.

L522: Is this assumption of vadose zone that is the entire thickness of the aquifer reasonable? I think that much more information about the vadose is needed. For instance, are there any wells that can constrain the vadose zone thickness?

We think that this comment/question relates to a misinterpretation of our statement. We do not assume here that the thickness of the vadose zone is equal to the full Luxembourg Sandstone thickness, but we state that the Luxembourg Sandstone thickness can be considered as a "proxy" of the thickness of the vadose zone (which is different). In statistics, a proxy is a variable that is not in itself directly relevant, but that serves in place of an unobservable or immeasurable variable. For a variable to be a good proxy, it must have a close correlation (either positive or negative, and not necessarily linear) with the variable of interest. The validity of our hypothesis can be supported to some extent with the graph shown hereafter which compares our estimates of the mean Luxembourg Sandstone thickness for each recharge areas from Table 1 (x-axis of the graph hereafter) to the mean vadose zone thicknesses that can be derived from the tritium results (subtracting H_{exp} values indicated in Fig. 8 from the mean Luxembourg Sandstone thickness values in Table 1; y-axis of the graph hereafter).

However, we agree with Referee #1 that it would have been preferable to rely on robust direct information to assess the vadose zone thickness for each recharge area. Unfortunately, as mentioned in lines 148-149, hydrogeological drillings in the Luxembourg Sandstone are too sparse and poorly distributed (one can also see the "Hydrogeological drillings" layer in the Luxembourg platform for governmental geodata and service https://map.geoportail.lu/). For instance, only one borehole is available in the KRD recharge area and none in the KRG, M, P and D ones (excluding shallow exploratory boreholes drilled in the very near vicinity of spring catchment facilities). However, it is worth noting that Farlin et al (2013a) indicate that the saturated and vadose zones are about 10 m and 45 m thick, respectively, at the level of the observation borehole in the KRD sector, which is consistent with the average numbers we assessed in our study (see Fig. 8).



References used in our response to Referee #1 that were not already cited in the manuscript:

Hissler, C., Gourdol, L., Juilleret, J., Marx, S., Leydet, L., and Flammang, F.: Pedotransfer functions for predicting soil hydrological characteristics in Luxembourg: literature review and reliability tests for predicting the soil maximum water-holding capacity (Fonctions de pédotransfert pour la prédiction des caractéristiques hydriques des sols au Luxembourg : analyse bibliographique et premiers tests de fiabilité pour la prédiction de la réserve utilisable maximale des sols), Report drafted on behalf of the Administration des services techniques de l'agriculture (in French), 2015.

Hissler, C., and Gourdol, L.: Assessment of soil maximum water-holding capacity in Luxembourg at national scale: a first estimate based on recent datasets (Évaluation de la réserve utile maximale en eau des sols au Luxembourg à l'échelle nationale : une première estimation basée sur des jeux de données récents), Report drafted on behalf of the Administration de la gestion de l'eau (in French), 2015.

Stewart, M.K., Morgenstern, U. and McDonnell, J.J.: Truncation of stream residence time: how the use of stable isotopes has skewed our concept of streamwater age and origin, Hydrological Processes, 24, 1646-1659, https://doi-org.proxy.bnl.lu/10.1002/hyp.7576, 2010.

Gusyev, M. A., Morgenstern, U., Stewart, M. K., Yamazaki, Y., Kashiwaya, K., Nishihara, T., Kuribayashi, D., Sawano, H., and Iwami, Y.: Application of tritium in precipitation and baseflow in Japan: a case study of groundwater transit times and storage in Hokkaido watersheds, Hydrology and Earth System Sciences, 20, 3043–3058, https://doi.org/10.5194/hess-20-3043-2016, 2016.