This study is an impressive assemblage of field, laboratory and modelling techniques to determine the spatio-temporal variability in DOC export (concentration and molecular composition) in a riparian zone. The abundance of different techniques make the manuscript quite dense and it is sometimes difficult to follow the details of the Material and methods, but I would not recommend providing more technicalities (see some exceptions in the detailed suggestions below).

My main suggestion to improve the paper is to rework the introduction and the discussion to 1) identify a clear research question or hypothesis, the introduction is lacking a "problem to solve". It was not clear to me why such a detailed study would improve management, because the resolution is far higher than any management action, or large scale modelling; 2) develop a discussion that put the results into perspective, while the current discussion is still very similar to the result section and does not contain implications for future research or management.

We appreciate your constructive, detailed evaluation of our Manuscript (MS). We realized that our research question/hypothesis was not formulated clearly enough, which is also reflected in the reviewers statement that there is a lack of a "problem to solve" in the introduction. This will be changed and addressed in the introduction and discussion sections of the MS. In line with the proposal of Referee #1 (comment on L80), we will outline our main hypothesis, that a small-scale, dynamic TWI can explain the main characteristics of the DOC export regime of a given RZ, more precisely. This hypothesis will be tested by a combination of field measurements and detailed hydrological modelling. We will furthermore carefully rework the introduction, discussion and conclusion with regard to implications for management and will put our research into a broader perspective of literature (e.g. references given in the review of Referee #2). As this rewriting needs to be done carefully and requires some work we are not yet able to propose specific changes at this time.

I also found that the authors made too little use of the different sampling dates, especially those during storm events. I did not understand why several analyses in the manuscript only consider April and December dates, while many other dates are available (I may have missed something here…). Similarly, a high-temporal resolution sampling was performed during selected storm event but the infra-storm events dynamics is not described.

Regarding the riparian water sampling, only 2 more samples are available for July (cf. Figure 3b). Generally groundwater sampling in summer turned out to be difficult due to low groundwater levels. Most wells, especially those screened closer to the surface, could not be sampled in summer. To ensure proper comparability, we decided to focus on April and December, when groundwater and surface water sampling was possible. Regarding the high resolution event sampling, all samples were used (cf. Figure 4a), but inter-event variance of DOC properties is higher than intra-event variance. Therefore we considered bulk sample properties of one event to be satisfactory in information content. We will state this more explicitly in the MS.

Some work would be necessary to improve the clarity of the text: shorter sentence, less and better use of conjunctions, correct some poor phrasing. I have identified a few examples of sentences to improve but note that English is not my native language either.

We will work through the MS to identify sentences similar to those identified by the reviewer and change them accordingly.

Detailed comments:

Title : “from a riparian zone of a” -> “from the riparian zone of a”?

We agree, the title will be changed accordingly.

L12 “but poorly understood component”: what specifically is not understood. Identify a “problem to solve”, a research question or hypothesis in the abstract.

We agree, the sentence will be changed to:

The mechanisms of dissolved organic carbon (DOC) export from riparian zones (RZs) is an important, yet still poorly understood component of the carbon budget in catchments with a temperate climate.

A “problem to solve”, a research question or hypothesis will be more concisely stated in the abstract. In line with the proposal of Referee #1 (comment on L80), our main hypothesis is that a small-scale, dynamic TWI can explain the main characteristics of the DOC export regime of a given RZ. This hypothesis is tested by a combination of field measurements and detailed hydrological modelling.

L15 “high spatio temporal resolution”: what is the resolution of the DEM?

This information is given in the following sentence (1m). Therefore this sentence will not be adapted.

L15 “Stream water DOC samples from differing hydrological situations”: describe these situations, number of sampling dates, study period, etc in the abstract.

We agree, the sentence will be changed to

From May 2017 until July 2019, stream water DOC samples (n = 73; five runoff situations and five grab samples) were compared to riparian DOC groundwater and surface water samples (n = 66) …

L18 “were then simulated”: avoid passive voice throughout the manuscript

We agree, the sentence will be changed to

We used a physically-based, fully-integrated numerical flow model (HydroGeoSphere) to simulate explicit water fluxes from the resulting riparian DOC source.

L20 “two distinct DOC pools (DOCI and DOCII)”: describe what make them different in the abstract

We agree, the sentence will be changed to

Chemical classification revealed an aromatic, oxygen-rich DOC pool with high concentrations (DOC_I) and a microbially processed, mobile DOC pool with lower concentrations but larger compositional variability across seasons (DOC_{II}) in the RZ.

L22 “high-resolution topographical wetness index (TWIHR)”: specify resolution in abstract.

We agree, the resolution will be specified accordingly.

L27 “should be considered in DOC export models”: any implications for management? Should large-scale models really consider this fine-resolution heterogeneity?

Considering fine-resolution heterogeneity is likely not applicable to large-scale models, but understanding the export mechanisms of riparian zones at fine scale allows to better estimate overall DOC export potential of catchments as a function of climatic variability and general topographic structure. However, the relationship between riparian zone structural heterogeneity and DOC export presented in this study suggests that knowledge about riparian topography and structure, easily derived from DEMs, may be useful for the development of more parsimonious models for the prediction of hydrologic and DOC export response by e.g. implementing a threshold-based surface runoff module. Measures of riparian zone source connectivity (like the presented TWIHR) provide an integrated measure of riparian zone surface runoff generation and the associated DOC export behavior that is – when integrated accordingly – scalable to catchment level. We therefore believe that proxies of fine-resolution structural heterogeneity can improve large-scale catchment models, which cannot represent topographic (riparian) relief at very fine scale. We agree with the referee that our detailed, fine-resolution modelling effort will generally not be feasible for day-to-day management operations. However, it helped us to demonstrate the usefulness of drone-based DEMs to gain process understanding. With the help of the model we could illustrate the importance of the TWIHR for DOC export. A practical consequence of our work could be that the availability of a high-resolution DEM allows one-time targeted earth-moving measures to manipulate the TWIHR to create a more favorable DOC export regime.

We will address this in the Introduction and Discussion of the MS and thus will modify this sentence.

L27 “But despite”: don’t start a sentence with “but”

We agree, please see next comment.

L33 “but could” second but in this long sentence

We agree, the sentence will be changed to

Changes in land use, climate and biogeochemical boundary conditions have increased DOC concentrations in surface waters and changed the quality of the exported DOC (Larsen et al., 2011; Chantigny, 2003; Wilson and Xenopoulos, 2008). Routine management of DOC could help to comply with water quality directives and lower the cost of drinking water purification (Matilainen et al., 2011), but is currently almost non-existent (Stanley et al., 2012).

L36 “Especially riparian zones (RZs) of lower order streams are potential targets for…” poor phrasing

We agree, the sentence will be changed to

Lower order streams make up a large fraction of the total river networks worldwide (Raymond et al., 2013) and their riparian zones (RZs) represent a main source for terrestrial DOC export (Ledesma et al., 2015; Musolff et al., 2018). Therefore RZs of lower order streams – as terrestrial-aquatic interfaces – constitute a general control unit qualifying them as potential targets for DOC export management.

L41 “Here, DOC …” add reference

We agree, the following reference will be added:

Luke, S. H., Luckai, N. J., Burke, J. M., and Prepas, E. E.: Riparian areas in the Canadian boreal forest and linkages with water quality in streams, Environmental Reviews, 15, 79-97, 10.1139/A07-001, 2007.

L46 “This leads to a stronger accumulation of DOC close to the soil surface…” I did not understand the link with the previous sentence

We agree, the sentence will be changed to

On the other hand, the rate of DOC accumulation and ultimately export is also dependent on hydrological connectivity of DOC sources to the stream. Between events, water cannot mobilize existing DOC pools close to the soil surface. Consequently, hydrological connectivity also contributes to DOC accumulation close to the soil surface in comparison to deeper soil layers that are more frequently connected with the stream.

L49 “led to concepts like variable source zone activation (Dick et al., 2015; Werner et al., 2019), the dominant source layer (Ledesma et al., 2015) and transmissivity feedback (Bishop et al., 2004)” explain these concepts and their limits. Listing them is not enough in an introduction

We agree, the concepts will be explained accordingly. The sentences will be changed to

Several attempts have been made to acknowledge the vertical variability of lateral DOC transport to streams. The dominant source layer concept (Ledesma et al., 2015) focuses on depth-dependent differences in DOC pools in different soil layers. Transmissivity feedback (Bishop et al., 2004) accounts for depth-dependent differences in hydraulic conductivities of soils and the resulting changes in the transmissivity of the soil profile under changing groundwater levels. Variations in the lateral hydrological connectivity of RZs were coneptualized by variable source zone activation (Dick et al., 2015). These concepts still describe a heterogeneous system in terms of an integrated response without acknowledging the distinct spatio-temporal variability in DOC export of single landscape units (Ledesma et al., 2018a; Ploum et al., 2020; Dick et al., 2015).

L55 “a strong focus on vertical heterogeneity” I my understanding the variable source area concept is more about horizontal heterogeneity.

We agree, the text will be changed to (please also see comment above)

Variations in lateral hydrological connectivity of RZs were conceptualized by variable source zone activation (Dick et al., 2015).

L57 “Moreover RZs are highly dynamic and heterogeneous with micro-topography” the role of micro topography is central to the hypothesis of this work and should be better highlighted.

We agree, the role of micro-topography will be better highlighted. We propose to change the text to

Moreover micro-topography in RZs can induce hot spots of biogeochemicall activity (Frei et al., 2012) that contribute disproportionally strong to nutrient turnover. Furthermore, micro-topography focuses drainage (Frei et al. 2010, Scheliga et al., 2019) and consequently solute export to the stream when nutrient hot spots get hydrologically connected to the stream (during hot moments). Therefore, micro-topography in the RZ is considered a fundamental organizing structure of soil chemistry (Diamond et al., 2019) and hydrological connectivity (Scheliga et al., 2019) that induces high spatio-temporal heterogeneity.

Additional references:

Diamond, J. S., McLaughlin, D. L., Slesak, R. A., and Stovall, A.: Microtopography is a fundamental organizing structure of vegetation and soil chemistry in black ash wetlands, Biogeosciences, 17, 901-915, 10.5194/bg-17-901-2020, 2020.

Scheliga, B., Tetzlaff, D., Nuetzmann, G., and Soulsby, C.: Assessing runoff generation in riparian wetlands: monitoring groundwater–surface water dynamics at the micro-catchment scale, Environmental Monitoring and Assessment, 191, 116, 10.1007/s10661-019-7237-2, 2019.

L65 “Model conceptualizations that are able to bridge those scales” with this sentence it seems that the paper will deal with this question of scales, but it is not the case.

We agree, the sentence will be removed from the manuscript.

L80 “We argue that a smaller-scale, dynamic assessment of the TWI…” should be the hypothesis of the paper. Please give a response to his hypothesis/question in the discussion/conclusion.

We agree, a dynamic assessment of the high-resolution TWI will be added in the regarding discussion/conclusion section. We therefore hypothesize that a small-scale, dynamic TWI can explain the main characteristics of the DOC export regime of a given RZ. This hypothesis will be tested by a combination of field measurements and detailed hydrological modelling.

L97 “In this paper we…” intro long enough, no need for a summary of the methods here. Develop problem to solve instead.

We agree, the methodological shares of this paragraph will be deleted and instead replaced by a problem to solve (see comment above).

L100 “More specifically, (1)…” it would be better to list specific research questions than summary of the methods.

We agree, we will replace the summary of the methods by specific research questions/hypotheses.

L128 “Electric resistivity tomography (Resecs DC resistivity meter system, Kiel, Germany) was applied at two transects” show the transects in figure 1?

We agree, transects of the conducted electric resistivity tomography will be shown in figure 1b. The captions will be adapted accordingly.

L134 “Two PCM4 portable flow meters (Nivus, Germany) measured discharge in the Rappbode stream at a chosen inlet…” show inlet and outlet in figure 1?

We agree, locations of in- and outlet will be shown in figure 1b.

L151 “To have maximum ability in capturing the magnitude and direction of this slope…” poor phrasing

We agree, the sentence will be rephrased to

Therefore we decided to install a piezometer network aligned on a square grid, with one principal axis oriented in parallel to the stream and the other perpendicular to the stream to capture the temporal dynamics of the groundwater level in both principal directions of this slope.

L156 “In addition 3 more wells were installed at 0.3 m depth inside the rectangular grid for surface near sampling.” I did not understand this sentence.

We agree, we will change the sentence to

We installed three additional wells with screens at 0.3 m depth inside the rectangular grid for sampling near the surface.

2.2.3. I found it difficult understand the maximum depth and the screening height of the different piezometers and wells. Please rework this section to improve clarity.

We agree, the section will be reworked with focus on improved clarity.

L166 "Biweekly routine samples…" it is never clear whether biweekly means twice a week or every second week. Please use a less ambiguous term. Please also add the number of sampling dates and the number of dates when FT-ICR mass spectrometry was used. Is it only two dates?

We agree, the number of sampling dates will be added. We further realized that the referee was confused about the description of FTICRMS samples. We therefore decided to additionally rephrase the whole paragraph as follows:

At the monitoring site along the Rappbode stream, we collected 68 stream event samples during five events, 66 riparian samples and 5 stream samples during 5 occasions (one stream sample per occasion), which were analyzed for DOC concentration. The molecular composition of the DOC was determined via FT-ICR MS (see Table S1 for detailed information). Auto-samplers (6712 Full-Size Portable Sampler, Teledyne ISCO, US) were triggered by the rate of water level increase to sample stream water during discharge-generating events at least once per hour. Auto-sampler bottles (PP) were soaked for 48 h in 0.1 N HCl prior to use. We prepared process blanks with deionized water to correct for eventual contamination during field work and sample processing. Due to the remoteness of the study site, we collected auto-sampled stream water samples within 4 days after the triggered event sampling. Samples were stored in the dark inside the sampler and air temperatures were below 10°C during that time. We collected riparian zone shallow groundwater samples from 3 to 18 out of the 28 installed piezometers depending on hydrological conditions during the five sampling dates. Before sample collection, we replaced water in the wells one to three times (based on the responsivity of the wells) through pumping. We rinsed the flasks and the pump with sample water prior to sample collection and subsequently transferred 100 mL sample into acid-rinsed (0.1 N HCl) and baked (500 °C, 4 h) glass bottles. Additional stream water was collected for each riparian sampling date. Moreover, we collected 38 routine samples in the Rappbode stream (every two weeks) and processed them accordingly to determine DOC concentration. Samples were stored dark and cool until further processing in the laboratory.

L181 "samples were filtered using 0.45 µm membrane filters" did you filter the samples in the field or back in the lab?

We agree, the sentence will be changed to

Samples were filtered (0.45 µm membrane cellulose acetate filters, Th. Geyer, Germany) and acidified to pH 2 (HCl, 30 %, Merk, Germany) on site. Subsequently samples were stored cool until DOC measurement and extraction in the laboratory.

L280 "2.4.3 Calibration" is it possible to provide the objective function of the calibration? I understood that the model aimed to simulate both the stream discharge and groundwater depth in several wells, with a weighting scheme giving a high importance to the groundwater, but it would be interesting to see the equation of this objective function.

We agree, the objective function will be provided in the SI:

$$\text{Multi} - \text{Objective function} = \sum_{i=1}^{i=nq} w_q (O_q^i - S_q^i)^2 + \sum_{i=1}^{i=nl} w_l (O_l^i - S_l^i)^2$$

Where O_q^i and S_q^i are the observed and simulated discharge. nq is the number of number of the discharge observation (611). O_l^i and S_l^i are the observed and simulated groundwater level. nl is the number of groundwater level observation (110140). w_q and w_l are the weights for the two observation groups, both being assigned with the value of 1 in the calibration. Because the observation number of groundwater level was significantly larger than that of the discharge, this multi-objective function highlight the importance of the groundwater levels, such that the 94% of the multi-objective function for the calibrated best-fit was attributed to groundwater levels.

L322 "The DInf algorithm was used" please explain what it is.

We agree, the DInf algorithm will be explained.

We applied the DInf algorithm to calculate a realistic hydrological routing (Tarboton, 1997). The DInf algorithm determines flow direction as the steepest downward slope on eight triangular facets formed in a 3x3 cell window centered on the cell of interest.

L335 "Discharge shows event-type, erratic variability" poor phrasing

We agree, the sentence will be changed to

Discharge showed typical erratic variability at the event-scale. At annual scale, discharge expressed a clear seasonal pattern, with lowest values in late summer and highest values in spring (Figure 3a).

L360 "DOC in riparian water samples was in general of highly unsaturated and phenolic composition, typically found in lignin and biomass type compounds" can we see this in a table or a figure?

This can be derived from Figure S10. We have noticed that the colors in this Figure were wrong and will include a corrected version of Figure S10:

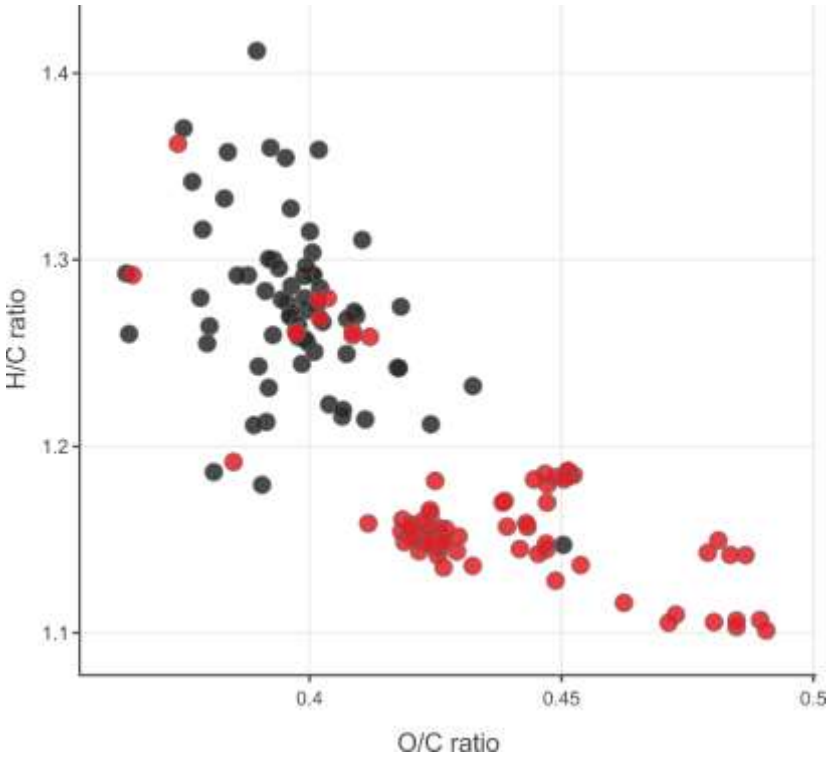

Fig. S10: Aggregated van Krevelen plot of all FT-ICR-MS sample of stream (red) and riparian (black) origin. Data represent the intensity weighted average of the molecular H/C and O/C ratios considering all valid MF in these samples.

Besides, the sentence will be changed to

DOC in riparian water samples was in general of unsaturated and phenolic composition (wa_{HC} = 1.27 ± 0.05; wa_{OC} = 0.40 ± 0.01; n = 66), typically found in wetland surface soils (LaCroix 2019). However, stream event samples significantly differed ($p < 0.001$) from riparian samples and were more unsaturated (wa_{HC} = 1.17 ± 0.05; n = 76) and more oxygenated (wa_{OC} = 0.43 ± 0.03) as shown in Figure S10.

L395 "Note that wells, sampled during different occasions throughout the year occur in both DOC clusters and according TWIHR values can thus occur in both clusters" it is unclear to me to what extend a given piezometer belonged to the same cluster throughout time. This sentence suggests that the cluster can change, but a quantitative assessment of how many piezometer remain in the same cluster or change clusters would be interesting here.

We agree, a quantitative assessment of how many piezometer remain in the same cluster or change clusters will be added.

L415 "The significant difference in TWIHR median values of DOCI and DOCII wells" I did not understand how you could classify wells as DOCI-well or DOCII-well if a given well could change clusters in different dates.

We agree, the correct term in this sentence is DOCI and DOC II *samples*, not *wells*. We will change this in the manuscript and apologize for the inconvenience.

L416 “using the median TWIHR value of the DOCI group (9.66) as a threshold.” I did not understand this choice; please explain the rationale behind this.

The rationale was to map DOC source zones of different DOC concentration and chemical properties. We found an overlap of TWIHR values between both groups although their median was found to be significantly different. Using the median of DOCI (9.66) as a manually chosen threshold we can separate both groups capturing 50% of all cases of group I in one class while only allowing 25% of group II.

We realized that the rationale behind the selection of the threshold was not addressed clearly enough in the MS. Therefore we will better describe and clarify the rationale.

L418 “Also note that different samples of one well can appear in both DOC groups” please give numbers.

We agree, numbers will be added in the manuscript.

L435 “Fig. S7, Table S4 for according water fluxes” -> “corresponding water fluxes”?

We agree, this will be changed in the manuscript.

L454 “During the model period, DOCI source wells had a median DOC concentration of 5.8 mg L-1 which was 2.3 times higher than for the DOCII source wells” it would be interesting to remind the mean+/- sd of the two types of wells. Do deeper wells match with the DOCI cluster?

We agree, ± sd will be added:

During the model period, DOCI source wells had a median DOC concentration of 5.8 mg L-1 (mean ± sd: 6.2 ± 2.7 mg L^{-1}) which was 2.3 times higher than the median for the DOCII source wells (2.7 ± 1.2 mg L^{-1}).

Certain deeper wells match with the DOCI cluster (e.g. A1-E1, B4, C4, see also comment below).

L487 “as typically found in deeper soil layers” what influences the difference between DOC I and DOC II more: the TWI or the sampling depth? (or both are related?).

There is no statistical significant relation between sampling depth and well classification in our samples (see Figure below). On the other hand, we presented significant differences in TWI_{HR} between the clusters. We therefore conclude that TWI (as postulated in the MS) controls the difference between DOCI and DOC II (more).

However, the possibility of a bias exists since samples were predominantly taken in deeper soil layers - also due to the fact that there often was no surface near water available when

groundwater samples were taken. Moreover, surface near samples as well as deep samples appear in both clusters.

Also with regard to referee #2 we will include this discussion and present respective statistical test results in the MS.

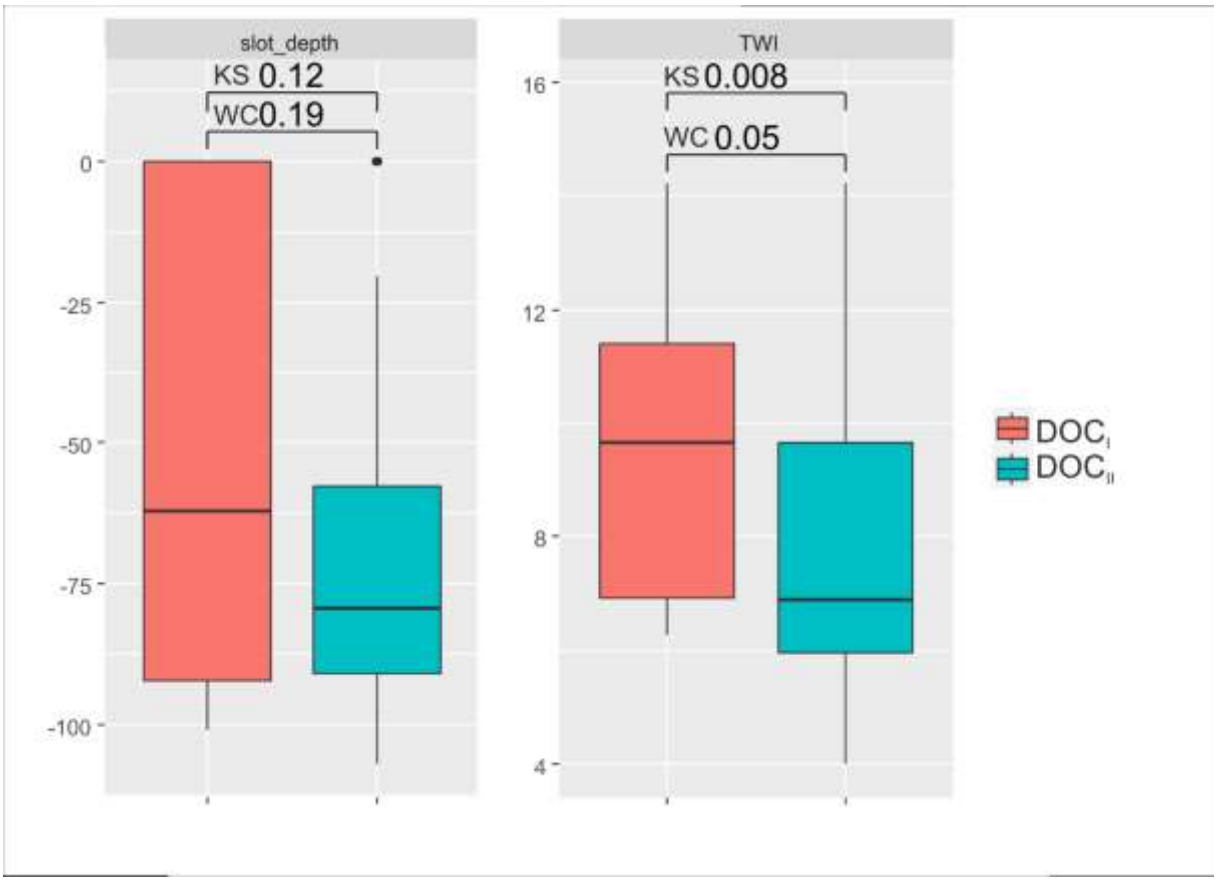

Figure: Boxplots of piezometer slot/sampling depth [cm below ground] and TWI [-] value for the DOC_I and DOC_{II} clusters. Horizontal brackets above describe the Kolmogorov-Smirnoff (KS) and Wilcoxon rank sum (WC) test statistics. Values were min-max normalized to values between 0 and 1 for better illustration.

L491 “indicating a replete DOC pool with constant contribution to the overall DOC quality in the stream” unclear sentence

We agree, the sentence will be changed to

In addition, the DOCI quality was similar between April and December indicating a large DOC pool which is less affected by seasonality and hydrologic conditions. Therefore DOCI can be regarded as constantly contributing to the overall DOC quality in the stream.

L493 “indicating the influence of seasonality on this pool.” It is difficult to make such a conclusion with only two dates.

We agree, the sentence will be changed to

In contrast, the DOCII composition was reflected in the stream water composition in December but not in April, suggesting a depletion of this pool (e.g. induced by seasonality) during high flow periods.

General comment on “4 Discussion”: this discussion is too similar to the result section, many conclusions are specific to the study site while readers would expect to see the results put into perspective, with more implications for management and research, more key messages and more references to the literature.

We agree, the Discussion section will be changed accordingly. In consistence with the Introduction (see general comment and comments on L12, L27), results will be put into a broader perspective in reference to other studies reported in the literature. Implications for management and research will be given more weight, key messages (see also comment L80) will be better worked out. The suggestions provided in the detailed comments by both referees will offer useful guidance for this. Because this rewriting needs to be done carefully and requires some work we are not yet able to propose specific details of these changes at this time.