Response to Referee 1

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 was changed and addressed in the introduction and discussion sections of the MS. In line with the proposal of Referee #1 (comment on L80), we outlined our main hypothesis, that both DOC production and transport are predominantly controlled by the micro-topography of the RZ (lateral variability), and by the depth of the riparian groundwater level (temporal variability). , more precisely. This hypothesis was tested by a combination of field measurements and detailed hydrological modelling. We furthermore thoroughly reworked the introduction, discussion and conclusion with regard to implications for management and put our research into a broader perspective of literature (e.g. references and discussions given in the review of Referee #2, see also general comment on discussion of Referee #1).

The extensive rewriting removed several segments on which the reviewers commented, tendering these comments moot. We therefore cannot provide a detailed explanation about our response to these comments. In these cases it is implicitly understood that the detailed comments were addressed by the rewriting of enter sections of the paper.

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.

R1GC2a, b: 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, were dry 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 stated 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 worked through the MS to identify unclear or long sentences similar to those identified by the Referee and changed them accordingly.

Detailed comments:

Please note that, due to the Referees suggestions, we entirely reworked the introduction, discussion and conclusion sections. Therefore answering some of the Referees detailed comments can deviate from the original response.

Title : "from a riparian zone of a" -> "from the riparian zone of a"?

R1C1: We agree, the title was changed to

Small-scale topography explains patterns and dynamics of dissolved organic carbon exports from the riparian zone of a temperate, forested catchment

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

R1C2: We agree, the abstract was changed to:

Export of dissolved organic carbon (DOC) from riparian zones (RZs) is an important component of temperate catchment carbon budgets, but export mechanisms are still poorly understood. Here we hypothesize that a spatially highly resolved topographic analysis of RZs allows to characterize and delineate DOC source zones for catchment scale DOC exports

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.

R1C3a,b,c: We agree, the information was implemented in the abstract.

L18 "were then simulated": avoid passive voice throughout the manuscript

R1C4: We agree, the passive voice was changed to active where appropriate.

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

R1C5: We agree, the sentence was changed to

The chemical classification by Fourier-transform ion cyclotron resonance mass spectrometry revealed revealed an aromatic, oxygen-rich DOC pool with high concentrations ($DOC_I$) and a microbially processed, mobile DOC pool with lower concentrations in the riparian groundwater and surface water samples ($n = 66$).

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

R1C6: We agree, the resolution was 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 appropriately – 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.

R1C7: We addressed this in the Introduction and Discussion (section 4.4) and Conclusions of the MS.

L27 "But despite": don't start a sentence with "but"

Please see next comment (R1C8).

L33 "but could" second but in this long sentence

R1C8: We agree, the sentence was changed to

Changes in land use, climate and biogeochemical boundary conditions have increased DOC concentrations in surface waters and altered the quality of the exported DOC in the last decades (Larsen et al., 2011; Chantigny, 2003; Wilson and Xenopoulos, 2008). Beside the ecological impacts this alteration may also affect safety and costs of drinking water production (e.g. Wang et al., 2017). Routine management of DOC could therefore 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

R1C9: We agree, the sentence was 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

R1C10: We agree, the following reference was 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

R1C11: We agree, the sentence was changed to

On the other hand, the amount of accumulated DOC and ultimately export is also dependent on hydrological connection of DOC sources to the stream, because water can only mobilize DOC pools if these contribute to riparian runoff generation.

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

R1C12: The concepts were explained. We reworked this paragraph to

Several attempts have been made to characterize and quantify the dynamics of runoff generation in RZs and the associated variability of DOC transport to streams. However, to date model conceptualizations have mainly focused on the vertical distribution of DOC sources in the subsurface and to a lesser degree on horizontal heterogeneity induced by topography. For instance the dominant source layer concept (Ledesma et al., 2015) focuses on depth-dependent differences in DOC pools in distinct soil layers, which are assumed to be uniform across the RZ. 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. This concept is taken up in the riparian profile flow-concentration integration model (Rim, Seibert et al., 2009) to model stream solute variability as a function of a non-linear vertical distribution of pore water solute concentrations in riparian soils. Frei et al. (2010, 2012) were able to simulate the complex effects of riparian micro-topography on runoff generation and the formation of biogeochemical hotspots in the subsurface, but their explorative model was computationally expensive and did not explicitly consider DOC transport. Variations in the lateral hydrological connectivity of a RZ to a stream have been conceptualized in a spatially lumped catchment DOC export model by defining different source zones with variable activation (Dick et al., 2015). In essence, most model conceptualizations for DOC export from RZs describe a heterogeneous system in terms of spatially lumped integrated functional relationships without explicitly acknowledging small-scale spatio-temporal variability in DOC export from individual, small landscape units (Ledesma et al., 2018a; Dick et al., 2015).

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

R1C13: We agree, the text was changed to (please also see comment above)

Variations in the lateral hydrological connectivity of a RZ to a stream have been conceptualized in a spatially lumped catchment DOC export model by defining different source zones with variable 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.

R1C14: We agree, the role of micro-topography was better highlighted. We changed the text to

Micro-topography in RZs can induce hot spots of biogeochemical activity (Frei et al., 2012) that contribute disproportionally to nutrient turnover. Temporary, hot spots of DOC production in the shallow soil layers of the RZ can become hydrologically connected to the stream (during hot moments), when the groundwater levels intercept the surface and micro-topography focuses drainage (Frei et al. 2010, Scheliga et al., 2019) and can consequently be a source for solute exports to the stream. Therefore, micro-topography in the RZ is considered a fundamental organizing structure, not only for soil chemistry (Diamond et al., 2020) but also of hydrological connectivity (Frei et al. 2010, Scheliga et al., 2019) that induces high spatio-temporal heterogeneity of DOC exports. Riparian topography and the dynamics of groundwater levels in the RZ thus are the key drivers of the spatio-temporal patterns of DOC export from RZs.

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.

R1C15: We agree, the sentence was 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.

R1C16: We agree, in our reworked MS, we hypothesize that both DOC production and transport are predominantly controlled by the micro-topography of the RZ (lateral variability), and by the depth of the riparian groundwater level (temporal variability).The hypothesis was tested by a combination of field measurements and detailed hydrological modelling. The discussion and conclusion were adapted accordingly to respond to this hypothesis.

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

R1C17a: We agree, we reworked the entire paragraph. The first methodological parts of this paragraph were deleted and instead replaced by a problem to solve.

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

R1C17b: We agree, please see comment above (R1C17a).

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

R1C18a: We agree, transects of the conducted electric resistivity tomography are shown in Figure 1b. The captions were 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?

R1C18b: We agree, locations of in- and outlet are shown in Figure 1b.

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

R1C19: We agree, the sentence was rephrased to

We installed 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.

R1C20a: We agree, we changed the sentence to

In addition we irregularly installed three wells with screens at 0.3 m depth (but no pressure transducers) inside the piezometer network 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.

R1C20b: We agree, the section was 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?

R1C21: We agree, the number of sampling dates was added. We further realized that the referee was confused about the description of FTICRMS samples. We therefore reworked the entire paragraph.

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

R1C22: The sentence was changed to

Samples were filtered (0.45 µm membrane cellulose acetate filters, rinsed with 20 mL of sample water to avoid bleeding; Th. Geyer, Germany) and acidified to pH 2 (HCl, 30 %, Merk, Germany) on site. Subsequently samples were stored cool (4 °C) and dark until timely 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.

R1C23: We agree, we clarified this in the MS and provided the objective function in the SI (S2)

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

Where $O_q^i$ and $S_q^i$ are the observed and simulated discharge. *nq* is the number of number of the discharge observations (611). $O_l^i$ and $S_l^i$ are the observed and simulated groundwater level. *nl* is the number of groundwater level observations (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.

R1C24: The DInf algorithm is explained in the revision.

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

R1C25: We agree, the sentence was changed to

Discharge showed high 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 (Fig. 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?

R1C26: This can be derived from Figure S10. We have noticed that the colors in this Figure were wrong and included a corrected version of Figure S10:

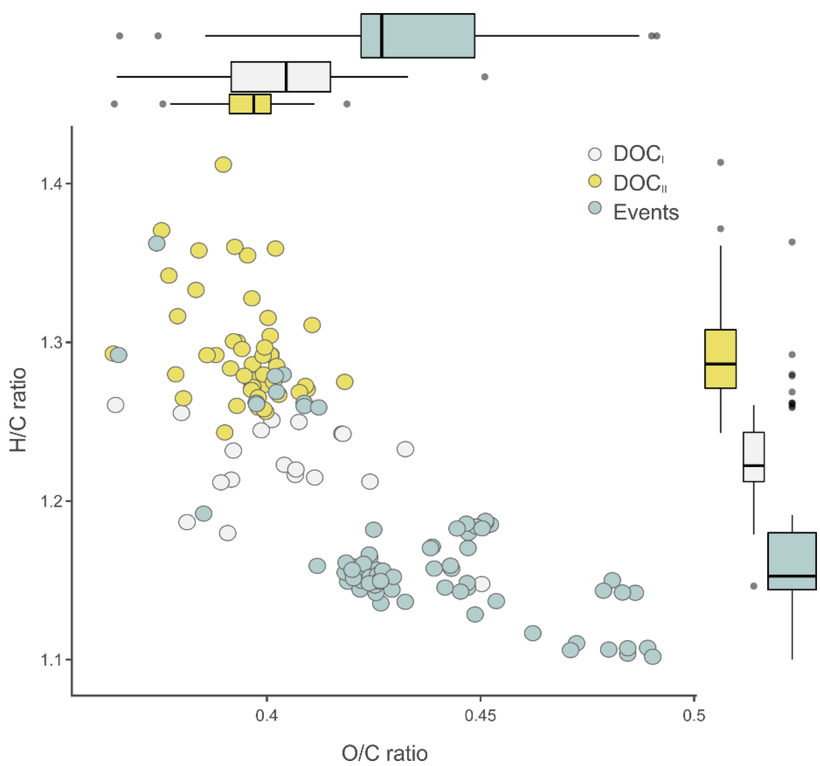


**Figure S10: Aggregated van Krevelen plot of all FT-ICR-MS sample of stream (blue) and riparian origin with type DOCI (grey) and DOCII (yellow). Data represent the intensity weighted average (wa) of the molecular H/C and O/C ratios considering all valid MF in these samples (n = 142). See also Table S5 for individual values.**

Besides, the sentence in L360 (R1C26) was changed to

In general, DOC in riparian water samples was of unsaturated and phenolic composition (waHC = 1.27 ± 0.05; waOC = 0.40 ± 0.01; n = 66), that is typically found in wetland surface soils (LaCroix 2019). Stream event samples significantly differed ($p < 0.001$) from riparian samples and were more unsaturated (waHC = 1.17 ± 0.05; n = 76) and more oxygenated (waOC = 0.43 ± 0.03) as shown in Fig. 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.

R1C27: We agree, a quantitative assessment of how many piezometer remain in the same cluster or change clusters was added:

Note that 7 (i.e. 6 wells and one surface pond sample) out of 15 locations occur in both DOC clusters as DOC quality varied over time. According $TWI_{HR}$ values also contribute to both clusters (Fig. 4d)

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.

R1C28: We agree, the correct term in this sentence is DOCI and DOC II *clusters*, not *wells*. We changed 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.

R1C29: 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 changed the sentence to… by using the median $TWI_{HR}$ value of the well locations of the $DOC_I$ cluster (9.66) as a threshold. Using this manually chosen threshold allowed to allocate the samples of both DOC clusters to two distinct $TWI_{HR}$-based groups. In this way, more than 50% of samples contributing to the $DOC_I$ cluster constitute one group while allowing less than 25% (15% in April) of the samples contributing to the $DOC_{II}$ cluster in that group

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

R1C30: We agree, numbers were added in the manuscript.

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

R1C31: We agree, this was 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?

R1C32: We agree, ± sd was added:

During the model period, $DOC_I$ 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 $DOC_{II}$ 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?).

R1C33: 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 (R2C4a, b) we included this discussion in the MS..

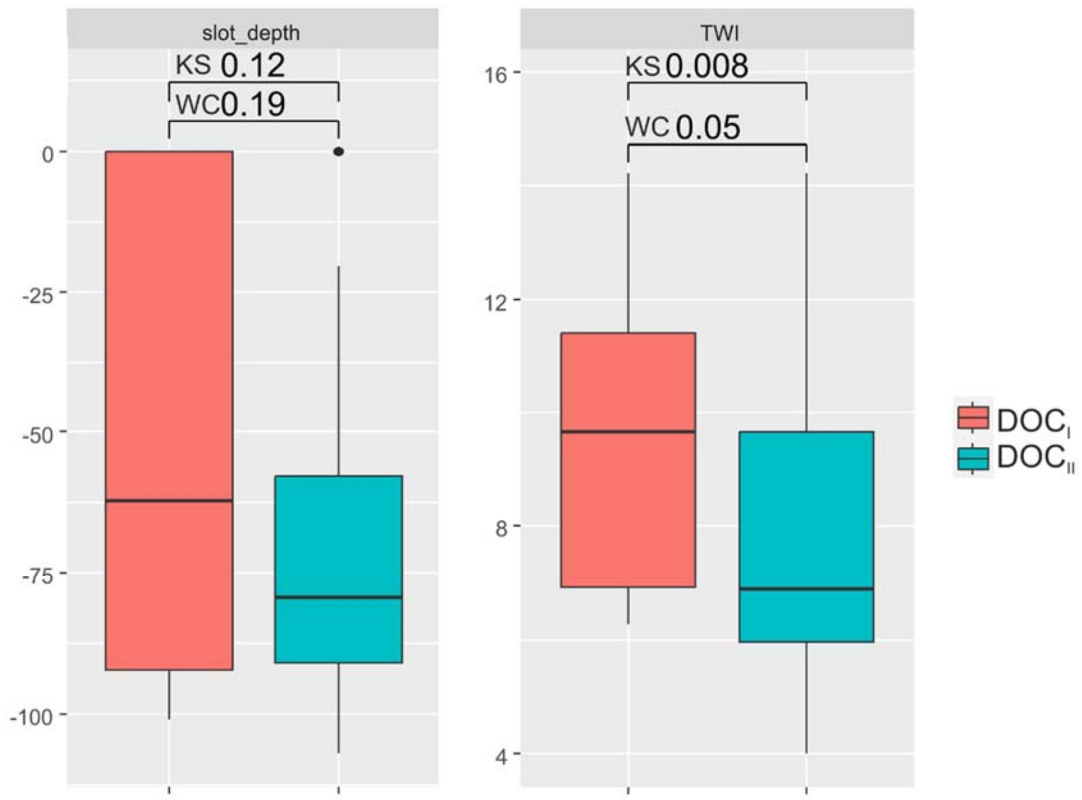


Figure: Boxplots of piezometer slot/sampling depth [cm below ground] and TWI [-] value for the $DOC_I$ and $DOC_{II}$ clusters. Horizontal brackets above describe the Kolmogorov-Smirnoff (KS) and Wilcoxon rank sum (WC) test statistics. Values were min-max normalized to values between 0 and 1 for better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

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

R1C34: We agree, the sentence was changed to

In addition, the $DOC_I$ quality was similar between April and December indicating a DOC pool which is not strongly affected by seasonality and hydrologic conditions (i.e. steady export during the wet and cold state). Therefore $DOC_I$ can be regarded as a permanently available source of stream water DOC.

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

R1C35: We agree, the sentence was changed to

In contrast, the $DOC_{II}$ composition was reflected in the DOC composition of stream water in December but not in April, suggesting a connectivity of this pool during high flow periods but a potential depletion over time.

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.

R1GC on Discussion: We agree. In consistence with Referee #2 we reworked the introduction, discussion and conclusions section (see general comment and comments on L12, L27). Here primary focus lied on the following tasks:

- The new hypothesis "We therefore hypothesize that both DOC production and transport are predominantly controlled by the micro-topography of the RZ (lateral variability), and by the depth of the riparian groundwater level (temporal variability)" was carefully established, discussed and ultimately answered and more clear key messages were worked out.
- We ensured that drawn conclusions were applicable to similarly shaped riparian zones.
- Further, results were better put into perspective and referenced to the literature e.g. by answering Referee #2's comments on horizontal vs. vertical heterogeneity (R2C4a).
- We added a new discussion section "4.4 Potential for future work and implications" and a paragraph in the conclusions to better promote implications for management and research.

Response to Referee 2

Werner et al. examines export patterns and dynamics of dissolved organic carbon from the riparian zone in a temperate, forested catchment. The paper used an array of different approaches to relate DOC source zones within the RZ to their dominant DOC export mechanisms. Stream DOC samples from different hydrological conditions were compared to riparian DOC groundwater and surface water chemistry. They also characterized DOC chemically (via Fourier-transform ion cyclotron resonance mass spectrometry) and used topographic analysis (at a resolution of 1m). Water fluxes were simulated using the code HydroGeoSphere. The paper concluded that surface runoff from zones of high TWIHR values, which occupied about 15% of the total area, exported about 1.5 times the load of DOC from the remaining 85 % of the area, and that "this study highlights that surface DOC export from the riparian zone plays an important role for lateral DOC export from hydromorphic soils with overall low topographic relief."

The work is interesting and collated an array of approaches from chemical analysis to modeling at high spatial resolution. The current manuscript is phrased from the angle of horizontal heterogeneity / landscape topography, which I think is actually already well studied [Herndon et al., 2015; Jencso et al., 2009; Ledesma et al., 2018; McGuire and McDonnell, 2010; Pacific et al., 2010]. But I find the conclusion is not particularly surprising.
On the other hand, it seems to me that this work presents a rare opportunity to dig deeper to think about the relative influence of vertical versus horizontal heterogeneity. The relative importance of vertical versus horizontal heterogeneity in doc export is poorly understood. In particular, there has been quite some interests in understanding the solute export from different subsurface depths, for example, [Seibert et al., 2009; Zhi and Li, 2020; Zhi et al., 2019]

We appreciate your evaluation of our Manuscript (MS). Your comments as well as those by the other Referee made it clear that we needed to phrase our conclusions more crisply and clarify the contribution of our work to the existing body of knowledge and its practical implications (see also Referee #1, General Comment on "4 Discussion"). We discussed the references you mention and put our results into perspective (see section 4.3, 4.4 and R2C5).

Please note that we reworked the introduction, discussion and conclusion sections to a greater extent. Therefore it was to some (small) degree not possible to answer the Referees comments explicitly anymore. In those cases, the questions and remarks were implicitly addressed in the rewritten text.

R2GC1: Please note that most of the references you mention (Herndon et al., 2015; McGuire and McDonnell, 2010; Klaus and McDonnell, 2013; Zhi and Li, 2020; Zhi et al., 2019) are concerned with much larger scales than our work, which studies an individual riparian zone (RZ) in detail. The scale of our experimental site is the smallest on which a hydrological landscape unit (in this case, the RZ) can be studied in its entirety and the largest that still allows extensive monitoring in the field during several seasons within realistic constraints on resources and personnel. The paper's contribution lies in the combination of multi-sensor field monitoring backed up by detailed hydrological modelling and high-resolution chemical analyses of DOC quantity and quality. This allowed us to determine the size and contribution of hydrologically different DOC source areas, something that so far has only rarely been quantified (Bernhardt et al., 2017). We better clarified this in the text (section 4.3) and improved the explanation of the practical relevance of some of our findings (section 4.4).

The relative influence of vertical versus horizontal heterogeneity was discussed in our MS (R2GC1, see also Referee #1, L487). As mentioned above, we found that surficial DOC export from high TWI zones dominates DOC export from riparian zones to the stream at our study site suggesting that horizontal heterogeneity predominates vertical heterogeneity in low relief catchments with hydromorphic soils. With our work showing that surface flow is an important carrier of DOC stemming from localized source areas, it contrasts the concepts of a dominant source layer in the subsurface (Ledesma et al., 2018) or a riparian integration model (Rim, Seibert et al., 2009) that both hypothesize predominantly horizontal subsurface flow as the main transport mechanism for DOC. This suggests – as also assumed by Jensco et al., 2009 – a dependency of DOC export on morphologic, climatic and topographic conditions. Given the nature of our study site (steep hillslopes, level riparian zone, low conductivity soils), overland flow is to be expected as a significant export pathway connecting riparian DOC sources to the stream, whereas subsurface DOC export heterogeneity seems to be relatively unimportant. We discussed this in section 4.2 and 4.3 (R2C5 and R2GC1, resp.).

The data from this work have depth profile (top 100 cm) of doc, and flow calcination from different depths. These two can be combined to calculate at what depth most doc was exported, and how the export varied with depth in high flow events. At a minimum, it would be nice to see some discussion along this line of vertical heterogeneity.

We agree, some discussion along this line has been added to the MS (see also: comments below, specific response to comment of Referee #1, L487, comment above). We point out that the top few decimeters of soil was very organic and contained many rhizomes. Somewhat deeper, the soil contained so many rock fragments that digging with a spade was no longer possible. At roughly 1 m (with considerable variations), fractured bedrock occurred. To study solute export from various depths in the field, tracer experiments are needed. However, taking samples at well-defined locations in this soil proved very difficult, even in the dry season when the groundwater was not near the soil surface. We do not believe it is feasible to excavate the soil in intervals of 5 or 10 cm to study the distribution of a dye tracer. Given the difficulties associated with tracer experiments to identify flow paths in this soil that we outlined here, this discussion was based entirely on the numerical model and therefore has to rely on a schematized conceptualization of the subsurface.

I also find “Surface export” is a confusing term. Is this really surface runoff, or does the water mostly flow through top soil? Unless in extremely large events, most forests do not see significant amount of surface runoff. In many places, stream water comes from “old” water from the subsurface, not surface runoff “new” water [Klaus and McDonnell, 2013].

R2GC2: There is no ambiguity in our use of the term. The RZ on which this paper focuses was not forested, although the surrounding slopes were. Overland flow in the RZ was quite common during wet periods (R2C5). With “surface export” we refer to water that has been on the surface at least once on its way to the stream – as indicated by respective exchange fluxes. We clarified

this in the text (R2GC2). Surface flow can reinfiltrate and flow to the stream/ boundaries through the subsurface.
The steep hillslope but mild slope and hydromorphic soil in the study site, the micro-topography, as well as scale and climatic setting of our field site all are in stark contrast with the features of the sites reviewed by Klaus and McDonnell (2013). Also, they view water contributions from a watershed/catchment-scale perspective, whereas we focus on the RZ itself. Given the nature of our study site, more overland flow is to be expected (although Klaus and McDonell also reviewed studies that had up to 100% "new" water contributions). We explicitly mentioned that our findings hold for a low relief riparian zone with hydromorphic soils to account for these differences (R2C5).
As discussed below (comment on L525 and following) we therefore do not think that the concept presented in Klaus and McDonnell holds for our field site.

Line 52-54, "a strong focus on vertical heterogeneity". Interesting thoughts but maybe not accurate. My impressions is that existing literature has focused much more on landscape hillslope - riparian heterogeneity. As I mentioned earlier, papers in hydrology and ecology have emphasized a lot on hydrological connectivity from hill to streams. In fact, the management practices related to riparian zones originated from our understanding of differences between hill and riparian and their connectivity.

Referee #1 and #2 both pointed to several studies that cover lateral variability. Therefore we changed this sentence as described in our response to Referee #1, L49 (R1C12&13).

Figure 3: also draw doc in this figure to help viewing when doc coming out most?

R2C1: We agree, DOC measurements (riparian, stream routine and event auto samples) were added to Figure 3.

Figure 4: this figure is busy. What is ns, hc, oc, … please explain in caption or provide legend. Why not show doc vs depth data. It would be cool to see that data. We rarely have subsurface solute depth profile. Also, these depth data, together with the modeling work for subsurface flow, provide rare opportunity to assess the relative importance of vertical heterogeneity vs horizontal landscape heterogeneity, as I I mentioned earlier

R2C2: We added the requested abbreviations to the captions.
Regarding the depth data, we are a bit in limbo. It is clear that this referee has a keen interest in the vertical distribution of DOC transport (see Figures below for DOC depth profile), but because of the nature of the soil and the bedrock, we were unable to explore this in the field in a meaningful way, as explained above. In principle, it is feasible to interrogate the model results to tease out the variations of DOC movement along the vertical dimensions but it has to be understood that the usual limitations to detailed interpretations of modeling results apply: the model is elaborate and fully 3D, but it remains a schematization of the real world, and we are not sure that an analysis of the numerical results at the level of detail desired by the referee is justifiable.

Furthermore, a discussion of the fine details of the model results will distract from the more practical aspects that referee 1 requested us to address. To us, these appear to be of more interest to the HESS readership.

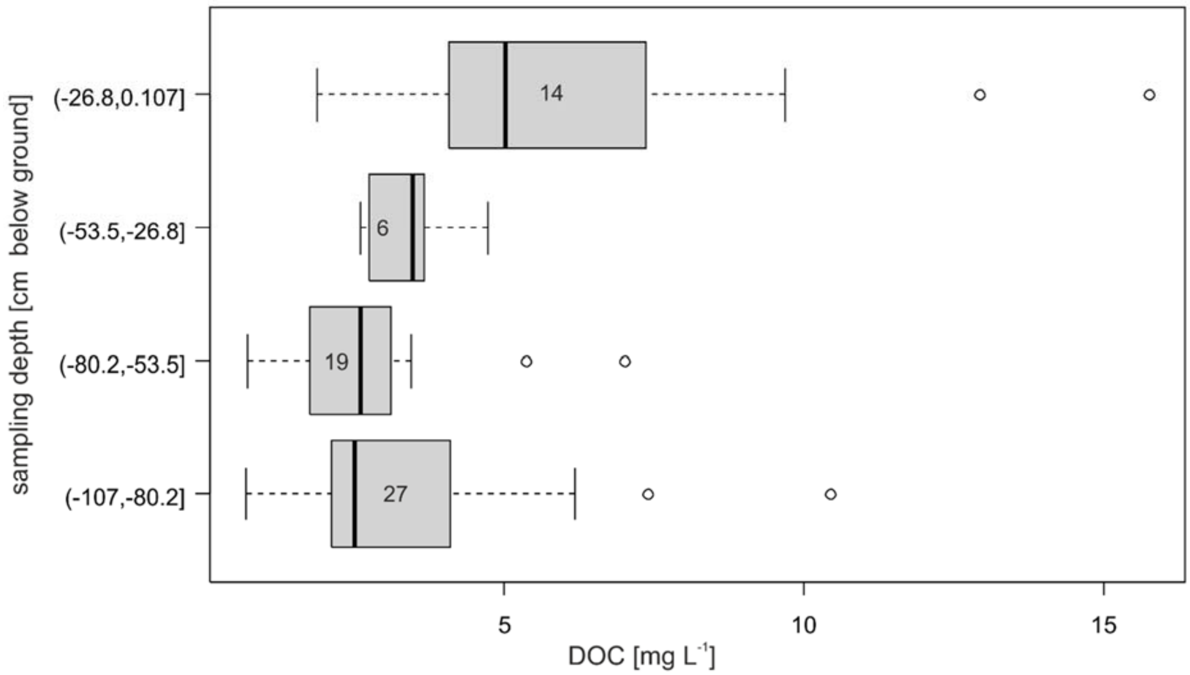


Figure: Boxplots of DOC concentration vs. sampling depth (negative is below ground). Numbers in boxplots indicate sample size.

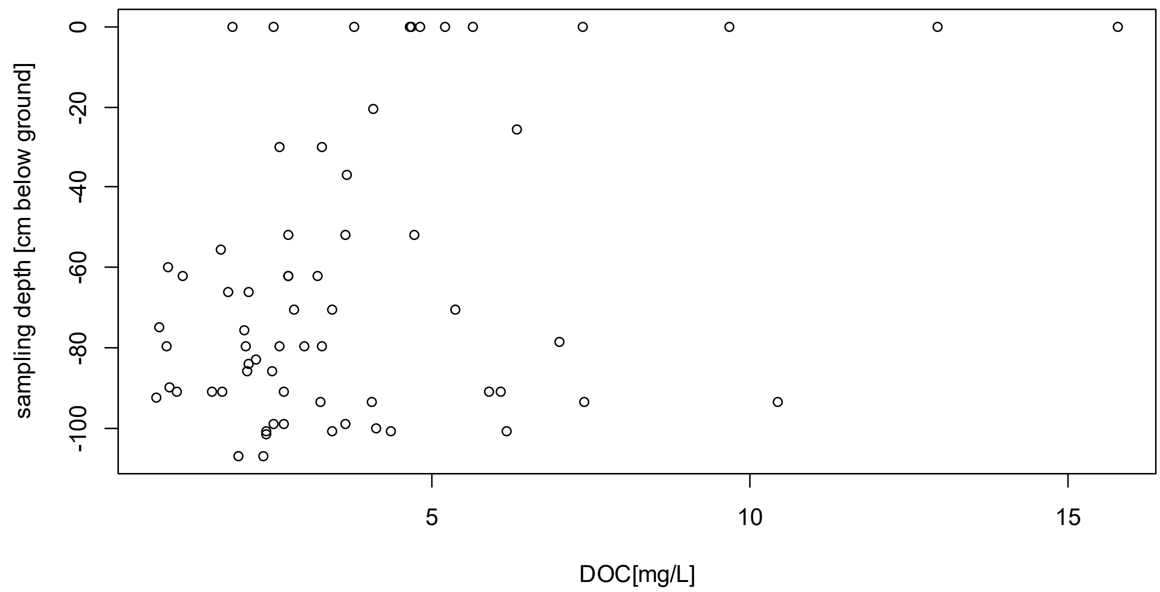


Figure: Plot of DOC concentration vs. sampling depth. Surface pond/soil solute sampling depth was set to 0 cm.

Figure 7: can the discharge data be added here? Would it be easier to understand the time series of doc export?

R2C3: We agree, discharge data was added to Figure 7. Figure captions were adapted accordingly.

Line 525-530: it seems that there is some mis-understanding about "lateral export". Lateral export means doc export via surface water (streams and rivers). Stream water can come from the surface runoff and subsurface (soil + gw).

R2C4: We apologize for the misunderstanding, we changed the sentence to

In this regard, we found that surficial DOC export dominated overall DOC export to the stream at our study site.

In fact, in many places, stream water comes from "old" water from the subsurface, not surface runoff "new" water (Klaus+McDonnell 2013). While I agree that surface runoff can be important during events, it may be misleading to present these numbers without mentioning the temporal scale (event scale). At the annual scale, these numbers might be quite different.

R2C5: At the annual scale, the median contribution of surficial runoff to total runoff generation was 61 % (± 12 % standard deviation), but at the event scale surface contributions increased up to 99 % during event situations (L424, Figures 6 and 7 in the MS). Note that the stream runoff is erratic (Q ranges between < 0.01 in dry summer and > 1.1 $m^3 s^{-1}$ in wet winter). Most of the runoff generation in our study site occurs during events under wet, saturated conditions. Furthermore the riparian study site has an overall low topographic relief and consists of hydromorphic soils of typically low hydraulic conductivity. The high groundwater level and the soil properties both increase the probability of surface runoff generation under the given conditions.
The generated surface runoff might be mixing of "old" (pre-event) water that exfiltrates from the subsurface (as indicated in our study and suggested by Frei et al. (2012)) with "new" (event) water, but there is no isotope data or StoreAge Selection function (SAS) at hand to investigate the age distribution or preferred selection of the generated runoff. Klaus and McDonnell (2013) arrived at a different conclusion because their spatial perspective was very different from ours (see above), and possibly they were interested in much larger spatial scales. We further want to note again that the range of surface contributions ("new"/event water) in our catchment is in line with studies mentioned in Klaus and McDonnell (2013).

We discussed and emphasized the different conclusions from Klaus and McDonnel in the MS accordingly.

References:

Herndon, E.M., Dere, A.L., Sullivan, P.L., Norris, D., Reynolds, B. and Brantley, S.L. (2015), Landscape heterogeneity drives contrasting concentration-discharge relationships in shale headwater catchments, Hydrology and Earth System Sciences, 19(8), 3333-3347.

Jencso, K.G., McGlynn, B.L., Gooseff, M.N., Wondzell, S.M., Bencala, K.E. and Marshall, L.A. (2009), Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale, Water Resour. Res., 45(4), W04428.

Klaus, J. and McDonnell, J.J. (2013), Hydrograph separation using stable isotopes: Review and evaluation, Journal of Hydrology, 505, 47-64.

Ledesma, J.L.J., Kothawala, D.N., Bastviken, P., Maehder, S., Grabs, T. and Futter, M.N. (2018), Stream Dissolved Organic Matter Composition Reflects the Riparian Zone, Not Upslope Soils in Boreal Forest Headwaters, Water Resour. Res., 54(6), 3896-3912.

McGuire, K.J. and McDonnell, J.J. (2010), Hydrological connectivity of hillslopes and streams: Characteristic time scales and nonlinearities, Water Resour. Res., 46(10).

Pacific, V.J., Jencso, K.G. and McGlynn, B.L. (2010), Variable flushing mechanisms and landscape structure control stream DOC export during snowmelt in a set of nested catchments, Biogeochemistry, 99(1-3), 193-211.

Seibert, J., Grabs, T., Köhler, S., Laudon, H., Winterdahl, M. and Bishop, K. (2009), Linking soil- and stream-water chemistry based on a Riparian Flow-Concentration Integration Model, Hydrology and earth system sciences, 13(12), 2287-2297.

Zhi, W. and Li, L. (2020), The Shallow and Deep Hypothesis: Subsurface Vertical Chemical Contrasts Shape Nitrate Export Patterns from Different Land Uses, Environmental Science & Technology, 54(19), 11915-11928.

Zhi, W., Li, L., Dong, W., Brown, W., Kaye, J., Steefel, C. and Williams, K.H. (2019), Distinct Source Water Chemistry Shapes Contrasting Concentration-Discharge Patterns, Water Resour. Res., 55(5), 4233-4251.

Additional References:

Bernhardt, E. S., Blaszczak, J. R., Ficken, C. D., Fork, M. L., Kaiser, K. E., and Seybold, E. C.: Control Points in Ecosystems: Moving Beyond the Hot Spot Hot Moment Concept, Ecosystems, 20, 665-682, 10.1007/s10021-016-0103-y, 2017.

Frei, S., Knorr, K. H., Peiffer, S., and Fleckenstein, J. H.: Surface micro-topography causes hot spots of biogeochemical activity in wetland systems: A virtual modeling experiment, Journal of Geophysical Research-Biogeosciences, 117, n/a-n/a, 10.1029/2012jg002012, 2012.