Response to **Anonymous Referee #4**

General comment

I argue that the authors have over-emphasized the trend of increasing DOC flux trends in the abstract and introduction. This is an important reason to study this subject and this work could certainly be used to better understand the mechanisms driving this trend, but the paper includes neither a report on this increasing trend or evidence for a mechanism for this trend. I think that most of the parts of the introduction are there, but I suggest that the text focus more on the aspects of DOC export reported on in the paper (transport of DOC from watersheds across hydrologic regimes and antecedent conditions). One thing that is missing from the introduction is any mention of antecedent moisture conditions. I think that discussing the role of antecedent conditions, discharge-normalized temperature, and their potential role on DOC quantity and quality should be discussed before the methods section since these are a major focus of the paper and the conceptual model discussed later (figure 6).

(R4GC1)

We agree with Referee#4, we will reorganize the introduction and add a section of antecedent (moisture) conditions to it, since they are a central focus of our findings. The introduction will be more focused on how event-scale DOC quality dynamics in headwater streams are linked to DOC mobilization processes in the riparian zone and how high-frequency measurements of C_{DOC} and especially spectral properties can be utilized to identify and quantify the key controls of DOC export from event to seasonal time scales. See also our response to the general comments of Referee #2 (R2GC1) and #3 (R3GC1), who similarly raised this concern before.

Specific comments:

Page 3; Lines 14-16: I highlight this sentence because I think that it does an excellent job of encapsulating this study. I suggest the authors reorganize the introduction to better emphasize concepts related to this idea.
We agree, this describes the general claim of the study. As written above, we will reorganize the introduction towards a sharper focus on these claims (see also R4GC1).

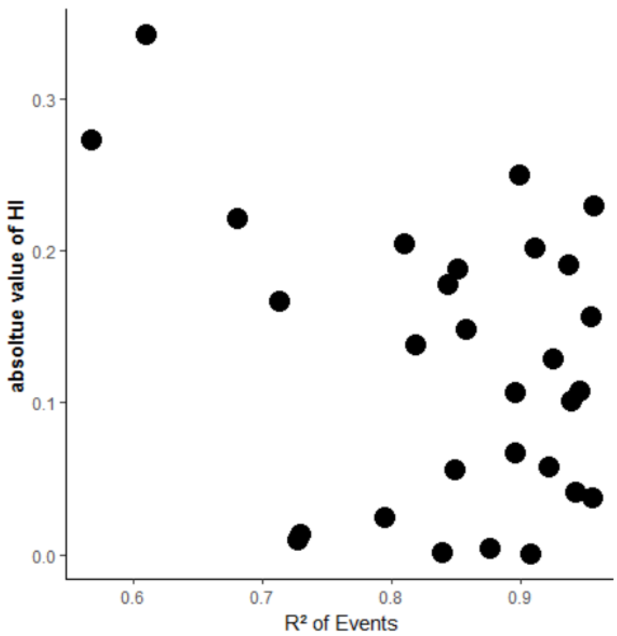
Page 4; Line 27: Were grab samples also run for spectral slope values in addition to UVA_{254}? I also suggest that the authors provide some additional detail about sample collection (e.g., filter size, sample handling).
(R4C1.1) No, there were no grab samples run for spectral slope values. This was written in the results section (P. 8 L. 11-12) but a description will be added in the Methods section in the manuscript (MS). Details about samples collection are provided in the S1 section (referred to at P.4 L.25).

Page 4; Line 29: Some brief information about methods for DOC analysis (e.g. acidification level) would be of use for the reader.
(R4C1.2) We agree. The requested information will be added to the MS.

Page 5; Line 6: I suggest the authors report how closely SUVA values obtained from the sensor match the grab sample values.
(R4C1.3) Respective information can be found in the results section (P.8 L.7-10). We will reference to this in the MS.

Page 6; Line 21: I'm concerned that hysteresis loop size is biasing regression slopes obtained from this method. I would appreciate some support for application of this method to events with varying degrees of hysteresis.

Although hysteresis of C-Q relationship potentially could explain some deviations of our hydrological event models we did not take hysteresis into account. However the high overall R^2 values of our event models (Figure 3) indicate that the influence of hysteresis on the R^2 should be minor. Evaluating hysteresis index (HI) after Lloyd et al. (2016) against R^2 of events (Figure below) indicated a negative, but non-significant effect of magnitude of hysteresis (depicted as absolute value of HI) on R^2 (method of linear regression: DOC ~ Q, [DOC~log(Q) was used where appropriate). Overall, Pearson correlation of HI~R^2 of Events was $r^2 = 0.12$ ($r_{Pearson} = -0.34$, $p = 0.07$), supporting the application of our method without explicit consideration of hysteresis effects.
The discussion and Figure will be analogously implemented in the MS and SI, respectively.



Page 7; Line 27: Reporting an actual mean AI or similar number would be helpful for the reader.
(R4C3) We agree. The median AI60 will be added in the MS.

Page 8; Line 28: I think that there are better ways to present this information. As written, it is hard for the reader to tell what the readers should take away from this paragraph.
(R4C4) We agree. This paragraph will be clarified in the MS.

Page 9; Line 16: Presenting this information here is repeatitive. I argue this authors should move some of this material to the methods section where similar methods are already covered.
(R4C5.1, R4C5.2) We agree. Repeated information will be moved to the methods section.

Page 9; Line 30: I am concerned about the interpretation of individual regression coefficients from a multiple regression of observational data due to issues of multicollinearity. In other parts of the analysis, partial least squares regression is used to address this issue, but for this analysis it appears that multiple linear regression was used instead (Page 6, Line 18).

(R4C6) This is true. We used multiple regression analysis. However, predictors (variables and interaction terms) were tested for multicollinearity (by looking at the variance inflation factor, c.f. P.6 L.29 – P.7 L.4) and excluded from the models if there was severe multicollinearity between the predictors. This will be remarked earlier in the method section of the MS.

Page 10; Lines 9-20 and Table 3: Values of a change dramatically depending on whether a 15 or 30 day lag is used. For example, a (C_{DOC}) is negatively correlated to T_{15}, but positively correlated to T30. The same is true for Q_{15} vs Q_{30}. This pattern is reportead for a ($SUVA_{254}$) and a ($S_{275-295}$). It would be interesting to see if the correlation between a and DNT_{15} is negative. This would seem to change the implications of the study substantially.

We agree. Unfortunately when checking the correlation table again, it turned out that there was an error in the script for T_{15} and Q_{30}, correlating the model parameter to some other/wrong variables instead. True correlation of 15 and 30 day aggregations fit together as it would be expected, hence no substantial implications and changes have to be expected from the (more in line) new correlation table. We apologize for this mistake and thank the Referee#4 for his/her thoughtful review. The mistake will be changed in the MS by replacing Table 3 by

Model Parameters	T_{15}	T_{30}	Q_{15}	Q_{30}	AI_6	AI_{14}	AI_{60}	DNT_{30}	Q_{hf}	Q_b
z (C_{DOC})	*0.05*	0.05	0.02	-0.02	0.05	0.07	-0.09	0.03	0.15	-0.12
a (C_{DOC})	*0.55******	0.52***	-0.48**	-0.43**	-0.52**	**-0.65*****	**-0.66*****	**0.63*****	-0.55***	**-0.71*****
b (C_{DOC})	*0.25*	0.25	-0.31	-0.31	-0.19	-0.33*	-0.15	0.32	-0.38*	-0.25
z ($SUVA_{254}$)	*0.07*	0.06	0.04	-0.06	-0.10	0.04	-0.10	0.04	0.01	-0.09
a ($SUVA_{254}$)	*0.50****	0.51**	-0.50**	-0.40*	-0.42**	-0.56***	**-0.64*****	0.58***	-0.54***	**-0.60*****
b ($SUVA_{254}$)	*0.21*	0.18	-0.32	-0.22	-0.10	-0.34*	-0.14	0.25	-0.29	-0.23
z ($S_{275-295}$)	*0.00*	-0.02	0.21	0.11	-0.09	0.23	0.04	-0.10	-0.02	0.07
a ($S_{275-295}$)	***0.62******	**0.63*****	-0.54***	-0.41*	-0.28	-0.47**	-0.56***	**0.62*****	-0.47**	**-0.64*****
b ($S_{275-295}$)	*0.13*	0.11	-0.31	-0.18	-0.12	-0.45**	-0.14	0.19	-0.20	-0.24

Additionally, DNT_{15} is in line with DNT_{30} (corr. With coefficients z (C_{DOC}) until b ($S_{275-295}$) (from top to bottom): 0.03, 0.69***, 0.33*, 0.03, 0.65***, 0.30). Yet DNT_{15} does not add any new variance to the proposed models in the paper. When replacing DNT_{30} by DNT_{15}, R^2 of C_{DOC}, $SUVA_{254}$ and $S_{275-295}$ for the complete models drops to 0.64, 0.59 and 0.59, respectively.

Page 10; Line 28: Referring to this analysis as seasonal-scale is somewhat confusing to me because there is no specific season analysis conducted here (rather variables that AI that typically change with season are used). I would also like to know if any data were held out for model validation or if the R^2 statistics in Table 4 are for the model with reference to the training dataset alone.

The expression "seasonal-scale" describes the time-scale in which parameters change. This is in the time-scale of seasons (roughly 3 months) and not to be mixed up with spring, summer, fall and winter. When we speak of seasonal-scale analysis, we argue that according variables describe our hydroclimatic "seasons" in terms of "warm & dry", "cold & wet" and "intermediate". We will clarify our definition of seasonal-scale in the MS.

The R^2 in Table 4 refers to the complete data set models (as depicted in the referred Equation 3) for modeling DOC concentration and quality. The complete data set models were five-fold cross validated to estimate the prediction error (c.f. P.7 L.5-9). A trainings data set was only used for the PLS regression in order to derive DOC concentration from absorption spectra (see section 2.2.1).

Page 13; Lines 9-12: The different weather scenarios make sense, but I think that stating that there are three discrete states is a bit arbitrary and is not supported by any sort of data or analysis. I think the general framework is right and it makes sense for the authors to highlight certain scenarios. That said, I don't see evidence in the data for discrete states but for a continuum without jumps from one state to another. I recommend the authors clarify the nature of their conceptual model.

(R4C7) We agree. The three discrete system states were chosen for the conceptual model to highlight certain, typical scenarios out of a continuum (see also comment above). We will clarify this in the MS.

Figure 6: How baseflow DOC concentration changes with season was not supported with particular numbers in the results, but appears to be important for the conceptual framework. It is discussed briefly in a qualitative fashion, but the degree of seasonal differences in DOC concentrations are hard for the reader to infer.

We agree. We want to point out that baseflow levels under cold and wet conditions are usually higher than baseflow levels during the warm and dry phase (see Fig 5). Thus, during the cold and wet situation, higher layers of soil, more enriched in DOC get activated, but at the same time, there is also a tradeoff between amount of water (see also Referee#1, R1C31) and available DOC in the respective soil layers which can account for lower DOC concentrations. Particular median C_{DOC} values were 4.13 mg L^{-1}, 3.72 mg L^{-1} and 3.16 mg L^{-1} for the warm & dry, intermediate and cold & wet state, respectively. Both warm & dry and intermediate state differ highly significant (Kruskal-Wallis test, $p < 0.001$) from the cold & wet state. According to the significance of the different hydroclimatical situations, initial C_{DOC} values of warm & dry and intermediate will be adjusted to a higher level than the cold and wet situation. Particular numbers will be integrated in the MS.

References cited in response to Reviewer #4

Lloyd, C. E. M., Freer, J. E., Johnes, P. J., and Collins, A. L.: Technical Note: Testing an improved index for analysing storm discharge-concentration hysteresis, Hydrology and Earth System Sciences, 20, 625-632, doi:10.5194/hess-20-625-2016, 2016.