Answer to the review by Dr Bechtold

We thank Dr Bechtold for the thoughtful review of the paper. We agree with all the points raised and we believe we can address them all in order to improve the quality of the paper. We explain below how we plan to address each of them in details.

Review of manuscript "Hydrological control of dissolved organic carbon dynamics in a rehabilitated Sphagnum–dominated peatland: a water-table based modelling approach" submitted to HESS by Bernard-Jannin et al.

The paper deals with the effect of peatland rewetting/restoration on DOC dynamics. The study combines field monitoring (DOC pore water concentrations [DOC] and water table depth) with hydro-biogeochemical modeling. Two monitoring sites were installed within the same acidic fen peatland, one defines the control site (disturbed, drier conditions) and one the rewetted case due to recent (3 years) blocking of a ditch. Field data indicate a strong effect of the rewetting on the water table depth (>20 cm shallower water levels during summer) and higher [DOC] in summer in the rewetted location compared to the control. Also the quality of DOC seems to change due to rewetting. Modeling results indicate very contrasting fluxes at the two sites, fast runoff with very low evapotranspiration at rewetted sites vs. slow "deep" drainage at control site, which is interpreted to control DOC composition and export.

The basic scope of the paper (topic, field data, model approach) is interesting, relevant for the community and suitable for publication in HESS. The authors dealt with a challenging site because fluxes were not observed. Calibration of a hydrological model only on water table depth data with the objective to interpret resulting fluxes is questionable although maybe not impossible when setting sufficient constraints on calibration scheme and parameters, and when providing reliable uncertainty estimates. However, as the modeling, and in particular the calibration, has been conducted and presented now, one cannot trust the highly questionable simulated hydrological fluxes (e.g. extremely low ET at rewetted site). Fluxes are crucial for the [DOC] dynamics, e.g. ET causes a 'physical' accumulation of DOC in the pore water, thus one can also not rely on the following implications for DOC concentrations and exports. In my major points, I thus focus on the hydrological part of the paper. The DOC part is commented with less detail, although it must be also fundamentally revised because I assume that results will change considerably with an update of the hydrological modeling.

In general, as mentioned before, fluxes of a hydrological model calibrated only on water table depth should be interpreted carefully. I present my major concerns and give some additional recommendations afterwards.

Major points

I see several major problems in the hydrological modeling part of the study. The hydrological model is of fundamental importance for the results of the manuscript. The authors intent "to identify the main hydrological processes and factors controlling DOC dynamics" with their model which is based on a conceptual approach based on reservoirs. Modeled are two monitoring sites that lie in the same disturbed fen peatland. I assume the two sites were very similar before 2014 (same vegetation, 60 cm peat thickness, … same peat properties?). The only that changed in 2014 was the boundary condition for the rewetted site, which is effectively a higher drainage level due to ditch blocking.

Separation into boundary-condition dependent and boundary-condition independent parameters:

Having repeated the field situation, it is only possible to rely on the simulated fluxes when all model parameters that do not depend on changed boundary conditions (i.e. here the changed drainage level) are the same for both monitoring sites. As far as I understood the three discharge coefficients (overland, lateral runoff through peat, and "deep" drainage) are dependent on the raised drainage level due to ditch blocking. As the conceptual model does not include a drainage level as a boundary condition, these parameters have to be tuned that simulated water levels match observed ones. (*This is actually a major weakness of the presented conceptual model, as for a more process-type of model only a single boundary condition would need to be changed and/or calibrated if not known precisely. Emphasize this disadvantage*.)

It is true that only one boundary condition would need to be changed for a process-type model, but additional(s) parameter(s) would be required to describe the process. Process-type models at the interface between surface and peat waters, including exchange between these two compartments could be complex and difficult to calibrate. The idea to use a more conceptual model is to have a model with few input parameters needed (at the cost of increase number of calibrated parameter indeed). This disadvantage of conceptual model over process based model will be stated in the ms.

In contrast to the three discharge coefs, I consider the ET crop coefficients and infiltration rate from the peat-macro to the peat-matrix reservoir (Imax) as "boundary-condition independent". They should be thus calibrated to the same value for both sites. In the study, however, all these parameters were calibrated separately at both sites resulting in extreme differences for the crop coefficients and Imax at the two sites. As a consequence, e.g. ET is extremely low at the rewetted site. The authors recognized that this flux is unphysical and explain it by additional open water reservoirs (not modelled) that are laterally coupled to the modeled 1D profile and buffer water table depth dynamics during high water levels. I think this might be an important process and a valid interpretation, but a dramatic reduction of ET for the rewetted site is an unacceptable result for the scope of the paper. If this process (open water storage) is really relevant, it needs to be accounted for. To further prevent such an "overfitting effect", I recommend that authors think about reliable constraints during the calibration. For example, I suggest to calibrate all parameters simultaneously for both sites keeping the "system boundary condition independent" parameters the same for both sites ('multi'-site calibration approach).

We agree with the comments of Dr Bechtold and redid the calibration by keeping ET and Imax coefficients similar for both sites in order to perform a multi-site calibration. In addition we decided to improve the choice of the calibration period by using 2014 and 2017 that are the driest and wettest years. 2015 is now used as a validation. In addition, we present additional efficiency criteria to assess the quality of the model (br^2 and RMSE). The new calibrated parameters are presented below (table 1, Fig. 1). We constrained the ET coefficients values to be in a range in agreement with observed values (Lafleur et al., 2005). The objective function is now the sum of NS for the two locations (control and rewetted) and two validation periods (2014 and 2017). We can see from the results that, despite being less performing than in the first version of the paper, the model is still able to reproduce water table for both site when taking into account a multi site calibration approach, with little differences in the coefficients found in the first version of the ms (except for evapotranspiration downstream). NS coefficients are all positives; br2 all greater than 0.2 and RMSE are between 1 and 9 cm. The model performed better for the control site in the dry year than in the wet year and the opposite can be observed for the rewetted site. Finally, we don't observe a drop of performance of the model for the validation period.

Table 1: Updated calibrated parameters and performance criteria based on a multi-site calibration approach. Range of parameters calibration are also indicated)

	REWETTED	CONTROL	Range
Kcd	0.37		0.01 – 0.5
Kcg	0.40		0.4 – 0.8
Imax	0.84		0.2 -5
αp	1.6E-05	1.9E-04	0 -0.01
αr	0.20	0.37	0 -0.5
αsr	0.20	0.27	0 -0.5
NS calib wet (2014)	0.1	0.61	
br2 calib wet (2014)	0.52	0.67	
RMSE calib wet (2014)	0.01	0.01	
NS calib wet (2017)	0.25	0.16	
br2 calib wet (2017)	0.26	0.24	
RMSE calib wet (2017)	0.065	0.080	
NS valid (2015)	0.10	0.30	
Br2 valid (2015)	0.54	0.39	
RMSE valid (2015)	0.02	0.09	

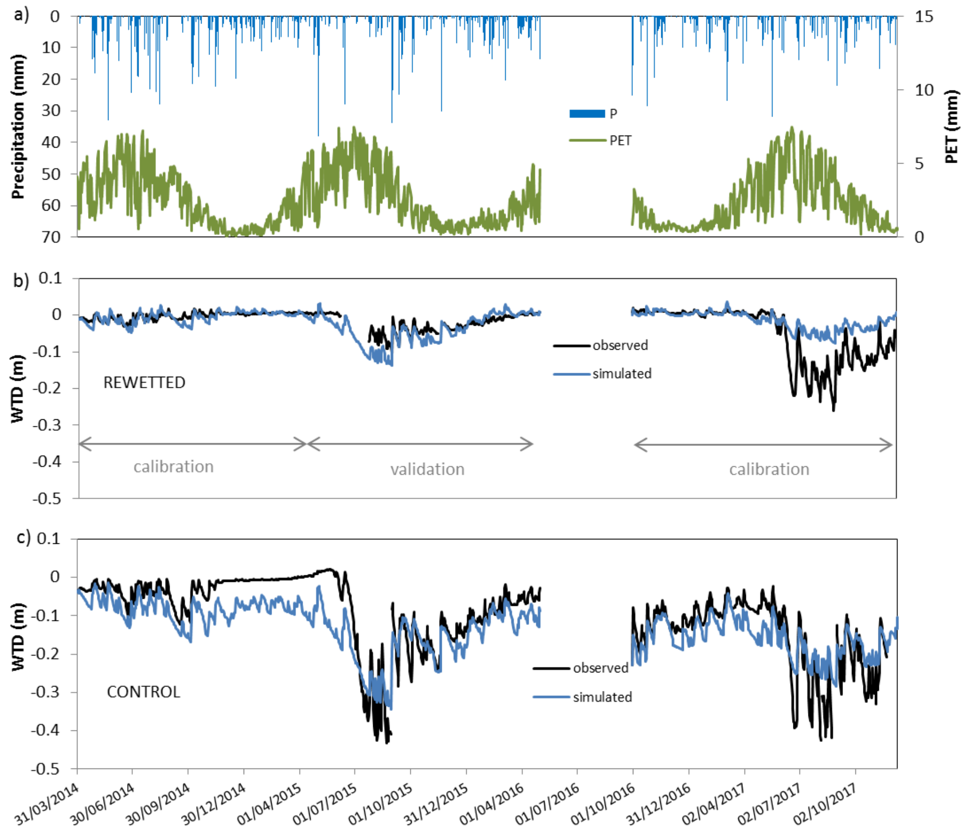

Figure 1: (a) Time series of meteorological data (PET, potential evapotranspiration and P, precipitation) used as input data in the hydrological model, (b) simulated and observed WTD in the rewetted site and (c) simulated and observed WTD in the control site. Calibration and validation periods are also indicated.

Reliability of validation period:

Validation with independent data is crucial. The authors used the third year of data for validation. This is basically ok. However, Nash-Sutcliffe efficiency (NS) drops from 0.8 (calib) to 0.3 (valid) for one of the sites. The enormous drop of performance is a strong indication of overfitting. Authors argue that it might be related to the exceptional flooding event between the second and the third year. If they are convinced that this is the reason, then the whole validation period is useless because we learn nothing about the reliability of the calibrated model from it. In fact, from what I can read and see in the paper, I have the impression that the third year can be used for validating the hydrology, because the initial water table depth of the validation period seems to be quite consistent between model and data. Or were the initial condition of the model set to the observed data? If yes this must be stated clearly, and the validation period should be shortened as validation NS would be overestimated by this manual adjustment of the state variable.

As stated in the previous point, we modify calibration and validation period. Calibration is performed on the driest and the wettest year (2014 and 2017) and validation is perfomed on an intermediate year (2016). For the simulation of 2017, the starting H is set as the observation, this will be made clear in the ms. With the new calibration strategy (multi sites) and the change of calibration/validation periods we don't observe a drop of performance between calibration and validation.

Definition of reservoirs

Sm is 0 at the lower peat boundary, why? Did authors determine in the lab that there are no macro-pores at 80 cm? At the surface the total porosity (Sm+Se) of the peat is defined to be 1. Justify also this decision, as it is not obvious how a material can have a porosity of 1. Did authors want to prevent a discontinuity between soil and surface storage? But why

The conceptual model was build in a way to obtain a good compromise between the number of parameters needed (2 parameters, Hmax and ϴinf) and a solid description of the flows at the interface between soil and surface. Sm reservoir was conceptualized to be a transitional reservoir between 100% Se peat at the lower boundary and a 100% Sr surface at the upper boundary. We thus made the following assumptions: macroporosity is null at the bottom of the peat layer and the total porosity is one at the surface, which is in agreement with observed values (Bourgault et al., 2017).

DOC measurements

The definition of the reservoirs is of particular importance also for the DOC interpretation. Were the sampling pipes emptied before taking samples? If not, it is not clear what the [DOC] in the pipes actually is (mixture of recent rain water, [DOC] of soil pore water X days/weeks before sampling minux the losses due to being 'open water' in the pipe afterwards, etc.) . If yes, mainly the macro-pore water was refilling the empty pipe afterwards, i.e. authors need to compare the field data with modeled Sm [DOC], if I understood the approach correctly. Regarding the issue of sampling different soil pore spaces see also Zsolnay, 2003, Dissolved organic matter: artefacts, definitions, and functions, Geoderma.

Indeed, the sampling pipes were emptied before sampling, so we can assume that we measure mainly [DOC] in Sm. We propose to calibrate production and loss constant on [DOC] in Sm. New results are shown in Fig 2. The model still perform well for after the changes except fot the year 2017 in control were concentration are systematically over estimated. This can be related to the difficulties of the model to simulate water table for this period. However the model simulated lower concentrations in rewetted than in control as observed. Calibrated parameters (only 2 parameters now as only DOC in Sm is simulated) and efficiency criteria are shown in table 2. Despite a relatively high RMSE, the model is able to represent the trend for most of the simulated periods. This is in agreement with the fact that hydrology plays a key role in controlling [DOC] dynamics through the water table depth as stated in the first version of the ms.

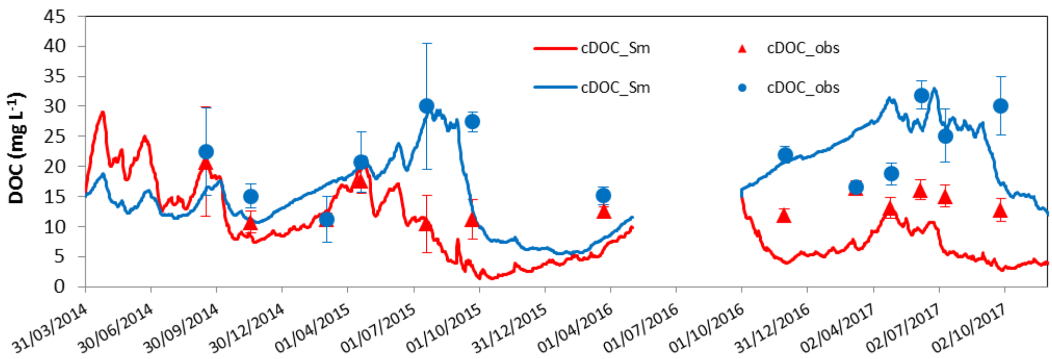

Figure 2: Simulated [DOC] in Sm and observed [DOC] after new calibration

Table 2: Calibrated parameter and efficiency criteria of the DOC model

		Rewetted	Control
	Kprod Sm	5.00E-08	9.00E-07
	Kloss Sm	5.00E-04	1.20E-02
Calib (2014)	RMSE (mg.L^{-1})	5.4	1.6
	Br2	0.003	0.89
Calib (2017)	RMSE (mg.L^{-1})	8.6	8
	Br2	0.18	0.03
Valid (2015)	RMSE (mg.L^{-1})	8.9	10.8
	Br2	0.34	0.31

Effect of flooding event on DOC validation period

What is the [DOC] in the river water? And do authors have any knowledge about how the flooding event affected the [DOC] of the peat profile in the third year? Authors need to discuss this. Till now, the third year is in particular questionable for validating the DOC model.

The DOC from the river during the flood was measured and is equal to 12 mg.L^{-1}. We think that this 50 years return period flood affected only Sr reservoir because the peat profile was already fully saturated with rain water when the flood reached the peatland. Therefore the impact of the flood on peat water and on DOC is expected to be negligible.

Recommendations:

Authors must establish much more confidence for their model. Besides taking into account my major and minor comments, I recommend that authors elaborate a sophisticated uncertainty analysis that provides uncertainty estimates for the flux components, i.e. that shows how variable the predicted flux components are for similarly performing models (similar fits of water table depth). Use probabilistic approaches like e.g. Bayesian modeling or GLUE. Authors can only generate sufficient confidence into the model results when authors are very critical with the calibration. For that, consider also the effect of optimizing the parameters for different objective functions, not just the daily NS or RMSE (e.g. different levels of aggregating the time series, daily, weekly, monthly averages) and add plausible errors to the forcing and water table depth data. These are ideas for directions authors need to look for.

Following the recommendations of Dr Bechtold, we performed an uncertainty analysis to better assess the confidence of the simulated fluxes. We ran a GLUE analysis with 300 000 runs for each location and using a criteria of NS>0.2 to select behavioral simulations. The ranges of each parameters used in the analysis are the same than in table 1.10th and 90th percentiles of each fluxes are presented in table 3.

Table 3: 10th and 90th percentiles of the fluxes resulting for the GLUE analysis (300 000 runs, behavorial runs for NS>0.2)

	Rewetted		Control	
	10th	90th	10th	90th
ET ($mm.yr^{-1}$)	307	384	349	468
R ($mm.yr^{-1}$)	94	294	24	242
D ($mm.yr^{-1}$)	2	33	37	349
O ($mm.yr^{-1}$)	199	390	0	190

Uncertainty on fluxes is high (as we can expect from the use of a conceptual model) but we can see significant differences concerning the partition between drainage and overland flow for the two locations (slow deep drainage dominated upstream and rapid surface drainage dominated downstream). These findings are in agreement with results presented in the first version of the ms. We plan to include the results of the uncertainty analysis (included a detailed description in the method) in the paper to strengthen our conclusions.

All the changes on the hydrological model calibration have been transfer to the DOC model as discussed above. The DOC balance has been recalculated including uncertainty analysis results (table 4). Results and discussion parts will be modified according to the new findings.

Table 4: DOC release for control and rewetted location over the studied period. Best indicates values for the best set of parameters. 10th and 90th percentiles resulting of the uncertainty analysis are also indicated.

	rewetted			control		
	Best	10th	90th	Best	10th	90th
DOC release ($gC\ m^{-2}\ yr^{-1}$)	1.9	1.6	5.6	2.4	0.9	9.0

Minor comments:

Page 4 - L15: “identify the locations” – I do not understand the sentence.

It will be rephrase: “to identify the significant differences between the factors”

Page 5 – L4: What is Smax? Not defined after the equation but two paragraphs later.

Smax will be defined just after the equation.

Page 6 – L32: Why t+1 for PDOC and LDOC? Isn’t it dependent on the concentration of time step t. Clarify this also in eqs 7 and 8

This was a mistake in writing the equations. Produced and consumed DOC are indeed dependent on time step t.

Page 7 – L14: How can the maximum porosity of the peat layer be equal to 1?

The maximum porosity represents the porosity at the surface, which is one.

Page 7 – L14: “flooded river” … wording. The river itself cannot be flooded

It will be change to : “the water coming from the river during floods”

Page 7 – L20: Give both NS and RMSE for both the hydrological and the DOC model. Not clear why different skill metrics have been chosen for the two variables. Would be interesting to know NS of the DOC model. It seems to be barely better than just the mean of the observed concentrations (i.e. NS~0), in particular for the validation period.

The reason we didn’t add the NS for the DOC is that NS is especially relevant when considering continuous time series with high variation (peaks in discharge or water table data). As we don’t have many DOC measurements and the variations is not as high as for water table data, we think that NS is not relevant to correctly describe the efficiency of the DOC model and that is why we just presented the RMSE that give quantitative information on the model performance. It is not unusual to not present NS when quality data are scarce, e.g. Garneau et al. (2017). We propose to present NS, Br2 and RMSE for the hydrology and RMSE and Br2 for the DOC model.

Page 7 – L22: calibrated over the calibrated hydrol. model. …. improve wording.

We will change to:”the DOC module was calibrated after the calibration of the hydrological model”

Table 1: correct table, reservoirs and discharge description are confused !

We will improve the description of discharge and reservoirs.

Table 1: How can Imax as a rate have the unit [mm]? A rate is always per time. Table 2: Format table correctly. Michel Bechtold, KU Leuven, BE; 6th Dec 2017.

This will be change. Imax is defined in mm in the original paper. The time unit was implicit and corresponds to the model time step. It will be change to $mm.day^{-1}$ (as the model uses a daily time step).

<u>References:</u>

Bourgault, M.-A., Larocque, M. and Garneau, M.: Quantification of peatland water storage capacity using the water table fluctuation method, Hydrol. Process., doi:10.1002/hyp.11116, 2017.

Garneau, C., Sauvage, S., Sánchez-Pérez, J. M., Lofts, S., Brito, D., Neves, R. and Probst, A.: Modelling trace metal transfer in large rivers under dynamic hydrology: A coupled hydrodynamic and chemical equilibrium model, Environ. Model. Softw., 89, 77–96, doi:10.1016/j.envsoft.2016.11.018, 2017.

Lafleur, P. M., Hember, R. A., Admiral, S. W. and Roulet, N. T.: Annual and seasonal variability in evapotranspiration and water table at a shrub-covered bog in southern Ontario, Canada, Hydrol. Process., 19(18), 3533–3550, doi:10.1002/hyp.5842, 2005.