Anonymous Referee #2

Received and published: 31 August 2015

Review of the manuscript “Map-based prediction of organic carbon in headwaters streams improved by downstream observations from the river outlet” by Temnerud and colleagues. This manuscript describes an attempt to model the DOC concentration in headwaters (catchments smaller than 2 km2) from nine boreal catchments (from 30 to 235 km2) combining GIS-landscape information with DOC observations from the downriver outlet of each catchment. Authors consider this study a step forward with respect to a previous similar study (Temnerud et al., 2010). In this new manuscript the step forward consist to: i) integrate into the analysis the landscape catchment properties and; ii) the implementation of complex statistic tools. Finally, the modeling effort helps to explain up to the 52% of the TOC variance in headwaters. Authors recognize that the proportion of the explained variance is not satisfactory. However also they remark that it is better than the previous work (Temnerud 2010). Therefore the main conclusions are that: i) DOC information from outlet alone is insufficient for predicting DOC (median and variability) in headwaters and ii) that, at least in these systems, GIS based catchment data is useful to improve partially the DOC prediction in headwaters. The manuscript is well written and objectives are well stated. Tables are appropriates however figures are difficult to understand. In any case, it is extremely arduous to follow and understand the modeling approach and results description. Overall, this contribution is interesting especially in a context of water quality monitoring and management. In a scientific context this study reveals that, although the GIS provide valuable information, it is a limited tool to model accurately DOC in small catchments. This suggests that important potential explanatory variables are missing in the analysis.

My most relevant comment pivots around the selection of the potential explanatory variables. Without being and expert on PLS and mixed model and being conscious of my limitation in understanding these sophisticated approaches, it surprises to me that some explanatory variable that does not emerge in the PLS are, a posteriori, included ad hoc in the mixed model. This is the case of the “proportion of lake surface”. As point out by authors, this variable is considered important for DOC in boreal rivers (see references in the manuscript). Authors reveal that some explanatory variables are not included in the PLS analysis as “consequence of ” large number of zero values” (pag 9015). Is this the situation of “proportion of lake surface”? According to figure 1 and Table S1 most of the catchments have lakes in their drainage network. Therefore this variable should not have a “large number of

zero values". Then the question is: Why the "proportion of lake surface" disappear from PLS output? If this apparently important variable cannot be included into the PLS analysis does it suggests that the PLS is an inappropriate tool? If Lake surface coverage is important and it emerges as significant variable in MM I wonder if the model calibration should to include an additional fourth version: "OutLsc": DOC outlet + lake surface coverage but no map information. This additional model run might help to weight the effective importance of the landscape parameters included in table S1. Is the GIS information overrated? Moreover, the importance of the "proportion of lake surface" also suggests that morphological structure of the river network (and the terrestrial zones surrounding the river network as well, i.e. riparian strips) might have some importance on DOC in headwaters. This comment leads inevitably to wonder why the list of potential explanatory variables do not include any parameter that might incorporate the hydro-geomorphology properties of the study streams/rivers (average main stem longitudinal slopes, river length of confluences, drainage densities).

Finally, I found anomalous the absence of some basic hydro-climatic parameter. At the discussion the authors affirm that sets Cal07 and Cal08 are measured during "different flow situations and seasons". It exists a very rich and abundant literature form the authors that explicitly explore the importance of discharge, winter climate/snowmelt and antecedent hydro-climatic biogeochemical conditions on DOC variability at the Vastrabacken catchment (see Agren et al., 2010 for an example). This headwater stream drains into the larger Nyanget catchment which is included in the present manuscript. In these studies it appears clear the importance of these hydroclimatic parameters on DOC concentration in these boreal headwaters. Therefore, having in mind this knowledge, I strongly suggest that some hidro-climatic parameter (although approximate and coarse) should be included in the analysis otherwise it will be really improbable to obtain satisfactory DOC estimation with GIS information only.

Reference cited: Ågren, A., Haei, M., Köhler, S. J., Bishop, K., and Laudon, H.: Regulation of stream water dissolved organic carbon (DOC) concentrations during snowmelt; the role of discharge, winter climate and memory effects, Biogeosciences, 7, 2901–2913, doi:10.5194/bg-7-2901-2010, 2010. Temnerud, J., Fölster, J., Bu_am, I., Laudon, H., Erlandsson, M., and Bishop, K.: Can the distribution of headwater stream chemistry be predicted from downstream observations? Hydrol. Process., 24, 2269– 2276, doi:10.1002/hyp.7615, 2010.

We appreciate both the queries in this review, as well as the finding that this paper is of interest in the context of water quality monitoring and management. For indeed it is just the challenge of living up to the EU Water Directive of protecting all water, which includes a myriad of headwaters with relatively little systematic monitoring, that is the motivation for our overall objective of finding ways to predict the situation in individual headwaters from more readily available GIS data, supported by water monitoring data from downstream sites. We note several concerns in the reviewer comments that we will reply to:

1. **Why isn't lake surface area included as a factor in mixed models: Our focus is on predicting the headwaters, and few of these headwaters have any lake area. Figure 1 shows that most headwaters lack lakes. The lakes are usually found a little bit further downstream in the investigated catchments. In the updated Table S1 the median lake surface coverage is stated, please note that this is not the number of lakes in headwaters. So including lake area would not be of much use in predicting the water quality of most (median) headwaters. And precisely as the reviewer understood from our text, it is the issue of many zero values in a PLS that led us to leave this variable out. In future work it might be appropriate to find a statistical methodology that can accommodate the few headwaters with some lake area, but given our ambition to predict the situation in all headwaters as well as possible, we have a situation where some lake area does not come through as a significant parameter. If our main goal had been to predict the TOC in larger watercourses where upstream lakes will be more common, the lake area would have most likely been selected by our statistical method (PLS).**
2. **Morphology and catchment structure: A second concern is that information on the morphology and structure of the catchment was not included. There are indeed great possibilities for constructing map information from maps. In the spirit of objectively choosing map information we have worked through the map information directly available from public data bases. The digital network of watercourses in Sweden are in scale 1:100000, which means that most headwaters is not found on this map and is not correct drawn. If other finds the MM approach useful, then the possibility is open to explore other sources of information. But given the focus on presenting a relatively sophisticated**

modelling approach, we have chosen not to add a new dimension of complexity in the construction of map information that may be helpful.

3. **Complexity of the figures and model presentation: This brings us to another point of the referee, and that is the difficulty of following the tables and figures, even though the referee found the text as a whole generally well written with clear objectives. In this revision we have sought to be more pedagogical in integrating the tables and figures into the text. We have rewritten the captions to better explain the figures and link them to the relevant sections in the text.**
4. **The influence of hydroclimatic factors: It is true that weather conditions do influence TOC, with both season and flow rate exerting different combinations of influence on different waters (Winterdahl et al., 2015). More intriguingly, "memory effects" from the antecedent conditions in the preceding year have been identified by Ågren et al. (2010). These memory effects were of secondary importance to the flow conditions and season at the time of sampling. Ågren et al. (2010) required a focused modelling effort to bring forth these memory effects. We did not seek to incorporate hydroclimatic data into our analysis for two reasons. The first is the issue of routing responses through the channel and lake network of weeks to months or years that would need to be included when relating the catchment outlet to headwaters (also explored in Hytteborn et al., 2015). There are some indications that TOC at river outlets correlates (albeit non-significantly at $p<0.05$) with weather related data (median of 30 Julian days before sampling air temperature, precipitation and discharge) for each survey (Table 2). The second was the need for more accurate flow and weather related data (hourly-daily) for each headwater (<2 km^2) for it to be scientifically sound to include weather related data in the modelling. Recent work from the boreal region showing the great variability of specific discharge in the boreal landscape (Lyon et al., 2012). Even without weather data, but with river outlet TOC, we could explain up to 52% of the variation in headwater TOC. We think this is satisfactory considering the small size of the catchments.**
5. **The lack of any clear explanatory variables from the map information. Hytteborn et al. (2015) was able to make significant statistical models of daily TOC using season and flow rate. They then sought to link map information to predict the variation in the sensitivity of specific water course to season and flow rate. Very little (ca 20%) of the large variation in catchment sensitivity to season**

and flow could be explained from map information. This is consistent with our finding that GIS information did not do better at predicting headwater TOC.

References

Ågren, A., Haei, M., Köhler, S. J., Bishop, K., and Laudon, H.: Regulation of stream water dissolved organic carbon (DOC) concentrations during snowmelt; the role of discharge, winter climate and memory effects, Biogeosciences, 7, 2901-2913, doi:10.5194/bg-7-2901-2010, 2010.

Hytteborn, J. K., Temnerud, J., Alexander, R. B., Boyer, E. W., Futter, M. N., Fröberg, M., Dahné, J., and Bishop, K. H.: Patterns and predictability in the intra-annual organic carbon variability across the boreal and hemiboreal landscape, Sci. Total Environ., 520, 260–269, doi:10.1016/j.scitotenv.2015.03.041, 2015.

Lyon, S. W., Nathanson, M., Spans, A., Grabs, T., Laudon, H., Temnerud, J., Bishop, K. H., and Seibert, J.: Specific discharge variability in a boreal landscape, Water Resour. Res., 48, W08506, doi:10.1029/2011WR011073, 2012.

Winterdahl, M., Laudon, H., Lyon, S. W., Pers, C., and Bishop, K.: Sensitivity of stream dissolved organic carbon to temperature and discharge: implications of future climates, J. Geophys. Res. Biogeosci., Submitted, 2015.