Anonymous Referee #1

The study addresses the scientifically relevant question on how to predict water chemistry of small headwater catchments that are not regularly monitored. The study combines two different approaches on how to predict headwater stream total organic carbon (TOC): (1) A map/GIS based approach including mainly land use, soil and geological data in combination with (2) river outlet TOC data of intermediate catchments that include the headwater catchments. The study is within the scope of BG and the combination of the two above mentioned approaches is clearly novel. Nevertheless, the manuscript suffers from several major shortcomings that make it difficult to judge about the quality of the results and the conclusions. Following, I will outline my concerns that I think need to be addressed before I might recommend the manuscript for publication:

Major concerns:

i) Loosing scope of what is promised in the title and the hypothesis. The title clearly states, that the study shows how downstream organic carbon observations can improve map-based predictions of organic carbon in headwater streams. One sentence in the abstract and two paragraphs in the text outline the goal of the study: (1) Testing whether river outlet chemistry can be used as an additional source of information to improve the prediction of the chemistry of upstream headwaters, relative to models based on map information alone (P007L3ff – Abstract); (2) finding out whether the combination of map and river outlet chemistry give better prediction than either one used separately (P9008L29f - Introduction); (3) determining whether models based on geographical data can be improved by adding concentrations measured at the river outlet (P9014L7ff). It is more or less clear, what the focus of the study should be. Unfortunately the focus of the study gets lost in the course of the manuscript. One of my major concerns is that only a small paragraph in the results sections (3.3.4 Evaluation of river outlets) is dedicated to the focus of the study. Moreover, the tables and figures (Table 5 and Figures 3-5) that are linked to this paragraph are too detailed and are not able to direct the reader to the focus of the study / the results of the study. Additionally, the above mentioned table and figures are not well described in the text. I would suggest a table or figure with reduced details and a clear emphasis on the focus of the study, that outlet measurements can improve map based predictions. Unfortunately the result that outlet TOC measurements can improve map predictions is not adequately discussed in the discussion section. The first two paragraphs in the discussion section (P9021L6-18) should be rather shifted to the results section. In the discussion section, the result that the OutMap version gave 5-15% better prediction than Map only, should have been put in a broader context. What is the interpretation of this improvement? Additionally, it should have been discussed in more detail, what are potential explanations that including OutletTOC is leading to an improvement. In contrast to that, the major part of the discussion is about how the unexplained variance could be explained; this is important to be discussed, but not to such an extent that the discussion about the focus of the study is marginalised.

Thank you for these valuable comments. We understand that the referee is concerned that the discussion spends too little time on the main focus of the paper – the ability to improve predictions of headwater TOC by combining GIS information from the headwaters with outlet TOC. Furthermore, the complexity of the figures and tables makes it unclear how they relate to these main goals. We have used these concerns to guide our revision of the text. Now the discussion deals more thoroughly with the overall goals of the paper as announced in the title. This includes exploration of why the outlet information can lead to improvement of headwater predictions. The figures and tables have also been given more pedagogical captions that are better integrated into the text.

With regards to why the outlet can provide information to improve headwater predictions. There is a great deal of variability in TOC between headwaters (km^2 scale) in the same meso-scale catchment (10s of km^2). Many studies have shown that map info can go a long way to describing that TOC variability, but these studies are generally of many sites within one small area (10-100 km^2), or single sites spread across very large areas (100-1000 km between sites). Here we are trying to achieve a model that captures small-scale landscape variability (numerous sites in a meso-scale catchment) across a large scale region (100s of km between the mesoscale catchments). The catchment outlet is a summary of a particular meso-scale catchment. As such the outlets "normalize" the overall average TOC from one particular region to another. While this can seem intuitively correct, finding a way to achieve this is not easy using many common approaches. But this is just what mixed models (MM) are well adapted for, and why they have found use in a variety of application. This paper has adapted the mixed model approach to the issue of predicting headwater chemistry, in this case TOC and found that the outlet does add useful information. It should be born in mind that when using landscape information together with outlet chemistry to predict headwater chemistry, TOC was the chemical parameter for headwaters that was least well modelled (Temnerud et al., 2010). Therefore applying the MM approach here to other parameters is likely to yield greater improvements.

The conclusion section contains an additional major shortcoming: It is concluded that the mixed models approach is improving predictions compared to the predictions that are solely based on outlet TOC (which was done by the author in a previous study). This is contrary to what is written in the title and the abstract (OutletTOC is improving map based predictions).

We appreciate that the reviewer noted that our conclusion was stated somewhat ambiguously, making it possible to interpret it the wrong way. As the reviewer noted from the title and abstract, what we intended to say in the conclusion was that the PLS/mixed model approach on individual headwaters using both OutletTOC and map information, gave better performance than the attempt by Temnerud et al. (2010) which used only OutletTOC on headwater median TOC and TOC IQR. This should not contradict what we wrote in the title and the abstract. However, the original sentence was indeed open to misinterpretation.

Therefore we have clarified the conclusion with the new sentences: "The PLS/mixed models approach, using both river outlets TOC and map information of the headwaters, could explain up to 52% of the variance in TOC among individual headwater streams. This is much better performance than the attempt by Temnerud et al. (2010) in which only one of seven different leave-one-out attempts gave a significant model, using river outlet TOC but no map information, for headwater median TOC and none gave significant models for headwater TOC IQR."

Moderate concerns:

ii) The experimental setup and the environmental conditions during the sampling are not well described.

We have revised the methods section to better describe the experimental setup by including information on sampling strategies and we have also been more explicit about the experimental conditions during sampling. The new text is now: "The synoptic surveys used in this study were designed to provide a snapshot of the water chemistry in

stream networks (Table 1 and Fig. 1), the aim was to sample most stream junctions, lakes inlets and outlets and the river outlets. In total there were data from 17 synoptic surveys conducted between 2000 and 2008 in nine catchments distributed across Sweden (Fig. 1). This data set amounted to 938 stream samples of which 420 were from headwaters. The catchments span a north-south gradient of 800 km through the north-temperate and boreal zones. All sampling during a given survey was carried out during a one to three day period, except for R. Krycklan in winter 2005 (two weeks due to cold weather and difficulties finding the streams in deep snow, but discharge was stable winter base flow). Which year and month each river were sampled can be found in Table 2. 30 days median of air temperature, precipitation and flow conditions before sampling are stated in Table 2, with daily values in Fig. S1."

In the methods section, it is not well described, how many headwater catchments were tested.

We have revised the Method section to be clearer about the numbers of tested headwaters. This can be found in Chapter 2.5.1, the two last paragraphs, and in the second paragraph of Chapter 2.5.2, In the Results section we clarified the cross-reference to Table 5 where the numbers of tested headwaters are also stated.

How many headwater catchments are within each of the 9 larger catchments?

The number of sampled headwaters differs slightly between the different synoptic surveys for each river, and the numbers (n_{HW}) are presented in Table 2. In the original manuscript we did not cross-reference to this table in Chapter 2.2 Study sites. We have now added that cross reference to the text.

How do the headwater catchments differ from each other in land use, soil, topography and geology?

Thank you for a great idea. In Table S1 median values of map information for each river are stated. We have now added box-plots of map information as well as TOC for each synoptic survey, see Fig. S2-S8.

Moreover, the naming of the larger catchments is not consistent.

We have checked the manuscript thoroughly and we could not detect where the naming of the rivers (the largest catchments) is not consistent. But we can see that there was a possibility to be clearer about the naming conventions. In Chapter 2.1 the full names of the nine rivers are first mentioned in the text as well as cross-referenced to Table 1. In the text the abbreviations for each river were not added, but were stated in Table 1. We have now added the abbreviations to the text as well. In Table 2 the same abbreviations are used with a suffix to designate the sampling year. What could be experienced as inconsistent is that Cluster A has the abbreviation as River Anråse å (A). We have now renamed the clusters from A, B, C to c1, c2, c3 to remove the risk for confusing Cluster A with River A.

Are intermediate sites/catchments congruent with the nine investigated catchments in figure 1?

The last paragraph in the Method section read: "As an additional step in the evaluation of the models the most successful MM from the nine MM calibrations was tested on the sites between the headwaters and the river outlets, the intermediate sites (n = 501)." So yes, the intermediate sites are congruent with the nine rivers.

Additionally, information of the meteorological conditions shortly before the sampling and discharge information during the sampling would be helpful in the interpretation of the results. I am aware, that it is not possible to get detailed information about discharge anymore (although it could be of great importance to have discharge information, as TOC variations are often closely linked to discharge variations). Nevertheless, it would be helpful to have at least the information, whether there was rainfall before the sampling or whether we have high flow or low flow conditions.

Records of river flow are maintained by the Swedish Meteorological and Hydrological Institute (SMHI). So it is in fact possible to recover this antecedent flow information. At this scale much of these flow records are modelled, but the modelling has been shown to be reliable. In Fig. S1 weather and flow conditions 30 days before, and including each sampling occasion are shown for each river outlet and synoptic survey. Information about these data can be found in Chapter 2.2 Study sites. For each headwater upstream from one of those outlets, we expect that the antecedent patterns are reasonably similar as the weather is similar at that scale, with the possibility that individual convective storm events may vary at the scale of a few km. The lacks of detail flow data for the individual headwaters are discussed in the Discussion section.

iii) Figures are not self-explanatory, not clearly laid out and not well explained. Figures 2 to 5 are not directly able to show the message that they should transport. This is a combination of several points: They are not clearly laid out and the labels are too small. Moreover, figures 3 to 5 (including their captions) are not self-explanatory or just understandable after studying them for a long time (at least when it comes to the point of understanding the message they want to transport regarding the focus of the study).

We have now increased the size of the labels. While the layout is harder to improve, there is indeed much information in the figures, especially figures 3-5, so we have expanded the captions.

For example the caption for Figure 3 now reads: "Scatterplots of measured headwaters with total organic carbon (TOC in mg L^{-1}) on the x-axis, and the three different versions of the mixed models $Cal_{MM}00_{c1}$ on the y-axis: Out version on the left panel, OutMap version on the right panel and Map in-between. Data for year 2000 ($Cal_{MM}00_{A}$) on the top row, followed by Test data; second row 2002 & 2005 data, third row is 2007 data and the last row is 2008 data. R. Anråse å indicated by circles, R. Danshytteån by diamonds, R. Getryggsån by rectangles, R. Krycklan by triangles (winter 2005 by upside-down triangles), R. Lugnån by squares, R. Mangslidsälven by multiplication sign, R. Ottervattsbäcken by up-side-down triangles, R. Vänjaurbäcken by right tilted triangles and R. Viggan by plus sign. The black line is the 1:1-line."

Minor concerns:

P9007L25f: Structure of the sentence: "The headwaters also combine to provide..."

The new sentence reads: "The headwaters also provide much of the water and solutes to downstream locations (Person et al., 1936; Leopold et al., 1964)."

P9008L10-16: These two sentences are contradictory to a certain extent; much of the small-scale heterogeneity is averaged out at larger spatial scales vs. monitoring of downstream sites might provide information about headwaters upstream. This needs further explanation and would be also an interesting question to be discussed in the discussion section.

We appreciate that there was an apparent contradiction in saying that downstream sites could provide useful information about individual headwater even though the variation is averaged out at the downstream sties. We have restated this to point out that while much of the small-scale heterogeneity is averaged out at larger spatial scales, the generally better monitoring of downstream sites still might provide information about headwaters. The measurements at the river outlet were necessary to reproduce more correct average headwater TOC levels. Excluding the OutletTOC measurements leads to the assumption that average TOC levels in the headwaters were similar in different catchment stream networks if the map information is similar, which is not always true (cf sampling 2007 and 2008) when looking at networks that are spread out across a large landscape. It might sound contradictory that the large small-scale heterogeneity was correlated with river outlets, but the mixture of all headwaters ends up in the river outlets, and the transition time from when the water enter headwaters until it reaches the river outlet is rather small (hours to a day) in many of these systems, allowing conservative mixing of headwaters to be reflected in the river outlet chemistry."

While we wrote in the Discussion about how well this approach worked out, we overlooked highlighting the value of coupling it to the scale issue of small-scale heterogeneity. We do this now by changing the second paragraph of the Discussion section to: "In 25 of 27 tests, including OutletTOC resulted in lower errors in predictions of the TOC for individual headwaters and intermediate sites, compared to using map information alone.

P9009L27: "kNN": needs explanation when firstly introduced.

We omit the term kNN from the Introduction.

P9009L27: Wasn't also TOC_Outlet used to derive TOC median and IQR values?

That is correct, we missed the most obvious candidate! We have now added that to the sentence.

P9011L4: I would not call spruce the dominant tree species. In several catchments, pine is the tree species with the largest volume and in most of the other catchments, spruce is not dominating but has just moderately higher volumes then pine.

That is correct. We have now rephrased the sentence to: "All catchments consisted mainly of forest (> 80 %) with a dominance of coniferous forest made up of Norway spruce (Picea abies) and Scots pine (Pinus silvestris) (Table S1 in the Supplement)."

P9011L2: “All catchments”: all headwater catchments? Is this paragraph only about the headwaters?

All catchments mean both headwaters, intermediate and river outlets, in other words all 938 sampled catchments. We have now added, in brackets: “All catchments sampled (including headwaters, intermediate and river outlets)…”

P9011L5: “Mires and small lakes made up most of the remaining parts”: Clear-felled areas are larger than the sum of mires and lakes in many catchments. I would not consider clear-felled areas as forests, especially when it comes to a study about TOC.

Clear-felled areas could have another impact on TOC than the land-use class forest. In this sentence we consider clear-felled areas as part of the forest. Should clear-felled areas have been treated as a separate land use class? In the future this might be worth treating this as a separate class, but in this study clear-felled areas have been treated as forest, since management practices require regeneration of the forest, meaning that what shows up on a particular edition of a map as clear-felled is at some stage of regeneration when sampled.

P9014L25: Abbreviations like Cal_PLS_00_A are not helpful to read the manuscript in a fluent way. Maybe you can come up with a better solution, although it might be difficult to find a better solution.

As the reviewer points out, these are long abbreviations. We have considered this, and have not found a better solution, even though we have considered different approaches. We consider the one we have in the manuscript as the best way to include all necessary information, even if it is a bit long.

P9014L27: Were the mesoscale catchments sampled or the headwater catchments?

All 938 catchments were sampled, both headwaters, intermediate and river outlets (the mesoscale catchments). OutletTOC is based on the measurements.

P9015L11ff: The assumption, that headwaters within the same catchment are more similar to each other than to headwater from other catchment, needs further explanation. I guess this depends on the similarity are heterogeneity of headwaters within a catchment and on the driving factors that control TOC behaviour.

Good point. We have tried to clarify the basis for this assumption. The sentence was changed to: “When modelling individual headwaters we want to reproduce individual values for the different headwaters in all catchments. As an effort to improve these simulations, we make an assumption that headwaters within the same mesoscale catchment are more similar to each other than to headwaters from other mesoscale catchments due to subtle combinations of physiographic, weather and other factors which combine to influence the TOC levels in ways which are not readily apparent from the available map information, but might be reflected in differences between the average TOC levels in the different meso-scale catchments. This assumption leads to a new data structure, where we need to assume that.”

P9018L4f: Sentence structure needs to be revised.

The sentence was changed to: "The first principal component (PC) of the PLS-model of median headwater TOC was significant for both PLS calibration sets year 2007 and 2008, but not the second PC."

P9018L19: "that is all other data than 2007": Sentence structure needs to be revised

The sentence was changed to: "Verification based on the evaluation data of CalPLS07c2, that is all contrasting data to 2007, had lower PRESS for TOC IQR than similar data for CalPLS08c2 (Table 3)."

P9020L24: "Out of 27 different combinations": Isn't there something missing? Perhaps: In 25 out of 27 different combinations.

The sentence was changed to: "In 25 out of 27 different combinations of MM (three different calibration data sets, three versions of each calibration (Out, Map, OutMap) and three different test data sets), the OutMap version (OutletTOC and map information combined) gave the best performance with the lowest PRESS, while two Map versions (map information only, no OutletTOC included) gave the lowest PRESS (Table 5 and Figs. 3–5)."

Table S1: What is the difference between no value and 0.00? Is 0.00 just a rounding effect or does it also mean no value?

0.00 is just a rounding effect.

References

Temnerud, J., Fölster, J., Buffam, 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.