Dear Editor Seibert

We would like to thank you reviewers for constructive comments and suggestions that have helped to improve the manuscript. We have answered for all comments and suggestions of reviewers. We have clarified the objectives of this study, improve the structure of introduction as suggested by the reviewers and highlighted the novelty of this study. We have addressed the targets in discussion and removed some parts of the conclusions. We have added the water quality monitoring data into the manuscript to justify the vulnerability aspects of this manuscript. We have also improved the figures and tables as suggested. We hoping to raise the clarity of our study.

Sincerely
Anne Rautio and co-authors.

We would like to thank the reviewers for making the comments and suggestions on the paper and for recognizing the potential of a manuscript. We appreciate their insightful and comprehensive comments that helped us to make changes and clarifications in the way how our data presented and discussed. The suggestions and comments helped to improve the manuscript considerably. We followed large majority of the reviewer's comments and suggestions when preparing the revised version of the manuscript. However, we have maintain our views and opinions in some points. General and detailed comments addressing all questions/comments/suggestions of the reviewer #1 are listed below.

Anonymous Referee #1
General comments:

Introduction I am missing a short review of potential thermal detection methods in the introduction that can be used on catchment scale (10th km-scale) and on study site scale (km-scale) (e.g. fiber-optic cables (DTS)) to identify the groundwater discharge zones to streams.

We have added a short review of thermal detection methods in the introduction.

Novelty of the work is not justified satisfactory. Make it more clear what is your contribution to the AIR survey literature that are not already published by e.g. Torgersen et al (2001); Conant and Mochnacz (2009) ?. I suppose it is the integrative approach of AIR survey and hydrogeochemical data that are new ?

Yes, it is the integrative approach to apply AIR survey and hydrogeochemical data on a catchment scale that is novel to achieve a reliable and comprehensive understanding of this exchange, and its commonness and potential deteriorate effects for water quality. We have highlighted the novelty of this work in the of the introduction chapter.

Specific objective (p. 2438, L25-26) must be formulated more precisely. E.g. is it an objective to examine DSi as a potential tracer to estimate GW to RW (p. 2444, L 10-11) ?

We have specified our additional objective. The additional objective was to assess the applicability of the used thermal method in boreal catchment by verifying the identified GW–SW interaction locations using site-specific thermal and hydrogeochemical methods.

Specific comments

p. 2439, L 8. Add: 20 to 120 meters above sea level (m a.s.l.)

Has been added

p. 2439, L. 21. Cover-moraine sheets and end-moraine ridges need to be explained in Figure 1a. These two geomorphological elements are not shown on the legend of Fig 1a ?

We have added references in the text followingly: bedrock terrain and glacigenic deposits forming cover-moraine sheets (glacigenic till in Fig. 1a) and end-moraine ridges (glaciofluvial sand and gravel in Fig. 1a) (Tikkanen, 1989) (Fig.1a).

p. 2440, L. 2-5. Delineation of aquifers is not given in Fig 2b and not in Fig 1.

Yes, it is true that the delineation of 29 aquifers are not given in Fig. 2b or Fig 1. However, we have referring the municipal water companies which are placed in the Fig. 1a.

p. 2440, L. 21-29. What is Tuusula artificial GW plant ? Is it the same as a water intake plant ? Artificial GW need to be defined.

Yes, Tuusula artificial GW plant is water intake plant using both the artificially recharged GW (70 %) and natural GW (30 %) from the Jäniksenlinna aquifer. We have added more specific definition for the artificial GW: Water from Lake Päijänne (9370 $m^3 d^{-1}$) is conducted to the infiltration site through a water supply tunnel and is artificially recharged into the aquifer by pond infiltration through the permeable esker deposits. This artificial GW is accounting for 70 % of water intake from the Jäniksenlinna aquifer (Kortelainen and Karhu, 2006).

P. 2440, L. 25. Jäniksenlinna aquifer is not shown in fig 3b.

The classified aquifer in Fig. 3b is Jäniksenlinna aquifer. We have added followingly: The recharge of natural GW in the shallow and unconfined Jäniksenlinna aquifer is approximately 4000 $m^3 d^{-1}$ (Hatva, 1989) (classified aquifer in Fig. 3b).

p. 2441, L. 10-14. What was the flying height? You need to repeat essential details on specifications defined for the AIR work. I dont think it is enough just to refer to Korkka-Niemi et al (2012).

We have added details (sensor and camera type, flying high, ground speed) of AIR survey of 2010 into the text.

p. 2442, l. 8. Explain why is most of the AIR surveys conducted in upstream direction ?

We have added the explanation why most of the AIR surveys in this study were conducted in upstream direction. The upstream direction was used due to the facility to follow main stream in upstream direction, the exceptions were mainly due to the logistical and economic reasons to save flight time.

P. 2442, L. 11. Add: meter above ground surface (m a.g.)

We have added meter above ground surface (m a.g.s.) to previous chapter dealing the AIR survey in 2010.

P. 2442, end of L. 18. Text/words are missing?

We did not notice the missing text/words.

P. 2442, L. 18-19. Need to explain what Tk and Tr is used for.

We have added explanation followingly. The T_k were compared to T_r to define the average absolute temperature difference between the reference measurements and remotely measured with TIR sensor.

p. 2443, L. 10-11. Two longitudinal profiles were collected. In what figure are they shown ? Unclear how they are measured ?

We have added in which figures two longitudinal profiles of RW temperature near the sediment–water interface are shown. The measurement devices are mentioned in following sentence. All RW temperature and EC measurements were collected with a YSI 600 XLM-V2-M multiparameter probe and the sediment temperature measurements with a stainless steel sediment temperature probe (Therma Plus, Electronic Temperature instruments Ltd, Worthing, West Sussex, UK, accuracy ±0.10 °C).

p. 2448 L.1-2. The AIR method cannot at some places detect the temperature anomalies close to the river bed due thermal stratification. How deep in the water column can the AIR pictures normally detect an anomali of discharging groundwater through the river bed if no thermal stratification exist ?

The thermal infrared (TIR) camera measure only the skin layer < 0.1 mm, and only if the GW contribution reach the surface of the water body it can be detected with TIR. We have added following sentence for clarification. It is a well-known limitation of the thermal infrared (TIR) technique to detect the surficial temperatures (“skin” layer < 0.1 mm), and only substantial subsurface GW contributions to SW bodies that reach the surface can therefore be detected (Torgersen et al. 2001).

p. 2448, L8-9. Sentence unclear. Change Fig 6 to Fig. 6b.

Has been changed and the sentence has been clarified. The bottom RW temperatures were mainly relatively equal and constant during the continuous water temperature monitoring period (Fig. 6b).

p. 2448, L13. Surges of GW. What is that?

We have change “the surges of GW” to “GW discharge”. The surges of GW described the pulses of discharging GW when the RW levels declined.

p. 2448, L16-18. What figure is the sentence referring to ?

The sentence is referring to Fig. 6b.

p. 2449 L7-14. Not sure how you have summarized this important EC information in figure 7. Explain more.

We have added references in the figures that clarified the sentences in text. In Fig.7, the lower EC values (indicating artificially recharged GW) observed concurrently with cold RW temperatures that can see in Fig. 7, e.g. in cross-sections P–PP’ to S–SS’, U–UU’, V–VV’ when the temperature dropped simultaneously with EC values.

p. 2450. L 1 The suggested ranked order is not logical based on the development in O18 and Deuterium (see suggestion to table 3).

We have changed the order suggested more logical both in text and Table 3.

p. 2450 L15-16. Is the AIR data collected hand-hold in 2010 and mounted in 2012 ? You need to give this information in section 3.1

No, the data was collected mounted in 2010 and hand-hold in 2011. The sentence has been changed and the collection method has added in section 3.1.

p. 2451, L28 EC in river varies considerable in time and space. Do you have any explanation why ?

We have added flowingly. However, in the river systems EC values range widely both temporally and spatially due to variable load from sewage treatment plants and urban areas, including residues of purified waste water and deicing chemicals (Vahtera et al. 2014).

<u>Figures</u>

In general I am not really happy with the layout of the figures. Be sure that the font of the text and size of symbols are readable. Especially Fig 5, 6 and 7 are difficult to read. The figures are too small to cover all the compiled informations.

We have increased the font sizes in figures and hoping the figures are more readable.

Fig1a. The legend is a mixture of lithological units and geomorphological units Fig 2b. Show the extension of the classified aquifers with a more pronounced signature. The dashed line used to outline the classified aquifers is difficult to see. I will suggest that you use oblique hatching as you use in Korkka-Niemi et al (2012, Figure 1). Explain “classified aquifer” – not explained in text. In the legend: IR flights 2011 and IR flights 2012 should be changed to AIR flights 2011 and AIR flights 2012

The classified aquifers have been explained in figure caption of 2b. However, we have left the classified aquifer symbol in previous for as it is the official symbol used by Finnish environmental agencies. We have changed IR flights 2010 and IR flights 2011 to IR flights 2010 and AIR flights 2011.

Fig 3. Indicate the area with the classified aquifers with another signature. It is impossible to see the position of the municipal intake plant and wells as they are grouped too close in Fig 3a and 3b.

We have left the symbol in previous form. Unfortunately, production well are close to each other and the municipal water intake plant well symbol is covered by them. We have added arrow to pointed out the location of intake plant.

Fig 6c. Scale on Y-axis missing on all cross sections. The geographical orientation of each cross section is not specified in the inserted map above Fig 6d. B in B-BB is missing. JJJ is missing. Fig 6d. what is the orientation of F-FF. Is the vertical T profiles measured in the middle of the river ? Dashed lines in the right side of the figure – what do they show ? How comes that it can be spring water. Where is the bottom of the river ?

We have made improvements in Fig. 6. The EC and temperature are two Y-axis which can be confusing as the graphs are turned. We have added the missing cross-sections in the inserted map above Fig. 6d. The F-FF’ cross-sections has similar orientation as the other cross-sections in Fig. 6c (from W-shore to E-shore). Yes, the vertical temperature in Fig. 6d has been measured in the middle of the river and we had added that also in caption text of Fig. 6d. Dashed lines in Fig. 6d shows the possible subsurface preferential GW flow paths into the lower part of the river channel. and part of cold and dense water can originates from springs sinking down by the river bank. The explanations of dashed lines and grey line in Fig. 6d has been added in the caption. The bottom of the river is the solid black line.

Fig 7c. Scale on the Y-axis

In Fig. 7 the principle of the graphs is same as in Fig. 6. We have clarified the figure by increasing the font side.

<u>Tables</u>

Table1. River bed width: is it the width from side to side or is it the periphery of the river along the river bed ? River bed depth: what does the interval represent (max and min depth ?)

We have been clarified the table by changing the river bed width to river width and river bed depth to river depth. Yes, the interval represent the minimum and maximum depth.

Table 3. In accordance to the delta O18 and D values. Reverse order of: R Lepsämänjoki and R Vantaa Reverse order of: R Palojoki and R Tuusulanjoki

Suggested changes has been made.

Table 4 Reverse order of: R Vantaa and R Palojoki

The order has been changed.

References I would suggest that you in a parenthesis indicates the English translation of the purely Finnish-language reports

We have added the English translations of reports in parenthesis.

We would like to thank the reviewers for making the comments and suggestions on the paper and for recognizing the potential of a manuscript. We appreciate their insightful and comprehensive comments that helped us to make changes and clarifications in the way how our data presented and discussed. The suggestions and comments helped to improve the manuscript considerably. We followed large majority of the reviewer's comments and suggestions when preparing the revised version of the manuscript. However, we have maintain our views and opinions in some points. General and detailed comments addressing all questions/comments/suggestions of the reviewer #2 are listed below.

Anonymous Referee #2
General comments:

Intro is not well in line with the actual study. Authors address flooding and water quality, which are not central to the methods or outcomes of the study. I would like to see more text related to methods in the intro.

We have reorganized the introduction and added chapter about thermal methods. We have also added water quality monitoring data (nitrate, dissolved organic carbon and turbidity) that will be more in perspective with vulnerability aspect. We have removed some sentences about the flooding.

Authors stress the possible contamination of GW with RW in the discussion section (P2456). Their data shows clearly GW discharge to streams, but they do not provide any evidence of flow in the opposite direction. Fact that there is interaction, doesn't tell us that there is a clear risk of contamination due to flow direction reversal at a given site. Contamination by bank infiltration is of course a well-known risk in general, but in my opinion the authors do not have convincing enough data to highlight this issue so much in the manuscript. Perhaps some simple estimates of the hydraulic heads in between the river (low flow and flood) and the location of water abstraction could be used to justify risk of contamination. In this regard on P2438 L12-14 authors address the first main aim well with this study, but in my opinion fail to prove much insight to the latter.

We have added the water quality monitoring data from high flow season. This data showed the RW infiltration into the production wells on high flow season and potential vulnerability of production wells. The abstraction wells are locating in proximity of main stream channel (Fig. 3) which increase the risk of contamination.

This is a nice case study with multiple methods used, but unfortunately I do not much novelty in any of them individually. I see, that authors could focus their discussion on what benefits do use of multiple methods bring to understanding of GW-RW interactions, instead of trying to force a management angle to the paper. Perhaps the authors could further expand their toolbox of methods by including stream flow measurements, referred to in the discussion, to this manuscript? And discuss more thoroughly the weaknesses and strengths of the methods based on their data.

We have added following sentence to address what is new in this study. Many previous studies have used the TIR to identify, classify thermal anomalies and modelling the stream water temperatures but not with the hydrogeochemical variables to explore the connection between anomalous stream water temperatures and GW–SW interaction indicative geochemical variables to assess the potential vulnerability of intake plants in proximity of main stream channels. We have maintained the management angle in paper due to the water quality monitoring data showed the contamination risk of production wells and proved that GW-SW interaction zones (discharge zones/low flow season) can be potential risks for water intake plants during the high flow season.

Specific comments:

Flow of intro, the first chapter is not related to the second one. You could use text currently on P2438 L4-L9 onwards to create link from GW-SW interaction to study catchment. P2438 L8 – It would be interesting if you could set you results of GW-RW interaction in this context. Do your field studies give any insight on the effect of GW on water quality and quantity?

We have reorganized the introduction as suggested. Our results (AIR, stable isotopes and DSi) indicate that in smaller tributaries, the water flowing in the streams is predominantly GW originating from the headwater aquifers in the low-flow period GW (e.g. Palojoki). Also Brander (2013) observed with river flow measurements both GW discharge into the river and RW recharging the aquifer.

P2444 L16 number of samples per site is not very clear, perhaps create a table for this?

We have added the numbers of samples per site in the text.

P2444 L23 – How were the spring identified, and were they considered to represent GW quality?

Springs (sampled springs) were identified based on their anomalously low temperature with field measurements. The springs were considered to represent the GW quality.

3.3 Statistics – You state that ": : :in order to assess the GW component. Please clarify what are you assessing exactly? Does the population taken within GW area differ from outside samples? Do you differentiate up- or downstream form GW area, for the "no

GW effect group"? I would expect the GW signal to be seen also downstream

We have added followingly: To test if the GW input could also be seen in RW quality inside the classified aquifers, the non-parametric Mann-Whitney U-test for two unrelated or independent populations (Rock, 1988; Ranta et al., 1991) were performed using IBM SPSS Statistics 22 on RW samples ($n = 36$) in order to assess the GW component in RW. Yes, there was a statistically significant difference ($p < 0.05$) between the "GW effect" sites and "no GW effect" sites e.g. inside or outside the mapped GW area in the measured DSi and d-excess values. No, we did not differentiate the up- or downstream directions.

Section 4.1 until P2446 L12 in description of methods, in my opinion

In our opinion the number of thermal images are fitting into the results and classification of anomalies is results and partly interpretation of results.

P2447 L11-16 – This paragraph is difficult to understand. I read figures in a way that point 0 is the river outlet, so the river flows from 20km (left) to 0 km (right). This doesn't agree with you wording telling that artificial GW cools the river water. Instead we see a sudden warming. Please clarify what you mean by distance upstream and to which direction does the river flow?

This was unclearly marked in figures and we have corrected that. We have added distance from upstream on the X-axis and arrow to clarify the river flow direction.

P2448-2449 – interpretation of figs. 6 and 7 is quite confusing. In many occasions I'm not sure to which of the subplots the authors are referring to. Perhaps consider dividing the data into more plots, instead of trying to fit in too many subplots. Or at least please refer to the letter of subplot more actively.

We have made improvements into the figures (increased font size) and referred more preciously to the figures.

P2451 L22-26 - Can we see this stratification in your plots? And furthermore, is the location lacking AIR observed GW discharge? It would be interesting, if you can explicitly demonstrate with your dataset that stratification is hiding GW discharge.

Yes, we can see clearly the stratification on figures 6c and 7c. The river surface water temperature is quite constant and close to 23-24 Celsius degrees in all cross-sections, e.g. the bigger the difference between the surface and bottom the stronger the vertical stratification in water column. In Hyvinkäänkylä study site, the river surface water temperature were warm and the AIR missed this subsurface GW discharge. However, there were large springs in side of the river which were clearly seen in AIR images. The cold water in lower parts of River Vantaa at Hyvinkäänkylä study site would missed without the field measurements.

P2453 L8-L10 – To me this is not very obvious in the data. I would interpret stratification also downstream the pool in cross-sections OO' and PP' section 5.3 – much of this regarding chemical tracers and M-W U test should be moved to results section, as they are central evidence of GW-SW interaction in the streams.

The stratification decreased as the water depth decreased. We have changed the word disappear to decrease. We have kept the M-W U test in discussion as it include the interpretation.

P2458 – L4 – “This research provided new insights for water management.” I don’t agree with this statement, what are the new insights here. Please also note the supplement to this comment: http://www.hydrol-earth-syst-sci-discuss.net/12/C788/2015/hessd-12-C788-2015-supplement.pdf

We have added followingly to clarify the new insights. This research provided new insights for water management, and the results could be used in evaluating the possible effects of GW and RW exchange on water quality in the identified exchange zones. Based on the results of this research potential GW quality deterioration during peak-flow periods has been acknowledged at several waterworks. Infiltration of RW through permeable strata was observed to affect GW quality in some water intake wells installed into sand and gravel deposits in the vicinity of river bed. In order to avoid disruption in the drinking water supply new locations of groundwater intake wells and intensified monitoring of hydraulic heads as well as quality of GW between river bed and wells have been considered at these water intake areas.