

REVIEWER 1

General comments

The authors investigated snow Hg dynamics in the interior arctic tundra at Toolik, Alaska. They compared their results to a temperate snowpack in the Rocky Mountain, Colorado and conclude that photochemical Hg(0) losses from the snowpack in models should be adjusted and treated differently in arctic and temperate snowpacks. I think that this research provides some very interesting results and the manuscript is fairly well written. However, to my opinion, it lacks a thorough comparison to other studies performed in polar regions and a discussion on instrumental limitations. I think that the manuscript will be appropriate for publication in The Cryosphere after the authors address the major comments discussed below.

Major comments

The conclusion of the manuscript (see above) derives from two main observations. Firstly, low concentrations of Hgtot and Hgdiss were measured in surface snow. Secondly, the photochemical formation of Hg(0)gas in the snowpack was largely absent.

1. I am quite surprised by the very low Hgtot concentrations in surface snow samples reported in this study, especially during depletion events. This raises concern about the potential influence of drifting snow. While the authors admit that the Toolik snowpack is subject to significant drifts and changes in snowpack height, there is no discussion on potential consequences and uncertainties.

We added more discussions about that. See below for more details. However, since we did not observe Hg concentration trend in snowpacks with depths or differences between lake and tundra snowpacks, we largely focus our discussion on the importance of spatial redistribution on snowpack Hg pool sizes.

2. While I agree that the photochemical formation of Hg(0)gas in the snowpack is most likely low compared to a temperate snowpack, I would like to see a comparison to other studies performed in polar regions. To me, the Toolik snowpack is more similar to a polar than a temperate snowpack (e.g., given permanent darkness/sunlight periods).

We added additional comparisons and discussions with respect to other polar snowpack (see below).

However, we maintain our discussion and comparison to temperate snowpack as well to highlight the dramatic difference between temperate and arctic snowpacks.

3. Additionally, there is no discussion on instrumental/analytical uncertainties. I would for instance very much like to know how the results collected with the two Tekran instruments compare. I expect a 10-20 % difference and wonder if the conclusions remain valid taking that analytical uncertainty into account. I also wonder whether the upper inlet in the snowpack was too far from the surface in spring, explaining why the authors did not observe any photochemical production of Hg(0)gas in the upper layers.

Please note that the Tekran used for the snowpack tower had one inlet line in the atmosphere which allowed us to adjust the two Tekrans used in this study based on atmospheric measurements (see detailed comment below). So, the differences between atmospheric and snowpack Hg concentrations cannot be explained by instrument variability.

The uppermost inlet in snow was always ≤ 10 cm from the atmosphere (note that was the spacing between snow inlets). Also, the absolute inlet depth (from the top of the snowpack) is variable based on the snowpack height. We added a discussion point that a possible reason for not detecting Hg^0_{gas} production may be that if this process was limited to a few cm in the top snow, our measurements would have missed this. We also determined using a webcam and snowpack sticks that the upper snowpack inlet was on average 7 cm from the atmosphere during the Apr. 2016's AMDE, 5 cm in Mar. 2016, and 6 cm in Mar. 2015 (i.e., during for the three biggest AMDEs). Also, we added that when using the exactly same measurement system, detecting Hg^0_{gas} enhancements in temperate snowpack (Fain et al., 2013) apparently were not problematic using this system.

The following line by line comments should be useful to fully comprehend and address the major comments.

Line by line comments

Page 2, lines 21-22: "In the Arctic and Antarctic, Hg cycling also is affected by atmospheric Hg depletion events (AMDEs) which are observed primarily in the springtime along coastal locations (Dommergue et al., 2010; Schroeder et al., 1998; Steffen et al., 2008)." The authors should also cite Angot et al. (2016a), latest review paper about

mercury in polar regions which includes a discussion on the occurrence and frequency of AMDEs in recent years.

Thanks for this reference. We added it.

Page 2, line 28: Typo; McMurod should be McMurdo.

Done, thanks!

Page 4, lines 6-8: "Atmospheric air sampling was performed using the top snow tower air inlet (. . .), as well as on a nearby micrometeorological tower at a height of 3.6 m above ground." How do the results compare? Did you collect data with two different Tekran instruments? If so, it would be an easy way to answer the following comment.

We collected data from two different Tekran instruments. The first one was connected to the micrometeorological tower (data not used in this paper, but in Obrist et al., 2017) and the soil wells (used in this paper). The second one was connected to the snow tower (snowpack measurement by 0, 10, 20, 30, and 40-cm inlets, as well as atmospheric levels measured at the top inlet of the snowpack (110-cm above ground, hence always located in the atmosphere). For this study, we hence always compared the snow data to atmospheric data using the same Tekran analyzer. For further comparisons, we were also able to adjust atmospheric measurements of the Tekran connected to the snow tower with atmospheric gradients: (we found on average 7% of differences, which could be adjusted to eliminate this difference). We added some information in the text to clarify this on page 4 lines 14-19.

Page 4, lines 10-11: "Gaseous Hg(0) concentrations were measured using two Tekran 2537B analyzers, one for interstitial snow air measurements and the other shared for soil gas and atmospheric measurements." Did you check how the results from the two instruments compare? What is the associated uncertainty? According to several

studies, the analytical uncertainty is about 10-20 % (e.g., Slemr et al., 2015) and this should be taken into account when comparing data acquired with different instruments. See major comments.

See responses above.

Page 4, line 12: "Air sampling was alternated between different snowpack heights every 5 min so that a full sequence of air extraction from the snowpack (six inlet heights) was achieved every 30 minutes." If I understood correctly, you just collected one data point per inlet height. Could there be any significant sampling-induced snowpack ventilation influencing the results?

We apologize that there was a mistake in this description. We clarified that snow tower measurements were based on switching heights every 10-min corresponding to two individuals of 5-min measurements per inlet (to avoid Tekran trap bias), resulting in 1-hour measurements per full gradient. The reviewer raises an important point, that is if and to what degree the active sampling of air from the snowpack induces artificial ventilation and dilution of the natural gas profile. The dual inlet sampling at a given depth, the 90° rotation of inlets at adjacent depths, and the only intermittent sampling (10 min every 1 h) were all selected to minimize this effect. The higher resolution trace gas (CO₂, O₃) were used to assess this artifact. Concentration changes within the 10-min interval were relatively minor (<10–20%) which suggests that mostly air from within the vicinity of the inlets was sampled. We have thoroughly evaluated this in other previous descriptions of this snowpack air sampling approach and refer the reviewer to those publications for a more in depth assessment (Seok et al., 2009). Moreover, our continuous and automatic sampling (i.e., resulting in 24 full snowpack profiles each day when data coverage is complete), limits any systematic bias that would be introduced by ventilation.

Page 4, line 27: "The top 3 cm of the snowpack was collected in triplicate."

1. What was the approximate distance between the replicates? According to lines 3-5 on page 6, the Toolik snowpack is subject to significant drifts and changes in snowpack height. Did you take that into account when interpreting the transect data? See major comments.

The triplicate samples were collected within a distance of 5 m (we added this in the text). We do not focus our discussion on surface snow patterns (since there was no significant difference to snowpack-averaged data with the exception of active AMDEs periods), so we did not further evaluate/discuss how drifts in surface snow may affect results.

2. According to lines 30-31 on page 5, average data are shown as mean $\pm$ standard deviation. Were the replicate surface snow samples pooled together for analysis, similarly to snowpit samples? If so, what does the standard deviation reported in Table S2 stand for?

Surface snow samples were analyzed separately. Surface snow SD is based on the triplicates (n=3) while snowpack SD is based on the two replicate pits (n=2). The two snowpack samples pooled together were collected from the same pit, but from different walls. We added this information in the Table S2 caption and clarified this in the methods (from page 4 line 25).

Page 5, lines 7-9: "The detection limits, determined as 3-times the standard deviation of blank samples, averaged 0.08 ng/L. For statistic purpose, values below the detection limit (DL) were included as 0.5xDL. Recoveries, determined by 5 ng/L standards analyzed every 10 samples, averaged between 93 and 107 %." If I understood correctly, the analytical uncertainty is of 7 % at 5 ng/L. Given that most of the concentrations you analyzed are < 1.0 ng/L, did you check what the analytical uncertainty at that concentration is? I expect it to be much higher (>50%). Additionally, is the analytical uncertainty the same for Hg_{tot} and Hg_{diss}? I am asking because you say later (page 13, lines 1-2) that Hg_{tot} concentrations were much more variable than Hg_{diss} concentrations.

This is an excellent point. We indeed determined analyzer performance using recoveries of 5 ng L⁻¹ samples every 10 samples, and report a recovery between 93 and 107% for these samples. We did not assess the stability of the analyzer specifically at concentrations <1.0 ng L⁻¹, and this could be more variable. However, it is unlikely that higher variability in Hg_{tot} can be explained by this since concentrations are higher in Hg_{tot} than Hg_{diss}.

Page 6, lines 24-25: "The transect between Toolik and the Arctic Ocean performed in March 2016 showed snowpack height ranging between 30 and 66 cm". How can you explain the difference? Could it be due to drifting snow? If so, how can you compare Hg concentrations at various depths and locations? See major comment.

We believe (and clarified) that such differences reflect a high degree of variability in snowpack distribution across the area, but our dataset is not dense enough (or frequent enough) to explain differences in depth across locations in this study. We state, however, that we observed that snow height is higher at Toolik (between 40 and 50 cm) compared to coastal sites (<30 cm). Secondly, we state that highest snow height was observed where the vegetation was composed of shrubs which may have facilitated snow accumulation.

We also focused our spatial discussion on pool size of Hg stored in the snowpack because a detailed comparison of depth patterns would require a frequent and detailed observations of snowpack drifts which is not possible in our study. We clarified this point.

Page 7, line 3: "compared to the literature from temperate snowpacks". It seems to me that your reference is the study by Fain et al. (2013) in Colorado. "compared to the literature from a temperate snowpack" would be more appropriate here. Additionally, while I understand why you compare your results to those obtained in the Rocky Mountain (same instrumental setup and so on), I would like to see a more thorough comparison to other studies performed in polar regions (e.g., Angot et al., 2016b; Steffen et al., 2014).

We perform a detailed comparison to Fain et al. (2013) because that study used the same instrumental setup. However, we added a sentence for comparing Hg⁰_{gas} data to other polar regions (see page 8, lines 3-7).

Page 7, lines 10-12: "The Hg(0)_{gas} measurements consistently showed strong concentration gradients in the atmosphere-snowpack-soil continuum with highest concentrations in the atmosphere (on average, 1.18 and 1.09 ng/m³, respectively) and lowest concentrations in soils (mostly below the detection limits for both years, i.e., <0.05 ng/m³)."

1. Could you please add the standard deviations?

Done!

2. Is this gradient significant given the large error bars in the snowpack (see Fig. 4)?

We clarified that the high error bars shown in Fig. 4 are mainly due to the temporal variability. The gradient is well defined for each given measurement sequence (see Fig. 1 in Obrist et al., 2017).

3. It's just a detail, but I think that the detection limit for the Tekran 2537 is 0.10 ng/m³.

Thanks for this remark, indeed, it is $<0.10 \text{ ng m}^{-3}$ for the Tekran 2537. We corrected that.

Page 7, lines 23-25, referring to AMDEs: “During one of these periods shown in Figure S2, $\text{Hg(0)}_{\text{gas}}$ concentrations in the snowpack showed variable $\text{Hg(0)}_{\text{gas}}$ levels generally following $\text{Hg(0)}_{\text{gas}}$ concentration changes in the atmosphere above”.

1. What do you mean by “generally following”? Is there a correlation between concentrations in the atmosphere and upper layers of the snowpack?

Hg^0_{gas} concentration variations in the upper layers (as well as for deeper layers) are linked to the Hg^0_{gas} concentration variations in the atmosphere (due to the snowpack diffusivity). We clarify this point in the manuscript (page 8, lines 27-31).

2. Could you please add the following data on Figure S2?

a) O₃ in the atmosphere b) Snow height (is the inlet at 20 cm above the ground far from the surface?) c) Hg_{tot} in surface snow samples d) Hg(II) concentrations in the atmosphere (the data do exist according to Obrist et al. (2017)).

O₃ and Hg^{II} concentrations were added in the figure. The snow height was between 25 and 28 cm, i.e., 5 to 8 cm far from the surface (information added in the caption). Hg_{tot} concentrations in surface snow samples were $1.00 \pm 0.07 \text{ ng L}^{-1}$ for Mar. 25th and $1.46 \pm 0.16 \text{ ng L}^{-1}$ for Apr. 2nd (only concentrations that we have, available in Table S2).

3. How can you explain the peaks at 10 cm above the ground (e.g., daytime on Mar 28th, 29th and 30th)? Was the temperature in the snowpack or the sample line stable? Could it be Hg(II) released from the sample line and analyzed as Hg(0) ?

In general, we found similar diurnal patterns in the atmosphere as well as in all snowpack levels. We attribute such diurnal variability to differences in daytime and nighttime boundary layer mixing. However, we cannot specifically explain why inlet depth at 10 cm was more variable during this time. In general, atmospheric Hg^0_{gas} concentrations, which then also affect snowpack concentrations, are most variable during these periods of AMDEs.

4. According to table S2 you collected surface snow samples during the AMDE. I am very surprised by the very low concentration (1.46 ng/L). This is rather unusual during a depletion event. How does it compare to other studies (e.g., Steffen et al., 2014)? If we do a back of the envelope calculation based on Hg(0) and Hg(II) concentrations in the atmosphere, what should be the concentration in surface snow to have a coherent Hg budget? With 1.46 ng/L in surface snow and $\sim 0.4 \text{ ng/m}^3$ of Hg(II) according to Obrist et al. (2017), I have the feeling that there is Hg missing in the budget. If so, a) How reliable are the Hg(II) measurements? What is the analytical uncertainty? Is an underestimation of Hg(II) concentration conceivable? b) Since Toolik snowpack is subject to significant drifts and changes in snowpack height, are you 100 % sure that you collected surface snow samples? If you collected deeper layers due to drifting snow that could explain the unusually low concentrations during the depletion event.

Indeed, the concentrations are low compared to Steffen et al. (2014) data. We revised this discussion and now state that the coarse resolution of our snow sampling campaign during AMDEs is not sufficient to closely track the fate of Hg deposition and subsequent Hg reemissions during AMDEs. So, while we feel that the reviewer's points are excellent, we strongly think a mass balance approach is above what we are comfortable to address in this study with a very coarse snow sampling activities during AMDEs. However, if we were to perform a back-of the envelope mass balance, it would amount to the following: First, please note that during the weeks of the most intense AMDEs (last week of March and first week of April), we observed highly variable daily average atmospheric Hg^{II} concentrations, ranging between non-detectable levels to a daily average of 310 pg m^{-3} based on the dataset presented in Obrist et al. (2017) (Extended Data Figure 2). Hence, peak Hg^{II} concentrations of $\sim 400 \text{ pg m}^{-3}$ as the reviewer refers were only observed during very few hourly measurements and cannot be applied over long time. In fact, we estimated a total Hg^{II} deposition of between 0.8 and $2.8 \text{ } \mu\text{g m}^{-2} \text{ a}^{-1}$ (based on a deposition velocity of 1.5 cm s^{-1}), of which most of this occurred during periods of AMDEs. At the same time, measured Hg^0_{gas} fluxes (re-emissions) showed a similar magnitude of Hg^0_{gas} re-emissions, e.g., of $1.5 \text{ } \mu\text{g m}^{-2}$ in ~ 20 day during AMDEs (Figure 1 of Obrist et al., 2017). Hence, we believe in fact that re-emission of Hg^0_{gas} after AMDEs largely accounted for the estimated Hg^{II} deposition, a finding we support with stable Hg isotope measurements in Obrist et al. (2017).

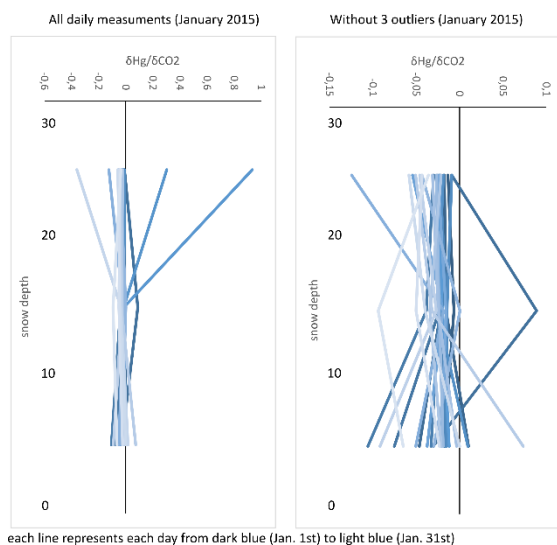
Now, looking at surface snow (3 cm, which has a water equivalent of 10 L m^{-2} considering a snow density of 0.3), an increase in snow Hg (difference of Hg_{diss} from 0.26 to $1.15 = 0.89 \text{ ng L}^{-1}$) would represent a very small input of only $\sim 9 \text{ ng m}^{-2}$ during the AMDEs. So, in fact, what would find that snow indeed only represent a very low amount of the potential Hg^{II} deposition and Hg^0_{gas} re-emission. It would be interesting to perform a high-resolution mass balance campaign to address this issue further during AMDEs.

Page 7, lines 32-34: “This pattern was consistent for two independent soil profiles measured at this site, one mainly representing an organic soil profile and one profile dominated by mineral soil horizons”. I do not understand this sentence. Was the experimentation carried out at two different locations with different soil composition? There is no mention of this in the Materials and Methods section.

We clarified that of our soil measurements, we had measurements in different soils types (present within 5 m), some were more organic soils a some were more mineral soils.

Page 8, lines 25-27: “Analysis of $\delta \text{Hg}/\delta \text{CO}_2$ ratios showed no statistically significant differences from the top to the bottom of the snowpack”. It seems like the authors used the mean concentration at each height to calculate the $\delta \text{Hg}/\delta \text{CO}_2$ ratios. Given the quite large error bars for $\text{Hg}(0)_{\text{gas}}$ in the snowpack (Fig. 4), it is not really surprising that such a calculation yields insignificant differences. On the contrary, Fain et al. (2013) calculated the ratio for each day. Did you try to do it this way?

Actually, the ratios presented in boxplots of the Fig. 4 were in fact calculated from daily Hg^0_{gas} and CO_2 measurements, as performed by Fain et al. (2013). See below the detail on the individual days. On the left panel, each dataline represents each day (from darker to lighter). On the right panel, the same plot without 3 outliers. No apparent trend was observed. What we have done in Fig. 4 was to apply boxplot for each snow height difference (0 to 10, 10 to 20, and 20 to 30 cm).



Page 9, lines 7-10: “No consistent temporal trends in Hg_{tot} or Hg_{diss} were observed with increasing duration of winter in both seasons, and no correlations were observed with air temperature (red line). One noticeable period of enhanced surface snow Hg concentrations was April 2016 when both Hg_{tot} and Hg_{diss} concentrations exceeded 1 ng/L .”

1. Is there an anti-correlation between Hg_{tot} and $\text{Hg}(0)$ in the atmosphere?

No, there is no anti-correlation between Hg_{tot} and Hg^0_{gas} , nor for Hg_{diss} and Hg^0_{gas} , during most of the time. One exception was during AMDE where we observed low Hg^0_{gas} and higher snow Hg concentrations.

2. See previous comments regarding the magnitude of the concentrations. How do a maximum of 1 ng/L compare with other studies performed in polar regions?

We indicated that snowpack $\text{Hg}_{\text{tot}}/\text{Hg}_{\text{diss}}$ measured in literature were between 0.14 and 820 ng L^{-1} . This range mostly includes Hg_{tot} concentrations (and so, potentially higher concentrations) when no filtration was applied. We state that our observed concentrations were in the lower part of the Arctic data set: the lowest values were observed in Greenland (Ferrari et al., 2004), in Canada for Hg_{tot} (St Louis et al., 2005), and in Alaska for Hg_{diss} (Douglas & Sturm, 2004).

Page 9, line 12: “Due to the low snow height on the frozen lake.” Why is the snow height lower on the frozen lake? Drifting snow?

We assume that both drifting snow and surface roughness (brush vs ice) are responsible to the difference of snow height, as already discussed page 11, from line 4.

Page 9, lines 15-16: “Measurements of Hgtot and Hgdiss across a large North slope transect (about 200 km) in March, 2016 showed concentrations of 0.70 μC^{-1} s 0.79 and 0.24 μC^{-1} s 0.20 ng/L, respectively.” Are you referring to surface snow or to average snowpack concentrations?

We referred here to average snowpack concentrations. Surface snow sampling was not performed for the transect. We clarified that.

Page 9, lines 16-19: “Concentrations of Hgdiss of the 5 northernmost stations were statistically significantly higher compared to those measured in the 4 stations located in the interior tundra which included the Toolik site where the mean Hgdiss concentrations were 0.33 and 0.11 ng/L for the same period, respectively”.

1. Please add the standard deviations (in the text and on Figure 7)

The standard deviations were added. The standard deviations of the Fig. 7, however, were not added since only a composite sample from both replicates of the same snow height was measured.

2. I would like to see a critical discussion of these results, notably in light of drifting snow issues and analytical uncertainty (50 % at such low concentrations?).

We discussed the influence of drifting snow in each part of the discussion (absence of vertical patter, absence of difference between the lake and tundra snowpacks...). Note that we reorganized the manuscript merging Results and Discussion.

Page 9, line 23: “similar to concentration of Hg”. Where is that discussed in the manuscript?

We clarified this sentence. Major cations and anions concentrations in snowpack were lower at Toolik compared to coastal locations, which we also observed for Hg as discussed in the previous section.

Page 11, lines 11-12: “Hg(0) concentration profiles in the arctic snowpack are inherently different to patterns observed in lower latitude snowpacks”. Now may be a good time to compare your results to other studies in polar regions (see comment page 7, line 3).

We added the comparison with other polar snowpacks, as discussed above.

Page 11, lines 23-25: “We speculate that a reason for the general lack of Hg(0)gas formation and volatilization in snow includes substrate limitation due to very low total snow Hg concentrations, several times lower compared to concentrations in temperate snowpacks.” While I agree that this is a possible explanation, I would like to see a more thorough discussion of other conceivable hypothesis, including instrumental limitations. A few ideas:

1. The comparison between your results and those by Fain et al. (2013) is based on data acquired in winter and early spring. I would expect Hg(0)gas formation and volatilization to start in spring, when the sun is back. What is the difference in UV load between Toolik and Colorado at that time of year? That could explain a lower photo-reduction in the upper layers of the snowpack at Toolik.

Indeed, the solar radiation is a little bit lower at Toolik. We clarified that solar radiation at Toolik was about 400 W m⁻² in Mar. and 600 W m⁻² in Apr. at Toolik versus ~700 W m⁻² in the end of Feb. in the Rocky Mountain (Fain et al., 2013). The solar radiation data were added in the text (page 8, lines 17-19). However, we do not believe this can explain the dramatically different behavior between these sites.

2. Atmospheric and snowpack Hg(0) concentrations were collected using two different Tekran instruments. Could a 10-20 % difference between the instruments explain why Hg(0)gas in surface snow is not higher than in ambient air (especially Figure 4c)?

See our comment above. In this study, we always compared snow and atmospheric measurements using the same Tekran instrument.

3. 85 % of photo-reduction occurs in the top two e-folding depths (King and Simpson, 2001). Fain et al. (2013) observed a diurnal cycle 0-60 cm below the surface. But the density/nature of the snowpack is most likely very different at Toolik.

a) You recorded the physical properties (e.g., density) of the snowpack. Could you compare your results to those obtained by Fain et al. (2013)? According to Durnford and Dastoor (2011), the range of chemically active depths may be explained by physical differences in the snowpacks.

b) The e-folding depth is about 5 cm at Alert (King and Simpson, 2001) while Poulain et al. (2004) reported that photo-reduction occurred in the first 3 cm of snow. I am worried that the upper inlet in the snowpack (the one at 40 cm above the ground, Fig. 4c) might be too far from the surface. This is more obvious on Fig. 3 if you compare the distance between the upper inlet and the surface in your case and in Fain et al. (2013).

We added the measurement depths of the surface snow inlet which was between 5 and 7 cm. We also discuss the possibility that our surface measurements may have missed Hg^0_{gas} production. However, at the same point, we point out that in Fain et al. (2013), we used the same measurement system and that Hg^0_{gas} concentration enhancements in that study were well measurable (with concentrations up to 8 ng m^{-3} and detectable to a depth of $>90 \text{ cm}$).

4. What about the occurrence of fresh snow at Toolik vs. Colorado? Fresh snow provides a new reservoir of photoreducible Hg(II) and highest surface snow Hg(0) levels are linked to the deposition of new snow (Faïn et al., 2013).

We performed plot as done in Fig. 5 in Faïn et al. (2013) but no relationships were observed.

Page 12, lines 1-2: “This illustrates that the Hg(0) gas uptake occurs in soils rather than in the snowpack.” The way the ratios are plotted, I am not sure I understand why you can conclude that uptake occurs in soils rather than in the snowpack.

We clarified the results as follows: the constant ratios and the fact that CO_2 is largely non-reactive in snowpack indicates that Hg^0_{gas} also was not subject to snowpack chemical reactions. The constant, and negative ratios, between CO_2 and Hg^0_{gas} ratios are hence indicative that both profiles are affected by underlying soil processes, i.e., soil sources for CO_2 and soil sinks (for H^0_{gas}).

Page 12, line 10-11: “which we attribute to higher variability in upper snowpack concentrations due to variable atmospheric Hg(0) gas levels”. Isn’t it in contradiction with error bars (Fig. 4 and 5)? I actually don’t understand why (Fig 5) a ratio based on highly variable concentrations (e.g., 0 cm) can be less variable than a ratio based on less variable concentrations (e.g., upper inlet). Am I missing something?

You are correct. we cannot only attribute the high variability to the Hg^0_{gas} atmospheric fluctuation. We think that the source of variability is due much lower concentration differences between 20 and 30 cm inlets for both CO_2 and Hg^0_{gas} which cause increases variabilities in ratios. We modified the text (page 10, lines 15-17).

Page 12, line 23: “relatively weak and infrequent AMDEs”. I quibble but the AMDE reported on Fig. S2 does not look weak to me. About the frequency, how many AMDEs did you observe? You can use a 1.00 ng/m^3 threshold to calculate the frequency of occurrence (Angot et al., 2016a; Cobbett et al., 2007; Steffen et al., 2005).

We changed this to infrequent and generally weaker AMDEs. We think it is important to note that the frequency and magnitude of AMDEs indeed is lower than those along the coast. While we see several occasions with air masses with Hg^0_{gas} levels below 1.0 ng m^{-3} (4 to 5 per year), it is very rare to find stronger depletions (e.g., one per year with concentrations $<0.5 \text{ ng m}^{-3}$). This is dramatically different to studies along the coast where Hg^0_{gas} frequently drops below detection for pronounced time periods.

Page 12, line 31-32: “Total Hg concentrations in all snow samples collected were always much higher than Hg_{diss} levels”. Can we really say that concentrations are “much higher” given the range observed ($0\text{-}1 \text{ ng/L}$)? By the way, I don’t see anywhere Hg_{tot} concentrations from the transect.

Done, we removed “much”.

Page 13, line 10: “(. . .) and few studies include inland sites such as Toolik”. What is the range of concentrations at these inland sites? Which studies are you referring to?

We clarified that in Douglas & Sturm (2004), snowpack Hg_{diss} concentrations were between 0.5 and 1.7 ng L^{-1} for the interior sites. While our values are even lower, they are not that different from Douglas and Sturm (2004).

Page 14, line 27: “Fresh surface snow”. There is no mention of the fact that you collected fresh surface snow in the Materials and Methods Section.

We removed “fresh” and only keep “surface” (i.e., top 3 cm).

Figure 2: Why don't you report the standard deviation for each green bar?

The absence of standard deviation may be due to either narrow differences between the two replicates (i.e., two snow pits), or different snowpack height between the two snow pits (top measurements of Jan. 2015 and Mar. 2016 for example, illustrating the spatial heterogeneity). See Table S2 for more detail about standard deviation values.

Figure 4: What does “soil: organic/mineral” refer to?

This information is now added in the site description in ‘Materials and methods’ (page 4, lines 8-9).

Figure 7: Please add standard deviations.

As described above, the two samplings were pooled together to constitute a composite sample. Only one sample was analyzed per layer and per site.

Figure 8: I really like this Figure but aren't the various concentrations (<DL, 0.25, 0.50) in the range of the analytical uncertainty?

No, values of 0.25 ng L⁻¹ and 0.5 ng L⁻¹ are significantly above the detection limit of our method (i.e., above 3 × standard deviation of blank samples).

Figure 9: The colors are really difficult to read. Can you use something else than shades of blue? Maybe a gradient from blue to red.

Done

References

- Douglas, T. A. and Sturm, M.: Arctic haze, mercury and the chemical composition of snow across northwestern Alaska, *Atmos. Environ.*, 38(6), 805–820, doi:10.1016/j.atmosenv.2003.10.042, 2004.
- Faïn, X., Helmig, D., Hueber, J., Obrist, D. and Williams, M. W.: Mercury dynamics in the Rocky Mountain, Colorado, snowpack, *Biogeosciences*, 10(6), 3793–3807, doi:10.5194/bg-10-3793-2013, 2013.
- Ferrari, C. P., Dommergue, A., Boutron, C. F., Jitaru, P. and Adams, F. C.: Profiles of mercury in the snow pack at Station Nord, Greenland shortly after polar sunrise, *Geophys. Res. Lett.*, 31(3), L03401, doi:10.1029/2003GL018961, 2004.
- Obrist, D., Agnan, Y., Jiskra, M., Olson, C. L., Colegrove, D. P., Hueber, J., Moore, C. W., Sonke, J. E. and Helmig, D.: Tundra uptake of atmospheric elemental mercury drives Arctic mercury pollution, *Nature*, 547(7662), 201–204, doi:10.1038/nature22997, 2017.
- Seok, B., Helmig, D., Williams, M. W., Liptzin, D., Chowanski, K. and Hueber, J.: An automated system for continuous measurements of trace gas fluxes through snow: an evaluation of the gas diffusion method at a subalpine forest site, Niwot Ridge, Colorado, *Biogeochemistry*, 95(1), 95–113, doi:10.1007/s10533-009-9302-3, 2009.
- St. Louis, V. L., Sharp, M. J., Steffen, A., May, A., Barker, J., Kirk, J. L., Kelly, D. J. A., Arnott, S. E., Keatley, B. and Smol, J. P.: Some sources and sinks of monomethyl and inorganic mercury on Ellesmere Island in the Canadian high Arctic, *Environ. Sci. Technol.*, 39(8), 2686–2701, doi:10.1021/es049326o, 2005.

REVIEWER 2

General comments

The manuscript describes a detailed study of Hg in air, snow and soil at an Arctic site, which unlike almost all previous studies is a significant distance from the coast. Given that vast amounts of tundra are inland this study begins to fill in some of the gaps in our knowledge of Hg cycling in these remote regions. Of particular interest are the differences seen between the processes seen at this site when compared to coastal sites. The fact that tundra soils are a sink for atmospheric elemental Hg has important repercussions for future multimedia modelling studies and hints at the potential remobilization of large amounts of Hg from Arctic soils in a warming climate. This should be emphasised more in the Abstract and the Conclusions, in the Abstract particularly the comment on this is brief and hidden in the middle.

Thank you for your comment. We agree with you that the fact that Arctic tundra soils constitute a sink of Hg⁰_{gas} is an interesting finding for the global understanding of Arctic Hg cycling. This is, however, not the

focus of this paper since we already published that in a previous study (Obrist et al., 2017). We slightly edited both the abstract and conclusion in order to highlight and clarify this point.

The manuscript is however rather long-winded. I think that both the Results and the Discussion section could be shortened significantly, and quite usefully (from the reader's point of view) combined. Just as an example, the discussion of the major ions and the O and H isotope signatures, repeats parts of the pertinent results section. Conversely the results section rather leaves the reader with a sense of 'and what do these results imply?', which is only answered six pages later. I would recommend combining these sections as it will most likely lead naturally to a more concise and less prolix article. If some of the detail in the methods section has already been published perhaps it could be shortened by including more references, if not maybe some of the detail could be moved to the Supplementary material.

We reorganized the manuscript by merging Results and Discussions.

The previous reviewer has comprehensively addressed a number of technical issues, and for me only a few real problems remain.

1. The issue of blowing snow, and where the snow that is being sampled at Toolik comes from.

Indeed, the entire snowpack is subject to important movements that generate uncertainties mentioned by the first reviewer and limiting the observation of trend in the snowpacks. We mention this in regard to vertical Hg concentration patterns, in response to the comments of reviewer 1.

2. The fact that the paper is interesting and adds an important contribution to polar Hg research but is unfortunately not very well presented and at times rather heavy going.

We hope that the key messages are now better presented in this reorganized manuscript version.

3. The importance of atmospheric elemental Hg effectively being sequestered (for the moment) by tundra soils, is not emphasised sufficiently from my point of view.

We previously mentioned the influence of tundra soil in Arctic Hg cycling in Obrist et al., 2017 and highlighted this in the abstract and conclusions.