

Final response to interactive comments of Referee #2

A dual **inlet**, single detector relaxed eddy accumulation system for long-term measurement of mercury flux

S. Osterwalder, J. Fritsche, M. B. Nilsson, C. Alewell, J. Sommar, G. Jocher, M. Schmutz, J. Rinne, and K. Bishop

General comments:

The manuscript entitled “A dual, single detector relaxed eddy accumulation system for long-term measurement of mercury flux” by Osterwalder et al. 2015 is a very well written easily-readable article. The authors describe a system that they adapted to attempt to measure gaseous elemental mercury surface-atmosphere fluxes via the micrometeorological relaxed eddy accumulation method. These additions include use of dual inlets to be able to collect updrafts and downdrafts during the same time periods and a single detector for analysis of GEM to eliminate potential differences between analyzers that could introduce systematic error. They have also added an automatic calibration source.

Response: We appreciate the reviewer’s suggestions to further improve the manuscript and that he/she sees the importance of the significant improvements in our system compared to other MM methods.

In my initial quick review I felt that the authors deserved a chance to prove to me that they have made an entirely new breakthrough, pushing the science forward and convincing me to never again use other micrometeorological techniques for GEM. Unfortunately, they have fallen short of that. I fail to see the significance of this development that warrants publication of just the method alone. Although the paper is well-written, most of this information would typically go into supplementary info of a manuscript that reports the findings of flux studies that use this system. Simply “tweaking” an existing method does not warrant publication as an entirely new atmospheric measurement technique and appears a little like an attempt to just add to publication numbers. Also, the valid data collected by this system is very low (32 – 38%), thus I would argue it is not much of an improvement over existing systems/techniques.

Response: We appreciate the effort given this manuscript by the reviewer. However, we do not agree that the methodological developments are such that they “would typically go into supplementary info (SI)”. The design we present here overcomes the major source of uncertainty in the current state of the art to measure land-atmosphere exchange of Hg using relaxed eddy accumulation as presented by Zhu et al. (2015b), namely the lack of the simultaneous up and down draft sampling on. We would like to also highlight the potential of the system for long-term measurements since it runs fully automated and has reduced staff effort compared to previous systems. Thus we believe that the changes and validations are substantial enough to merit publication of a method paper by itself. With respect to the central concern about the proportion of valid data collected (32-38%), that low yield was by no means representative for the innovative new design we are presenting here. The absolute majority of the data loss was due to substandard performance in a standard piece of equipment. Poor data yields are by no means an inherent feature of the new design. For subsequent deployments of the system, including the most recent measurement campaign at Degerö in July 2014 we had data yields of over 90%. These data will be presented in an upcoming publication that will focus on the features controlling evasion – if we succeed in getting the presentation of the new method accepted in this paper. To exemplify that average data coverage we have added a figure showing diurnal patterns of the flux between 02.07.2014 and 16.07.2014 (Fig. F3). Within the field of biogeochemical research on GEM dynamics and exchange, unbiased, high precision, measurements are even more critical than for many other constituents. Thus, we believe that detailed descriptions of developments and system validations are crucial in advancing such a research field. We also do not believe this publication is an “attempt to just add to publication numbers”, since this is the first article our research group has made on

relaxed eddy-flux accumulation, and it comes after four years of full-time development work by the main author. Our aim is to break into a complex field of research with a novel technique to overcome the major sources of uncertainty in the current state of the art to measure land-atmosphere exchange of Hg using relaxed eddy accumulation as presented by Sommar et al. (2013a) and Zhu et al. (2015b). Our effort is also consistent with the call by Agnan et al. (2015) for further improvements of methodology and establishing standards for GEM micrometeorological measurements.

Specific comments:

Title: Should be "A dual inlet, single : :"

Response: Thanks! We have changed the text accordingly.

1) Pg 8115 Line 9: should be " : :difficult due to technical: : :"

Response: We followed the suggestion.

2) Pg 8116 Line 10: Statement " Over an annual cycle: : : " This is stated as a fact, but we actually do not know for sure, so citations are needed or it needs removed.

Response: We reviewed the paragraph accordingly.

Suggested changes in the manuscript: "Gustin et al. (2008) suggest that today a substantial amount of Hg deposited on soils with natural background concentrations of Hg ($< 0.1 \mu\text{g g}^{-1}$) is reemitted back to the atmosphere and that over a year deposition is largely compensated by re-emission resulting in a net flux close to zero."

3) Pg 8116 Ln 16: should this be "...of GEM flux measurements."?

Response: We changed that in the manuscript.

4) Pg 8116 Ln 22: Cobos et al. 2002 did not use flux chamber so why is it cited here?

Response: Cobos et al. 2002 discuss pros and cons of the application of flux chambers as well as micrometeorological methods. Indeed we refer now to Gillis and Miller, 2000, Wallschlager et al. 1999 and Eckley et al. 2010 who presented different effects which bias flux chamber measurements.

Suggested changes in the manuscript: "But DFCs alter the enclosed environment of the volume and surface area being studied by affecting atmospheric turbulence, temperature and humidity (Gillis and Miller, 2000; Wallschlager et al., 1999; Eckley et al. 2010). Also the concern about influencing plant physiology, means that DFCs are restricted to short term measurements and studies comparing the relative ..."

5) Pg 8116 Ln 27: " : :for deducing regulating mechanisms: : : " I would argue that MM methods are applied over a large source area, making them difficult to use for mechanistic studies.

Response: It is certainly true that MM techniques do represent the net exchange over a certain area, the footprint area and thus the interpretation needs to be connected to the appropriate system borders. However, we do not understand why this would make it principally difficult to make mechanistic interpretations. Within closely related fields mechanistic interpretations of MM results are common, and also well validated. Indeed there are a number of publications that have made the connection between peatland processes and the exchange of other scalars such as CO_2 , CH_4 or energy fluxes using MM techniques. Examples of such publications from the same footprint we report here are (Nilsson et al. 2008, and Peichl et al. 2013). So we believe that the statement "...they

provide flux data valuable to characterize ecosystems as sinks or sources of atmospheric Hg and to interpret seasonal flux patterns.” is scientifically justified.

6) Pg 8121 Ln 6: *Please indicate to the reader that (A1, A2) refer to Figure 1.*

Response: We followed the suggestion.

7) Pg 8121 Lines 12 – 21: *What is the approximate “dead volume” of the system? What are the implications of the dead air volume in the system?*

Response: One of the advantages of the dual inlet system is that with two separate lines the dead volume is a much smaller source of error than for single inlet systems. The only dead volume that has importance in a dual-inlet system is the one before the sampling valves which has been minimized in our system. Its effect has been accounted for in high frequency attenuation calculations (discussed in Sect. 3.3). Flux damping at high frequencies is negligible for the data from Basel but the integral damping factor for the REA flux can be significant at high frequencies at Degerö especially during stable and strong wind conditions.

8) Pg 8122 Lines 2 – 9: *What gold traps are used? How were they prepared? Or were they bought from a supplier? If so what supplier? What specifications were used to identify matched pairs of gold traps?*

Response: The gold cartridges were purchased from Tekran Inc., Toronto, Canada as matched pairs. The supplier uses internal criteria that ensure less than 5% difference in sensitivity – often it is less. The cartridges were tested extensively during the manufacturing process (personal communication with Tekran Inc.). After the field campaigns the system was calibrated manually. The results of these injection procedures are presented in Figure S1.

Suggested changes in the manuscript (Pg. 8121 Line 16): “Conditionally sampled GEM is subsequently accumulated on the two matched pairs of pure gold cartridges (Tekran Inc., Canada, difference between cartridge sensitivity < 5% according to manufacturing tests by the supplier).”

9) Pg 8122 Line 6: *For a 1.2 ng m⁻³ calibration standard in a 5 min sample (0.003 m³ volume) then this is a loading of 6pg of mercury. This is a difficult thing to accurately load the incredibly small amount. If this paper were to be revised, considerably more information would be needed on how this is possible and the data provided on the reproducibility. If this has already been published then a citation is needed.*

Response: We appreciate this comment and corrected the inexact statement in line 6 which led the reviewer to believe that we tried to load a ~6 pg GEM standard. As the reviewer notes this would provide a very low level of accuracy for injections using a permeation device (Dynacalibrator Model 150, Valco Instruments Co. Inc.). As stated in the manuscript we used this permeation device to generate a precise GEM concentration to detect potential sensitivity loss of the gold-cartridges and to adjust for temperature sensitivity of the Tekran detector. The average recovery of the GEM standard (back calculation from manual calibration of Hg detector) of cartridge pair 2-4 corresponded to 27.2±1.1 and 22.2±1.3 pg in Basel and 32.1±2.1 and 32.1±2.3 pg at Degerö.

Suggested changes in the manuscript: “The temperature controlled Hg permeation source provided precise GEM concentrations in a constant stream of dry, Hg-free air. The average recovery of the GEM standard was determined by back calculation from the manual calibration of the Hg detector. The average loading on cartridge pair 2-4 corresponded to 27.2±1.1 and 22.2±1.3 pg in Basel and 32.1±2.1 and 32.1±2.3 pg at Degerö.

10) Pg 8122 Line 18. *“The air samples were analyzed alternately every hour.” To me this sounds like you’re defeating your purpose of being able to continuously monitor updrafts and downdrafts for the same 30 min flux periods. If that is not the case then this needs rephrased. If it is the case, then this is an additional weakness of the system.*

Response: We appreciate this comment and have simply deleted this sentence. In fact the idea was to refer to the automated calibration source and not to the up-and downdraft air samples which were measured every 30 min and allowed continuous measurements. So, using our dual-inlet REA, updrafts and downdrafts are continuously monitored by being continuously captured on gold cartridges. During the automated calibration procedure the cartridges were loaded with GEM reference air and Hg-free air alternately every 30 min, just after up-and downdraft air samples were analyzed. The procedure is illustrated in Fig. 2.

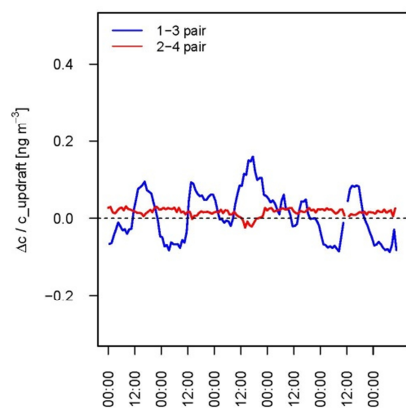
11) Pg 8123 Section 2.4: *The actual criteria for rejecting data are not presented here or in Section 3.1.4. The only thing that is reported is the percentage of data that was rejected, but I think the reader should be allowed to make the decision whether the criteria used were valid. Sometimes data rejection thresholds can be too restrictive unnecessarily rejecting good data. The reader should be given this opportunity. For instance, exactly what percentage difference between the gold cartridges was used as a threshold to invalidate data.*

Response: We agree with the Referee that the cartridge-pair offset criterion was presented insufficiently in the manuscript. Therefore, we have added information about the bias tests to determine the rejection criteria and to derive the system detection limit. The tests were performed after the campaigns during 5 days in Basel and over 28 h at Degerö, respectively. After the standard correction for temperature sensitivity of the detector the ratio of the Hg concentration difference and the Hg concentration in the updraft were investigated (Fig. F1). The cartridge pair 1-3 showed a significantly higher offset variability in Basel (0.009 ± 0.06) and Degerö (0.17 ± 0.06) compared to cartridge pair 2-4 in Basel (0.016 ± 0.01) and Degerö (-0.004 ± 0.02). If up- and downdraft lines sample the same air, the offset between these should be constant, independent of air Hg content. Line pair 1-3 gravely deviates from this criterion and thus needs to be rejected (Fig. F2). Regardless, we have derived two thresholds to identify malfunction of a cartridge pair:

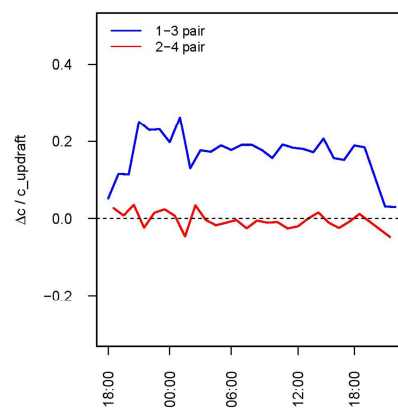
- 1) a maximum of 10% difference between the gold cartridges
- 2) The maximum standard deviation of the acceptable offset is 0.05.

The other rejection criteria listed in Table 1 are already sufficiently described in Sect. 3.1.4.

In the manuscript, Fig. 6 (=Fig. F1a) will be presented together with Fig. F1b, Fig. F2a and Fig. F2b and moved to the Supplement.

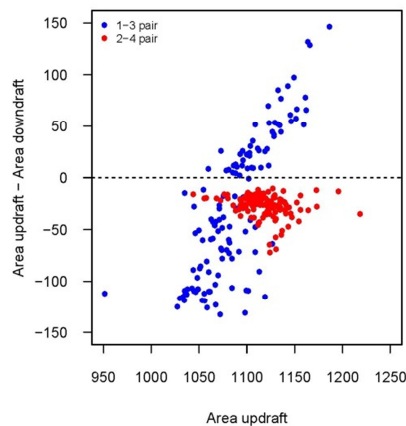


a)

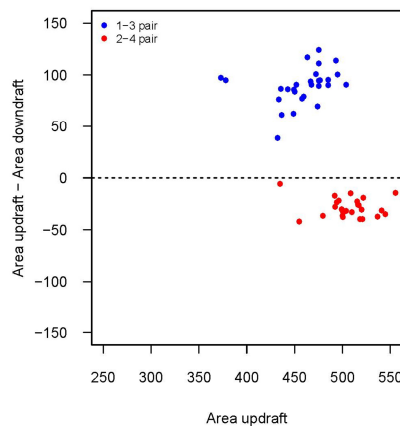


b)

Fig. F1: Precision in concentration difference measurements of the same GEM concentration between cartridge pair 1–3 (blue) and 2–4 (red) in Basel (a) and at Degerö (b).



a)



b)

Fig. F2: Ratio of the difference between the updraft and downdraft integrated air sampling to the updraft signal for both cartridge-pairs during the bias test in Basel (a) and Degerö (b).

Suggested changes in the manuscript (Sect. 3.1.3): “The instrument detection limit of the Hg detector was $< 0.1 \text{ ng m}^{-3}$ and allowed discernment of GEM peaks from the baseline noise for all measurements. The gold cartridge-pair offset criteria and the method detection limit were derived in the field from sampling the same air through updraft and downdraft lines. For this study we defined two strict rejection criteria for 1) maximum standard deviation of the offset of 0.05 and 2) maximum difference in gold cartridge response of 10%. The assessment of the offset between the sampling lines during the Basel measurements was 0.009 ± 0.06 ($\pm$ SD) and $0.016 \pm 0.01 \text{ ng m}^{-3}$ for gold cartridge-pairs 1-3 and 2-4 respectively (Fig. F1a). At Degerö the offset was 0.17 ± 0.06 and $-0.004 \pm 0.02 \text{ ng m}^{-3}$ for 1-3 and 2-4 (Fig. F1b). If up- and downdraft lines sample the same air, the offset between these should be constant, independent of air Hg content. Line pair 1-3 gravely deviated from this criterion (Fig. F2) Therefore, measurements with cartridge-pair 1-3 were discarded for both campaigns due to the above threshold variability in Basel and the large cartridge-pair offset at Degerö.

12) Pg 8124 Section 2.4.3: *How much bias between the lines was found? If the number is large then then this could indicate a contamination in the system, which may be causing some reactions within the system and impacting measurements. Did the bias change over time? This could indicate a systematic problem and needs to be reported.*

Response: In Basel the bias between the lines was 0.9 % and 1.6% for cartridge-pair 1-3 and 2-4 respectively (in the manuscript now). At Degerö the bias accounted for 17% and 0.4% respectively. The reason for that offset and its fluctuation was not due to poor performance of the gold cartridges themselves but rather due to a system artefact caused by Teflon valve restrictions within the 1-3 cartridge-pair sampling path. In between the Basel and Degerö campaigns we tested the system thoroughly by swapping sampling lines, exchanging and turning Teflon valves and checking heating wires. We thought that we had solved the problem, but discovered afterwards that we had not. After that Degerö 2012 campaign we replaced the valves that seem to have been creating the problem, and we also now check the gold-cartridge bias on a regular basis.

13) Pg 8125 Line 22 – 26: *Why were different footprint models used? I would think that to actually do a comparison between the two sites, then the same footprint model should be used. If there is a valid methodological reason then that should be stated. Or if it doesn't affect then that needs to be stated and discussed.*

Response: Given the very different conditions above a city and a mire, the footprint models were used which best fit to the specific environments. Each model has also been used at the specific site to determine the source area for other gases. The footprint model used at Degerö for instance was made for measurements in the surface layer up to 20 m height and for sites where measurement height >> canopy height. This model would not be suitable above a city. Generally, the characterization of the footprint area within the extremely heterogeneous urban environment required footprint models adapted more specifically for such conditions. At the Basel site the Kormann and Meixner (2001) approach was already implemented for the evaluation of continuous long-term (10+ years) measurements because of its minimal requirements on computing resources. We do not feel that the qualitative comparison of the two very different sites is impaired by the specifics of the footprint model used at each site. Anyhow, general patterns of the footprint area should be reproduced by both models. The footprint modelling is of major importance when addressing spatial or temporal variability within each of the sites. We agree that using the same model would be much more important when comparing fluxes at two sites with similar source properties

Suggested changes in the manuscript: "To predict the size of REA flux source areas during the campaigns the analytical footprint model of Kormann and Meixner (2001) was applied in Basel and a Lagrangian stochastic forward model following Rannik et al. (2003) at Degerö. The footprint models were chosen in order to fit best the specific requirements as defined by the source areas at each site."

14) Pg 8128 Line 5: should be "was determined to be 18 ms : :"

Response: We changed that in the MS.

15) Pg 8128 Line 20: *Which sampled volumes? I take it this is combined updraft, downdraft, and deadband. I would not consider the dead band as sampled because in this sense it sounds like there is lots of loading of air onto the gold traps but later we find out that the percentage of time that there is actually loading is quite small.*

Response: That's true, we mentioned the combined volumes for updraft, downdraft and deadband. The air "sampled" by the deadband is not analyzed. The analyzed volumes of updrafts and downdrafts were 32% and 33% in Basel. In Degerö they were 28% and 27% respectively.

Suggested changes in the manuscript: "The total volumes drawn over updraft, downdraft and deadband lines averaged 30 ± 0.09 in Basel and 45 ± 0.01 l ($\pm$ SD) at Degerö. The proportion of the air not analyzed accounted for 10.5 l in Basel and 20.3 l at Degerö."

16) Pg 8128 Line 27: *" : : high resolution voltage data were logged." So what was found in this data? Were there times when peaks were irregular? What could this indicate?*

Response: No, we did not find any irregular peaks. The peak integration procedure worked perfectly. We suggested to delete this sentence in the manuscript because it doesn't give any additional information.

17) Pg 8130 Line 12: *" : : also in Basel : : " Also what? I think there may be something*

Response: Yes it's confusing. We have deleted this sub-clause and refer to Sect. 3.1.2 where the application of the two different deadbands is discussed. (from Page 8119, line 25):

18) Pg 8130 Line 15 – 21: *So this paragraph means that the system actually did not work as advertised. To me, this means that the system does not actually continuously measure, since the purpose of a 4 cartridge system is continuous measurement. So instead of having updrafts and downdrafts separated in time, they are now only measured every other hour, missing important information. Gold cartridges should have been thoroughly tested before deployment and data on their performance presented as part of the QA/QC of the data. Also the performance of the gold traps should be monitored over time and reported. When using a commercially available analyzer such as a Tekran 2537A/B/X then there are precise protocols that the company follows before selling the analyzer. This is not necessarily reported all of the time, even though information on the performance of the gold traps is often reported in the supporting info. So this information really needs to be here if this is to be published.*

Response: We concur entirely with the points made here about the need for quality control and agree that we should have been testing for the gold trap bias more frequently than we did in our initial deployment, as this would have quickly revealed the inadequacy in one of the gold-cartridge pairs. We do, however, disagree that the system presented is only going to be able to measure up and down drafts for just ½ hour out of every hour. The system measures continuously and it was an unfortunate situation that the bias test for gold-cartridge pair 1-3 led us to a relatively low data yield. As pointed out in our earlier remarks, this is by no means a level of performance to be expected from this system due to some inherent flaw. As also outlined earlier, we do not believe that this data loss should be a reason for refusing publication of this novel system. We now have data yields of the sampling campaign 2013/2014 from Degerö where both gold-cartridge pairs worked well for months at a time. The data coverage for a sampling campaign of a comparable duration from July 02 to 16 was 93% (Fig. F3). That means just 7% of the data were rejected due to the same rejection criteria summarized in Table 1. We could substitute two weeks of data from that new measurement period in this article, though we would prefer not to since there were slight differences in the system that would need to be explained and thus expand the method section to describe two slightly different systems.

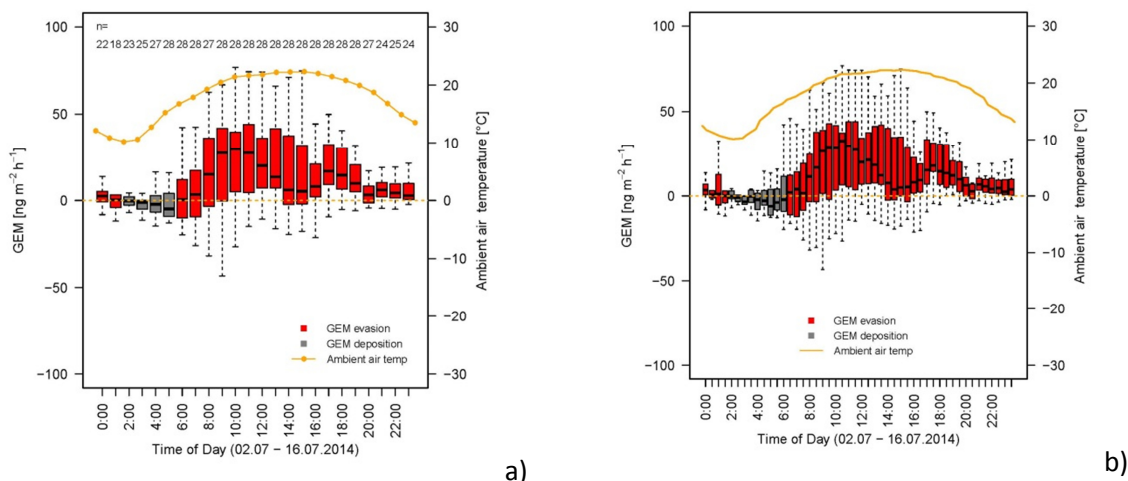


Fig. F3: Diurnal patterns of GEM flux during a two-week campaign at Degerö in 2014. The hourly (a) and half hourly (b) averages of the data presented in accordance to Fig. 10 in the manuscript.

With regard to the referee's concern that *"Gold cartridges should have been thoroughly tested before deployment and data on their performance presented as part of the QA/QC of the data. Also the performance of the gold traps should be monitored over time and reported."* - we agree. The supplier, Tekran Inc., ensures delivery of well-matched gold-cartridge pairs with a maximum offset of 5%. As described in the manuscript (2.4.2) the GEM standard measurement was implemented for routine monitoring of the gold-cartridge performance. During the campaign in Basel the offsets between gold-cartridge pair 1-3 and 2-4 were 1.6% and 9.9%, respectively. At Degerö the offset of pair 1-3 was 4.2% and for 2-4 it was 4.9%. The linear relationship between GEM reference air for up- and downdraft and ambient temperatures are illustrated for both sites in the manuscript (Fig. 4). We are of the opinion that Figure 4 gives enough information about the performance of the gold trap pair in question (2-4) during both campaigns. However, the time series of the performance of all gold traps is given below and we could report in the Supplement if needed (Fig. F4.1 and Fig. F4.2).

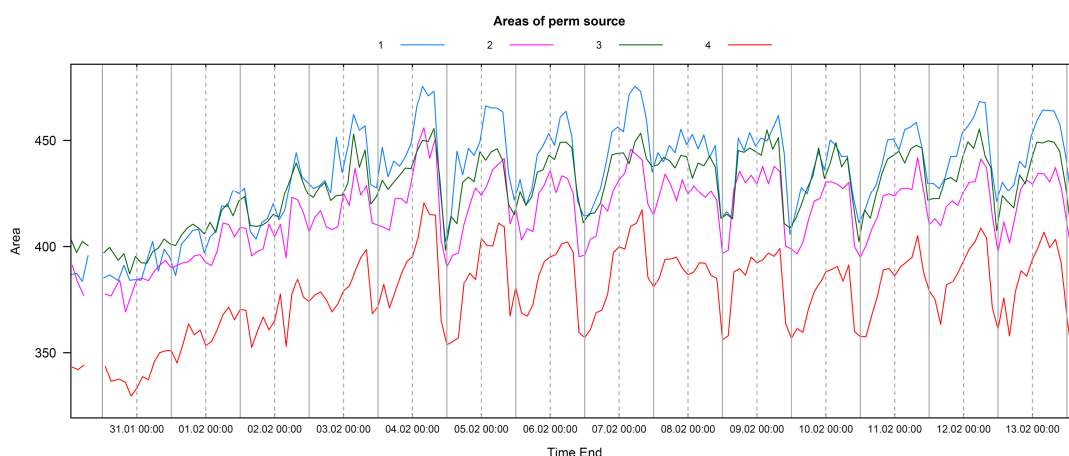


Fig. F4.1: Time series of the analyzed GEM reference air samples from all four gold-cartridges during the campaign in Basel.



Fig. F4.2: Time series of the analyzed GEM reference air samples from all four gold-cartridges during the campaign at Degerö.

19) Pg 8131 Line 15 – 17: *I hate to say it, but I am not sure a new system that only accomplishes 32 – 38% data coverage may not be much of an improvement on current systems.*

Response: As we have noted in several earlier responses (comments 11, 12, 18), level of data yield is not inherent in the design of the system. We have also shown in response to comment 18 that the system can yield over 90% good data during a deployment of similar length to that report here. And while we do not see 35% data loss as anything to be expected in future deployments of the system, we would like to point out that 16 half-hour measurements per day (out of 48 possible) could still be considered a useful number of values for many other measurement methods. The major problem would be if we have preferential loss or coverage of certain time periods during the day, e.g. if all losses occur during the night. This is actually a common problem with classical CO₂ eddy flux measurements.

20) Pg 8134 Line 8: *Need to have some citations for high Hg in vehicle exhaust. This is not exactly a straightforward, easy statement to make.*

Response: We have added two citations to support our statements. Civil heating is known to increase GEM concentrations in the air within cities (Fang and Lee, 2004). Won et al. 2007 found that Hg in the exhaust gas of motor vehicles was mostly GEM. Hg²⁺ and HgP was not detectable. At driving mode exhaust GEM concentrations were 3.8–16.8 ng m⁻³ (gasoline), 2.8–8.5 ng m⁻³ (diesel), and 20.0–26.9 ng m⁻³ (LPG). Our statement that vehicles elevate GEM conc. in the urban surface layer is therefore confirmed.”

Suggested changes in the manuscript: “It might be possible that during the exceptionally cold period in Basel gas and oil-fired thermal power stations within the dense urban source area contributed to enhanced GEM concentrations. In urban areas total gaseous Hg concentrations were highest during heating season (Fan and Lee, 2004). Highest GEM levels in Basel were observed during periods of low wind velocities ($u^* < 0.3 \text{ m s}^{-1}$) and southern wind directions (Fig. S3a). Most likely additional GEM emissions from vehicular traffic along a highly frequented road contributed to observed elevated Hg concentrations during southerlies since GEM concentrations in the exhaust of vehicles in driving mode are elevated and range from 2.8 to 26.9 ng m⁻³ depending on fuel types (Won et al., 2007). The road runs in a north/south direction and is the major source of CO₂ (Lietzke and Vogt, 2013).”

New literature we refer to:

- Agnan, Y., Le Dantec, T., Moore, M., Edwards, G. C., Obrist, D.: New constraints on terrestrial surface-atmosphere fluxes of gaseous elemental mercury using a global database. *Environ. Sci. Technol.*, in preparation.
- Nilsson, M., Sagerfors, J., Buffam, I., Laudon, H., Eriksson, T., Grelle, A., Klemedtsson, L., Weslien, P., Lindroth, A.: Contemporary carbon accumulation in a boreal oligotrophic minerogenic mire - a significant sink after accounting for all C-fluxes. *Glob. Change Biol.* 14, 2317–2332, 2008.
- Peichl, M., Sagerfors, J., Lindroth, A., Buffam, I., Grelle, A., Klemedtsson, L., Laudon, H., Nilsson, M.B.: Energy exchange and water budget partitioning in a boreal minerogenic mire, *J. Geophys. Res.-Biogeosci.* 118, 1–13, 2013
- Won, J.H., Park, J.Y., Lee, T.G.: Mercury emissions from automobiles using gasoline, diesel, and LPG, *Atmos. Environ.* 41, 7547–7552, 2007
- Fang, F., Wang, Q., Li, J.: Urban environmental mercury in Changchun, a metropolitan city in Northeastern China: source, cycle, and fate, *Sci. Total Environ.*, 330, 159–170, 2004.