

Reviewer: 1

I have read this paper again. I am still not convinced of the utility of this method as described, and the authors did not really address all my previous comments. This paper is just not suitable for ESTL as is. The authors should consider their data within a new framework of thinking that GOM concentrations are much higher than expected and the chemical form varies across space and time.

I suggest the authors first address whether these samplers measure GEM or TGM. Given new knowledge that GOM can be much higher than previously realized the authors should assess what these samplers are measuring. The samplers will take up GOM, especially since passive samplers with a lesser reactive surface, have been demonstrated to take up GOM. It is very likely that these samplers take up GOM. However, the configuration with the outer shield may influence sampler uptake please see Huang and Gustin 2015 EST.

Please see the following paper to update the framework of your discussion. GOM can be 25% of Hg in the boundary layer. Gustin, M. S. , Amos, H. A, Huang, J., Miller, M.B., Heidecorn. 2015 Measuring and modeling mercury in the atmosphere: A critical review, Special Issue of Atmospheric Chemistry and Physics. Atmospheric Physics and Chemistry, 15: 5697-2015. doi: 10.5194/acp-15-5697-2015.

Over the last few years, Gustin et al. have become convinced that GOM constitutes a much larger share of atmospheric mercury than previously believed and they try with much missionary zeal to convince the rest of the atmospheric mercury community. A close look at the evidence presented to support this claim reveals that it is rather circumstantial and speculative and far from an established fact. Most of the atmospheric mercury community is quite skeptical as to the merits of this suggestion. While we acknowledge that this is an interesting hypothesis worthy of further examination, an unsubstantiated hypothesis cannot be the basis for rejecting any study that does not conform to it. A number of additional points:

1. The question of whether a method samples GEM only or TGM (= GEM and GOM) applies not merely to our new sampling technique, but to essentially all techniques for gaseous mercury species.
2. The fact that a gaseous mercury species has to diffuse through the tortuous path of the Radiello diffusive barrier in order to reach the sorbent of our sampler, suggests strongly that GOM will not be sampled.
3. Ongoing deployment of our sampler in Alert, Nunavut, where atmospheric mercury depletion events, characterized by oxidation of GEM, are regularly observed will help us to establish whether GOM is sampled or not.
4. Our sampler will be useful irrespective of whether we can confirm that it samples GEM only or if it turns out to be a TGM sampler.

The word level though out should be replaced with concentration or value. This still has not been totally done. We can make that change if it is deemed important.

I think that long deployment times are not very useful. The shorter the deployment time the more useful the data for both rural and contaminated areas.

This is a useless comment because it is completely substantiated. It is, of course, also completely wrong. The literature is full of countless examples of applications of passive air samplers recording concentrations averaged over time periods of weeks to years. We have ourselves published more than 20 papers relying on such long sampler deployments. Limited and/or costly site accessibility is one of the key reasons for the need for samplers that can integrate concentrations over long deployment periods. In any case, one of the strength of our sampler is that it is suitable for both, short (i.e. a few days at background concentrations, much shorter at elevated levels) AND long deployments (i.e. at least one year).

It would be useful to point out that we do not know whether these samplers are collecting GEM or TGM.

The paper explicitly states this.

However, activated charcoal would certainly be collecting GOM.

Yes, if it reaches that charcoal. No, if it is prevented from doing so, because it “sticks” to the surface of the pores within the Radiello diffusive barrier.

You need to consider the uncertainties associated with passive sampling methods and GOM (reference above) and Aspino K, Gauchard P-A, Steffen A, Temme C, Berg T, Bahlmann E, et al. Measurements of atmospheric mercury species during an international study of mercury depletion events at Ny-Ålesund, Svalbard, spring 2003. How reproducible are our present methods? *Atmospheric Environment* 2005; 39: 7607-7619.

Ebinghaus R, Jennings SG, Schroeder WH, Berg T, Donaghy T, Guentzel J, et al. International field intercomparison measurements of atmospheric mercury species at Mace Head, Ireland. *Atmospheric Environment* 1999; 33: 3063-3073.

Costs can be quite high for PS if you include shipping, and analyses. If you are going to discuss costs all the different aspects of the costs should be included. Usually costs are not the focus of a paper. But if you are going to consider costs you must consider all the costs.

We do not discuss the cost of the sampler in this manuscript, so we don't know what this comment is about.

Method detection limit of the Tekran presented will not allow for measurement of any trends in areas with background concentrations. The MDL of this passive method is only useful for making measurements in contaminated areas and not work across background areas. Precision can be good, but the MDL is too high to make this useful across rural areas.

1. The Tekran is used routinely around the world to make measurements of Hg in air at background concentrations, so its MDL is clearly sufficiently low.
2. As we show in the manuscript, our sampler collects Hg above the MDL within a few days of being deployed at background concentrations, so the reviewer's assertion is completely without basis.

In fact you can see from Figure 2 A that the Tekran gives the same general concentration indoors relative to outdoors, but is much more variable outdoors.

The variability is somewhat (certainly not “much”) larger outdoors than indoors, which is what one might expect.

Looking at Figure 2B and 2D adsorbed Hg is much higher outdoors than indoors. This makes sense if you consider that this sampler would measure GOM and the indoor GOM would be much lower. In fact concentrations are almost 4 times higher and this is not demonstrated by the Tekran (how was this instrument calibrated??)

The uptake curve for the samplers deployed outdoors reaches levels that are indeed approx. 4 times higher than the uptake curve for the samplers deployed indoors. The reason is that the deployment period was four times longer (1 year instead of 3 months)! It is necessary to look at the time-axis when interpreting plotted data.

It will be very interesting to see what the samplers measure during a Hg depletion event when GOM can be very high. The authors might want to get their field data and combine it with these and adapt their thinking. This would produce a more realistic result and a better paper.

Line 148 should be was.... Concentration of Tekran “GEM” has too many significant figures

How did they calibrate their Tekran? What is the MDL in ng/m³ of their passive sampler? They do not know because we have not been accurately measuring air concentrations outdoors. It is likely the passive sampler is better for measurement of TGM and is measuring RM.

It is bit difficult to respond to these stream-of-consciousness ramblings, partly because they are repetitive of what was stated earlier, partly because they reveal a blatant misunderstanding of passive sampling concepts. For example, expressing a MDL for a passive sampling method

in units of ng/m^3 is nonsensical, as it is a function of deployment length. We write in the manuscript: "Assuming a background GEM concentration of $1.5 \text{ ng}\cdot\text{m}^{-3}$ and using the calibrated SRs listed below this MDL translates to a minimum deployment time for the PAS of approximately two days." Of course, if you sample longer than 2 days the MDL is smaller than $1.5 \text{ ng}/\text{m}^3$, if you sample less than 2 days it is higher than $1.5 \text{ ng}/\text{m}^3$.

The configuration of the Tekran inlet needs to be described, otherwise we do not know what they were measuring and the effect of the inlet configuration relative to GEM has not been adequately tested (Gustin et al., 2015)

So, Gustin et al. now expect that no more papers are published that rely on a Tekran to measure GEM?

Reviewer: 3

In my initial review of this manuscript, I stated the following as one of my primary concerns: Hopefully, though, more of these passive collectors will get deployed at sites with Tekrans so that further calibrations can be done [for calculating sampling rates (SRs)]. This is exactly what is described in the "Looking Forward" section. In fact, this would be a MUCH stronger manuscript if the authors waited for the following data prior to publishing: "We are currently measuring uptake curves in PASs deployed at a number of sites with ongoing active sampling across a wide range of climactic conditions." These data would give me much more confidence in the range of SR that can be expected in the field. SRs are essential for calculating GEM concentrations at each site, and if they vary from site to site, then these passive samplers are going to be difficult to use. However, if further testing shows that the SRs calculated from field data don't vary greatly, or differ from the ones empirically determined, then all should be fine. However, just having one field site might make the publication of this manuscript a bit premature.

The response from the authors was this:

We do not necessarily disagree with the sentiment expressed by the reviewer. In fact, we have considered waiting with submission of a manuscript on this sampler until the results of the field tests are available. Eventually, we decided against such a course of action based on the following reasoning:

1. Our field sampling described in the "Looking Forward" section is very ambitious in scope and involves the determination of uptake over year-long deployments at around 20 sites with ongoing TEKRAN sampling around the world. This implies that it will be more than one and a half years before those data become available. (Note that while a field evaluation over a short time period at one location would obviously be possible, the results of such an effort would be largely meaningless, because the performance of the sampler needs to be evaluated under a wide variety of environmental conditions (e.g. wind, temperature, relative humidity, presence of interfering compounds) and mercury concentration (absolute concentrations, speciation.)
2. We feel that we cannot wait that long with a publication containing a first description of our sampler, as (i) we have presented the sampler design at several conferences (SETAC Vancouver, ICMGP Jeju), (ii) the sampler design is simple and easy to emulate, and (iii) we have shown that it works better than any of the sampler designs previously proposed. The danger of being "scooped" is simply too high.

The fear of being "scooped" is not even remotely a valid reason for pushing through the publication of this manuscript, so I'm just going to ignore their second line of reasoning.

Maybe this reviewer is one of the lucky scientists who has not had unfortunate experiences and thus can afford to ignore such a line of reasoning. Some of us have been burned because we waited to write a "better", more complete paper only to see a shoddy imitation to be published first (in ES&T, no less).

Instead, I would like to concentration on their own admission that more testing is required to fully understand the sampling rates (SRs). The passive collectors themselves only provide a quantity of Hg that they absorb over the period of time that they were deployed. To turn that quantity of Hg into anything meaningful requires that the users know the rate at which the collectors were sampling the air (i.e., the SRs). As shown in this preliminary study, these SRs differ between indoors and outdoors, and between the beginning of deployment and later (Figure 2). This difference may "only be ~23%" as stated by the authors, but to me that is fairly significant.

The key phrase here is "to me that is fairly significant". The reviewer is plainly wrong in this belief:

1. The conditions of deployment were extremely different in the two experiments (indoors, no housing, stable temperature vs. outdoors in an extremely seasonal climate, with housing). Nevertheless, the difference was only 23 %. If the reviewer believes this to be fairly significant, (s)he is clearly not familiar with passive air sampling methodology.
2. We can explain most of that discrepancy. A result of identical sampling rates in the two deployments would have been problematic, because it is well-known that molecular diffusion of GEM in air is temperature dependent and we therefore know that diffusion in the Canadian outdoors has to be different from diffusion indoors.

The authors also state that the sampler needs be evaluated under a wide variety of environmental conditions (e.g., wind, temperature, relative humidity, presence of interfering compounds) and mercury concentration (absolute concentrations, speciation). With this in mind, one application of these passive collectors that I think would be really interesting is their deployment out in the open and under nearby forest canopies to examine foliar uptake of GEM (and possibly GOM). Under this simple scenario, wind, temperature, relative humidity and presence of interfering compounds (e.g., volatile organic compounds) would all come into play.

We appreciate the suggestion, but it is not helpful in improving the paper. Incidentally, the conditions of the two deployments the reviewer proposes (forest vs. clearing) are far more similar than the conditions of the two deployments we compared (indoors vs. outdoors).

Furthermore, most of these parameters are not incorporated into Equation 3, which is supposed to calculate theoretical SRs.

Yes, they are, although that may not be readily apparent:

- temperature affects the diffusivity D of mercury in air
- wind speed affects the thickness of the air boundary layer ($r_a - r_{dout}$)
- water (i.e. relative humidity) and other interfering compounds would affect the sorption coefficient of GEM onto the carbon. Equation 3 is based on the assumption that uptake is air-side controlled, which implies that the resistance on the sorbent side is not rate-limiting. In that case, the sorption coefficient plays no role in determining the uptake rate. We are well equipped to quantitatively address the estimation of sampling rates if it turns out that there is a resistance to uptake on the sorbent side. We were the first to have proven the existence such a resistance in passive air samplers for organic compounds (Zhang, X., M. Tsurukawa, T. Nakano, Y. D. Lei, F. Wania. Sampling medium side resistance to uptake of semi-volatile organic compounds in passive air samplers. *Environ. Sci. Technol.* 2011, 45, 10509-10515) and we have developed numerical approaches to describe it (Zhang, X.M., F. Wania. Modeling the uptake of semi-volatile organic compounds by passive air samplers: Importance of mass transfer processes within the porous sampling media. *Environ. Sci. Technol.* 2012, 46, 9563–9570 and Armitage, J.M., S.J. Hayward, F. Wania. Modeling the uptake of neutral organic chemicals on XAD passive air samplers under variable temperatures, external wind speeds and ambient air concentrations (PAS-SIM). *Environ. Sci. Technol.* 2013, 47, 13546-13554).

I feel bad because the lead author is a graduate student and publications are important. But this study is just not ready for publication, especially in ES&T Letters. If they can tone down their convictions, and clearly state that there is a lot more work that needs to be done, maybe this study would be better suited for a less impactful journal. However, without a better understanding of the full variation possible in SRs, the collectors themselves are actually pretty useless.

This last statement again betrays a lack of understanding of passive air sampling practice. Let us hypothetically assume that there is unexplainable variability in SRs during different outdoor

deployments that is double the 23 % we observed between indoor and outdoor deployments (which is highly unlikely). We could clearly not discriminate concentrations differences between true background sites. However, we could distinguish reliably places with concentrations at background levels (1.5 ng/m^3) from those that have only slightly elevated levels ($>3 \text{ ng/m}^3$, as are for example prevalent throughout most of China!). We could likely even detect the occurrence of ADMEs. And we could certainly use it to map the concentration variability around sources of GEM to the atmosphere, which result in much higher concentration gradients. This would be useful for source characterization and quantification, as well as compliance monitoring. The reviewer really has not given this much thought, if (s)he suggest that such a sampler would be “pretty useless”.

The sampling rate of the commonly employed PUF disk-double bowl passive air samplers for organic compounds easily has an uncertainty that is on the order of a factor of 3 to 4, but judging from the number of publications based on studies using it, which is easily more than a 100, it somehow still manages to be pretty useful to a lot of folks.

Reviewer: 4

I have re-examined the revised version by McLagan et al. as well as the rebuttal letter (upon request of ES&TL). I find that the authors have adequately addressed the reviewer's concerns; the rebuttal letter also sheds more light on the context, i.e. the difficulty in developing a precise passive Hg sampler and its potential application in light of the Minamata convention, which will involve a lot of monitoring. I recommend the revised version for publication in ES&T Letters.