REVIEWER 1

4. Another explanation for the absence of a correlation between local Hg and thrush-Hg in this dataset is that Hg is often plotted logarithmically. So a change in an order of magnitude in atmospheric deposition may be needed before any observable difference in fauna Hg concentrations.

Transforming two variables (e.g., via a log-transformation as suggested here) will not change the estimated coefficient of correlation.

5. Pg 11, Line 8+: The argument for why Bicknell's thrush makes a good bioindicator is tenuous. It seems the rest of the paper makes the opposite argument. Either provide a stronger argument for why Bicknell's thrush are a good bioindicator and should continue to be used when they are not able to show local Hg deposition, or remove. It may be that Bicknell's can provide part of the story for montane Hg cycling, but they shouldn't be the key fixture as suggested. A fauna with a strong correlation to local Hg would be a better bioindicator (maybe invertebrates with a small home/life range?).

This is a good point, and one that you and I had discussed. I personally don’t see any reason to think BITH is a good bioindicator.

REVIEWER 2

There is not much to criticize, but in my opinion the authors have not fully informed on inter-annual trends. I like their suggestion of sampling blood after July 1, to "avoid the confounding influence of seasonal carry-over effects." How about performing a statistical analysis to test for inter-annual variation in mercury concentrations in samples collected after July 1? Likewise, an analysis of inter-annual variation in early Julian dates (e.g., before Julian data 175) may help to understand whether mercury exposure is changing in their tropical habitats.  
  
My philosophical answer to this is that pursuing additional tests after a finding of no significance is poor practice. This is what leads to spurious findings: on failing to find a relationship, we subdivide the data in different ways until we get a result is statistically significant and interesting. The problem is that these results are often meaningless. By “trying again” with a slightly different data set, we have introduced our own bias into the analysis and increased the likelihood of finding a significant result by chance alone. Imagine the worst-case scenario: we found no relationship between blood mercury and atmospheric mercury, so we decide to look at only blood drawn after July 1. Again, imagine we find no relationship, so we re-analyze with only blood drawn after July 2, and we repeat this process until the analysis yields a significant result. The problem here is that we can expect, by chance alone (i.e., commit a Type 1 error), to “discover” a significant relationship between two variables if we simply conduct enough tests.

The bottom line is that this kind of data-dredging does not yield reliable inference about the nature of the world, and should be avoided at all costs.

If we want to pursue this question, it should ideally be done using an independent data set. Re-analyzing a subset of the data is just not a good idea.

Reviewer 3  
  
3. I admire the authors efforts to try to explain the seasonal patterns in blood mercury over the long-term, but this is a high bar. Have they considered trying to obtain some weighted annual value of blood mercury and compared these against annual deposition and other more long time scale drivers. Just a thought. It might be a lower bar.

I do not understand what the reviewers is suggesting.