Review ACP Custodio et al

Overall, it is not clear to me the purpose of this paper and how it is providing new information. The datasets presented are rich and have the potential to provide interesting new results, but as of now there is little in the paper that is new. A lot of previous work related to what is presented is not discussed. I elaborate on the main issues first and then present line-by-line comments.

Major and General Comments

Discussion of the Tekran instrumentation and uncertainty should be improved, related to whether TGM or GEM is being measured.

Line 134: "usually" – which of these stations' instruments have this filter operating and which do not?

Line 137: "probably" – what does this mean quantitatively? Slemr et al. (2016) describes the CARIBIC aircraft measurements in the upper troposphere/lower stratosphere which are specifically in a low-humidity but high ozone environment. How do the conclusions regarding GOM compound collection from Slemr et al. (2016) apply to the surface stations, especially coastal ones, in this study?

Line 139-140: "soda-lime filter.. is known to capture GOM" – please add citations here. Gustin et al. (2021) state: "The Global Mercury Observation System standard operating procedure states that a soda lime trap in front of the Tekran 2537 removes GOM, though this has not been adequately tested."

In particular, it is not actually known whether >95% of TGM is GEM outside of polar depletion events (the discussion at the end of the sampling sites part of Section 2, lines 143-146). Please add a reference to at least one of the papers and reviews discussing the major uncertainties measuring gaseous oxidized mercury (e.g., Marusczak et al., 2017; Lyman et al., 2020; Gustin et al., 2015, etc.) as it relates to uncertainties in the operationally-defined TGM Tekran measurements. I find that Lyman et al. 2020 summarize it better than I could myself and quote from them here: "it remains unclear whether Tekran 2537 and similar mercury analyzers without upstream processing equipment measure total gas-phase mercury or only elemental mercury. Thus, these measurements are also operationally defined. We use the acronym GEM (gas-phase elemental mercury) to describe them, though they likely include GEM and some portion of Hg^{II} compounds that exist in the atmosphere, and the amount of Hg^{II} they include likely depends on the sampling configuration and the chemical and physical conditions of the atmosphere. Also, when KCl denuders are used upstream of elemental mercury analyzers, some of the captured Hg^{II} is reduced to elemental mercury and measured in that form (Lyman et al., 2010). Some have used the acronym TGM (total gaseous mercury), rather than GEM, but, to our understanding, no information exists about the percentage of gas-phase Hg^{II} that is analyzed by elemental mercury analyzers with an upsteam KCl denuder or without upstream sample processing. Many have asserted that this issue is inconsequential, because atmospheric Hg^{II} concentrations are low

relative to Hg^0 (Ci et al., 2011; Fu et al., 2012a), but this assertion has been shown to be inaccurate in some environments (Fu et al., 2015; Obrist et al., 2011; Swartzendruber et al., 2006; Weiss-Penzias et al., 2009), especially when the low bias in KCl-denuder based GOM measurements is taken into account (Huang et al., 2013; Lyman et al., 2016)."

Overall, I am confused when the methods state that it is concluded only GEM is being measured but then all data is treated as TGM. Should it be GEM and not TGM? Interpreting the observations as GEM is consistent with other studies (e.g., Travnikov et al., 2017) but a more thorough discussion of the caveats is warranted.

Clarify and provide more information on the methods used. While I understand not wanting to repeat too much from previous work, as the sub-sections "Back-Trajectory Analysis" and "Concentration-weighted trajectories (CWT) and probability mass function (PMF) models" are written now the reader cannot understand what is being done. More information must be provided so that what is in the current paper becomes sufficient, and only the reader who wants to learn more specific details needs to look to Custodio et al. (2020).
Back Trajectory analysis: I found this section very confusing. Perhaps a schematic (potentially in the supporting information) could be added showing each step of this method. Please provide more information on when the trajectories are initiated. Is it every 12 hours for all days and years of data or a single year of data? If it is a single year, which specific year is it? Please explain and justify all choices made for the HYSPLIT simulations carefully including the arrival times of midnight and 12pm (is this local time or UTC?) and arrival heights of 50 and 500m.
Combining the concentration-weighted trajectories for all 5 sites, rather than analyzing them on a site-by-site basis, seems like it waters down the results and adds uncertainty especially as the number of sites with data is different for different years.
Concentration-weighted trajectories are described in the "Back-Trajectory Analysis" section, yet they are in the title of the following section.
Line 169: I am much more familiar with PMF standing for the EPA positive matrix factorization method. In Custodio et al. (2020), PMF is used as an abbreviation for positive matrix factorization, not "probability mass function".
CWT and PMF section: these two paragraphs are quite confusing with some very long sentences that are difficult to follow.
The methods in Section 3.2 are also in need of additional explanation, including "Kernel-regression", "LSQF" (including defining this acronym in the main text), how the annual cycle was removed from Figure 3, and how confidence intervals were calculated (or what the trend ± values actually mean: are the ± values the standard deviation, standard error, 95% confidence interval from bootstrapping?). It is not clear from the description here or in Figure 3 that the regression is being performed on monthly medians. The standard deviation or confidence intervals reported for trends in this paper are much narrower than in previous studies cited (e.g., Cole et al., 2013). Why is this, and how do the methods in calculating the confidence intervals in Cole et al. (2013) differ from this study?

Improve statistical analysis of the observations and distinguish what is new in this paper, including relative to Custodio et al. (2020).

In the opening section of the results (Lines 194 – 215), please provide quantitative metrics of the skewness and kurtosis of the distributions, and differences between sites or average vs. median rather than qualitative descriptions. Can you quantitatively define pollution events, such that it is clear that "Events with local and regional pollution are also missing at Andøya"? I also am confused about how Weigelt et al. (2013) is a reference for "frequent events with local and regional pollution" at Waldhof. As far as I can tell they mention a single time that a local combustion plume was observed and overall put Waldhof on par with other rural background sites. Again, a quantitative definition of pollution events would be helpful here.

The 3.1 seasonal variation section does not seem to add any new information as seasonal patterns at these sites have been published in other studies previously. What would be new is if you could add more detailed and rigorous comparison of the seasonality between sites (they are "similar") or from the longer time record if the seasonal cycle has been changing over time at any individual site. An example of a more detailed analysis of observed seasonality at Arctic sites including statistical tests of the significance of differences between seasons and between sites' seasonality is presented in Angot et al. (2016). The amplitude of the annual cycle at each site is only referenced in the caption of Figure 3. This should be in the main text along with an explanation of its calculation. In section 3.2, the comparison of the trend at each site to Mace Head because it is the "baseline trend" is not explained. Table 1 is confusing as it currently stands, please add more explanation in the text and in its caption. Please also explain why and how the different time periods of trends are calculated for Mace Head. Similarly in this section, the comparison of trends between sites are described qualitatively (e.g., "similar", "somewhat higher", "substantially higher") rather than quantitatively.
The paper would be stronger if the PMF analysis was performed at other sites besides Mace Head, where it has been done before for a subset of the time period presented.

The conclusions drawn in the paper do not seem to be supported by the analysis.
The discussion in Lines 281-288 is speculative. I do not have enough information to evaluate this argument.

- While the back-trajectory analysis may show transport from Greenland, how do we know from observations that the atmospheric mercury there was elevated?
- I would need to see a timeseries of the AO and NAO plotted against the mercury concentrations in order to evaluate how extreme the AO was for those 2 particular years and its relationship to mercury concentrations over the full time period.
- Hawkings et al. (2021) is reporting measurements from 2018 as far as I can tell. How is this related to 2010 and 2013?

The discussion in lines 316 – 334 referencing specific policy actions and their relationship to observed trends is lacking evidence. Some specifics:

1) Minamata convention signing in 2013: The assumption that policies were already implemented at this point before the treaty went into force, and that this would be observable at Mace Head, ignores previous work done showing the difficulty in linking observations to the Minamata Convention and other policies (e.g., Selin, 2014; Giang et al., 2018). Whether emissions may have already changed at the timing of the treaty

signing could be investigated directly by examining emissions inventories over this time period (e.g., Streets et al., 2019 global inventory for 2010-2015; GMA2018 global inventory for year 2015 (AMAP/UNEP 2019); Leclerc et al., 2019 – 2000 – 2014 for the EU;)

2) air pollution policy in the United States: I assume this is referencing what eventually became the MATS rule (https://www.epa.gov/mats) which was tied up in the courts for a number of years (and potentially still is as far as I can tell). Zhang et al. (2016), while the inventory itself is only presented until 2010, did analyze changes in Hg emissions from US coal-fired power plants from 2005 to 2015 (see Figure 1) and attributed a significant portion of the decline to mercury-specific control technologies resulting from the planned MATS rule coming into force. However, after 2016, when the MATS rule became tied up in the courts, I would not have expected coal-fired utilities to continue operating expensive mercury-specific controls if they did not need to. It would be possible to investigate this by applying similar techniques of Zhang et al. (2016) to later years of data for US power plants.
3) The 2007 policy banning mercury in non-electrical measuring devices in the EU could be quantified based on existing inventories (e.g., Horowitz et al., 2014; Streets et al., 2017, 2019; GMA from AMAP/UNEP 2013 or AMAP/UNEP 2019 reports) rather than speculated. From these studies, this sector is not a large portion of the mercury emissions and thus I doubt it would be able to be observable in atmospheric trends.

The paper discusses downward trends and states them as such in the conclusions in lines 426 – 427, but at four out of the six sites the trends shown in percentage form are not statistically significant (confidence interval includes 0% yr^{-1}). In Table 1 the trends are all statistically significant – how is this the case? As I mentioned earlier, the standard deviations(?) or confidence intervals of the trends shown in Table 1 are much narrower than in previous studies like Cole et al. (2013). Much past work has discussed how Arctic sites do not have a decreasing trend unlike mid-latitude sites. Skov et al. (2017) performed a more detailed statistical analysis at Villum from 1999 – 2017, finding no significant annual trend in GEM, and their comparison with a model allowed for more convincing attribution of emissions to the mercury concentrations observed.

The “seemingly non-monotonic downward trends with inter-annual ups and downs are not well-explained” also ignores previously published work. The attribution of trends to anthropogenic emissions is not convincing. Other potential mechanisms are not discussed and natural sources and sinks are referenced only briefly. A detailed review of observed trends and potential mechanisms including anthropogenic emissions, natural emissions/re-emissions, sinks, and oxidative capacity is also provided in Lyman et al. (2020) and references therein. Declines in anthropogenic emissions are referencing studies who only estimated anthropogenic emissions until 2008 (Streets et al., 2011) or 2010 (Horowitz et al., 2014; Zhang et al., 2016). There is no reference to a more recent study (Streets et al., 2019) that estimated mercury emissions as well as modeled their impact on concentrations from 2010 to 2015, including discussions on a regional basis. Another recent study that is not referenced here (Wang and Mao, 2021) specifically examined how trends in emissions from the EDGARv4.tox2 inventory

(1970 through year 2012; Muntean et al., 2018) and the location of Arctic fronts interacted to affect transport of emissions to the Arctic. Finally, interannual variability in Arctic atmospheric mercury from 1979 to 2008 was also investigated in Fisher et al. (2012) and was related to specific environmental and climatic factors.

Figures are not adequately described prior to when interpretations based off them are given.

- Figure 2: interpretation is given in Lines 218 – 220, but it is not until Line 237-238 that it is explained what is specifically in the Figure. This explanation is also not linked to the methods section where more detail is given on the density map. However, there is still not enough information – for example, what is the size of the grid cells in Figure 2? how was this chosen?
- Figure 4: Interpretations of Figure 4 are given in Lines 282, 297-299, 302-306, and 312-313, but the figure is introduced / described in Lines 352-353. Please move the description of Figure 4 from Lines 352-353 before any interpretations of the Figure are given.

<u>Line-by-line comments</u>

line 31: is there only a single probabilistic source contribution factor?

Line 35: Is this superscript 1 a citation or footnote? Please clarify whether these are results you have found or you are citing other esults.

Line 38: I would appreciate if the “other atmospheric trace gases” are listed here or at least a few examples are given.

Lines 37-39: This sentence is confusingly worded. Please clarify and potentially split into multiple shorter sentences.

Line 40: Instead of “accessed” I believe you mean to say “assessed”

Line 42, Line 68: I similarly would like to know what the long-lived anthropogenic species are, or at least a few examples.

Line 71: In this discussion of regional and hemispheric trends, the mention of global trends doesn’t seem to follow. Is there a reference that hypothesizes or inventory that estimates that global anthropogenic mercury emissions are decreasing? Perhaps this should be replaced by “hemispheric” instead.

Lines 72-73: “still not unknown sources” is unclear. I thought it would make more sense to say “still unknown sources”, but how could one do a time series analysis of unknown sources? Overall this sentence is confusing.

Line 80: “On this raw” – I do not know what this means.

Line 81: “baseline factor” – the meaning of this is not explained. Perhaps this should not be mentioned in the introduction, but explained instead in the results and also methods. Otherwise, please explain what this means in a way that does not require reading the methods first.

Line 84-85: I would add “observed” in front of “atmospheric mercury temporal variability” to distinguish it from models that you are trying to constrain. This sentence is also a bit vague – what kind of temporal variability (seasonal? interannual? or other analysis) and what kind of cycling predictions by the model? (e.g., concentrations, deposition?). I have not again seen a reference to constraining models outside of the introduction.

Section 2 – Sampling Sites: A map with the stations labeled would be helpful. Although the latitude and longitude are given in individual paragraphs, having this information in a table or on the map would be helpful so it is all in one place. Sometimes the altitude of the stations is not given. Similarly the altitudes could be included in a table or map label.

Line 96-97: At this point, we do not have the latitudes of the stations and it would be helpful to have seen these (or the locations on a map) for the comparison of Andøya and Amderma. Is it known from previous analyses or your own analysis that the mercury concentrations at Andøya are more similar to those observed at other mid-latitude sites? (Angot et al., 2016 found this, for example, in their analysis of seasonality of Arctic sites). It seems your own analysis from the seasonality in your Figure 2 also supports this, but at this point in the paper the classification of Arctic vs. Midlatitude sites is not well-motivated. Explain why this classification will be useful.

Line 109, 115, 120: Altitudes are not given for the Waldhof, Villum, or Andøya sites.

Line 120-122: To me, it does not add anything to mention ALOMAR here.

Line 130: At the end of every paragraph introducing the measurement sites, I was wondering what instrument they were operating. It would be helpful to give the reader a heads up (perhaps at the end of the first paragraph) that all sites operated Tekran instruments and more detail will be given at the end of the section.

Line 152: write out the acronym for HYSPLIT the first time it is used.

153: make clear that Stein et al. (2015) is the reference for HYSPLIT model. As it is written now it sounds like Stein et al. (2015) describes the method of choosing the trajectory heights and time initiations, but these are choices made by the authors here which must be explained.

Line 171-173: this sentence should be rewritten and split into multiple sentences for clarity. Right now it seems like some words are missing for it to make sense.

Lines 179-181: this is also a very long and hard to understand sentence that should be split up for clarity.

Lines 188 – 192: rather than list the number of days of data separately from the station names, include this information in a table. The number of months are listed in Table 1. What information does the number of days add? is the data averaged daily? Are daily averages what is summarized in Figure 1?

Figure 1. please state in the caption the temporal frequency of the measurements presented (e.g., daily averages?).

Line 195 – 199: This sentence is also very long and should be split into multiple sentences for clarity. Also, what is the detection limit of the Tekran instruments at these sites? Are minima of 0.0 vs. 0.1 ng m^{-3} real and different? It would be more useful to know the number of depletion events in each year rather than just the descriptor "frequently". Referencing studies of depletion events in and around these particular sites should happen sooner in this paragraph. A detailed analysis of changes in the frequency and timing of depletion events as well as comparison with previous studies is presented in Angot et al. (2016), referencing in addition to the studies already cited here, Berg et al. (2013) and Chen et al. (2015).

Line 202 – 203: I am confused how the altitude of the Zeppelin station is relevant here as the area around the station and higher-altitude areas near it can also be snow-covered.

Line 261: What does "regional emissions" mean? Is this anthropogenic emissions? Natural/legacy re-emissions from land?

Line 279-280: I do not understand the meaning of this sentence and how it relates to the previous sentence. Changing the discussion from qualitative to quantitative might help.

Line 289-291: Please add a citation for this sentence.

Line 298-299: Please show and cite specific anthropogenic emission estimates for this region, otherwise this sounds like speculation.

Line 300: Bringing up uncertainties in the HYSPLIT modeling approach is critical and there is not enough explanation of what you mean here. I do not understand the shift to the South that is being described.

Line 301-302: Please add a specific citation showing the "geolocation of TGM sources in Europe." A figure showing the locations of major point sources or gridded anthropogenic Hg emissions inventory for the region would be helpful.

Line 303-304: The explanation given in Lines 281-288 seems to conflict with what is shown and described in Figure 4 here. The earlier explanation seems to be specific to 2013 and potentially

2010-2014, but similarly high levels are shown in Figure 4 over Canada and Greenland for 1997-2000 and 2005. Please clarify.

Line 307: This is the only time sinks are mentioned as affecting the concentration of mercury in the atmosphere. This should be expanded in more detail.

Line 311: “This phenomenon can be explained only by reductions in global atmospheric mercury sources” : I am not yet convinced. Evidence must be presented to back up this claim. More discussion of specific anthropogenic emissions inventories and estimates of emissions from natural sources, legacy re-emissions, as well as sinks, chemistry, etc. is needed.

Lines 310-311 and 312-313: A quantitative assessment of the trends in the concentration-weighted trajectory would be more helpful than qualitatively referring to the colors on the figure.

Lines 314-315: The trend at Villum seems very different from the that at Zeppelin. What are they affected by if not anthropogenic emissions from continental Europe? Please reference here prior work that has been done in this space (e.g., Fisher et al., 2012; Skov et al., 2017; Angot et al., 2016; Wang and Mao, 2021).

Line 321-322: Landfills are not a significant source of mercury. Horowitz et al. (2014) reviewed measurements of emissions of mercury from landfills and found they can be treated as a long-term sink of mercury on centuries-long timescales. as their emissions of mercury are so small the lifetime is on the order of 20,000 years.

Line 341-342: This may be true, but there have been recent studies on other natural processes, e.g., Hg uptake to land (Zhou et al., 2021; Jiskra et al., 2018) driving seasonality and trends.

Lines 348-350: Where is this shown?

Lines 357-358: How do you know that the North Atlantic a sink of anthropogenic pollutants? The North Atlantic is a hot spot of net evasion in Soerensen et al. (2010) and Zhang et al. (2019).

Line 359-361: This hypothesis seems like it would be testable by performing chemical transport modeling experiments.

Lines 364-365: I would like to see analysis separating out these two processes, transport and the concentration in transported air. This might be similar to what was done in Wang and Mao (2021).

Line 368: The discussion of factors is a surprise. Please introduce the PMF analysis here before going into the factors. It might need to be a separate section.

Lines 369-373: I don't understand how the trends in a factor from 2013 – 2018 can be explained by declines in emissions that were estimated only until year 2008 (Streets et al., 2011) and year 2010 (Zhang et al., 2016). Please explain.

Line 376: there is not enough information on the baseline factor given.

Line 377: What anthropogenic species?

Line 379: This is an extremely high correlation coefficient. I need to know more about what went into this factor and what we can learn from it if it is explaining 94% (r^2) of the variance in observed concentrations at Mace Head.

Line 380-381: More explanation is needed as to how Figure 4 has to do with the anthropogenic species factor.

Lines 383 – 397: I am very confused by this section. Explaining the baseline factor in detail would help with that.

Line 399-400: A different reference for the residence time of mercury in the ocean would be more appropriate (e.g., Amos et al., 2014?)

Lines 400-402: Additional references about the oceanic mercury influence on atmospheric mercury due to anthropogenic activity should be cited here like Soerensen et al. (2012); Sunderland and Mason, 2007; Sonke et al. (2018); Cossa et al. (2018); etc.

Lines 403-305: I don't think this adds much to the paper, unless you add references to studies that have examined some of these processes for mercury specifically.

Lines 410-411: This conundrum has been explained by other studies. The analysis presented here does not include any estimates of anthropogenic emissions to corroborate that the current study also shows this.

Line 426-427: I think there is a typo, is the trend actually 4 ±16 % per year or should it be 4 ±1.6% per year? Also, typo Waldholf -> Waldhof

Line 421-422: This is not appropriately caveated given other studies on how difficult it is to observe the impact of the Minamata Convention and other policies (e.g., Selin, 2014; Giang et al., 2018).

Additional references cited in this review, not already cited by Custodio et al. (2021):
Gustin et al. (2021): https://www.mdpi.com/2073-4433/12/1/73

Lyman et al. 2020:
https://www.sciencedirect.com/science/article/pii/S0048969719355706#bb1685

Marusczak et al. 2017: https://pubs.acs.org/doi/10.1021/acs.est.6b04999

Gustin et al. 2015: https://acp.copernicus.org/articles/15/5697/2015/

Travnikov et al., 2017: https://acp.copernicus.org/articles/17/5271/2017/

Angot et al. (2016): https://acp.copernicus.org/articles/16/10735/2016/

Selin (2014): https://setac.onlinelibrary.wiley.com/doi/full/10.1002/etc.2374

Giang et al. (2018): https://pubs.rsc.org/en/content/articlehtml/2018/em/c8em00268a

Streets et al. (2019): https://www.sciencedirect.com/science/article/pii/S1352231018308884?via%3Dihub

Leclerc et al. (2019): https://www.sciencedirect.com/science/article/pii/S0160412018330101

AMAP./UNEP 2013: https://www.amap.no/documents/download/1265/inline

Streets et al. (2017): https://pubs.acs.org/doi/abs/10.1021/acs.est.7b00451

Wang and Mao (2021): https://www.sciencedirect.com/science/article/pii/S1352231020307603?via%3Dihub

Muntean et al. (2018): https://www.sciencedirect.com/science/article/pii/S1352231018302425

Fisher et al. (2012): https://www.nature.com/articles/ngeo1478

Berg et al. (2013): https://acp.copernicus.org/articles/13/6575/2013/

Chen et al. (2015) https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2015GL064051

Soerensen et al. (2010): https://pubs.acs.org/doi/10.1021/es102032g

Zhang et al (2019): https://pubs.acs.org/doi/10.1021/acs.est.8b06205

Amos et al. (2014): https://pubs.acs.org/doi/abs/10.1021/es502134t

Soerensen et al. (2012): https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2012GL053736

Sunderland and Mason, 2007: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2006GB002876

Sonke et al. (2018): https://www.pnas.org/content/115/50/E11586

Cossa et al. (2018): https://bg.copernicus.org/articles/15/2309/2018/

Zhou et al. (2021): https://www.nature.com/articles/s43017-021-00146-y?proof=t+target%3D

Jiskra et al. (2018): https://www.nature.com/articles/s41561-018-0078-8