Authors' Response to Reviewers

By Irene Cheng, Amanda Cole, Leiming Zhang, and Alexandra Steffen (Environment and Climate Change Canada)

Reviewer 1 (Danilo Custódio)

We appreciate the comments from this reviewer and have provided responses to all comments and improved on the manuscript.

The use of unsupervised methods, such as Positive Matrix Factorization (PMF), in the source apportionment of Total Gaseous Mercury (TGM) has proven to be highly insightful. PMF provides a data-driven, top-down approach to estimate mercury fluxes by breaking down the observed concentrations into potential sources and their relative contributions. The strength of PMF lies in its ability to reveal hidden patterns in large datasets without predefined assumptions about the sources. This makes it a powerful tool for identifying unknown or unexpected sources of mercury emissions and for quantifying their impacts on the environment.

However, despite its advantages, the PMF model—especially as implemented in the U.S. EPA's PMF software—has some significant limitations. One of the main drawbacks is that it operates as a black box, where users have limited control over the relationships between the variables loaded into the model. While this makes the tool remarkably user-friendly and accessible to non-experts, it can also introduce risks when interpreting the results. Since users have little insight into the inner workings of the model, there is a danger of not fully understanding the factors driving the apportionment. This becomes particularly critical in environmental science, where subtle changes in the data can lead to vastly "different source contributions".

Response: We hear your concerns about the PMF model. For meeting the objectives of this study, we consider the PMF model to be the most appropriate choice for the following reasons. First, the model identifies potential sources and quantifies their contributions to TGM concentrations using in-situ measurements. Second, the TGM source contributions can be analyzed over annual, seasonal, and daily time scales to gain a deeper understanding of source impacts. In this study, we are particularly interested in understanding the emission drivers of long-term TGM trends, which is a fundamental question of the Minamata Convention on Mercury. Third, the PMF model has existed for more than 20 years with hundreds of publications using the model for source apportionment. The principles and fundamental equations are detailed in the PMF user's guide (Norris et al., 2014) and scientific literature. The PMF model ingests a dataset of observations and estimates the species' composition in each factor and the factor contributions with the objective of minimizing the scaled residuals (model-observation differences). The user must derive the optimal number of factors in the final solution by examining the model performance statistics and the physical meaning of the factors. Caution must be taken in the interpretation of the factors. Although PMF is relatively less complex than chemical transport models, there is a lot of effort involved with the measurement data preparation, running the model, analyzing and interpreting the model results, and re-tuning the model runs based on results obtained in order to come up with the optimal

solution. The interpretation of the PMF results is not trivial and should be based on scientific literature and sound knowledge of potential sources of pollution and pollutant tracers.

The main PMF drawback is not the limited control over the relationships between the variables loaded into the model. The relationships between variables can be examined prior to running the model itself. Scatterplots between the loaded variables can be viewed within the PMF interface, which allows users to examine whether there are any bivariate relationships (Norris et al., 2014).

Norris, G., Duvall, R., Brown, S., and Bai, S.: EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide. Office of Research and Development, Washington, DC 20460. https://www.epa.gov/air-research/epa-positive-matrix-factorization-50-fundamentals-and-user-guide, 2014.

Moreover, a key issue with PMF is that it will always produce an output, regardless of the quality or representativeness of the input data. This brings about the risk of "garbage in, garbage out." If the input data is not carefully curated, or if the underlying assumptions about the sources and their relationships are flawed, the model can generate misleading or incorrect apportionments. This is particularly concerning in mercury studies, where TGM concentrations are influenced by various factors, including natural emissions, re-emissions, and anthropogenic activities. The complexity of mercury's behavior in the atmosphere makes it imperative for researchers to critically assess the outputs of PMF, and not blindly trust the results.

Response: We agree that the PMF results are strongly dependent on the input data. In this study, we only used quality controlled measurement data that are published online by their respective monitoring networks. The quality control and quality assurance procedures are available from ECCC Canadian Air and Precipitation Monitoring Network (CAPMoN), ECCC National Air Pollution Surveillance (NAPS) program, ECCC Canadian Greenhouse Gas Measurement program, USEPA Air Quality System (AQS), Interagency Monitoring of Protected Visual Environments (IMPROVE), and ECCC Climate Data. In addition to using data that have been reviewed rigorously, we thoroughly checked our data processing steps for data averaging and merging. From our experience, these errors can lead to erroneous PMF outputs. After correcting errors in the input data and re-running the model, the model output was significantly more reliable. Often, the selection of chemical species is based on the measurements available at the site. We agree that TGM concentrations are influenced by different activities and sources, and this is the most challenging part of understanding atmospheric mercury processes. Therefore, in this study, we included variables that are representative of different emission sources to better derive our conclusions. This ensures the source attribution is not limited to well-established anthropogenic Hg sources (like coal combustion, metal smelting), but also the less studied Hg sources like wildfires, Hg reservoir, and terrestrial and aquatic surface re-emissions which also deserve more attention. We also recognize that chemical transformation is not taken into account in the PMF model. Therefore, we have limited the number of reactive species into the PMF model except for the ones crucial for differentiating local combustion (SO_2) and regional emissions transport (SO_4^{2-}). These chemical species are frequently used in the PMF literature. There is a lot of work that goes into preparing the dataset for PMF modeling. The steps taken in this study ensure the PMF results are robust and consider all types of Hg sources where possible.

Therefore, while PMF serves as a valuable tool for source apportionment, particularly in a top-down framework for estimating mercury fluxes, it must be used with caution. Researchers should be aware of the potential pitfalls and ensure that the input data is thoroughly vetted. Additionally, incorporating other methods to validate PMF results could mitigate the risks of misinterpretation. This reflective and cautious approach will ensure that PMF's insights into TGM source apportionment remain robust and scientifically sound.

Response: See response to the previous comment regarding the input data. Additionally, the PMF results were verified independently using source apportionment results from chemical transport modeling studies (ECCC, 2016; Fraser et al., 2018). PMF and chemical transport models (CTMs) can both produce source apportionment estimates and their methods are very different. The TGM source contributions from the PMF model were consistent with the GRAHM model (ECCC, 2016) and GEM-MACH-Hg model (Fraser et al., 2018) as discussed in section 4 of the paper. Both PMF and CTMs suggest anthropogenic TGM contributions comprised only a few percent of the annual TGM and the largest proportion of the TGM was attributed to long range transport. Wildfire contributions from the PMF model were a few percent higher compared to GEM-MACH-Hg model. However, given the uncertainties in air quality models and the extreme temporal and spatial variability of wildfires, differences in the order of a few percent are reasonable.

Environment and Climate Change Canada (ECCC): Canadian mercury science assessment: report. Gatineau, Quebec. ISBN 978-0-660-04499-6. www.publications.gc.ca/pub?id=9.810484&sl=0, 2016.

Fraser, A., Dastoor, A., and Ryjkov, A.: How important is biomass burning in Canada to mercury contamination?, Atmos. Chem. Phys., 18(10), 7263-7286, 2018.

In the source apportionment presented by the authors, several systematic issues have been identified, ranging from variable selection to the model run setup. One notable concern is the inclusion of temperature as a variable alongside atmospheric tracers. While temperature is a fundamental environmental parameter, its use in this context introduces a significant risk of spurious correlations due to its high amplitude variations throughout the year. These fluctuations can heavily influence the eigenvector decomposition in the rotational factorization, constraining the results based on seasonal temperature trends rather than genuine emission sources.

The problem arises when temperature, with its pronounced seasonal patterns, overwhelms the underlying relationships between the atmospheric tracers and the sources of mercury emissions. This can lead to misleading factor identifications, where the eigenvectors are more reflective of the temperature's seasonal cycle than of the true emission dynamics of the tracers. When eigenvectors are primarily driven by trends or seasonality, there is a high risk that the source apportionment is dictated by external variables not directly linked to the emission processes themselves. As a result, the interpretation of the PMF output can become compromised, leading to inaccurate conclusions about source contributions.

Response: The decision to include temperature in PMF is based on our understanding of GEM surface re-emissions (volatilization), which is a process strongly dependent on air temperature as well as soil properties. This was discussed in section 4 with relevant papers cited. Unfortunately, daily data on soil properties are not available at the sites. We also note that temperature was

included in the PMF model for identifying GEM surface re-emissions in a previous study (Qin et al., 2020).

Hg sources, such as wildfire emissions and terrestrial surface re-emissions, are closely linked to seasonal temperature trends. The occurrence of wildfires and GEM re-emissions are expected to increase during the warmer months. These emissions are coupled with seasonal temperature cycle. Thus, temperature should not be viewed as an external variable; it drives many of the Hg emissions. Justification for including temperature in the PMF model is provided in the revised paper (section 2.2), which reads "Temperature was included in the model because it is an important driver of wildfire emissions and terrestrial Hg re-emissions (Zhu et al., 2016)."

Qin, X., Zhang, L., Wang, G., Wang, X., Fu, Q., Xu, J., Li, H., Chen, J., Zhao, Q., Lin, Y. and Huo, J.: Assessing contributions of natural surface and anthropogenic emissions to atmospheric mercury in a fast-developing region of eastern China from 2015 to 2018, Atmos. Chem. Phys., 20(18), 10985-10996, 2020.

Furthermore, it is crucial to emphasize the importance of sensitivity tests and residual analysis to ensure the robustness of the PMF solution. Sensitivity tests help to evaluate how the model responds to changes in variable selection and model parameters, offering insights into the stability and reliability of the source apportionment. Residual analysis, on the other hand, provides a valuable check on the quality of the model fit, indicating whether the factors identified by the PMF model adequately explain the observed data or if there are unexplained variances that need further investigation.

Response: The model fit results were shown in Supplement Figure S1 and discussed in section 2.2. From Fig. S1, it can be seen that the source apportionment results are reliable. The coefficient of determination (R^2) between the PMF modelled and observed TGM concentrations is above 0.7 for the entire time series for each site, indicating a good model fit based on the current variable selection and model parameters chosen. There is also strong overlap between the modelled and observed time series for 24 h mean TGM, indicating the model adequately captured the daily variability. One of the top considerations in deriving the final PMF solution is the justification of the factors and assignment to Hg sources. The sensitivity tests using 5 and 7 factors resulted in factors that were difficult to interpret. A 5-factor solution results in tracers loading on multiple factors. This leads to a scenario where multiple factors can be assigned to the same source. An ideal PMF solution is one where each factor is assigned to a unique source. A 7-factor solution contains all the results of the 6-factor solution; however, it includes an additional factor that cannot be definitively assigned to a source. Results from the PMF sensitivity runs are discussed in Supplement section S2 and corresponding Tables S5-S7.

The PMF residuals are within the recommended limit of three standard deviations and follow a normal distribution (Hopke et al., 2023). This confirms the modeled factors adequately explain the observed TGM data. The strong R^2 between modeled and observed TGM concentrations also indicates a good model fit (Fig. S1). The model could not reproduce a few elevated TGM concentrations at the KEJ site. There was a total of 7 data points in the entire 2005-2016 time series (out of 3118 data points) where the scaled residuals were beyond three standard deviations. The residuals analysis is discussed in Supplement section S3 and corresponding Fig. S12.

Hopke, P. K., Chen, Y., Rich, D. Q., Mooibroek, D., and Sofowote, U. M.: The application of positive matrix factorization with diagnostics to BIG DATA, Chemometrics and Intelligent Laboratory Systems, 240, 104885, 2023.

I find it puzzling why the authors chose to perform a separate PMF run for each individual year, as this approach undermines the potential insights that could be gained from analyzing the full, continuous time series together. Running the entire time series as a single dataset would provide a more robust and comprehensive analysis, allowing the model to capture long-term trends, interannual variability, and potential seasonality in a more holistic way. A year-by-year analysis may artificially constrain the factors identified, leading to fragmented or incomplete source apportionment, and it limits the ability to understand how certain sources or processes evolve over time.

Response: The reviewer's point is valid; there is some loss of statistical power by not using the entire time series in a PMF run. On the other hand, there is additional information gleaned from the separate annual runs. For example, the fact that the same factor profiles are reproduced each year provides additional confidence in the results, while the variability from year to year gives us an indication of the uncertainty in those profiles. As shown by the error bars in the PMF factor profiles (Figs. 2, 8, and 12), the species percentages vary from year to year. For example, in the case of the GEM re-emissions and biomass burning factor at SAT, the TGM percentage was 19%±10% (annual mean ± std. dev.). This temporal variability is captured by performing a separate PMF run for each year, but it would be missed if PMF was performed once for the entire time series. Source profiles can change over time, for example the fuel type and temperature of a forest fire can vary from one fire to another (and is influenced by climate variability); emissions control technology may impact the relative proportions of CO and TGM in the anthropogenic component of the Hg pool; Hg emissions from powerplants and industrial sources have decreased due to domestic regulations (Gov't of Canada, 2024; USEPA, 2024). Thus, a constant factor profile assumption over the long-term may not be valid in all cases. Performing a separate PMF run for each year requires more effort than a single PMF run for the entire time series, but it offers more insight into the consistency of the factor profiles year over year. The rationale behind performing separate yearly runs is discussed in the revised paper (section 2.2), which reads "Source profiles can change over time, for example the fuel type and temperature of a forest fire can vary from one fire to another (and is influenced by climate variability); emissions control technology may impact the relative proportions of CO and TGM in the anthropogenic component of the Hg pool; Hg emissions from powerplants and industrial sources have decreased due to domestic regulations. Thus, a constant factor profile assumption over the long-term may not be valid in all cases."

Government of Canada: National Pollutants Release Inventory: mercury, Canada's actions, https://www.canada.ca/en/environment-climate-change/services/national-pollutant-release-inventory/tools-resources-data/mercury.html, 2024.

USEPA: Mercury and Air Toxics Standards, https://www.epa.gov/stationary-sources-air-pollution/mercury-and-air-toxics-standards, 2024.

As suggested by the reviewer, we performed an additional PMF run for the entire time series. The results are discussed in the revised Supplement section S1, and the daily TGM source contributions are plotted in Fig. S11 (yearly runs are labelled as 'separate' and additional run for the full time

series is labelled 'all'). For the GEM re-emissions, biomass burning and Hg pool factors, the two runs produced similar seasonal patterns. However, there are noticeable differences for other sources. The separate runs show greater interannual variability in the daily TGM relative contributions compared to 'all' run scenario. For example, elevated relative contributions during 2010, 2011 and 2014 for local combustion. This is likely because the TGM percentages in each factor vary year to year as discussed above. We also see that the daily relative contributions from natural surfaces are elevated over the course of the time series for the 'all' run, whereas contributions from anthropogenic emissions were occasionally more important for the 'separate' runs. The 'separate' yearly runs capture the variability in the observations better than 'all' run. This is supported by the higher R^2 between modeled and observed TGM concentrations for the 'separate' case (0.75, Fig. S1) compared to 'all' case (0.68). The slope of the trendline for the 'separate' case is also closer to the 1:1 line (0.81 vs. 0.70).

The percentage contribution from natural surface emissions (wildfires plus re-emissions) from 'all' run was 70.6% on average, with a range of 66-76% depending on the year. This is higher compared with results from 'separate' runs; however, the conclusion that natural surface emission contribution exceeds anthropogenic contribution to TGM has not changed. Based on the 'all' scenario, the mean relative TGM source contributions were 50% from the Hg pool, 26.4% from terrestrial GEM re-emissions, 10.4% from shipping and SSA processing, 6.3% from oceanic evasion, 3.8% from wildfires, and 3.4% from local combustion. Most of the percentages were comparable to those obtained from 'separate' runs, except for higher terrestrial re-emissions and lower oceanic evasion. Secondary sulfate (regional Hg emissions and chemical transformation) did not contribute to TGM in the 'all' run case, whereas it contributed a few percent of the TGM in the 'separate' case.

Furthermore, I also do not understand why the authors did not merge all the sites into a single analysis. In a data-driven model like PMF, the number of receptors (sampling sites) is crucial for increasing the model's ability to discern distinct sources. By running the sites separately, the authors miss out on the advantage of having a broader spatial coverage and a larger dataset, both of which can significantly improve the resolution of source identification. Combining data from multiple receptors across different sites increases the model's power to detect and distinguish between sources, especially when there are overlapping emission signatures that might vary in strength and frequency across locations.

Response: The data size at each site and in each year is sufficient for PMF analysis. The three sites are very far apart from each other with the estimated distances (km) being: SAT-EGB (3290), SAT-KEJ (4355), and EGB-KEJ (1165) (see map below). The regional and local sources affecting the sites have potentially different emission signatures. For example, metal smelting and cement production are the primary industrial sources near SAT, while coal-fired power plants are near KEJ. Thus, it was preferred to derive the factor profiles unique to each site.

Increasing the number of receptors also enhances the model's capacity to address complex atmospheric dynamics, as it allows for a richer dataset that captures variations due to both local sources and regional transport processes. When multiple sites are analyzed together, the model has more information to work with, potentially identifying regional patterns that might be missed when each site is treated independently. The inclusion of multiple receptors also reduces the risk of overfitting the model to local conditions or short-term fluctuations at any single site, leading to a more generalizable and reliable source apportionment.
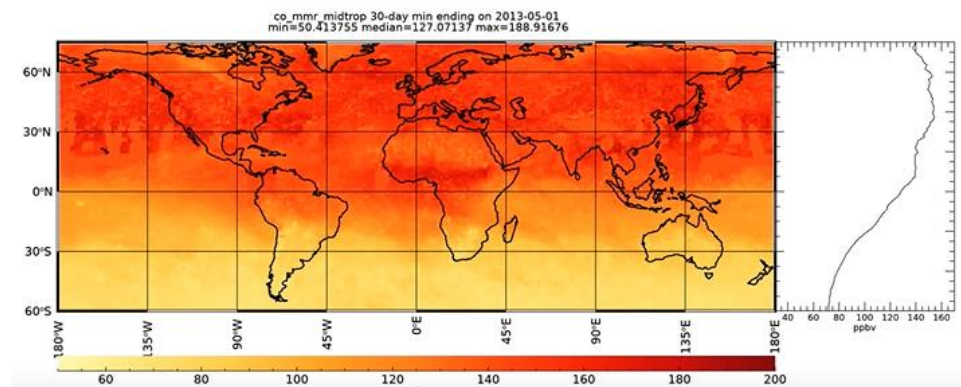
Response: As mentioned above, the data size at any single site is large enough. Each site has 8-14 years of 24-h data containing 12 variables. The long-term data coverage and high-temporal data resolution (Supplement Table S1) are sufficient for PMF modeling to capture variations in local and regional sources. This is clarified in section 2.2 of the revised paper. In this study, the PMF model captured both local source impacts (i.e., GEM re-emissions from land and nearby ocean, local combustion, local shipping, crustal/soil emissions) and regional source impacts (i.e., wildfires, regional combustion from U.S. northeast). The Hg pool was also inferred from the PMF model; this captures all types of Hg emissions mostly from the northern hemisphere and its impact on the receptor sites via long range transport.

It is not possible to include more receptor sites because some ancillary measurements are not available at other TGM sites. To clarify, a sentence was added in section 2.1, "The CAPMoN TGM sites were selected for this study to provide an update on current concentrations and patterns. The same set of ancillary measurements are also available at the sites to conduct PMF analysis." As mentioned above, the sites are very far apart; hence the data should not be combined considering the potentially different sources, meteorology, and transport patterns affecting the sites. There is minimal risk of the model overfitting to local conditions or short term variability because the sites are located in rural-remote areas infrequently impacted by large emission sources.

The number of species used in the factorization performed by the authors appears insufficient to properly apportion the sources they claim to resolve. The profiles of the factors presented do not

convincingly align with the expected "fingerprints" of the emission sources they attribute them to. For example, in Figure 12, the "local combustion" factor is primarily characterized by SO2 only, yet it lacks loading of CO, which is a well-known combustion tracer. Instead, CO is predominantly loaded in the "background" factor, which should typically be dominated by long-lived species. This misallocation raises concerns about the accuracy of the factor assignments and suggests that the model may not be adequately capturing the true source profiles.

Response: SO_2 is a short-lived pollutant (lifetime of a few days) and is emitted in large quantities from combustion sources, especially coal combustion and metal smelting which are also important anthropogenic Hg sources. This makes SO_2 a suitable local combustion tracer. In previous mercury source apportionment, SO_2 and Hg were typically used to assign profiles to local combustion (Eckley et al., 2013; Wang et al., 2013). While CO is also emitted from combustion sources, it has a relatively longer lifetime in the order of a few months, which allows for it to be transported over long distances (Jeffery et al., 2024). Given the relatively small magnitude of combustion sources near these rural and remote sites, it is not surprising that the CO emissions from these sources contribute a relatively small proportion of the total CO observed, which is dominated by the hemispheric background. This characteristic is similar to that of GEM and TGM. These are clarified in the revised paper (section 3.1.2). Instead of "background", we also renamed the factor to Hg pool, which encompasses all types of Hg emissions mostly from the northern hemisphere and its subsequent long distance transport. As shown in the global distribution of CO mixing ratios below, higher CO levels are broadly distributed across most of the northern hemisphere owing to fossil fuel combustion and biomass burning.

co_mmr midtrop 30-day min ending on 2013-05-01
min=50.413755 median=127.07137 max=188.91676

Paulo Penteado, NASA/JPL-Caltech, https://airs.jpl.nasa.gov/resources/239/airs-global-carbon-monoxide-over-20-years-2002-2022/

Eckley, C.S., Parsons, M.T., Mintz, R., Lapalme, M., Mazur, M., Tordon, R., Elleman, R., Graydon, J.A., Blanchard, P. and St. Louis, V., 2013. Impact of closing Canada's largest point-source of mercury emissions on local atmospheric mercury concentrations. Environmental Science and Technology, 47(18), 10339-10348.

Jeffery, P. S., Drummond, J. R., Zou, J., and Walker, K. A.: Identifying episodic carbon monoxide emission events in the MOPITT measurement dataset, Atmos. Chem. Phys., 24, 4253–4263, https://doi.org/10.5194/acp-24-4253-2024, 2024.

Wang, Y., Huang, J., Hopke, P. K., Rattigan, O. V., Chalupa, D. C., Utell, M. J., and Holsen, T. M., 2013. Effect of the shutdown of a large coal-fired power plant on ambient mercury species, Chemosphere, 92(4), 360-367.

Moreover, it seems that the authors constrained the number of factors to a level that exceeds what the available data can reliably cluster (with physical apportionment mean). The factorization results in some clusters that are difficult to justify from a source attribution perspective. For instance, there is a factor which unlikely represen a source, this suggests that the model may be overfitting, potentially driven by non-source-related variables.

There are also several indications in the presented profiles that attempting to resolve six sources from the limited number of species (or variables) used in the analysis constitutes an over-extrapolation of what is feasible through this factorization method. Some factors, for example, the cluster is basically loaded with Ca and Mg, other that is loaded with Cl and Na only. These elemental groupings suggest that the model is forming clusters based on chemical similarity only, rather than true source-specific emissions. This is further evidence that the factorization may be over-constrained, leading to artificial factors that do not accurately represent distinct sources.

Response: Twelve variables were included in the PMF model. The six model factors represent a variety of sources, which are all potential sources of Hg. The sources include not only well-established anthropogenic Hg sources (i.e. fossil fuel combustion), but also those that are less understood and less constrained in terms of the global Hg emissions budget, e.g. GEM re-emissions, wildfires, evasion of GEM from ocean, crustal/soil dust, and contributions from the Hg pool. Given that the sites are in rural-remote areas (infrequently impacted by large Hg point sources), the role of these less-studied sources are potentially important. Therefore, we used Cl and Na as tracers of a potential marine source of Hg, such as oceanic Hg evasion or shipping emissions (explained in section 3.1.2 of the revised paper). Ca and Mg are tracers of crustal/soil dust. In terrestrial ecosystems, Hg is derived either geologically or via atmospheric wet and dry (including litterfall) deposition, and is mainly bound to soil organic matter as oxidized Hg or Hg(II) (Eckley et al., 2016). Hg(II) also has a strong affinity for NaCl particles (Rutter and Schauer, 2007), which may explain the presence of TGM in the road salt factor. Hg can subsequently be re-emitted by wind erosion or land disturbance. This is explained in the revised paper (section 3.2.2).

The TGM percentage in the crustal/soil factor with high Ca and Mg and factors with high Cl and/or Na) are small; however, the values are not negligible (see Figures 2, 8 and 12). The annual TGM percentages are 8-10% on average and vary interannually. On a daily basis, TGM contributions can reach up to 90% for crustal/soil emissions and road salt (Fig. 10). In Figure 5, we can see another example where daily TGM contributions can reach up to 90% from shipping and SSA processing and oceanic evasion. The reviewer is concerned that these are artificial or non-significant sources of Hg; however, it is clear that their contributions to daily TGM can be important.

Eckley, C.S., Tate, M.T., Lin, C.J., Gustin, M., Dent, S., Eagles-Smith, C., Lutz, M.A., Wickland, K.P., Wang, B., Gray, J.E., Edwards, G.C., Krabbenhoft, D. P., and Smith, D. B.: Surface-air mercury fluxes across Western North America: A synthesis of spatial trends and controlling variables, Sci. Total Environ., 568, 651-665, 2016.

Rutter, A. P., and Schauer, J. J.: The effect of temperature on the gas–particle partitioning of reactive mercury in atmospheric aerosols. Atmos. Environ., 41(38), 8647-8657, 2007.

Based on the issues discussed, I recommend a major revision of the manuscript. It is clear that the authors have put significant effort into this work, but there are substantial improvements that need to be made, particularly in the source apportionment performing and analysis. I strongly encourage the authors to further explore the capabilities of source apportionment resource. While there is certainly a learning curve, catching up in this area will greatly enhance the robustness and accuracy of the study. I am confident that it will be worth the effort and will lead to more defensible results and insightful discussions.

Given the concerns raised about the current factorization and its interpretation, revisiting the apportionment process will likely result in significant changes to the manuscript's overall findings and discussion. I recommend that the authors reanalyze the data, addressing the over-extrapolation issues and ensuring that the profiles correspond more clearly to recognizable emission sources. Once the source apportionment is properly refined, I encourage the authors to resubmit the manuscript, as I believe it has the potential to make a valuable contribution to the field.

Response: The reviewer's concerns regarding the source apportionment analysis have been addressed in the responses and additional analysis and results have been incorporated in the revised paper. The major revisions to the manuscript are summarized below.

- Justification for including temperature in the PMF model (section 2.2)
- Clarification on the long term data coverage available at the sites (section 2.2)
- Justification for performing yearly model runs (section 2.2) and results of PMF model runs using the entire time series and comparison with the yearly runs (Supplement S1, Fig. S11)
- Discussion of PMF sensitivity runs using a different number of factors (Supplement S2, Tables S5-S7)
- Discussion of model fit and residuals plots for the final PMF solution (Supplement S3, Fig. S12)
- Renaming of background factor to Hg pool (throughout the manuscript)
- Clarification on using SO_2 as a tracer for local combustion and CO as a tracer of the Hg pool and long range transport (section 3.1.2)
- Intercomparison of model X and B TGM concentrations and related discussion (Supplement S4, Fig. S13)

In addition, I suggest that the authors expand on the differences between the two analyzers used (2537B and 2537X) as part of their revised manuscript. A thorough comparison of these instruments would be highly interesting, especially regarding any differences in performance or measurement outcomes. I would be very keen to see these comparisons and an error vector decomposition.

Response: Our assessment of the hourly TGM differences between Tekran 2537X and 2537B is published along with the quality controlled TGM dataset (ECCC, 2024). The monthly plots of TGM from model X and B are shown in Supplement Fig. S13. The model X and B analyzers at the EGB site operated side by side during Feb-Aug 2017. The model X reported slightly higher TGM than model B

with a mean hourly difference of 0.06 ng m^{-3} (3.9%). Monthly mean hourly TGM differences were in the range of 0.02-0.1 ng m^{-3} (1.4-6.3%). Model X and model B analyzers were also operated side by side at the same site during Mar-Jul 2018. The model X reported higher TGM than model B with a mean hourly difference of 0.08 ng m^{-3} (6.4%). Monthly mean hourly TGM differences were in the range of 0.07-0.09 ng m^{-3} (5.4-7.4%). Considering the differences were not significant and the concentrations showed similar trends, TGM concentrations from the model X and B analyzers were averaged during the periods when valid hourly data were available.

TGM was measured concurrently at KEJ and KEB from February to June in 2017. Note that KEJ and KEB sites are not co-located. KEJ was operating a model B analyzer; the site was relocated 3 km south of the original site in Feb 2017 (KEB) and the model X analyzer began operating at the new site. TGM was higher at KEB than at KEJ with a mean hourly difference of 0.26 ng m^{-3} (18.8%). Monthly mean hourly TGM differences were 0.20-0.29 ng m^{-3} (16.6-21.5%). These differences may be due to the different analyzer models and/or relocation of the monitoring site; the exact cause is inconclusive. Given the large TGM differences between KEB and KEJ, our decision was that the data from the two sites should not be combined into a single time series. Therefore, the 2017-2018 data at the new site were not used for PMF modeling and long-term trends analysis. We have added this discussion to Supplement section S4 and corresponding plots in Fig. S13.

Environment and Climate Change Canada: Canadian Air and Precipitation Monitoring Network (CAPMoN), Toronto, Ontario, Canada, Data files: AtmosphericGases-TGM-CAPMoN-AllSites-2017.csv, AtmosphericGases-TGM-CAPMoN-AllSites-2018.csv, https://doi.org/10.18164/e1df5764-1eec-4a9f-9c03-f515b396b717, 2024. **(Go to Downloads, views, and links > View ECCC Data Mart)**