Referee 1 Commentary

General Comments:

In this study, McKinnon et al. analyze the total columns of CO retrievals from MOPITT by applying varimax empirical orthogonal function (EOF). The study evaluates if the dominant CO sources associated with inferred modes of the variability at the regional to global scales can be disentangled and identifies the strengths and weaknesses of the EOF analysis. This paper is well-structured, informative, and highlights important science questions, but will need to be strengthened with more references throughout the manuscript (especially in the Introduction section), in addition to addressing some information gaps, and making some figure-related and technical changes before it can be accepted. This manuscript also has sentences that would need to be broken down for simplicity, along with the numbering of all equations used, for clarity, and would need to be scrubbed for missing punctuation/commas/grammar.

Specific Comments:

1. L1 – L3 Please state the objective more clearly and define MOPITT.
2. In general, the manuscript would need to be supported throughout with more references, particularly in the Introduction section, such as L27, L59, L63, L83, L1234-127 (for TES, AIRS, IASI, TROPOMI, and CrIS), L133, L152, L171, L409 etc.
3. L149 – Please define NH and SH.
4. L152 – Please split the sentence after the term “burning”.
5. L172 – “measurable in breath” – This is unclear.
6. L207 – “ENSO … AAO” Please expand each acronym and add citations.
7. “Southern Hemisphere” and “SH” (and “Northern Hemisphere” and “NH”) have been used interchangeably throughout the manuscript. Please define at the first occurrence and use the abbreviated form thorough the rest of the paper.
8. Section 2/Methods lacks enough citations.
9. L213 - Please provide a brief description of the term “EOF” before advancing to Section 2.4.1.
10. Please add equation numbers or reference numbers for all matrices and equations shown in the paper for the ease of reference/discussion in the manuscript.

11. L228 – Can the “F” matrix be identified differently to not confuse it with the original F matrix?
12. Please explain the “SVD” acronym in L253.
13. L259 – How is this approach different from the one shown in the manuscript?
14. L267 – Please define “PCA”.
15. L268 – Please explain the motivation behind taking this approach.
16. L271 – Please start a new sentence.
17. L272 – What are the potential drawbacks/ uncertainties associated with this approach?
18. L276 – Please remove the period next to where N* is defined.
19. L277 – “…not accurate for the case of Non-Gaussian red noise”. Please explain why.
20. L280 – 282 – Is there a way to identify which is the case for sure?
21. L285 – How do these impact the results? What are the introduced uncertainties and can they be quantified?
22. L324 – Please start a new sentence.
23. L358 – Please split the sentence from “for instance…”.
24. L367 - 394 – Please explain this part in points for clarity.
25. L395 – Please start a new sentence from “however…”.
26. L424 – Please write the equation outside of the paragraph for clarity.
27. L459 – “leads us to believe there …” This may be speculative. Please consider changing this to “suggests the possibility of”.
28. L462 – What could be the other possible causes?

Figure Commentary

Fig 1 - Please change the y axis magnitude to “molecules/cm^2”.

Fig 2 - Please state the value (molecules/cm^2?) shown by the colorbar next to it.

Fig 3- Please add subplot numbers/names such as a), b), etc., instead of identifying them as “top left”, “top right”, etc. The values in the figures could be grid-binned, and shown

using a consistent colorbar (ranging from -0.03 to 0.015?) for the bottom left and the two figures on the right. Please state the name of the value/variable/unit next to each colorbar.

Fig 4 - Please include subplot names/numbers, as listed for Figure 3. L451 – Can the data be grid-binned for better visualization?

Fig 5 - Please consider using brighter colors. The black and blue solid lines blend into each other at some points.

Fig 6 - Please add subplot names/numbers, grid-bin the data, use a consistent colorbar where possible, and list the name of the variable being identified by the colorbar next to it, same as the comments for Figure 3.

Fig 7 - Please add subplot numbers/names.

Fig 8 - Please add subplot numbers/names.

Fig 9 - Please grid-bin the data if possible, and list the name of the variable being identified by the colorbar next to it.

Fig 10 - Please add subplot numbers/names.

Technical Comments:

- L3 – Please change "Additionally, ……intention" to "This study aims to".
- L27 – Ozone has been defined as "O_3", but the term "Ozone" has been used throughout the paper (L28, L39, etc.). Please change this to O_3. This is also true for CO and other species.
- There are formatting errors with the citations in the paper. Please correct these. E.g., L36, L37 (missing parenthesis), L47, L49, etc. L72 – Are the citations arranged alphabetically? L77 – Please add a comma between the two citations. L123-124 – Please check the citation format.
- L215 – Are the citations listed alphabetically or in the order of publication year?
- L262 – "mth" Please use superscript "m^{th}".
- L85 – Please remove "we are interested". Please change "include the following" to "are the following".
- L89 – "will use" --> use
- L91 – Please remove "next section on Methods", and please "section" after "Methods".

- Please check and add commas where necessary in L67, L93, etc.
- L95 – performs --> performed, verifies --> verified
- L97 – Please remove “current”, and please change “will be on” to “is”.
- L140 – “so that we” --> ” to”, “be able to study” --> “study”
- L186 – “has shown” --> “showed”
- L240 – “,The” --> “,the”
- L293 – “very immediately” --> “an immediate”
- L349 – “plain English” --> “simpler terms”, “is named because” --> “is named so, as”
- L484 – “we have said” --> “discussed”
- L524 – Please remove the term “First”.
- L531- “Because” --> “Since”

Referee 2 Commentary

The manuscript provides a general tutorial for EOF and singular spectrum analysis, with an application to 8-day aggregates of near-global (60S to 60N latitude) carbon monoxide columns retrieved from the MOPITT instrument from 2005-2018. The authors demonstrate consistency in EOF patterns derived across different domains to confirm that the patterns revealed are a robust property of the dataset. Varimax rotation suggests two dominant EOFs in contrast to three emerging from the raw analysis (annual, semi-annual and 3-month variations). Singular spectrum analysis is then applied to the principal components to identify different time scales of variations. The authors initially hypothesize that the spatiotemporal patterns will reflect individual CO sources but find that this is not the case, and instead recommend future exploration of cyclostationary EOF analysis.

General Comments:

1. The tutorial-style sections may be more suited to a technical note. Given that the techniques themselves are described in textbooks such as Wilks, 2011 (note there is now a 4th edition published in 2019), is AMT a good fit for this work? Note that the NCAR Climate Data Guide provides some basic introduction, which may be worth pointing to as well: https://climatedataguide.ucar.edu/climate-tools/empirical-orthogonal-function-eof-analysis-and-rotated-eof-analysis

2. A stronger case could be up front for why an EOF analysis is expected to reveal source-specific signals in total column CO, such as the anthropogenic and natural variability noted in Question 1, or in global background in Question 2. How do the patterns from sources contrast to other factors contributing to spatiotemporal variations, such as weather variability or OH? To what extent do spatiotemporal patterns overlap between fires and biogenic VOC contributions to CO?
3. Do the authors have a recommendation as to whether to apply Varimax rotation?
4. The manuscript would benefit from additional editing throughout to improve the flow and clarity of messaging. Given the interest in identifying patterns form specific sources, it would be helpful to have a table summarizing current understanding of CO emissions from specific sources.

Specific comments:

1. line 87. Need to define global background. Is it all non-anthropogenic CO, and if so, how is this different from the natural CO in question 1?

2. Lines 90-100. The discussion is a nice summary but misses a key early application to air quality in the U.S., including: Eder, B. K., J. M. Davis, and P. Bloomfield, A characterization of the spatiotemporal variability of non-urban ozone concentrations over the eastern United States, *Atmos. Environ.*, 27, A(16), 2645–2668, 1993.

3. Line 136. Given the lifetime of CO, why is this resolution a problem? Wouldn't daily data be better suited to identify daily variations such as those due to fires?

4. Line 141. How is the data de-trended? Linear fit? At each satellite pixel? Does this remove the signal from long-term trends in anthropogenic or other sources?

5. Figure 2. Do these patterns change over time?

6. Lines 185-188. This is an example of discussion in need of editing. A recent review includes discussion of OH sources and sinks: Fiore et al. (2024), Climate and Tropospheric Oxidizing Capacity, Annual Reviews of Earth and Planetary Sciences, 52:321-349, https://doi.org/10.1146/annurev-earth-032320-090307

7. Lines 250-251. The choice here depends on the goals of the analysis. Examples of why one is better suited would be helpful here.

8. Lines 403-405. Is the data also area-weighted? Some discussion as to why standardizing is preferred here as opposed to giving more weight to regions with larger sources and thus larger absolute variations would be useful.

9. Lines 441-442. Some discussion is needed as to the spatiotemporal variations in the sources in contrast to weather-driven variations.

10. Line 459. What is oceanic variability in CO?

11. Line 476. Why is there a ~3-year time scale when the lifetime of CO is on the order of a couple months?

12. Figure 9. What processes might drive the sharp gradients in skewness? It would be stronger to frame the findings in terms of any hypotheses as to what drives the variations in CO emerging from this analysis. For example, what drives the annual/semiannual cycle noted in lines 534-535?

13. Appendix A1. Line 589. Normalization procedures can vary by software package.

Author's Response to Referee 1 Commentary

The authors acknowledge the lack of references in the introduction as well as in the methods section and agree to add in numerous citations throughout the manuscript to address these shortcomings. We ensure that all our quantitative statements, or observations which carry significant importance which are not our own have been cited throughout our analysis. We add discussion (with citations) to address all specific comments which point out information gaps. We address all specific comments which have scientific questions that are intended to facilitate discussion and add citations as appropriate. The authors agree to fix all formatting issues with existing citations, and to add equation numbers throughout the manuscript. We break down statements throughout the manuscript in cases where a statement is a run on sentence, is unclear, or requires additional explanation. We also address all other grammatical and technical issues which have been pointed out. The authors would like to thank Referee 1 for giving feedback that is extremely clear and easy to follow. Referee 1's comments about our gaps in literature highlight an important weakness in our manuscript and we believe our motivation will be much stronger by addressing them. We especially thank Referee 1 for asking us questions regarding uncertainties and the assumptions used in our approach as we believe this will address a crucial gap in our data analysis.

Author's Response to Referee 2 Commentary

The authors acknowledge that many EOF tutorials already exist in the literature and as such agree to reduce the amount of technical detail in the methods portion of our manuscript, moving these aspects to the appendix instead. The authors add in a table to our introduction section on source variability with an accompanying discussion to help us summarize our current understanding of the variability from specific sources of CO. We add additional discussion in our introduction to make a better case up front for why we had expected EOF analysis to have better success in being able to separate sources into independent signals. We also add in our Results section 3.2 a clearer recommendation for why in our specific case, it was better to use our unrotated modes rather than to use our varimax modes. We agree to incorporate all changes which have been requested in the specific and technical comments. The authors would like to thank Referee 2 for asking many general questions regarding our manuscript including its structure, motivation, and our approach to data analysis. We especially appreciate Referee 2's questions about area weighting in our data and our time scales, as we believe answering these questions will address a vital gap in our

methods. We also appreciate that Referee 2 was kind enough to give us 2 specific citations to help us address gaps in our literature.

General Changes Throughout Manuscript

1. Equation numbers have been added to all existing equations, and matrices are defined using new equations on separate lines for ease of reference and discussion. **(Referee #1)**

2. Many new citations have been throughout the introduction, as well as in the methods section. **(Referee #1)**

3. We add a motivation section to the introduction to make a case for why we expect EOF analysis to be successful in being able to separate sources into independent signals. **(Referee #2)**

4. We reorganize our Methods section, putting the more technical details in our appendix to reduce the emphasis of our paper being a tutorial on EOF analysis. **(Referee #2)**

5. Acronyms for Northern Hemisphere, Southern Hemisphere, and different species are defined at their first occurrence and then referred only by their acronym from then on. **(Referee #1)**

6. Citation formatting errors were double checked against AMT standard and corrected accordingly. **(Referee #1)**

7. Grammatical and technical fixes were implemented throughout in accordance with reviewer recommendations. Some sentences have been split or broken into simpler statements where the authors were requested, to aid with the flow and clarity of language. **(Referee #1)**

Changes Addressing "Specific Comments" with References to Line/Page Numbers

The following changes are implemented to address the "specific comments" from each referee, with line/page references. For each change we indicate in parenthesis which referee is being addressed in our proposed changes. Please note that the line numbers refer to the **original version** of the manuscript

Lines 1-3 Pg 1(Referee #1)
We define MOPITT and more clearly describe the objectives of the study.

Line 27 Pg 2 (Referee #1)
Citations added to describe how the concentration level of NOx affects ozone production.

Line 59 Pg 2-3 (Referee #1)
Citations added to emphasize retrieval inaccuracies and the lack of monitoring stations across Africa.

Line 63 Pg 3 (Referee #1)
Citations added to bolster arguments that changes in local trends are difficult to separate from background trends.

Line 83 Pg 3 (Referee #1)
Citations added to emphasize that the separation of secondary CO from surface emission aids source attribution and inverse modeling studies.

Line 87 Pg. 3 (Referee #2)
We clarify that in our specific context the global background refers to a mathematical, non-physical quantity and simply refers to an average value.

Lines 90-100 Pg 3-4 (Referee #2)
We add discussion regarding the application of EOFs to air quality monitoring in the US. We use citations including Eder, B. et. al. (1993), who used EOFs to study spatiotemporal variability of non-urban ozone.

Lines 124-127 Pg 4-5 (Referee #1)
Citations added for other satellite instruments (TES, AIRS, IASI, TROPOMI, and CrIS)

Line 133 Pg 5 (Referee #1)
Citations added regarding potential discontinuities of IASI-A retrievals, the smaller number of measurements for TES, and the cloud clearing algorithm used by AIRS.

Line 136 Pg. 5 (Referee #2)

We add discussion that daily data would not be better suited because MOPITT does not have daily global coverage. We must ensure global coverage in the analysis because EOF analysis fails to produce patterns if there are gaps in the data.

Line 141 Pg. 5 (Referee #2)

Discussion is added regarding the fact that our data was detrended by taking the globally averaged time series and then applying a Fourier fit to it. Because the trends in the total column data are a non-linear combination of the trends from individual sources, it is not immediately clear what sources may have had their long-term trends removed.

Line 152 Pg 6 (Referee #1)

Citations added to emphasize high contribution of anthropogenic emissions from China due to coal reliance and biofuel burning.

Line 171 Pg 7 (Referee #1)

Discussion added with citations to emphasize the relative contribution to CO due to oxidation from CH_4 as compared to the oxidation from NMHCs. Citations added regarding the temperature dependence of isoprene and terpenes. Discussion added with citations to clarify how the large annual contribution and high reactivity of isoprene makes it the most dominant hydrocarbon emission. The term "measurable in breath" has been removed and clarified (with citations) by describing the biological processes in animals where we can measure VOC emission.

Lines 185-188 Pg 7 (Referee #2)

We edit our discussion with appropriate citations including Fiore et al. (2024) to incorporate up-to-date information on the sinks and sources of OH.

Line 207 Pg 8 (Referee #1)

Acronyms for all 4 climate indices are expanded, with references included for each.

Line 213 Pg 8 (Referee #1)

We clarify the term "EOF" and add a brief description.

Lines 250-251 Pg 9-10 (Referee #2)

We add discussion regarding the fact that for our specific application, it makes very little difference in whether we choose to normalize our covariance since we are only

using one observation field. If we were to compare EOFs of different datasets, then we should normalize the covariance.

Line 253 Pg 10 (Referee #1)

We clarify the term "SVD" and add a brief description

Line 259 Pg 10 (Referee #1)

We clarify with references how Lanczos iteration is different from eigendecomposition and svd.

Line 267 Pg 10 (Referee #1)

We clarify the term "PCA" and add a brief description

Line 268 Pg 10 (Referee #1)

We discuss, with supporting citations, how our motivation for using asymptotic approximation is due to simplicity, unlike similar methods that rely on statistical analysis and resampling the dataset.

Line 272 Pg 10-11 (Referee #1)

We discuss, with supporting citations, the strengths and weaknesses of using an asymptotic expansion as compared with other methods.

Line 277 Pg 11 (Referee #1)

Significant discussion has been added, together with supporting citations, to explain why our estimation for the effective sample size is only valid for the case of red noise Gaussian processes where the true sample size N is very large.

Line 280-282 Pg 11 (Referee #1)

Discussion is added, with references, regarding tests that can be used to determine if eigenvalues are degenerate or potentially represent white noise.

Results 3.1 Pg 15-16 (Referee #2)

We respond to a comment that initially refers to Figure 2, however because it refers to multiple patterns, we assume the referee had intended to ask about how the EOF patterns may change over time in Figure 3. We clarify that the EOF patterns did not show significant changes when we segmented the data into yearly sets to compare them from year to year. It's possible this could be because the data had long-term trends removed.

Results 3.2 Pg 16-19 (Referee #2)

We add discussion to clarify that for our study, varimax rotation did not improve the representation of modes and that for further analysis with our seasonal decomposition we used our unrotated EOF modes. We note that decision should not be done in general for climate data. Our decision is because our unrotated modes were reproducible across multiple domains and had better representation with respect to variance.

Line 409 Pg 16 (Referee #1)

Citations added to emphasize that one potential benefit of varimax rotation is to make spatial patterns simpler and easier to interpret.

Lines 441-442 Pg 17 (Referee #2)

We add discussion to indicate that we have no reason to believe the patterns fit the behavior of source variability *or* that of weather patterns either large scale or local scale.

Line 459 Pg 18-19 (Referee #2)

We add discussion in the introduction as well as to Results Section 3.2 to describe sources of oceanic CO including photoproduction from colored dissolved organic matter (CDOM), direct production from phytoplankton, and dark production from CDOM in the absence of sunlight.

Line 462 Pg 19 (Referee #1)

Discussion is added, with references, to bolster our argument for why we do not see a connection between Residual mode 1 and oceanic variability from oceanic cycles. We mention potential issues such as the fact that oceanic CO budgets are very inaccurate because of changing oceanic environments.

Line 476 Pg 20-21 (Referee #2)

We clarify, using citations, that the use of time scales for CO makes sense even when the scales are longer than its lifetime. Some variability of CO is driven by large scale weather patterns, for example mixing time is about 6 months and hemispheric exchanges can happen across 1-2 years. It also makes sense from a perspective of studying the impact of climate drivers. (Bowman:1997 reference)

Line 534 Pg 25 (Referee #2)

We add discussion as to what physical processes drive the sharp gradients in skewness related to Figure 9. The annual cycle should be driven by seasonal changes in its sources and sinks and will differ depending on hemisphere and region. For example, in the northern hemisphere it's primarily driven by changes in fossil fuel burning usage i.e. it's largest in winter and smallest in summer because of changes in heating demands. In the southern hemisphere biomass plumes drive variability in central Africa and over part of Australia while biogenic and VOC emission mostly drives changes in South America. The large banding effect over latitude in the SH variability is probably because of a combination of regional transport as well as long term variability due to the North/South movement of the ITCZ which affects the strength of hemispheric exchanges through the Hadley circulation.

Appendix A1. Line 589. Pg 27 (Referee #2)

Here we clarify the reason for our normalization and mention other potential normalizations which could be used.

Changes to Figures

Fig 1 (Referee #1)

We change the y axis magnitude to "molecules/cm^2".

Fig 2 (Referee #1)

We label our colorbar with the proper units molecules/cm^2).

Fig 3 (Referee #1)

We add subplot names to our figures, and grid bin our data using a consistent colorbar. We label our colorbar with proper units. Just label as Magnitude

Fig 4 (Referee #1)

We add subplot names to our figures.
Delete extra parenthesis!

Fig 5 (Referee #1)

We alter our color scheme so that the lines do not blend in as much.

Fig 6 (Referee #1)

We add subplot names to our figures, and grid-bin them using a consistent colorbar. We label our colorbar with proper units. Just label as Magnitude

Fig 7 (Referee #1)
We add subplot names to our figures.
We capitalize axes.

Fig 8 (Referee #1)
We add subplot names to our figures.

Fig 9 (Referee #1)
We grid-bin the data and we label our colorbar with proper units.
Label as Skewness

Fig 10 (Referee #1)
We add subplot names to our figures.
We capitalize axes.