Review of Galytska et al 2025: "Causal inference for stratospheric chemistry: insights into tropical middle stratospheric ozone variability" submitted to ACP

December 3, 2025

The submitted paper uses methods from causal inference to examine short-term trends in ozone in the tropical middle stratosphere since 2004, with a particular focus on the difference between trends during the periods 2004-2011 and 2012-2018. Sophisticated methods are used that bring together causal inference, observations, prior physical understanding, and chemical transport models. The paper has a direct message about the the importance of transport-induced variability in N_2O and NO_2 and a 'meta-message' promoting the usage of causal inference methods in stratospheric photochemistry.

Note that this is the second time I have reviewed a version of this paper, having reviewed an earlier version submitted to a different journal. The paper has been revised since my first review, and I have had the opportunity to read it with fresh eyes. Although two major comments (and all of my minor comments) were addressed since my first review, others remain to be addressed. I also raise some new concerns upon this fresh reading.

As noted in my first review, broadly, this paper does a commendable job bringing together a range of methods to explore trends in mid-stratospheric ozone. A strength of the paper is that it clearly explains the causal inference mechanisms that it is deploying, including with the help of an illustrative toy model. These methods are used to support the key result of the paper, which is essentially a simple mechanism previously explored by some of the same authors (Galytska et al., 2019): in the tropical mid-stratosphere, where the dominant sink of ozone is catalytic destruction by NO_x, stronger residual upwelling has led to less NO_x and more O_3 (and weaker residual upwelling leads to more NO_x and less O_3). This could explain some decadal variability in ozone at these altitudes, which is well worth explaining. The paper supports this simple mechanism with sophisticated methods, which it illustrates as a proof of concept, and with some auxiliary nontrivial results. Although I am confident in the value of explaining ozone variability at these altitudes, and this mechanism seems quite plausible, I still have significant concerns with specific methodological choices made in this paper, which presently do not allow it to add as much value to current understanding as it should. One of my major concerns from the first review has been fully addressed, but the other three remain, and I must also introduce a fourth:

- A primary motivation of the study is a puzzling reduction of ozone from 2004-2011, yet this declining trend was not statistically significant, and it is excluded from the analysis by detrending in the data preprocessing step.
- The directed acyclic graph implies that low amounts of N_2O "cause" high amounts of NO_2 (a Simpson's paradox whereby an increase in reactant is associated with a reduction of reaction product), yet this seems ill-suited to capture the response to variability in the source of N_2O.
- The paper emphasizes that small (0-1 month) differences in the lags between variables constrains the upwelling, but it is not clear why.
- Is it appropriate for this method to omit a temperature-mediated pathway between upwelling and ozone? (Stronger upwelling can lead to cooling, which can increase ozone)

Because these concerns relate to fundamental methodological and interpretive choices in the manuscript, I will be recommending that this paper is Rejected with Encouragement to Resubmit. I see great promise in these methods and in the style of causal inference matched with observations and comprehensive models in this work, so I believe that, conditional on the resolution of these concerns, this paper has a bright future.

1 General comments

1.1 Challenges understanding the motivation and its relation to methodological choices

I see challenges in understanding the intended scope of the paper. What physical drivers of variability are intended to be included versus excluded? How general are the results intended to be, and how general are they? The concerns raised in this section are the most important, as I believe the paper needs to improve the alignment between its stated goals, its methods, and its conclusions.

To begin, I am concerned by how the paper treats ozone trends. It divides the period 2004-2018 into two subperiods (2004-2011 and 2012-2018), and it is stated (e.g., in the abstract at Lines 5-6) that ozone decreased during the first subperiod and increased during the second subperiod. But, Figure 1a appears to instead show a very noisy time series with no obvious trend within each subperiod. Any significant trend that was found would likely suffer from endpoint problems. The variability is dominated by the QBO, and during an 8-year period, the start and end phase of the QBO could be very important for trends.

It is further confusing how this questionable trend enters into the arguments of the paper. The Introduction is mostly dedicated to discussing decadal ozone trends, which gives the (perhaps mistaken) impression that the paper will be about processes that drive decadal ozone trends. However, in the data preprocessing step, the subperiod time series are detrended before performing causal inference (Lines 144-145). Thus, the method removes the 8-year trend that motivated the consideration of this time period, and focuses instead on residuals from that trend.

There is no guarantee that residuals from a long-term trend will exhibit the same relationship between the variables of interest as does the long-term trend. Specifically, these residuals are probably dominated by the QBO, as the paper analyzes later, and it is not guaranteed (nor likely) that the relationship between upwelling and ozone (or mediating variables) is identical in QBO-induced variability compared to other sources of variability that might be implicated in long-term trends, such as rising CO_2, rising N_2O, or declining ozone-depleting substances. This review will later return to some of these themes.

Indeed, in Section 4.4, the algorithm is tested on data that has been stratified by QBO phase, so that the variability is not necessarily driven by the QBO anymore. Despite this section discussing extensively how this is a robustness test for the methodology (Lines 353-357), my takeaway is that the DAG doesn't look very robust. It looks qualitatively different when QBO-driven variability is conditioned out (Figure 6). Comparing across the rows of Figure 6, it appears that the relationship between upwelling and N2O vanishes in both subperiods at a given QBO phase. Thus, it would seem to me that the (unconditioned) relationship between upwelling and N2O is dominated by QBO-driven variability. In 2012-2018, the DAG has no linkages in common during easterly wind shear as during westerly wind shear, and neither of these reproduces the full unconditional fit.

So, it seems that the QBO might be dominating the relationships between upwelling/N2O/NOx/ozone documented in this paper, and that these relationships might be in some sense particular to the QBO, rather than general to other mechanisms. Noting these limitations, I hope that the paper can be clarified on whether it is providing a general understanding of upwelling/N2O/NOx/ozone relationships applicable to trends vs. a specific understanding of upwelling/N2O/NOx/ozone relationships driven by the QBO. The scope of information provided by the methods should be aligned well with the stated goals of the paper, especially in the Introduction.

1.2 N_2O source variability?

During my review of the prior submission, I raised concerns about structure of the directed acyclic graph (DAG) and whether low amounts of N_2O can be understood to "cause" high amounts of NO_2. Aspects of the discussion about this question have been revised, which I appreciate, including a reference to the idea that the negative correlation between N_2O and NO_2 is an example of a Simpson's paradox (Line 242). I still have core concerns about how to understand the linkage between N_2O and NO_x in the DAG.

It seems that the linkage from N_2O to NO_x results from the superposition of (at least) two physical processes: (1) more source N_2O tends to causally lead to more NO_x (just as generally more of a reactant leads to more reaction product), and (2) more source N_2O tends to be statistically associated with less NO_x due to their mutual response to upwelling (stronger upwelling increases N_2O while decreasing NO_x). The fact that these methods find a negative relationship between N_2O and NO_x must be that upwelling-mediated co-variations in N_2O and NO_x have dominated over variability in the source of N_2O over the historical period of interest. But even if the upwelling is dominant, might there be some role for source variability in N_2O that affects the DAG structure or coefficients? How would this be represented in your method? Would it tend to make the linkage between N_2O and NO_x less negative? These questions relate again to the first question about the generality of the results: is the structure and coefficients of the DAG reflecting a mechanism-agnostic relationship between N_2O and NO_x or rather a QBO-specific relationship? (I suspect the latter.) These considerations are further relevant considering the stated motivation of the paper, because in the Introduction, the paper refers to possible trends in NOx from long-term changes in N_2O, but it is not clear that this method would be able to capture those, since the method isolates the negative relationship between N_2O and NO_x associated with QBO variability rather than the positive relationship that would be associated with an increasing trend in tropospheric N_2O.

1.3 Interpreting 0 vs. 1 month lags?

The paper ascribes significance to a seemingly small difference in the value of the lag between N_2O and NO_2 (0 months in 2004-2011 versus 1 month in 2012-2018). I am not sure whether this difference is significant or how it implies that upwelling has increased. I appreciate that a sensitivity test has been added in Appendix B, which shows that the 0 vs. 1 month lag difference is robust to some amount of variation in the hyper-parameters.

Robustness notwithstanding, I am also concerned about the claim that a change in the lag between N_2O and NO_2 implies that upwelling has increased. First, upwelling is an input to the algorithm by which the linkages are calculated, so I am not sure whether the results of the algorithm can provide an independent constraint on upwelling. Please comment on this. Second, I am confused about whether the target to explain here is the average upwelling during the subperiod (implied in Lines 286-288) or the subperiod trend in upwelling (implied by the comparison with the LOTUS subperiod trends).

1.4 Is it appropriate for this method to omit a temperature-mediated pathway between upwelling and ozone?

The paper explains some of the expert judgment required for formulating the DAG, including formulating a prior for the DAG and sometimes overruling linkages that are otherwise suggested by the algorithm. I am concerned about one such overruling, which is the direct connection between upwelling and ozone. At lines 178-9, it is stated; "*Note, however, that we included one link assumption in this study, namely that the residual vertical velocity w? does not directly influence O3.*" I assume that this overruling assumption is motivated by the fact that ozone around 10 hPa is in the photochemistry-dominated regime, and not the transport-dominated regime of the lower stratosphere. That is a good reason to block the advective connection between upwelling and ozone. However, I am concerned that this also blocks other connections between upwelling and ozone that are mediated by intermediate variables, in particular temperature. It seems plausible that an increase in upwelling can cool the stratosphere around 10 hPa, leading to an increase in ozone by changing the photochemical reaction rates. I am thinking both of the ozone source ($O+O_2+M \rightarrow O_3 + M$, which speeds up at lower temperatures) and the ozone sinks (e.g., the Chapman sink, which is not the dominant sink but is very temperature sensitive, $O + O_3 \rightarrow 2O_2$).

In the QBO context, the omitted effects of temperature are not dominant compared to NOx, but nor are they trivial. For example, Ming et al. (2025) found that NOx explained about 65% of the ozone QBO at 10 hPa, with the remaining 35% due to other processes, which likely included temperature-mediated effects (and perhaps direct advection).

It would be helpful to see a discussion of this temperature-mediated connection. On the one hand, it could be correctly argued that it is not dominant. But nor is it necessarily negligible, so it would be helpful to know whether it contributes to noise in the fit, or, less ideally, whether it would modifies the coefficients of the NOx-mediated pathway. To be specific:

- In the NOx-mediated pathway, strong upwelling leads to low NOx and high O3. Thus, this linkage in the DAG is negative.
- In the temperature-mediated pathway, strong upwelling leads to cooling that leads to higher O3. This strong upwelling also separately leads to low NOx. This would also lead to a linkage in the DAG from NOx to O3 that is negative.

Thus, the temperature-mediated pathway could be potentially amplifying the apparent strength of the NOx-mediated pathway. Is this the case? If so, this would not necessarily be a major problem, depending on the intended use of the DAG, but it would be helpful if the consequences could be discussed, or, better yet, constrained. It would be a problem to the extent that this method was claimed to have produced a generic characterization of the ozone sensitivity to upwelling perturbations, rather than a QBO-specific characterization.

2 Specific comments

- Note that it is a bit confusing that Figure 1a shows data from 2002-2011 and then 2012-2018, but these are not actually the subperiods analyzed, because the first subperiod is 2004-2011 (due to the availability of N2O data, Line 99).

- Is the end year of the subperiod inclusive or exclusive? If the end years are exclusive, then is all the data from 2011 ignored, as visually suggested by Figure 1a?

- Do the subperiods include the same number of months?

- Line 33-34: A small point, but it is stated that trends in stratospheric ozone from 2000 to 2020 are "*unrelated to changes in ODS abundances*". I think you might mean to say that ozone has declined despite ODS recovering trends, but these ODS trends could have still affected ozone (i.e., by reducing the magnitude of the decline).

- Line 67-68: "*Nonetheless, understanding the mechanisms that influence O3 behavior remains essential, as past trends do not preclude the possibility of their recurrence.*" This sentence is confusing. Do you mean that "trends that have occurred before can occur again"?

- Lines 228-229: The vertical shear of the zonal mean zonal wind is calculated as a proxy for the secondary upwelling circulation of the QBO, which is physically justified, but the shear is calculated over too narrow a vertical depth, only between 10 hPa and 12 hPa. Over this narrow a region, the shear is likely too noisy and would contain little information. I would recommend calculating the shear over a deeper layer, like 30 hPa to 10 hPa, to approximate the QBO vertical wind shear at 10 hPa. (Normally, I would recommend going above 10 hPa, but I see that 10 hPa is the top of the dataset, which is fine.)

- Line 245-246: "*It is important to note that $w^\star$ does not exert a direct causal influence on NOx, instead, its effect is mediated through N2O.*" This implies that NOx is not advected by upwelling. This would make sense if NOx equilibrates photochemically much faster than transport can generate anomalies, but this depends on whether one is thinking about NOx or NOy. Considering the long-lived chemical family of NOy (see for example the discussion in Brasseur and Solomon (2005)), of which NOx is a key constituent, transport can generate substantial anomalies. If NOy were increased due to some external perturbation, then this would increase the amount of NOy and this anomaly would be transported upwards with the flow, and it would bring along with it anomalies in NOx as well. This allows for a direct pathway for upwelling to affect NOx. This comment relates narrowly to the statement about the whether upwelling can affect NOx and broadly to the formulation of the DAG, which does not include a direct arrow from upwelling to NOx.

References

Brasseur, G. P., and S. Solomon, 2005: *Aeronomy of the Middle Atmosphere: Chemistry and Physics of the Stratosphere and Mesosphere.* Springer, Dordrecht, Netherlands.

Galytska, E., A. Rozanov, M. P. Chipperfield, S. S. Dhomse, M. Weber, C. Arosio, W. Feng, and J. P. Burrows, 2019: Dynamically controlled ozone decline in the tropical mid-stratosphere observed by SCIAMACHY. *Atmospheric Chemistry and Physics*, **19 (2)**, 767–783, doi:10.5194/ACP-19-767-2019.

Ming, A., P. Hitchcock, C. Orbe, and K. Dubé, 2025: Phase and Amplitude Relationships Between Ozone, Temperature, and Circulation in the Quasi-Biennial Oscillation. *Journal of Geophysical Research: Atmospheres*, **130 (4)**, e2024JD042 469, doi: 10.1029/2024JD042469, e2024JD042469 2024JD042469.