Review comments on egusphere-2025-4745

We thank the Reviewer for providing feedback on this manuscript. We have now fully revised the manuscript. While the methodology remains the same, the manuscript now focuses on main chemical–dynamical pathway affecting tropical middle stratospheric ozone (O_3) over the single long period 2004–2021, instead of examining shorter time periods. The revised manuscript quantifies the processes that control O_3 variability on monthly timescales, which can also influence trends over shorter intervals. To better reflect the focus of the revised manuscript, we now propose an improved title: "*Causal inference for quantifying chemical–dynamical pathways controlling tropical middle stratospheric ozone variability*". Below, we address the Reviewer's comments one by one in blue. We use the notation **P L** to refer to the changes in the revised manuscript on a specific page and line.

This study designs a framework of causal inference to diagnose combined chemical and dynamical drivers of tropical middle-stratospheric ozone variability since 2004, highlighting differences between two subperiods of 2004-2011 and 2012-2018. Satellite observations and chemistry-transport simulations, as the input of the LPCMCP model, are used to explore a physical-chemical mechanism for ozone variability in this region. This work uses a broad tools methodologically and explains the causal framework technically, and the topic and framework are promising especially in the data-driven science. But I have several concerns about how the scientific motivation aligns with the dataset preprocessing, the interpretation of key links from the causal inference results. I would recommend a major revision because of the following concerns.

General comments

- The major motivation is contrasting "decline ozone (2004-2011)" and "increasing ozone (2012-2018)". However, the analysis of causal inference process with two detrended subperiod, which removes low-frequency variability that may be part of the causal story. This leads to a mismatch between the aim of the study (interpretation of subperiod ozone changes) and what the method is actually applied to (detrended residual variation).

This has been fully revised. The analysis of two subperiods has been completely removed, and the focus is now on the full 2004–2021 period. The Introduction has been shortened and adapted to highlight the role of mechanisms controlling interannual ozone variability in the tropical middle stratosphere. We therefore removed from the manuscript Fig. 1a, which previously showed yearly ozone anomalies for the two subperiods. In the revised Introduction, we reduced the discussion of tropical ozone trends. However, the remaining discussion on tropical *middle* stratospheric ozone trends serves as the motivation to quantify the mechanism that governs interannual ozone variability.

On **P2 L49-50** we state: "*While previously discussed O_3 trends motivate this study, our objective is to quantify the mechanisms that control O_3 variability on monthly timescales and thus, that can modulate trends over limited time periods*".

- **Line 283-285:** The paper's one of the key findings is the 1 month lag for the path of $N_2O \rightarrow NO_2$ in 2012-2018, resulting from delayed NO_2 production due to the shorter N_2O residence time. One of my concern is that one month lag derives from monthly means dataset, which might be not reasonable. Because this lag could be caused by data preprocessing, like sampling, aggregation, or autocorrelation structure. Please add related arguments to clarify this point.

Since we do not analyze separate subperiods anymore, this small lag difference is no longer relevant and has been removed from the revised manuscript.

- **Line 171-172:** For the τ_{max} setting, you state that you ran sensitivity tests, which make the results more robust. But it ultimately use $\tau_{max} = 1$ in the discovery graphs, which limits the unfold of longer lag effect. Because the main narrative rely on lag interpretation, the main text should give more arguments on this point.

We thank the Reviewer for raising this point. Our revised manuscript now includes the sensitivity tests of the observed links to the choice of α_{pc} and τ_{max} for the analyzed period 2004-2021 in Fig. 5.
The graphs for the direct causal effect estimates for the observations and the TOMCAT CTM simulation shown in Fig. 4 are now based on α_{pc} = 0.05 and τ_{max} = 2. We would also like to highlight here that the detected observational graph is robust across tested α_{pc} and τ_{max} as shown in Fig. 5.

In the revised manuscript, we state the following on **P14 L308-318**: "*To ensure the robustness of the detected in Fig.4a connections in observations in the tropical middle stratosphere during the period 2004-2021, Fig. 5 demonstrates the results of the application of the causal discovery algorithm with different setups of τ_{max} (depicted on the x-axis) and α_{pc} (depicted with markers). τ_{min} set to zero to account for the contemporaneous connections. It should be noted that choosing a τ_{max} that is too low risks missing causal links with longer delays, which violates the assumption of causal sufficiency. However, choosing τ_{max} too high without mitigation can dilute the detection power of the causal algorithm. Specifically, a larger τ_{max} expands the search, as the algorithm tests more possible lagged pairs. This leads to larger conditioning sets, which can further reduce the effect size and detection power. Therefore, it is important to condition only on a few relevant variables that actually explain the* *relationship (Runge et al., 2019). Sensitivity tests from Fig. 5 show very similar results for different configurations of τ_{max} and α_{pc}. The chemical connections related to N_2O-NO_2 and NO_2-O_3 pairs are robustly detected as contemporaneous across all experiments. The dynamical connections w*-N_2O and w*-O_3 are robustly detected with a one-month lag*".

- **Line 294-298:** The paper states that LPCMCI does not robustly detect the expected chain in TOMCAT, and then uses the observation-derived graph as a fixed structure to estimate TOMCAT causal effects. Appendix C provides a useful diagnosis. This is a reasonable workflow. I confused by "given the observational causal template, does TOMCAT produce similar effect magnitudes?" Please highlight this point in the Result or Discussion.

We now add the discussion of the reasons why the causal discovery algorithm did not robustly detect the anticipated graph, in particular, the dynamical coupling of w* with the N_2O. This is now addressed on **P13 L295-304** as follows: "*Unlike the observations, in the TOMCAT CTM simulation, the causal discovery algorithm does not fully reproduce the expected chemical-dynamical coupling as outlined in the Introduction, Sect. 4.1 and as shown in Fig. 4a. In particular, the anticipated one-month lagged link from w* to N_2O is not robustly detected. We found that this occurs because O_3 exhibits very strong contemporaneous coupling with both N_2O and NO_2, such that conditioning on O_3 makes the one-month lagged w* to N_2O link statistically insignificant. As further demonstrated and discussed in Appendix B, this does not indicate a lack of dynamical coupling in the TOMCAT CTM simulation. Instead, it reflects the strong shared variability among the chemical tracers as the model uses a chemically consistent scheme for all the variables. To still assess the strength of the processes represented in TOMCAT, we did not rely on the TOMCAT-derived graph. Instead, by adopting a fixed graph structure derived from observational ground truth and expert knowledge, we can accurately estimate direct causal effects with the TOMCAT data (Fig.4b), providing a valid and pragmatic solution for quantifying model sensitivities*".

Minor Comments

1) When you say sensitivity tests were done for α_{pc} and τ_{max} consider briefly summarizing the outcome in the main text.

Given the updated flow of the revised manuscript, we have now included the sensitivity test in the Sect. 4.2 of the manuscript, see Fig. 5. We therefore include the discussion of the robustness of the detected connections on **P14 L308-318** (as addressed above).

We would also like to note that the sensitivity analysis pointed to an interesting feature related to an emerging positive link from N_2O to NO_2. On **P14 L319-323,** we discuss this as follows: "*A further analysis of the sensitivity experiments reveals an additional feature in the N_2O-NO_2 relationship. A positive one-month lagged link from N_2O to NO_2 is detected across all tested τ_{max}, but primarily at relaxed significance thresholds (α_{pc}=0.2). For τ_{max}=1, the link is also identified at α_{pc}=0.1 and 0.05. Such a positive lagged connection is physically plausible, as NO_2 is produced from N_2O, as discussed in Table 1, and a delayed response may emerge at the monthly timeseries. However, given its sensitivity to the choice of α_{pc}, this link cannot be considered a robust pathway*".