Review comments on egusphere-2025-4745

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

1) 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).

2) **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.

3) **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.

4) **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.

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.