The submitted manuscript undertakes an interesting analysis – an examination of day-of-week patterns in ozone concentrations across many US regions. The goal is to illuminate possible changes in regional photochemical environments due to transitions between NOX versus VOC sensitive ozone formation. The authors present a great deal of analysis of both observational data and model results, and reach important conclusions, e.g., "Nine (large US urban) areas have disappearing weekend effect … indicating .. that they are shifting to more NOX-limited conditions: …." However, there are shortcomings in the authors' analysis; below we discuss two of these that compromise the robustness of the author's conclusions if they cannot be adequately addressed.

Comment: First, a significant fractional "disappearance" in observed day-of-week ozone patterns must be expected simply due to the decrease in anthropogenic ozone formation driven by effective precursor emission controls implemented over the past decades. Parrish et al. (2016) show that between 1980 and 2015 in the Los Angeles urban area (i.e., California's South Coast Air Basin) the temporal dependence of the anthropogenic ozone contribution to the observed distribution of MDA8 ozone concentrations is well-described by an exponential decay with an e-folding time of 22 years. Decreases in anthropogenic ozone formation of similar magnitudes have been documented in other US regions (Parrish et al., 2022, and references therein). In the submitted manuscript, the authors consider 18 years (2002 to 2019) during which this decrease would have amounted to a factor of 2.3. Since the authors use an absolute measure of the day-of-week ozone pattern (i.e., mean difference in ozone in ppb between weekends and weekdays, their Equation 1), that measure would be expected to decrease by that same factor, or by 56% over their selected analysis period, even in the absence of any change in photochemical environment. Unless the authors can demonstrate that the disappearance of day-of-week ozone pattern is significantly greater (or lesser) than 56% of the anthropogenic contribution, those disappearances cannot be taken as evidence for a change in the photochemical environment of the ambient atmosphere.

Response: Thank you for this comment. We believe that the original "disappearing weekend effect" terminology used in our manuscript was confusing as it implied that the weekend effect became less pronounced over time when, in fact, for most areas the "weekend effect" (i.e. higher ozone on weekends than weekdays) turned into a "weekday effect" (i.e. higher ozone on weekdays than weekends) (for example see Figure 1 for Denver). Given that the sign of the WE-WD ozone difference changed for most areas, we don't think that adding the additional comparison to 56% suggested above would provide a meaningful change to the results or conclusions. We have however changed the terminology throughout the paper to "transitioning chemical regime" trend to help clarify the trend pattern we are describing.

Comment: Second, the details of the trend analysis require further discussion. The authors analyze 18 years (2002 to 2019) of observed MDA8 ozone concentrations. However, rather than analyze the 18 individual years, the authors choose to base their analysis on 5-year rolling periods (i.e., 14 different periods covering the 18-year time series). Importantly, there is a great deal of autocorrelation between the 14 different 5-year rolling means. In fact, the 18 years of observations gives fewer than 4 (i.e., 3.6) independent 5-year means. Attempts to derive trends with reliable confidence limits from such a limited number of independent data is quite uncertain. It must be fully appreciated that the use of 5-year rolling means does not improve the confidence limits of derived trends over those of trends derived from the individual years. In particular, use of the Mann-Kendall test to determine the statistical significance of the derived trends in WE-WD O3 differences will give overly optimistic results if the 14 different 5-year

rolling periods are considered to be independent data. The rolling means can give plots that apparently illustrate well-defined trends (e.g., upper right graphs in Figures 1-3), but these illustrations are misleading if the autocorrelation of the 5-year means are not fully considered. A complete discussion of the trend analysis and derivation of confidence limits based upon less than 4 independent observational data points is required; if the autocorrelation of the 5-year rolling means has not been adequately considered, then revisions are required.

Response: Thank you for this comment. We agree that the auto-correlation is an issue for the Theil-Sen trends using the 5-year rolling time periods. This is a challenge because the single-year values are subject to meteorological and other random effects which are minimized by using the five-year rolling windows. We do believe that the trends shown by the Theil-Sen method show real changes over time as they are consistent with patterns shown in Figure 10 which compares WE-WD differences in 2002-2005 vs 2015-2019 which do not have any overlapping data points. However, we think it is important to acknowledge this weakness and are thus de-emphasizing any language implying "statistical significance" of trends and rather are simply reporting P-Value ranges which still provide some indication of which areas have stronger trends. We note that reporting P-values while de-emphasizing "statistical significance" is also consistent with the recommendations of other reviewers and the approach taken in TOAR. We have added the following language to the methods section to acknowledge this issue: "Because we use a 5-year rolling window for each area, the individual data points in the trends analysis are correlated. While this should not systematically bias the calculated slopes, it will lead to lower P-values and narrower 95% confidence intervals than would be calculated if the data points were uncorrelated. However, the P-value is still informative to characterize which areas have the strongest trends. Therefore, while we do report P-values we do not rely on a strict threshold for determining statistical significance."