Comments on:

Temporal variation of northern midlatitude baseline O3: 48-year observational record challenges our understanding of tropospheric chemistry

David D. Parrish, Charles A. Mims, Richard G. Derwent, Ian C. Faloona, Henry Bowman, Tongwen Wu, Jie Zhang, Makoto Deushi, Naga Oshima

The manuscript addresses an important and timely topic: the long-term trend of baseline O_3 at northern midlatitudes and the ability of current Earth system models to reproduce observed changes. The observational record analyzed here is valuable, and the authors have assembled long-term O_3 datasets. However, I do not recommend publication in its current form because I have fundamental concerns about the conceptual and methodological framework supporting the main conclusions.

My main concern is the estimation and interpretation of baseline O_3. The manuscript defines baseline air masses as those not recently influenced by continental O_3 sources or sinks, on synoptic time scales. However, this baseline is a methodological construct that depends strongly on the selection of the sites, filtering criteria, representativeness assumptions, and temporal consistency of the data treatment. The manuscript combines datasets from different environments, including marine boundary layer sites, mountain sites, ozonesondes, and aircraft measurements, each with different sampling characteristics and filtering approaches.

This issue is particularly important because the manuscript uses the derived baseline to make broad hemispheric-scale interpretations. The argument that these sites can represent northern midlatitude baseline O_3 relies on a conceptual model of a relatively well-mixed free-tropospheric reservoir. While this may be a useful first-order approximation, I do not think the manuscript sufficiently demonstrates that this assumption is robust enough to support the strength of the conclusions. Processes such as STE, long-range transport, changes in precursor emissions, wildfires, dry deposition, and regional differences in chemical regimes can all affect the interpretation of background O_3. A more rigorous sensitivity analysis of the baseline definition and filtering procedures would be required before the central conclusions can be considered reliable.

My second major concern is the use of polynomial fits as the main framework for interpreting long-term O_3 trends. The manuscript relies heavily on cubic polynomial fits to infer the timing of the O_3 maximum, the post-2000 decrease, and the apparent slowing of that decrease. However, these polynomials are descriptive mathematical functions with no direct physical or chemical interpretation. As the authors well know, tropospheric O_3 is controlled by nonlinear chemistry involving NO_x, VOCs, methane, CO, photolysis, deposition, transport, and stratospheric transport. Given the existence of more explicit CTMs and Earth system model simulations, I find it difficult to justify why the core observational interpretation is built primarily on polynomial functions.

A related concern is that the manuscript moves from descriptive fits to causal interpretation without sufficiently linking the observed changes to explicit drivers. The discussion raises changes in anthropogenic emissions, methane, NO_x-rich *versus* NO_x-poor chemical regimes, and model biases, but the analysis does not explicitly test these mechanisms. Instead, the drivers are inferred largely from temporal patterns and comparisons with model simulations. In my view, this is insufficient to support some of the stronger claims, particularly the hypothesis that models simulate a background troposphere that is too NO_x-rich and consequently overestimate the role of some processes while underrepresenting others. This may be an interesting hypothesis, but it is not demonstrated convincingly by the analysis presented.

For these reasons, addressing individual technical aspects or adding clarifications would not be sufficient. My concerns are not mainly about presentation or isolated details, but about the methodological foundation of the manuscript. The conclusions depend critically on how baseline O_3 is defined and extracted, as well as on the use of polynomial fits to infer chemically meaningful long-term behavior. Without a more rigorous and physically grounded treatment of baseline representativeness, trend estimation, and chemical drivers, I do not consider the manuscript suitable for publication.