Supplement to Response to 2nd Referee Comment (https://doi.org/10.5194/egusphere-2026-1939-RC2)

We thank Dr. Collins for his constructive comments. We have posted an overall summary of our response to those comments, and follow here with point-by-point discussions (in plain text) of each referee comment (reproduced in *italic* text). The more important text revisions that we have made to our paper are indicated in **bold,** with direct quotes of added material in **"bold"** text within parentheses. For ease of reference, we have divided these comments into General Comments, which we number ***G1-G7***, followed by Specific Comments which we number ***S1-S17***.

General Comments

This paper presents long term in situ measurements of ozone and fits polynomials to these to characterise the trends. It further speculates on the reasons for chemistry climate models deviating from the observed trends.

G1. *The paper needs to be clearer as to what the advance on the 2020 paper is. Is it the seasonal cycle analysis?*

In our submitted manuscript, we ended the Introduction with two sentences, the penultimate of which enumerated five specific issues (i.e., advances) that we address in this paper, and the sentence before that describes these in contradistinction to the Parrish et al. (2020) manuscript. We have strengthened these sentences to read:

> "This paper extends that analysis to include up to 9 additional years of measurements in those records available by late 2025, and also includes the following additional analyses: 1) more accurate and precise quantification of the long-term ozone changes over the longer data records, 2) determination if baseline ozone continues to decrease after a maximum reached in the first decade of this century, since a continuing decrease would bring both climate and air quality benefits, 3) accurate and precise quantification of the change in baseline ozone concentrations during the period of decreased anthropogenic precursor emissions associated with the COVID-19 pandemic, 4) determination of systematic shifts in the phase and amplitude of the baseline ozone seasonal cycle, and 5) comparison of the observational analysis with similar analyses of ozone simulations from global models. **These observation-model comparisons show pronounced disagreements regarding long-term trends of baseline ozone at northern midlatitudes. We associate the observed long-term changes with concurrent precursor emission changes. As an initial approach to investigating the cause of the model shortcomings, we hypothesize that the discrepancies of the model simulations arise from shortcomings of model treatment of nitrogen oxide precursor emissions, and include some analysis of relevant cause and effect mechanisms.**"

We believe that these sentences provide a clear summary of the important advances included in our present paper, which extend well beyond the analysis in Parrish et al. (2020).

G2. *The use of polynomial fits needs much better justification and more discussion of the limitations of such an approach. A 2020 analysis by this group used quadratic fits and now acknowledges that this led to unrealistic extrapolations. Although a cubic term has now been added the graphs are not obviously cubic in form. The uncertainties in the fit parameters in table 1 are very large casting further doubt on the applicability. Forcing the data into an inappropriate function could introduce questions of validity of the results. This can be especially problematic towards the endpoints of the data (1978 and 2024). Fitting a 5th order polynomial to the model data could lead to even greater problems, the authors need to justify they have considered all these.*

Our use of polynomial fits to quantify long-term changes of atmospheric trace species is a well-established technique used in multiple publications, where it has been fully described. Therefore, for brevity, we include in the paper itself only a brief description of the equations and techniques employed in our quantitative analysis, with a more thorough discussion of these methods included in the Supplement. In order to provide the much better justification and more discussion of the limitations of our approach that the referee requests, we have added some details to that discussion in the Supplement, and provide additional discussion in this response, as detailed in the paragraphs below.

Importantly, the referee's comment gives an inaccurate description of our use of polynomial fits. First, this fitting cannot force the data into an inappropriate function. In fact the fitting process actually works in exactly the opposite

manner – the data themselves force the polynomial function to fit the long-term changes revealed by those data. Formally, fitting a polynomial function (i.e., a power series) to a data time series is akin to a Taylor series expansion of a continuous function about a reference point. This expansion provides a completely general approach to approximating the continuous function describing the long-term atmospheric changes underlying the time series with no assumption regarding its mathematical form; many calculus texts (e.g., Sokolnikoff and Redheffer, 1958, pg. 138-147) discuss this technique. We implicitly assume that there exists an unknown, continuous function that describes the long-term changes of baseline ozone in the northern mid-latitudes, and that the data time series provides information regarding that function. A polynomial fit to the time series accurately approximates that continuous function to the precision allowed by the scatter in the data with no assumptions concerning the unknown function.

The uncertainties in the fitted parameters in Table 1 can be considered large, but this does not cast any doubt on the applicability of the approach; expect for the *a* parameter value (which we expect to be zero due to the normalization) all parameter values are statistically distinct from 0 with high probability, which we interpret as providing significance to the fit. Importantly, there is a great deal of covariance between the derived parameter values. If one parameter were actually near the extreme of its uncertainty, then the other parameters would be adjusted toward the opposite ends of their uncertainty. As a result, the family of curves generated by the parameters varied properly through their uncertainty ranges would be much more similar than might be expected from the large uncertainty quoted for each parameter in isolation. The tabulated RMSD values in Table 1 of the paper give a proper evaluation of the applicability of our approach. The 3,220 monthly mean data points are fit by the overall analysis with a RMSD of 3.7 ppb, which represents the total influence of all sources of ozone variability (and any measurement error) on the monthly means beyond that captured by our seasonal cycle and long-term change analyses. This RMSD value corresponds to a variance of 13.6 ppb^2. The original monthly mean data (colored points in Fig. 1) have a variance of 97.1 ppb^2, so our analysis accounts for 86% of the total variance in the original data, firmly showing the applicability of our analysis. It is instructive to compare this scale of deviation of our very simple polynomial model from the observations (3.7 ppb) to typical values reported in the literature for full chemical transport models. For instance, Lin et al. (2024) report that summer mean MDA8 values across the Northern Hemisphere have a RMSE (root mean squared error from the observations) of 11 ppb in their latest AM4VR model (down from last version's 15 ppb). A recent meta-analysis of air quality modeling studies conducted by Sharma et al. (2017) reported a mean RMSE value from 50 studies to be 14.5 ppb (± 7.7 ppb). Even limited to 'background' sites such as in Zhang et al. (2020) the RMSE is seen to be similar for the GFDL-AM4 and GEOS-Chem models (10-15 ppb, as seen in Figs. 17 & S14). Furthermore, these comparisons exclude the typical model mean bias values of 5-15 ppb, and considers only the deviations from mean values, whereas values for higher statistical percentiles, representative of air quality standards, tend to be considerably larger.

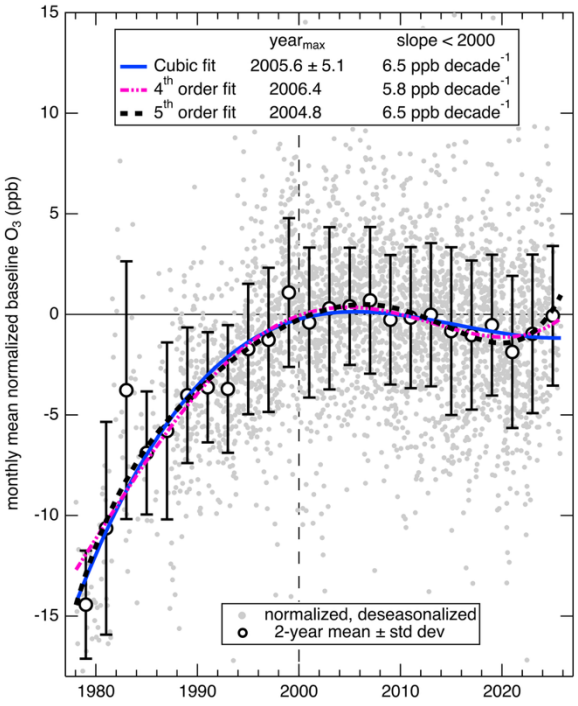
The referee is correct that fitting a polynomial to a time series of data can cause problematic effects at the beginning or ends of the data set. The figure below shows that attempts to fit a 4^{th} or 5^{th} order polynomial to the data in Fig. 2a of our paper. Problems at the end (but not the beginning) of the data are evident. However, three points are important to consider. First, the parameters of the 4^{th} and 5^{th} order terms are not statistically significant, i.e., distinct from 0 (see table below); for this reason we did not consider these fits. Second, the table shows that all common parameter values derived in cubic, 4^{th} and 5^{th} order fits agree within their confidence limits. Third, the quantitative information that we derive from the cubic fit (e.g., $year_{max}$ and slope before 2000 agree between all three fits; see annotation in the figure).

In summary, the referee's concerns do not reveal any problem in our analysis. The analysis techniques that we employ are robust, free from undisclosed artifacts, and accurately and precisely quantify the long-term changes in baseline ozone that have occurred at northern mid-latitudes.

For clarity, we have expanded the discussion of these issues in Section S3.1 of the Supplement.

Table. Parameter values (with 95% confidence limits) from polynomial fits to the time series of deseasonalized, normalized, mean ozone mole fractions from the eight baseline data sets.

parameter	Cubic	4th order	5th order
a (ppb)	-0.2 ± 0.8	-0.1 ± 1.0	-0.3 ± 1.0
b (ppb yr^{-1})	0.14 ± 0.10	0.18 ± 0.06	0.22 ± 0.16
c (10^{-3} ppb yr^{-2})	-15 ± 4	-20 ± 6	-15 ± 12
d (10^{-4} ppb yr^{-3})	3.3 ± 2.7	1.8 ± 2.7	-3.7 ± 11
e (10^{-5} ppb yr^{-3})	---	1.3 ± 2.3	-0.2 ± 2.7
f (10^{-6} ppb yr^{-3})	---	---	1.1 ± 1.8



G3. *There are similar issues with assuming a sinusoidal cycle with two harmonics as non-sinusoidal behaviour can alias onto the second harmonic. There is little evidence presented for the Gaussian functional form for the time of maximum amplitude or phase.*

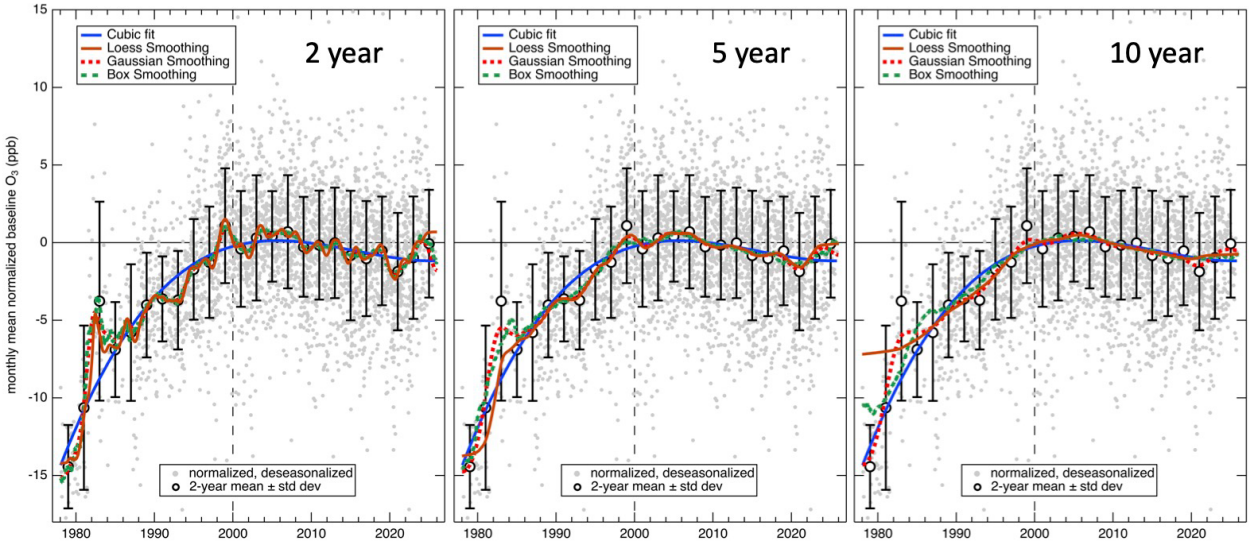
The first sentence of this comment about our Fourier series analysis of the seasonal cycle is similar to the above comment about our power series analysis, so our response is similar. Formally, fitting a Fourier series to a data time series containing a continually repeating term, such as a seasonal cycle, is fundamental to the treatment of many physical problems such as heat conduction, sound, electromagnetic waves, electric circuits, and mechanical vibrations. A Fourier series expansion provides a completely general approach to approximating the continuous function describing seasonal atmospheric changes, with no assumption regarding its particular mathematical form (beyond the formal Dirichlet conditions which are mathematically rigorous yet physically lenient); many calculus texts (e.g., Sokolnikoff and Redheffer, 1958, pg. 175-211) discuss this process. We implicitly assume that there exists an unknown, smooth, repeating function that describes the seasonal cycle of baseline ozone in the northern mid-latitudes, and that the data time series provides information regarding that function. A Fourier series fit to the time series accurately approximates that repeating function to the precision allowed by the scatter in the data with no assumptions concerning the unknown function. One advantage of a Fourier series over a power series is that the terms of the Fourier series are orthogonal functions, which means that the information in the fitted time series that determines the contribution of any one term is independent of the information that determines any other term. This means that if a Fourier series fit is extended from two terms to three terms, the analysis of the first two terms remains unchanged. In contrast, in a power series (see Table above) the terms are not orthogonal functions, so the derived parameter values differ to some extent when the number of terms in the power series changes. We do not understand what the referee means by his phrase "… *non-sinusoidal behaviour can alias onto the second harmonic.*", but since the fundamental and the second harmonic are orthogonal functions, we cannot envision how any such effect could create a problem for our analysis. In summary, the Fourier series analysis we use does not imply a sinusoidal seasonal cycle; however the theory of Fourier series analysis guarantees that any seasonal cycle can be expressed by a series of sine functions of varying amplitude and phase, as we have done.

The second sentence of this comment refers to our use of a Gaussian functional form to approximate the time variations of the amplitude and phase of the fundamental of the seasonal cycle. In our paper we reference the

description and successful application of this approach developed by Bowman et al. (2022), a reference that presents extensive evidence for the utility of this analysis approach. We believe that it is not necessary to duplicate that discussion here.

G4. *The authors should ideally use smoothing functions instead without assuming functional forms.*

The literature (and data analysis software packages) are replete with smoothing functions that assume various functional forms for smoothing, treatment of missing data, period of smoothing, treatment of data at end of record, etc. The referee does not suggest any specific smoothing approach so it is difficult for us to evaluate the "ideal" approach envisioned. In lieu of a recommendation, we have attempted to smooth the data of our Fig. 2 using 3 different smoothing functions (Loess, Binomial and Box) over 3 different time periods (2, 5 and 10 years) with other variables set to nominal values selected by our software package (Igor Pro 9 from WaveMetrics®). The Figure below compares those smoothing results with our cubic polynomial fit. In our judgement the cubic fit provides a far more informative representation of the overall long-term change of northern midlatitude baseline ozone than any of the 9 smoothing functions. In summary, we strongly disagree with the referee's comment for three reasons. First, as discussed in our response to comment **G2** above, we do not assume any functional form; the data themselves dictate the functional form of the power series cubic fit. Second, smoothing approaches do not provide any quantitative parameter values that can serve as benchmarks for comparisons of observed and simulated long-term changes of baseline ozone. Third, the graphs below show that 3 different approaches for 2-year period smoothing provide no information beyond that apparent in the 2-year averages that we analyze, and all smoothed curves would yield generally the same quantitative information that we have extracted from our cubic fit.



G5. *Some graphs (e.g. 3 (a-b), 4(b)) show the fitted curves rather than the full data which gives an overly precise impression. There will be a large spread in these.*

Please see responses to **Specific Comments S8** and **S9** below.

G6. *The reasons for the high bias model bias in the 1950s is attributed to the modelling of the transport of NOx out of the boundary layer. However no analysis of NOx is presented so this seems very speculative. How does this model comparison compare to that done in the 2020 paper? What is new here?*

We agree that our hypothesis is only poorly supported with evidence, although below we discuss additional evidence that we have added to our paper. Nevertheless, we believe that it is quite important to discuss the possible reasons that model simulations have such marked shortcomings. The poor simulation of historical changes of baseline ozone at northern midlatitudes by global models has been recognized for more than a decade (e.g., see Parrish et al., 2014).

To our knowledge, this model failure has not yet been successfully explained or corrected in any published study. Parrish et al. (2020) include no model-observation comparisons; thus, the comparisons based on the simulations of a new generation of global modes, which show very similar shortcomings as found by Parrish et al. (2014), is new in our present paper. Despite this known deficiency of models, their simulations are used in a great many studies of atmospheric composition, including attempts to understand the impact of present and future climate change. It is not an exageration to forcefully state that all conclusions of these many studies are compromised by this known model deficiency; thus, it is of great importance that the deficiency be corrected. We view our hypothesis as an initial indication of a fruitful direction in which to seek that correction. As noted in the Summary of this response, there are excellent reasons to propose this hypothesis. Importantly in our paper, we clearly indicate where we are discussing a provisional hypothesis, as opposed to the solid observational analysis that is the foundation of the paper.

Nevertheless, we thank the reviewer for this comment because it led us to seek further confirmation of our hypothesis in the extant literature. Over the past few years there have been several studies that support our hypothesis of excessive free tropospheric NOx in many chemical transport models, and we have now included a discussion of these studies in Section 4.6: "Hypothesis: Uncertainty in NOx concentration field limits current model simulations". This new text is reproduced below:

> "To our knowledge, similar issues remain inadequately investigated for the photochemistry of dispersing aircraft and lightning stroke plumes in the FT. **For instance, recent investigations have called into question absolute rates of lightning NOx production given the colocation of lightning HO_x also produced in electrified cloud systems (Mao et al., 2021) which can significantly decrease the lifetime of NOx, although large uncertainty remains due to the severely complicated details of turbulent mixing in the environment within which such emissions are embedded.**
>
> **Looking to recent comparisons of simulated free tropospheric NOx with observations we find four compelling examples of systematic overestimates of NOx by current chemical transport models. Travis et al. (2016) present median profiles of NOx on 4 flights over the Gulf of Mexico during the NASA SEAC^{4}RS campaign (Studies of Emissions, Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys) which shows a ~50% overestimation by GEOS-Chem in the region from 2-7 km (their Fig. 11). Ye et al. (2023) compare distributions of GEOS-Chem NO_2 versus those observed onboard the NCAR C-130 during 19 research flights of the Southeast Atmosphere Study indicating a factor of 2 overestimate of the median value in the free troposphere (their Fig. 2a). In both of these studies, unlike the free troposphere, the agreement within the boundary layers was very close. Using a different observational technique altogether whereby satellite measurements are compiled from cloudy pixels with varying cloud top heights to derive average NO_2 profiles (so-called 'cloud slicing'), Horner et al. (2024) found that GEOS-Chem overestimated mid-tropospheric NO_2 by 20–50 pptv in northern midlatitudes based on TROPOspheric Monitoring Instrument (TROPOMI) data (their Fig. 8). A similar technique applied to the geostationary satellite mission Tropospheric Emissions: Monitoring of Pollution (TEMPO) data by Dang et al. (2024) showed that GEOS-CF consistently overestimates NO_2 between 300 - 700 hPa across North America by ~30 % (27 ppt instead of 20 ppt, their Figure 1).** These considerations lead us to hypothesize that significant errors arise in model simulations of ozone due to inadequate simulation of the NOx concentration field, specifically an overestimate of NOx in the background troposphere."

G7. *The statement that the impact of methane is less than simulated by models is presented in the abstract, but there is very little evidence in the paper to back this up. I worry that if this paper were published this sentence could be taken and quoted repeatedly as a peer-reviewed finding. I urge the authors to remove this from the abstract.*

We appreciate the referee's concern. We have modified the final sentence of the abstract to read: **"We hypothesize that models simulate a background troposphere that is too NOx-rich, which could account for the deficiency in the model simulations of long-term ozone changes."**

Specific Comments

S1. *Line 21 "concentrations"->"mole fraction" and many times in the paper.*

Thank you. Generally, we correctly use "concentration" in the sense of the 3rd simple meaning given by Merriam-Webster (https://www.merriam-webster.com/dictionary/concentration; last accessed 28 May 2026) i.e., "the amount of a component in a given area or volume". The term "mole-fraction" is the particular quantitative measure of concentration that we use. We have carefully reviewed our usage of "concentration" in the manuscript and supplement. All instances where the use of "concentration" may cause some ambiguity, have been changed to "mole-fraction".

S2. *Line 29-30 This "lesser role" for methane is presented without evidence. Please remove this sentence from the abstract.*

As indicated in our response to **Comment G7** above, we have modified this sentence in the abstract as requested.

S3. *Line 83 It would be useful to show a map of the measurement sites*

Thank you for this comment. We have added a map of the measurement sites (Fig. S8) to Section S3 of the Supplement.

S4. *Lines 150-160 There is insufficient justification for these formula. Retrieving the maximum year from the fit parameters which are so uncertain seems very unwise.*

The techniques for observational analysis of tropospheric ozone measurements that we have developed have been described in multiple publications over the past 4 decades. References to several of those publications are given in our manuscript, and the references therein cite many others. Section S3 of the Supplement gives a more detailed justification for the formula we use in this manuscript. In our view, the formulae are sufficiently justified in the manuscript including the Supplement, and more details are available to the interested reader in the references.

Retrieving the maximum year from the fit parameters, despite their uncertainty, is indeed wise, because that retrieval is fully justified so long as the confidence limits of those fit parameters are duly considered in a propagation of errors treatment, which we have done. We quantify that uncertainty with 95% confidence limits quoted in Table S1. The retrievals of the maximum year from all data sets (see Table S1) are usually quantified within useful confidence limits, i.e., the 95% confidence limits included in Table S1 from a standard propagation of errors treatment. Of particular note is that the $year_{max}$ values from all baseline data sets in Table S1 agree with 2005.2 ± 5.1. That confidence limit is derived from the fit to the 2-yr averages of normalized, deseasonalized monthly means from all eight data sets, which we judge to represent the best treatment of autocorrelation in the data.

Finally, in our response to **Comment G2** above, we include a figure with an annotation that shows the $year_{max}$ values derived from cubic, 4th order and 5th order polynomials. All values, including that derived from a quadratic fit by Parrish et al. (2020) – see Table 1 of the manuscript - agree (within quoted confidence limits). Our analysis and the derived quantitative results are firmly supported in all respects.

S5. *Lines 206-207 "rate of decrease … accelerating" I suggest rephrasing as this could imply the second derivative of the rate – i.e. 3rd derivative of the mole fraction.*

Thank you for catching this issue. This sentence has been changed to "**This indicates that the decrease in baseline ozone concentrations, which is seen in all of the fitted curves, is no longer accelerating.**"

S6. *Figure 2. The uncertainty in the cubic and linear fits need to be shown. For the cubic, the slopes around the end points (around 1980 and 2020) will likely vary wildly according to the fit parameters. Similarly the range of slopes of the linear fits need to be shown to be able to judge between the linear and cubic formulations.*

Thank you for this suggestion. The uncertainties of all fits are given in Table 1 of our submitted manuscript. We have kept that table, but now we also explicitly quote them in the text. For the linear fits, the first two sentences in the paragraph before Fig. 2 have been revised to read: "The piece-wise linear fits included in Fig. 2 provide an alternate, but consistent, quantification of the long-term baseline ozone changes. Over the 22 years before 2000 a

marked increase occurred at an average rate **(i.e., slope)** of 5.9 ± 2.3 ppb per decade, followed over the next 26 years by a much slower decrease (**slope** = -0.53 ± 0.47 ppb per decade)." For the cubic fits, the following sentence now follows those two: "**The cubic fit in Fig. 2 indicates mean trends of 6.5 and -0.37 ppb per decade for the earlier and later periods, respectively, which agree with the linear fits within their confidence limits.**" Thus, the cubic and linear formulations are consistent, and any variation of the slopes around the end points of the cubic do not affect this analysis to a statistically significant extent.

Note that the figure included in our response to **Comment G2** above shows that the fitted curves around the end points (primarily around 2020) do vary, but as discussed in that response, that variation has no effect on any quantitative discussion in our paper.

S7. *Line 267 What is the justification for assuming sinusoidal behaviour? If the shape of the cycle is not sinusoidal (i.e. sharply peaking in summer, but a broader minimum in winter) this can introduce spurious coefficients in the second harmonic. What evidence (beyond the fitting) is there that there should be a secondary peak?*

As discussed in our response to general comment ***G3***, no assumption of sinusoidal behavior is made. Indeed, in some data sets, both the overall shape of the seasonal cycle, and the sum of sinusoidal terms of differing frequency and phase fit to the seasonal cycle are certainly not sinusoidal. The fundamental and the secondary harmonic are orthogonal functions, so no spurious coefficients can be introduced in the second harmonic.

In Section 3.3 (which began on line 267) we do not discuss any secondary peaks. The figure in that section (Fig. 3) does show that one site's seasonal cycle has a clear secondary peak - that derived from the Lassen NP data set. Fitting data to a Fourier series, keeping only statistically significant terms, provides excellent evidence (actually, we argue the best possible evidence from the available data) for any secondary peak in the seasonal cycles quantified from measurements. Another line of reasoning (which is inferior to the fitting to a Fourier series) can provide additional evidence by simply calculating monthly means and their confidence limits, as tabulated below for the Lassen NP data set.

month	Jan	Feb	Mar	**Apr**	May	June	July	**Aug**	Sep	Oct	Nov	dec
mean	37.6	40.1	41.3	**42.2**	41.0	40.2	43.1	**44.2**	40.9	36.1	33.7	36.1
Conf. limit	0.50	0.42	0.76	**0.64**	0.57	0.64	0.70	**0.76**	0.67	0.59	0.58	0.50

This table suggests a secondary peak in spring (April) and a primary peak in summer (August), which is in accord with the Lassen NP fit in Fig. 3. However, the confidence limits on those monthly means indicates that the magnitude of the secondary peak above the valley between the two peaks may not be statistically significant. Further, it must be realized that quantification of seasonal cycles by simply examining monthly means amounts to an over-fitting of the data. The Fourier series indicates that only the fundamental and second harmonic are statistically significant, which provide only 4 independent variables to quantify the seasonal cycle (actually 5 when the annual mean, i.e., our *a* parameter value is also considered). The 12 monthly means constitute 12 independent variables, but they are not statistically independent due to autocorrelation between monthly means. Determining if a secondary peak is a statistically significant feature of a measurement data set is a non-trivial effort, which we have not undertaken in our paper.

S8. *Figure 3: These are simply plots of the functional fits without any uncertainties. This gives an unrealistic impression of the precision of the fitting. The detrended observed seasonal cycles should also be shown to justify whether these fits are realistic.*

Thank you for this comment. Figures S7-S10 of the Supplement of Parrish et al. (2020) (https://agupubs.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1029%2F2019JD031908&file=jgrd56322-sup-0001-2019JD031908-SI.pdf) show the Fourier series fits to the detrended seasonal cycles of the 8 data sets considered in our manuscript. These figures do indeed justify that these fits are realistic. All derived parameter values annotated in those figures agree (within statistical confidence limits) with the corresponding values in Table S2 of our present paper. We see no need to repeat these figures; instead we have expanded the discussion of Fig. 3 to

read: "**Parrish et al. (2020) give a detailed discussion of the small numbers of Fourier series terms that are statistically justified for inclusion, and their Figs S7-S10 illustrate the fits of the fundamental and second harmonic terms to the detrended observed monthly means over the shorter time series of the same data sets examined here; all derived parameter values annotated in those figures agree (within statistical confidence limits) with the corresponding values in Table S2.**"

S9. *Figure 4: Panel (b) needs to represent the uncertainty somehow. It is claimed that the timing of the peaks moves, but is this statistically significant?*

Thank you for this comment. However, we believe that we have already adequately represented the uncertainty in the analysis of the seasonal cycle shift. Figure 4a shows the general increase in the amplitude of the seasonal cycle early in the data record, followed by the general decrease later in the record. The change in the amplitude of the fundamental (i.e., the value of the *r* parameter) is indicated in Table 4 as 3.2 ± 1.9 ppb, which is statistically significant at greater than the 3-σ level. Figure 4b illustrates the small change in the phase of the seasonal cycle. (i.e., the timing of the peak). The change in the phase of the fundamental (i.e., the value of the r_ϕ parameter) is indicated in Table 4 as 10 ± 13 days; it is statistically less certain, significant only at the 0.7-σ level; nevertheless, we believe that it is worthwhile to indicate this likely significant shift in the phase of the fundamental.

S10. *Line 355: The authors need to acknowledge the likely misleading end effects throughout the paper.*

To the best of our understanding, the influence of misleading end effects is limited to the discussion included in the line cited by the referee. Our response to **Comment G2** above shows that the end effects have no significant impact elsewhere in the paper, so no further acknowledgement has been added.

S11. *Line 394: The fit showing a shift in the harmonic before the data starts is just a fitting artifact and cannot tell us anything about what actually happened then.*

This is a minor point. We do not actually discuss the fit before the data starts, so there is no fitting artifact in our work. The fit we show does have a positive slope at the start of the time series of measured data, so our statement on line 394 ("This fit indicates that the shift in the fundamental harmonic began before the 1978 start of measurements, ….") is accurate and informative.

S12. *Line 395: Extrapolating the fit to pre-industrial times is completely unjustified.*

We have changed the wording of the sentence ending on line 395 to read: **"This fit indicates that the shift in the fundamental harmonic began before the 1978 start of measurements, maximized in 1990 ± 6 years, and by 2025 had nearly returned to unshifted seasonal cycle as determined in that fit."**

S13. *Figure 5: Again just showing the fits without uncertainty conveys greater precision than is justified. In panel (b) the year_max will be extremely sensitive to the model parameters even with a cubic fit.*

Figure 5b does not convey greater precision than is justified. The annotation in Fig. 5b gives the derived year_{max} values with 95% confidence limits propagated from the confidence limits of the parameter values from which those values are calculated; thus, the precision is directly quantified. Those confidence limits do indicate that the year_{max} values are sensitive to the model parameters; however, the important point with regard to these year_{max} values is stated in two sentences in our paper: "The year of the occurrence of the maximum of each fit to the simulations and observations (year_{max}) are annotated in Fig. 5b. The weighted mean of the model simulations is 2010 ± 6 years, which agrees with the observational-based estimate of 2006 ± 4 years derived from the fit illustrated in Figs. 1 and 2." Further, the figure included above in our response to **Comment G2** shows that the derived year_{max} values are insensitive to the number of polynomial terms included in our fits.

S14. *Line 466: "4th" order - the figure says 5th order.*

Thank you for catching this error. These are all 4th order fits. The figure caption has been corrected.

S15. *Line 517: This needs more analysis on the significance of the differences between the two locations in the models and the lack of difference in the observations. This is crucial given the inferences made later in this paper*

The spatial difference in the model simulations of the FT ozone seasonal cycle, and the lack of such differences in the observations, is significant in itself. This remains one of the shortcomings of the models which have yet to be resolved. Our purpose in this paper is to identify such shortcomings, and to discuss some of their implications - the discussion of which follows in the paragraph the referee cites. We agree with the referee that these differences need more analysis, but we are not prepared to include such analysis in this paper.

S16. *Figure 8: There needs to be discussion of the validity of the Gaussian forms for the time evolution of the amplitude and the seasonal max. The lines here do not show any uncertainty therefore it is not possible to tell whether the difference between model and measurements are real or a fitting artifact.*

For discussion of the validity of the Gaussian forms for the time evolution of the amplitude and the seasonal max**.,** please see the 2nd paragraph of our response to **Comment G3.** Briefly that discussion is given by Bowmn et al., (2022), which we reference. In the discussion of Figure 8 in the text, it is stated that Table 4 compares the parameter values derived in the fits to the simulations and measurements, so the statistical significance of the difference between model and measurements can be judged by the reader. Generally the simulations and measurements do show differences, but agree within their derived confidence limits.

S17. *Sections 4.6 and 5 are mostly speculative and make statements on the modelling of NOx without analysing this. They should be removed.*

Here we again strongly disagree with the referee and have kept both of these sections. Section 4.6 discusses the hypothesis that we believe is quite important for the reasons given in the 3rd bullet of the **Summary** posted separately. We have further supported this hypothesis by referencing four recent model/observation comparisons that all indicate a systematic overestimation of NOx in the northern midlatitude free troposphere. Section 5 notes that a hierarchy of models is required to fully understand tropospheric composition. Indeed, any scientific question cannot be considered to be understood unless it can be answered, at least qualitatively, through a simple physical model. This section makes that point for our understanding of atmospheric chemistry.

References not included in our revised manuscript

Lin, M., Horowitz, L. W., Zhao, M., Harris, L., Ginoux, P., Dunne, J., ... & Zhou, L. (2024). The GFDL variable-resolution global chemistry-climate model for research at the Nexus of US climate and air quality extremes. Journal of Advances in Modeling Earth Systems, 16(4), e2023MS003984.

Sharma, S., Sharma, P., and Khare, M.: Photochemical transport modelling of tropospheric ozone: A review. Atmospheric Environment, 159, 34-54, 2017.

Sokolnikoff, I.S. and Redheffer, R.M.: Mathematics of Physics and Modern Engineering, McGraw Hill, New York, NY, 1958.

Zhang, L., Lin, M., Langford, A. O., Horowitz, L. W., Senff, C. J., Klovenski, E., ... and Conley, S.: Characterizing sources of high surface ozone events in the southwestern US with intensive field measurements and two global models. Atmospheric Chemistry and Physics, 20(17), 10379-10400, 2020.