

Please find our comments (non-italicized) to the following author comments (italics) addressed below:

The manuscript describes a comparison of P(O3) measured using a second generation MOPS instrument and P(O3) modelled using a detailed chemistry mechanism (MCMv3.3.1) and lumped chemistry scheme (RACM2). A higher measured than modelled P(O3) is reported during the morning hours and the authors postulate that this discrepancy could be caused by a model under-estimation of HO2 under high NOx conditions and present a possible reaction which cycles OH to HO2 via NO that could rectify the discrepancy. The manuscript concludes with a discussion on how these model uncertainties may influence ozone reduction strategies in the region. From reading this manuscript in detail, I have not been convinced that the differences reported between modelled and measured P(O3) are not, in part, caused by instrumental issues (I have detailed my specific concerns below) and to a smaller extent, caused by uncertainties in the model constraints (e.g. the interpolation of the VOC measurements and the use of a constant, median VOC value for certain VOCs). I cannot recommend publication in ACPD before the specific comments I have raised below are adequately addressed.

We thank the reviewer for presenting several important points to be addressed in the manuscript. We note that we have adjusted the manuscript focus to address peroxy radical (HO₂ and RO₂) discrepancies at high NO_x and have performed additional model case studies exploring the plausibility of missing peroxy radical sources. Since we proposed the reaction OH + NO (+ O₂) to form HO₂ + NO₂ earlier in 2016, preliminary work in our laboratory and that of a colleague, as well as a theoretical study, have all indicated that this reaction is unlikely. We have thus decided to omit this reaction as an explanation to the model-data mismatch described in the main text. We have instead explored the plausibility of missing peroxy radicals and other chemical causes.

Throughout the manuscript, the authors argue that the higher observed P(O3) than modelled P(O3) at high NO is supported by observations of higher (than modelled) HO2 under high NO conditions. Only recently, however, have HO2 measurements been reported in the literature that have not been influenced by artefact signals induced by the fast decomposition (within instruments) of certain RO2 species. The authors do not discuss recent observations and model comparisons made in China (Tan et al., 2017) where interference-free HO2 observations agree well with model predictions under elevated NO conditions. The authors need to provide a more balanced discussion of modelled and measured HO2 comparisons under high NO conditions which includes a review of the possible problems relating to HO2 measurements as well as the more recent HO2 modelled and measured comparisons.

The introduction has been modified to include a) known HO₂ measurement interferences and b) an extended literature review of HO₂ measurement comparisons with models including more recent studies. This extension of the introduction includes a synthesis of comparisons made in high and low-NO environments in various studies (Stone et al. (2012)) as well as other studies

such as Tan et al. (2017) and Hofzumahaus et al. (2009), which show small and/or insignificant model HO₂ underprediction during morning hours. However, Tan et al. (2017) HO₂ measurements do not suffer from the RO₂ interference and yet still exhibits slight underprediction of HO₂ at high NO_x. The highest NO values reported for some of these studies was only a few ppbv as in Tan et al., 2017. For many studies, the separation between measured and modeled HO₂ is within uncertainties at these low NO levels but then grows at higher NO levels, particularly above 10 ppbv. Thus, even when accounting for HO₂ interferences, model underestimation of HO₂ at high NO_x appears to be robust. If the ratio of atmospheric RO₂ to HO₂ is used to predict a maximum HO₂ interference, this interference can be approximately a factor of two higher than HO₂. Thus, an HO₂ interference is not the sole cause for model-data HO₂ mismatch at high NO_x presented in the literature studies outlined in the introduction. These ideas have also been added to Pg. 3 lines 1-25.

We have also modified the introduction to describe model RO₂ underestimation, presenting this as a viable reason for model P(O₃) underprediction. Model peroxy radical underestimation at high NO is therefore outlined throughout the rest of the paper as a possible cause for model underestimation of P(O₃) and case studies are presented in the revised manuscript.

Pg 5, line 8 -16: The authors discuss the influence of humidity on the O3 analyser used in the MOPS instrument and cite previous papers which have reported large changes in O3 concentration as RH is rapidly changed. How fast and by what percentage did RH typically change during these ambient measurements made in Golden? Did the authors attempt to correct for this RH dependence?

Additional laboratory studies have been conducted to address this comment. Please see our comments for Reviewer 1 that address this issue.

Pg 5, line 23: 'some excess HONO was measured..' how much and how did this vary with RH, T, [NO2] and actinic flux? Would this excess HONO be expected to be greatest during the morning hours when NO2 concentrations were high? Details like this need to be provided so the differences reported between modelled and measured P(O3) during the morning hours can be fairly assessed.

We would expect the HONO bias to be greatest during the morning hours when NO_x, relative humidity, and actinic flux is high. However, we would expect that the P(O₃) bias due to excess HONO production in Golden, CO be considerably smaller as we have been able to decrease this bias ~ 30% in State College, PA (Baier et al. 2015) by covering the MOPS inlet face to avoid excess HONO production and photolysis, and as NO_x levels Golden, CO are considerably smaller than in Houston, TX. The relative humidity and NO_x levels in Golden, CO were significantly lower than in Houston, TX, while the actinic flux is similar.

Model calculations of the MOPS HONO bias using the measured chamber HONO to observed NO_x ratio in Houston, TX and scaling this ratio to the observed NO_x in Golden, CO indicate a bias of 3 ppbv/h. However, in this estimation, we did not consider that only a small fraction of

chamber-generated HONO would be photolyzed within the residence time in the chamber, which makes this potential positive P(O₃) interference less than 0.5 ppbv/h and thus insignificant. We expect that, since this bias is considerably smaller than the difference between measured and modeled P(O₃) shown in Figure 1, the model-measurement P(O₃) mismatch is not due to this artifact.

We have added lines Pg. 6 line 32 to Pg. 7 line 5 describing what we have done to decrease the HONO bias. We have added how we have calculated this potential HONO bias for Golden, CO on Pg. 11 line 30 to Pg. 12 line 8.

Pg 5, line 24: ‘..actions were taken’. Some details on the actions that were taken should be provided here.

Please see previous comment.

Pg 6, discussion of model constraints: I have some concerns over how well the model constraints represent the local conditions in Golden. How many WAC samples were collected between 0700 – 1200? I worry that interpolating these canister samples may lead to underestimations in peak VOC concentrations.

A total of 46 whole-air canister samples were taken between 0700-1200 LT. Therefore, over 64% of the WAC samples were taken during the morning hours. As continuous VOC measurements were not available in Golden, CO, it is not possible to determine median diurnal values of these VOCs in any other manner. We have added the following text to lines 18-24 on Pg. 7:

“In the absence of continuous ground-based VOC measurements, C₂-C₁₀ non-methane hydrocarbons (NMHC) and organic nitrates were measured from 72 total whole-air canister (WAC) samples that were collected in Golden and analyzed by gas chromatography (GC) and gas chromatography-mass spectrometry (GC-MS) in the laboratory. An average of five samples were taken daily over 16 days, with approximately 64% of sampling occurring between 0700 and 1200 local time (LT) to capture VOC mixing ratios during morning O₃ production hours, and sparser sampling in the afternoon between 1400 and 1800 LT to examine advection from sources east of Golden, CO such as the Denver metropolitan and Commerce City regions.”

In addition, Shirley et al. (2006) employ a similar approach to estimating VOC abundances, but use OH reactivity measurements to calculate VOC speciation abundances. Given several model sensitivity runs varying these VOCs, we expect that the estimation of VOC abundances in this manuscript can adequately capture diurnal variations -- in an average sense -- in order to compare hourly measured and modeled P(O₃) rates in Figure 2.

In table S1 the authors state that constant, median values of various species measured on board the aircrafts were used to constrain the model. How can this be a valid approach given the high variability of some of these species, e.g. the biogenics? Would this variability bias the model predictions as a function of time of day? I.e. were VOC concentrations highest

during the morning when the boundary layer was lowest? It is unclear if the sensitivity analysis conducted (presented in Fig S1) included varying VOC concentrations by the observed variability or varying by the uncertainty in the VOC measurement only?

Aircraft measurements in the boundary layer for areas surrounding Golden, CO were available after 9am LT for the P-3B and after 10am LT for the C-130. Thus, the concentrations of aircraft VOCs used in this analysis are limited to mainly after 9-10am LT and early morning concentrations were not known. Because of this limited aircraft data, the few morning VOC species that were measured on aircraft are set at constant, median values, while the majority of VOCs are taken directly from whole-air canister samples as described above. For those chemical species measured in common between the aircraft and the ground site, the agreement is within 30% on average. Further, biogenic chemical species do not have a substantial role in OH reactivity in Golden and so would have a smaller impact on modeled $P(O_3)$. Thus, the most important chemical species measurements for ozone production came from the whole air sampler on the ground. We have added Pg. 7 line 33 to Pg. 8 line 5 to clarify the few aircraft points that we have available to incorporate into the model:

“Canister VOCs were supplemented by boundary layer inorganic and organic chemical species measurements obtained on the NASA P-3B and NSF/NCAR C-130 aircraft and constant, median values were calculated for the limited times of the day when these aircraft were in the vicinity of Golden and used in the model (Table 1, Table S1). Aircraft measurements for Golden were available after 0900 LT on P-3B overflights, which occurred up to three times daily, while C-130 measurements were available after 1000 LT when this aircraft was within roughly 20 km of the measurement site. Airborne measurements of inorganic and organic species agree to within 30% on average.”

The analysis in Table S1 is conducted by varying the VOC concentrations by their uncertainty only in order to determine the total model uncertainty that is due to parameter uncertainties as in Chen et al (2010). This analysis was conducted to solely investigate the dependence of model uncertainty on model parameter uncertainties.

Pg 7, line 1: ‘..integrated for 24 hours..’ I would be surprised if the reactive intermediates and, more importantly, modelled HO₂ and RO₂ are in steady state after 24 hours. By what percentage did the modelled peroxy radical concentrations change from day 1 to day 2?

Additional MCMv331 model runs have been conducted that vary the integration time from one to three days with little difference seen between model HO₂, RO₂, or OH for these three different model integration times. Therefore, we do not expect these species to vary significantly whether we use a one-, two-, or three-day integration period for model runs. We have further explained this result on P8 lines 16-18 and deem a 24-hour integration period for these species acceptable:

“In addition, a one-day integration time is calculated to be sufficient for radical concentrations

and intermediate species to reach steady-state as a two-fold or even three-fold increase in this integration time period does not impact radical concentrations or the $P(O_3)$ results described below.”

Pg 8, line 19: How does the diel measured $P(O_3)$ change if the negative $P(O_3)$ points are not omitted? By omitting these negative points, does this not positively bias the $P(O_3)$ measurements in the morning when these negative points are most frequent? Were the modelled $P(O_3)$ calculated at the same time also removed? Could the authors be more quantitative when they discuss the O_3 analyser RH dependence – what do they class as a ‘drastic change’ in RH?

Please see our comments for Reviewer 1 on this issue.

Section 3.2 would benefit from Rate of Production Analysis. This would help to reveal what is driving the high modelled $P(O_3)$ on certain days. It would also reveal the cause of the differences that were observed between RACM and MCM at times rather than rely on speculation ‘..perhaps due to more explicit treatment of VOCs..’

RACM2 and MCMv331 produce essentially the same $P(O_3)$ for all but a few short time periods. This model difference is far from the focus of this paper. Therefore, we have dropped the discussion of these differences and their possible cause in order to keep the focus on the differences between the measured and modeled $P(O_3)$ which is unaffected by these infrequent RACM2 – MCMv331 differences.

Pg 9, lines 6-16: The authors give some reasons as to why they have confidence in the MOPS measurements during the morning hours when the model and measurements do not agree. I am not convinced, however, that the MOPS measurements are not positively biased at this time: By removing the negative $P(O_3)$ values which most frequently occurred during the morning hours, I would expect this to positively bias the measurements at this time even if a median rather than the mean value is used. Furthermore, I would expect heterogeneous HONO production to be greatest when NO_2 concentrations are high, and, this would also artificially positively bias the measurements during the morning most.

Please see comments for Reviewer 1 describing new correction techniques and providing evidence for the MOPS measurements being real in the mid-morning. Our additional analyses show that the MOPS diurnal pattern is rather robust for whichever correction to the raw $P(O_3)$ data that we apply.

Please also see our comments above as to why we believe that HONO production (whether gas-phase or heterogeneous) is not a major issue in the MOPS chambers causing the measured $P(O_3)$ to deviate from modeled $P(O_3)$ in the morning. Deviations of $P(O_3)$ from the average zero correction are seen in the MOPS raw data as stated in lines 15-17 on Pg. 6. These deviations from the MOPS raw data are greater than +3 ppbv/h: our calculated HONO bias from model

simulations. If we account for the amount of HONO that could possibly be photolyzed during the chamber residence time, this bias drops to less than 0.5 ppbv/h as mentioned above.

Pg 10, lines 1-5: The measurement bias reported due to photolytic HONO production is of the same magnitude as the discrepancy between model and measured P(O₃) observed during the morning (Fig 2). Although the authors argue that individual modelled and measured P(O₃) deviations are often larger than the bias caused by photolytic HONO production, there seems to be, from studying figure 1, a similar number of positive and negative deviations of the MOPS P(O₃) from the modelled P(O₃) suggesting that many of these deviations are caused by noise rather than missing or uncertain model chemistry.

Please see our comments for Reviewer 1 regarding the noise level on the MOPS data. By limiting the MOPS measurement time periods to those with ambient relative humidity below 70%, a majority of the observed P(O₃) noise is reduced. In addition, we average the MOPS data in Figure 1 in the main text to one hour. With this averaging, noisy data is also reduced and deviations from the hourly MOPS P(O₃) still exhibit differences relative to modeled P(O₃) that are larger than the HONO biases we estimate.

Section 3.3.4: A rate of $9-15 \times 10^{-11} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ for the postulated reaction 'OH+NO (+O₂) > HO₂ + NO₂' cannot be justified given that the known reaction rate for OH reacting with NO is two to three times slower. The impact of this highly speculative reaction is presented as positively influencing the modelled P(O₃) in line with the MOPS observations - this is misleading, however: Figure S2 highlights that although inclusion of this reaction, proceeding at a rate of $9-15 \times 10^{-11} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$, improves the modelled and measured agreement in the morning, it actually worsens the agreement between the modelled and measured P(O₃) during the afternoon; there is no comment about this in the text, however. How does this reaction influence other modelled species, e.g. OH?

The reviewer is correct that the formation rate of the OH+NO adduct is known, but it has not been measured in atmospheric pressure air. It is possible that O₂ could act in some way to speed up this reaction. However, preliminary laboratory studies by us and another group and theoretical work have indicated that this reaction is unlikely to be fast enough even in the presence of oxygen. We have omitted this discussion as a possible explanation for the model-data mismatch discussed in this manuscript.

Section 3.3.6: Some discussion of the modelled HONO sources is needed here. If only gas-phase sources, i.e. OH+NO are considered, the model likely under-estimates the actual HONO that was present.

It is possible for the model to underestimate HONO. However, studies that have used measured HONO to constrain the models HONO still cannot resolve the HO₂ discrepancy at high NO_x, providing some evidence that a further HO₂ source at high NO_x may still be needed to resolve

this issue. HONO levels needed to match the MOPS P(O₃) are a factor of two too high as discussed in comments to Reviewer 1.

We have significantly modified section 3.3.4 on Pgs. 14-15 to discuss model HONO studies and sources of HONO present in the models used here.

Reference Tan et al., Atmos. Chem. Phys., 17, 663-690, 2017

This reference has been added to Pg. 4 in the Introduction.