

### **Author's reply to Referee #3**

The authors highly appreciate the thorough review and the valuable comments for improving the quality of the paper. We also like to thank for the recommendation for publication. We revised the paper following the suggestions and comments and include below a detailed reply to the individual points.

*This paper aims to characterize chemical Boundary Conditions (BC) on the West coast of the US and evaluate its impact on regional air quality modeling. To do so, aircraft in-situ measurement taken during the ARCTAS campaign are compared to global model (MOZART) concentrations (as global models are used to provide BC to regional models). Comparison is further conducted with satellite data (TES) and MOZAIC data over 4 years. Once chemical BC are characterized and differences in observation and simulation are analyzed, authors evaluate the impact of those BC on regional model results.*

*Impact of BC on regional modeling is an important topic, related to long-range transport problematic. Several older studies focus on the difference between BC from different global model, and their impact on regional modeling. Here the originality is the characterization of BC using measurement, and the comparison with global model simulation. Several measurement (including relatively new dataset and also long time period dataset) are used and their comparison with model results are well explained with Figures but also Tables. The paper is, on general, very well written and clear. I would recommend this paper for publication in ACP once the following comments have been addressed:*

*Page 28923, line 26: "The combined variability of CO in AVGRegionTime is up to about 40 ppbv". I guess author used "variability" for "standard deviation". Is it a common used for "variability" ? If not, you should simply indicate what "variability" stands for.*

We use the term "variability" to describe the standard deviation and state this in the revised manuscript (Section 3.1.3).

*Page 28919, line 20: Here is mentioned for the first time, the exclusion of Californian forest fire air masses from the dataset. As its exclusion is further explained section 3.1.3, you can make here a reference to section 3.1.3.*

We acknowledge this recommendation and include an according reference in Section 3.1.1. of the revised manuscript.

*Section 3.1.4 : Table 2 shows statistics for the 10th, 50th and 90th percentiles. However 50th percentile is never mentioned when the 99th percentile is described in the text. This 99th percentile is useful as it probably represents long-range transport events with highly polluted plumes, as observed in the measurement. It may be interesting for the reader to see the 99th percentiles and not the 50th one in Table 2. By showing this 99th percentile, you can be more clear on your "CO discussion": an overestimation of CO background and low CO value, and an underestimation of long range transport event (as the 99th percentile should be strongly underestimated by the model).*

We like to thank the reviewer for this very valuable suggestion. We added the 99th percentile to Table 2 and include related discussions in Section 3.1.4.:

“Omitting the fire influence in the  $\text{AVG}^{\text{RegionTime}}$  model results has little effect on the lower end of the distribution, but reduces the high end. However, also in this case the model underestimates the observed tail (Table 2). The low model bias for the 99th percentile is more pronounced compared to the 90th percentile, which is an indicator for the even stronger dilution of intense pollution plume events in the model.”

We further refer to the 99th percentiles in Section 3.3.2. (see fourth comment below)

***Page 28929, line 17 : “Average FT CO concentrations over land are 95 ppbv and 104 ppbv for modeled and observed flight track data”: I think “modeled” and “observed” should be exchanged.***

Thank you for pointing out this mistake. We corrected for it in the revised manuscript (Section 3.2)

***Page 28931, line 22: it is said that model calculation indicates that the TES region is influenced by fresh plumes with high CO concentrations. How can you be sure it's fresh plumes ? It is clear that plumes are fresh when CO is very high and O3 very low (and often with a negative correlation between them). I did not see in the model data such plume (O3 is not lower than around 50 ppbv).***

The reviewer is correct in that we used the term “fresh plume” incorrectly in this discussion. The data points indicate plumes, but not necessarily fresh. We rephrased this paragraph in the revised manuscript (Section 3.3.1):

“ The model represents the main correlations seen in the TES data, but shows a smaller range, less variability and misses some of the retrievals with increased CO and O<sub>3</sub>. Overall the model, in agreement with the previous evaluation results, shows a mean positive bias compared to the TES CO retrievals of  $2 \pm 15$  ppbV ( $3 \pm 14\%$ ) and a mean negative bias compared to the TES O<sub>3</sub> retrievals of  $-8 \pm 10$  ppbV ( $-11 \pm 21\%$ ). The smoothing effect of the TES operator reduces the variability from 17 ppbV to 13 ppbV for CO and from 14 ppbV to 7 ppbV for O<sub>3</sub>. Comparing MOZART\_true to the sampling for the entire region (MOZART\_region) we see in the latter more elevated CO/O<sub>3</sub> plumes (Figure 7). In a statistical sense (Table 2), however, these points have small weight and the statistics for the TES sampling gives an overall good representation of the temporal and spatial average with comparable means and medians for CO, but somewhat enhanced O<sub>3</sub> mean and medians.”

In the course of updating the manuscript we also noticed that the submitted graph was an older version that missed showing points for MOZART\_region at altitudes between the surface and 850 hPa and submitted a revised Figure. This has no effect on the discussions or statistics. In the course of updating this Figure we further included labels for MOZART\_true and MOZART\_region for an easier understanding.

***Page 24, line 24: same question, it is difficult to be sure it's fresh plume when O3 is around 50 ppbv (but on those dataset, I agree that some plumes are typical of fresh plumes with negative CO/O3 correlation and O3 around 20-30 ppbv).***

Please see reply to above comment

***Section 3.3.2: it is a very good idea to look at data over 4 years. It put the measurement performed during the campaign in more context. However, I found that you did not enough use such data. Is it possible with this dataset, to evaluate the number of 'extreme' plume case ? You can for example choose a CO limit that would represent strong plume events (180 ppbv ? 200 ?), and evaluate the percentage of time when such event are encountered (this is just an example). As you suggest several time that the 22 June flight is atypical compared to other day, this would help to evaluate how often happened such case, and how significant is the error made by the model when it does not represent such event.***

The reviewer makes a very good suggestion. We looked into these statistics and added the following discussion to Section 3.3.2:

“A comparison of the very high CO data between the MOZAIC and the DC-8 ocean data sets confirms the exceptionally high plume values that were measured during the ARCTAS-Carb Boundary Leg Flight. The DC-8 Ocean 90th percentile is by about 50 ppbV higher and the data set also has an enhanced frequency of high plumes: 3 out of 101 (3%) DC-8 Ocean observations have CO values larger than 170 ppbv and one observation > 200 ppbv compared to the MOZAIC data set with 4 out of 2824 (0.1%) CO values > 170 ppbV and the largest value not exceeding 190 ppbV. High value frequencies for the DC-8 Land data set with fire influence removed are also smaller compared to the Ocean flight: the 99th percentile is lower by 30 ppbV and 7 out of 415 (2%) CO values are > 170 ppbV. For the land data set the maximum observed value is 175 ppbV.”

***Page 28935, line 25 : does Parrish et al, give explanation for this pattern ? If it is so, can you explain it here ?***

We added more details about the Parrish [2010] study and added the following paragraph (Revised Manuscript, Section 4)

“Parrish et al. [2010] found that measured surface O<sub>3</sub> concentrations correlate with the O<sub>3</sub> measured in the lower free troposphere by the Trinidad Head sondes after allowing for a time delay of approximately 20 hours. On days that surface O<sub>3</sub> exceeded the air quality standard of 75 ppbv, the transported O<sub>3</sub> as measured by the sondes averaged 11 ppbv higher compared to days with lower O<sub>3</sub> observed at the surface.”

***Page 28936, line 4 : can you explain why more photochemically driven species are more influenced than source driven species (I guess it's because source driven specie are dominated by local emission when other specie can be formed 'en route'). It may seem trivial, but it's better to explain it briefly.***

We added more explanation to Section 4 in the revised manuscript:

“CO and NO<sub>x</sub>, which are strongly source driven and as such highest near local emission sources, show smaller relative changes compared to the more photochemically driven species O<sub>3</sub> and PAN, which are formed photochemically during transport.”

***Page 28936, line 24: you may also mentioned that, as has been shown in Section 3.1, global model underestimate strong LRT events. It they were better represented in the BC, the influence of pollution inflow would be even larger.***

We included the following statement in Section 4 in the revised manuscript:  
“However, the estimated sensitivities might even be on the lower end during strong long range pollution events, which, as was discussed in Section 3.1, are strongly underestimated in a global model (Section 3.1).”