Reply to Referee #2:

We would like to thank referee #2 for detailed comments that helped us to improve the manuscript. We have carefully considered each of the reviewer's comments in our revision. Our responses are provided below (the reviewer's comments are shown inline in italics).

Much of the paper (Sections 3 4) focuses on comparing different features in the observations to the two simulations. However, these comparisons are largely descriptive (and often qualitative). Section 5 begins to address the causes for the differences, but doesn't go very far (particularly in terms of inter-model differences). As the authors point out, the two models are very similar because they use the same meteorology and emissions. This means the places where they differ would present a very nice opportunity to understand which processes contribute to the differences, but there is very little discussion of this. For example, why is GEOS-Chem generally higher at 100 hPa but lower at 215 hPa? Is it differences in the convective transport parameterisations in the two models? Differences in chemical production or loss in the UT, or loss in the LS? What else could be driving these differences? This would also require a more thorough accounting of the similarities and differences between the models (e.g. how do chemical schemes differ? How similar are convective parameterisations? etc.) Without this level of analysis, it feels a little like an opportunity to deepen our understanding has been lost.

Reply: We agree that understanding the processes contributing to the differences between MLS observation and model simulations, as well as the inter-model differences is important, especially for improving model parameterization and simulations in the future. However, it is beyond the scope of this study which aims to evaluate the CO concentration and its distribution and variation in the UTLS simulated by two CTMs using the latest version (V4.2) of Aura MLS data. The factors accounting for model-observation and inter-model differences can be quite complicated, including biases in direct surface emission, the fraction of CO emissions released above the boundary layer, biogenic NMVOC oxidation, horizontal advection of CO, and model parameterizations of convective transport. We do plan to study such processes and factors in another work, but not in the current paper. In our original manuscript, Table 2 did provide some insights on the inter-model differences in terms of CO budget, such as chemical production and loss rates of CO in the troposphere, which help to explain some of the discrepancies between the two model simulations. In the revised manuscript, we have added one sub-section "2.2.3 Differences between GMI and GEOS-Chem" and one new table (Table 1), together with Table 2 (original "Table 1") to highlight the major differences between the two models. The added section is as follows:

"To highlight the differences between the GMI and GEOS-Chem model run, we summarize their major differences in Table 1. In addition, we calculate the annual mean values and interannual standard

deviations of CO budget (including biofuel and fossil fuel emissions, biomass burning emissions, tropospheric chemical production, tropospheric methane oxidation, loss with tropospheric OH, and net transport from troposphere to stratosphere) for GMI and GEOS-Chem during the period 2004–2012, and the results are provided in Table 2. In general, CO emissions from fuel combustion and biomass burning are mostly the same, but the chemical production and loss rates of CO in the troposphere are quite different between the two models. Specifically, GEOS-Chem is 40%, 16% and 15% higher than GMI in tropospheric chemical production Of CO, tropospheric CH4 oxidation and CO loss with tropospheric OH, respectively. For the net CO transport from troposphere to stratosphere, GEOS-Chem is ~9.5% larger than GMI."

On a related note, I find Sections 3-4 long and hard to parse. Some sub-sections would help, especially in Section 3. There are a lot of qualitative descriptions of features in the figures, paired with phrases like "[Feature X] in the GMI simulation is more consistent with MLS observations than in the GEOS-Chem simulation" – but these are hard to judge from the figures and often not backed up with quantitative information. In many cases, (e.g. Figs 1-3, possibly 4-11 as well) it would be easier to follow the text descriptions if the figures showed for the model differences plots (e.g., GMI – MLS and GEOS-Chem – MLS) rather than absolute concentration plots. The absolute plots could go into a supplement as the paper is already long and contains many figures. It would also be nice if some of the statements could be quantified using e.g. regional or temporal averages, or even mean difference statistics over all grid squares.

Reply: Thanks for these helpful comments. In the revised manuscript, we have made major changes which include:

1. We divided Section 3 into three sub-sections: "Seasonal Distributions of CO in the UTLS", "Monthly Variations of CO in the UTLS", and "CO 'Tape Recorder'". Section 4 is divided into two sub-sections: "Monthly Variations of CO in the UTLS" and "Vertical Profiles of CO in the UTLS".

2. We plotted figures of the differences between the two models' simulations and MLS observations. Since we want to highlight the spatial and temporal patterns of model simulated CO, we think it is better to put the difference figures in the supplement (Figures S1-S11 corresponding to Figures 1-11 in the original manuscript).

3. We have added more quantitative discussion about the comparison between model simulations and MLS observations following the reviewer's suggestion. More statistic values are computed, such as correlation coefficient, maximum, minimum and mean values.

My final major concern is that the paper doesn't reference much recent literature. Of the 11 referenced papers published since 2011, 7 were led by authors from this paper. There is significant newer literature surrounding, for example, injection of trace gases to the UTLS in the Asian monsoon (e.g., Park et al., 2009; Randel et al., 2010; Randel and Jensen 2013). There is also newer literature on CO distributions, including in the upper troposphere, than the 2006 Shindell work cited here (e.g., Naik et al., 2013; Fisher et al., 2015; Zeng et al., 2015).

Reply: Thanks for this comment. We have searched and added more recent literature (including these mentioned by the reviewer) and discussions about related references in our revised manuscript.

Minor Comments (by line)

163: "climatological" CH4 files – are these year-specific, and if not do is a trend imposed?

Reply: Sorry for confusion. We have revised this sentence to "Surface methane is read from monthly mean distributions interpolated from NOAA flask observations, and allowed to advect and react".

200-201: Are biogenic emission calculations the same between models? GMI section references Guenther; GEOS-Chem section refers to MEGAN. Theoretically these are the same but the implementation could vary. In general it would be really nice to see what exactly is same vs. different between the models (see above).

Reply: The biogenic emission is similar between GMI and GEOS-Chem. For the model differences, please refer to our reply above.

224-225: Are MLS averaging kernels and a priori profiles time-varying or constant?

Reply: MLS averaging kernels are constant, they were obtained from MLS website (http://mls.jpl.nasa.gov/data/ak/). MLS *a priori* profiles are time-varying. In MLS Version 4 data, all the standard product files now include the *a priori* information used in the retrieval (as an additional "swath").

243-244: How do you know trans-Pacific transport from East Asia is weaker in the models? If just judging from the figure, couldn't it just be that the East Asian CO is lower to begin with? Can this be quantified? (e.g. relative difference between East Asia East Pacific?)

Reply: We agree with the reviewer that the simulated CO concentration at the beginning of the transport is much lower than the observation. We checked and confirmed that the models did reproduce the transport. Thus, we have revised the text to: "The trans-Pacific transport of CO from East Asia in MAM and JJA to North America is shown in the model simulations, but the CO concentrations are ~30% lower than the observations."

323: What does "well captured" mean here? To me it looks like models are quite a lot lower (relative difference would help)

Reply: Over Indonesia during 2006-07 El Niño, model simulated CO maxima are similar to MLS observation (relative difference < 5%). We have revised the text to "The maxima (~160–170 ppbv) over Indonesia during 2006-07 El Niño are well captured by the models (difference between model and observation < 5%)."

391: "Remaining two" isn't quite right here as only 3 of 6 regions have been discussed so far in this paragraph (no mention of East Asia).

Reply: This issue is fixed in our revised manuscript.

421: Should "less than" be "greater than" here?

Reply: Yes and it is corrected in the revision.

453-454: Stating that the MLS IWC and modelled convective mass flux have "good linear correlation" is unsatisfying. Given how importance the simulation of convective transport is for this region of the troposphere, it would be really nice to show this comparison to the reader (perhaps in the supplement), or at least quantify it.

Reply: We have added a scatter plot (Fig. S12) in the supplement showing the linear correlation between MLS IWC and model convective mass flux. The correlation coefficients are also added in the revised manuscript to quantify the relationship.

Table 1: Does "tropospheric chemical production" really mean tropospheric chemical production from NMVOCs, or does this include the CH4 contribution? Please clarify.

Reply: Tropospheric chemical production includes both CH4 and NMVOC oxidation. To avoid confusion, we have restyled the original Table 1 as below:

Model		GMI	GEOS-Chem
Biofuel + Fossil Fuel		20.6 ±0.16	19.6 ±0.29
Biomass Burning		11.9 ±1.9	11.9 ±2.0
Tropospheric Chemical Production		42.3 ±0.92	59.1 ±0.77
	Source from Methane Oxidation	30.3 ±0.95	35.2 ±0.42

Loss with Tropospheric OH	77.7 ±2.1	89.1 ±2.4
Net Transport to Stratosphere	1.37 ±0.49	1.50 ±0.47

Figs 12-13: Since error bars are interannual standard deviations (not measurement errors) and simulations cover same period as observations, why not show these for the models as well?
Reply: The error bars are added for the two models in the revision.

Fig 14: For the analysis, it would be helpful to also show (or in a separate figure) the full vertical profiles from the surface. This would help determine whether the differences seen starting at 215 hPa are there because the two models start with different surface values, or because they are vertically transporting the CO to different altitudes (e.g. maybe there is more GEOS-Chem CO at 300 hPa), or something else.
Reply: We agree that it would be helpful to show the full vertical profiles for comparison between models and observation. However, MLS is only sensitive to CO at 215 hPa and above, the CO profile below (i.e., pressures greater than) 215 hPa is not reliable. The two models have different pressure levels, and the model simulated CO profiles are interpolated to the pressure levels of the MLS observation by using MLS CO averaging kernels and a priori profiles. Thus, such comparison may be done in future work using satellite retrieved or ground-based measurement of full vertical profile.