Reply to Anonymous Referee #2

We thank the reviewer for investing valuable time into performing the review and greatly appreciate the constructive comments and suggestions. Below we give a detailed response to each individual point. Reviewer comments are in bold and our replies in regular font characters.

This is the modelling study of the long-living pollutant, which originates from several types of sources, including anthropogenic, natural, fire sources, and chemical transformations. In general, the paper is solid and quite well presented study. However, reading it I was missing a few significant items listed below. All-in-all, they sum-up to major revision because some simulations are missing. General comment I am somewhat puzzled by the stress in the paper. The authors have found a major deficiency of the input data, namely the fire emission being strongly under-estimated. However, no effort was made to deal with it or, at least, to study this problem (the sensitivity run did not help – see below). Instead, lengthy considerations were presented about the relative contribution of various sources regardless the fact that, if the fire emission is estimated properly, this budget would be different.

We are working on understanding the strong underestimate in the CO fire emissions for the California wildfires. From preliminary analysis we believe the cause is found in the MODIS LCT land cover dataset used in FINN. MODIS LCT assigns large areas to shrublands in California, where other land cover datasets assign forested regions. Shrublands are assigned lower CO emission factors and in addition, burn less fuel than forested land cover types. The combination of these two factors can lead to an underestimate of the CO emissions. A more detailed explanation for the low emissions has been added to Section 3.3.
We did not make attempts to correct for this underestimate in the current study for various reasons. Using the evaluation results, we will work together with Christine Wiedinmyer on improving the performance of the FINN model rather than "correcting" the specific emissions used in here. But the current estimates are what is provided to users.
A comparison of the FINN inventory to an independent fire inventory from Sonoma Technology for 15-30 June 2008 (personal communication) indicates that over this episode the FINN CO emissions are less than the Sonoma Technology inventory by nearly a factor 4, while other emissions, including NO_x, agree much better. Given that the CO chemistry is to first order linear, an increase in the CO fire emissions can be simulated by adding multiples of the CO fire tracer to the modeled total CO fields. We included estimates for an increase in COfire by a factor of 4 in the revised paper (Sections 3.3, 4.1 and 4.3), and with this provide a more realistic and at the same time upper bound to the budget.

Specific comments

Introduction P.3628. Line 25: the opposite is true. Lifetime from weeks to months means that the CO distribution does not resemble its source distributions. About a week is enough for the plume to cross the American continent mixing-up the emission of all sources there. Near the sources, CO can even be considered as a passive tracer, so that its distribution pattern would be a result of competition of emission flux and dilution due to transport. The sentence has to be rewritten or removed.

We changed this part to:
"Its global distribution at the surface reflects the location of large emission sources but with a tropospheric lifetime on the order of weeks to months, CO is also a useful tracer for atmospheric pollution transport [Staudt et al., 2001; Liang et al., 2004; Yashiro et al., 2009]."

Introduction
The technique of CO tracers should be explained. At present, the paper assumes that reader knows it, which is not very appropriate since the paper is significantly based on this technique.

We added more details about the CO tagging technique:
"CO tracers are artificial tracers that are added to a simulation and are produced or emitted from predefined sources and undergo the same transport, loss and chemical processes as the total simulated CO. CO tracers are additive and the sum of all tagged sources (direct emissions and photochemical production) equals the total simulated CO mixing ratios."

Section 3.1. P.3634, Line 19->. The good representation in the free troposphere is essentially due to the inflow from the boundaries. Does MOZART show the same quality? Does MOZART have the same low bias closer to the surface? Figure 1. Offset of 100ppbV is not justified and only overshadows the actual fire impact predicted by the model. Also, the MOZART data should be added at least to some panels – similar to Figure 2 Comparison with in-situ data (p.3637). This part turned disappointing. I strongly doubt the possibility of averaging over sites, even after splitting them into two groups. Before doing that someone has to prove that these sites have the same statistical features – at the very least. Since they probably do not, the "mean" time series shown in figure 5 and discussed in the text do not have much value. As an additional confirmation, the correlation coefficient for the mean time series not affected by the fires is zero. Is the model so bad that it cannot get the simple diurnal variation? If the problem persists for individual sites, it has to be discussed and measures taken but I would expect some "poor" and some "good" sites to show up when the analysis is done individually.

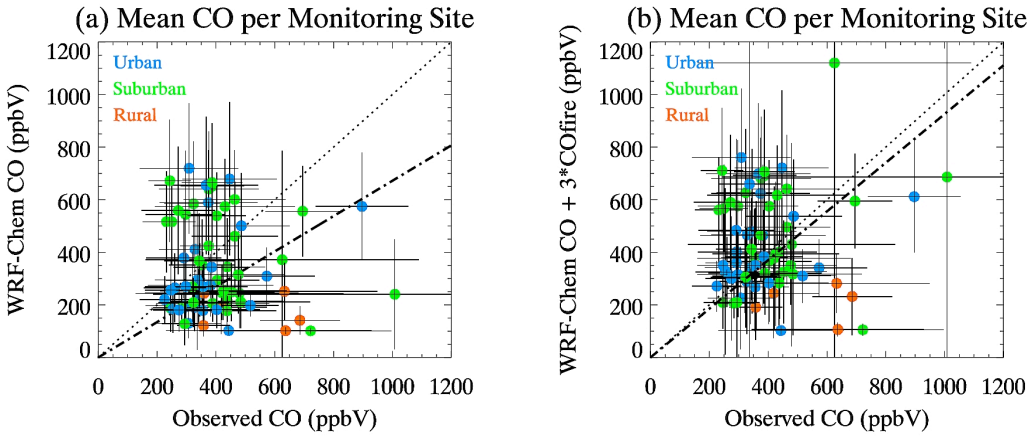
We revised Figure 1, which now includes also results for MOZART and took out the offset of 100 ppbV that was added to the COfire tracer. The following text has been added in Section 3.1:
"In the free troposphere WRF-Chem fields are strongly influenced by pollution inflow and both, WRF-Chem and MOZART-4 show a very similar behavior. At the lower altitudes the two models diverge more strongly due to localized influences. Largest differences are seen in the fire impacted data sets with MOZART-4 being higher. This is caused by MOZART-4 not considering plume rise but releasing all fire emissions at the lowest model level."

Section 3.3 describing the comparison to surface sites has been significantly modified.
Very few research studies make use of surface CO monitoring sites as they represent a big challenge for models for various reasons. We made these challenges clearer by adding the following statements:
"The majority of the sites (68 out of 72) are stated as being located in either urban or suburban environments. Since CO emissions are primarily associated with mobile sources, the monitoring sites are often located near strong localized pollution sources, such as busy intersections, creating a strong source-to-receptor relationship that is a challenge to duplicate in models with limited spatial resolution. Evaluation of the model is further impacted by the low resolution of the observations (multiples of 100 ppbV with a varying detection limit of 100 ppbV or 200 ppbV depending on the site) and by variable temporal coverage and large data gaps, especially for early morning hours due to nightly precision and span checks. Despite these limitations, the surface observations can provide additional valuable information for evaluating the overall model performance at the surface."

We spent great effort on finding proper ways to group sites as suggested by the reviewer. However, because of the localized nature of these sites, there is no easy way to group sites into categories with similar characteristics, even more so as wildfires added an additional highly variable and strong source over the considered time period. The most optimal way remained a grouping into Northern and Southern CA sites since this separates sites subject to large fire impact from sites with minor fire impact. Sites in Northern CA are generally also in less polluted environments compared to sites in Southern CA.

As additional information for the referee we include below a graph comparing modeled and observed mean and standard deviation surface CO at individual sites. The left panel shows WRF-Chem results for the reference simulation, the right panel shows results for a four-fold increase in fire CO. Different site categories are indicated by different colors.
Overall there is a decent correlation (which improves when fire CO is increased) and most sites are grouped around the 1:1 line. However, we see a group of urban and suburban sites for which the model overestimates the mean CO, while the model is biased low at some of the rural sites. It is quite interesting, and also reflects the local nature of the monitoring sites, that rural sites show some of the highest observed

CO, while the model represents them as low CO unpolluted environment. The large spread in the mean values and according large standard deviations show how highly variable CO concentrations are at individual sites and that a grouping into categories with similar site characteristics is extremely challenging.



In the revised manuscript we further included in the text and in Figure 5 a more comprehensive evaluation for surface sites, such as statistics for hourly values as well as for mean time series, and further added evaluation results for MOZART-4.

The low correlation for Southern CA sites is explained by differences in the diurnal cycle. We added the following discussion:
"The correlation between WRF-Chem and observations is poor and in addition to the above listed limitations in the observational data sets is explained by an overall discrepancy in the timing of peak concentrations. Both model and observations generally show an early morning and evening peak in the diurnal cycle, but while the observed peak is stronger in the morning, the evening peak in WRF-Chem is more pronounced. This might be caused by WRF-Chem incorrectly simulating the timing of the boundary layer evolution (e.g., a too rapid growth in the morning or a too fast shrinking in the evening) or by errors in the temporal allocation of CO emissions in the emissions inventory. Future model studies are needed to better understand the cause of the discrepancy in the timing of the modeled and observed peak in CO. The correlation between modeled and observed CO improves to 0.43 if instead of an 8-hour running mean daily averages are analyzed."

Figure 5. A technicality: the charts are essentially unreadable, especially the upper panel. They should be widened or split into several panels, the main lines should be made thinner.

Figure 5 has been revised (please see reply above)

Sensitivity simulations. I am greatly surprised by the fire sensitivity run. The authors contradict to themselves. Firstly, throughout the paper the red line is that the fire emission in underestimated – no matter whether near-surface of aloft observations are taken for the model evaluation. Secondly, emission under-estimation means that the fire intensity is under-estimated as well. Thirdly, the plume rise routine gave comparatively reasonable estimation of the injection height, may be, only slightly too high.
According to, for example, MISR analysis, the split between the ABL and FT is closer to 80-20 but varies greatly, so that for powerful Californian fires I am not too surprised with 50-50. Fourthly, increase of the estimated fire intensity would mean increase of the injection height. Nevertheless, the authors reduce the injection height and leave intact the emission! What was studied by this run? The problem is even admitted in the paper (p.3643, lines 20-30 and further) but no efforts to correct it were made. I think that the sensitivity study has to be rethought and the fire simulation redone.

As mentioned above, these issues have been addressed by presenting budget estimates for an increase in the COfire tracer representing increases in CO fire emissions.

The reasoning behind conducting the sensitivity simulation SENS_fire is that most chemical transport models either release fire emissions at the surface or use predefined injection heights. Our sensitivity simulation provides a range for how CO concentration fields change for different treatments in injection height. We included this justification in Section 6 (previously Section 5).
"In the first simulation (SENS_fire), we changed the treatment of fire emissions from having them distributed vertically through the online plumerise module to emitting them at the lowest model level only. Most chemical transport models either release fire emissions at the surface or use predefined injection heights. This simulation provides a range for how modeled CO fields might change for different treatments of injection height."

The sensitivity simulations further support the model evaluation and allow for separating different contributing factors.