Response to Anonymous Referee #1

We thank the referee for their valuable comments, which substantially improved the paper.

Referee comments are in *red italics*, our responses are in black text.

Observations from two different spectrometers are used in this work, the MkIV spectrometer covering a much longer period, but with sparser observations, and the Caltech TCCON spectrometer. I am a bit concerned with respect to the compatibility of the derived gas anomalies. The point is not primarily the spectral measurement itself, but the significantly different sampling strategies of the measurements (MkIV observations are constrained to local noon). It would be instructive to demonstrate that an analysis of a reduced TCCON dataset (local noon observations only) generates compatible gas anomaly values, or whether the sampling strategy can introduce a significant bias.

The figure below demonstrates that there should not be a significant bias induced into the tracer-tracer slopes. Shown below are a subset of the time series from MkIV and Caltech, filtered both datasets appropriately (removed plumes, cloudy data, etc.), and then subselected the Caltech data to points within 15 minutes of the MkIV measurements. The black dots below are all the filtered Caltech data; blue are the Caltech data time-matched with MkIV; red dots are the MkIV data themselves. Slopes of the tracer-tracer anomalies in the third panel below show a small bias between the Caltech and MkIV data that is well within the uncertainties in the slopes.
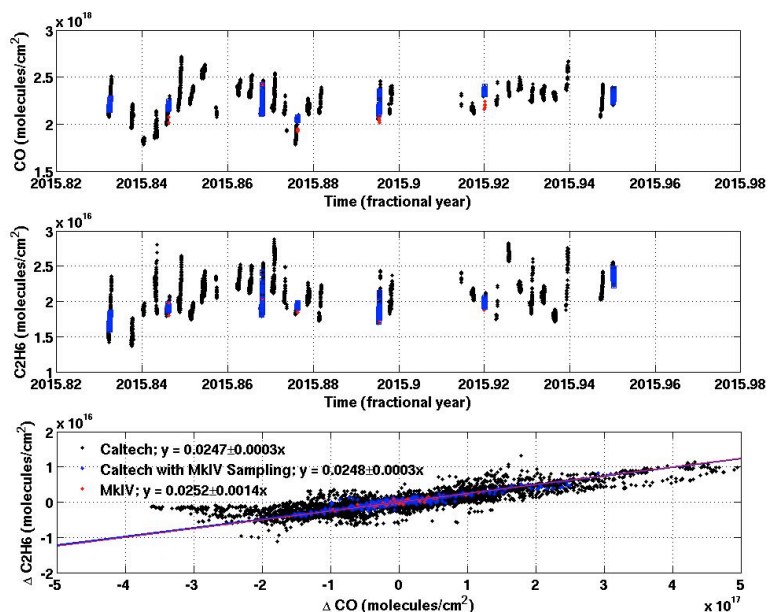

The determination of gas anomaly values from the difference of afternoon and morning values is in principle a convincing approach. However, as small changes are derived from differences of much larger column values, I wonder whether the heating of the boundary layer during the day might also mimic a gas anomaly contribution? Is the analysis performed assuming a constant temperature profile? Is the heating effect a significant disturbance?

The analysis uses a single *a priori* temperature profile throughout each day, that is representative of the local noon temperature profile, derived from the NCEP/NCAR reanalysis data. There is a systematic increase in surface temperature throughout the day; typically a 5K error between mid-morning and mid-afternoon at the surface (see histogram below); temperature changes aloft should be smaller and thus the integrated temperature error throughout the PBL should be smaller than 5K.

To minimize the temperature sensitivity of our retrievals, we chose windows in which the target absorption lines have average ground-state energies of around 300 cm-1. For example, we use the entire CO and CH_4 bands in the near infrared, which have roughly the same number of high-j and low-j lines, reducing the temperature sensitivity. C_2H_6 is measured in its Q-branches between 2976 and 2997 cm-1. Based on performing C_2H_6 retrievals using correct and incorrect (perturbed) temperature profiles under a range of different conditions (temperature, humidity), we know that the retrieved C_2H_6 amount will change by <1% for a temperature perturbation of 5K at the surface, decreasing to zero at 3.5 km altitude. Since a typical diurnal change between mid-afternoon and mid-morning in the retrieved C_2H_6 is ~20%, the temperature-induced affect is comparatively small.

A sensitivity study for CH_4 was performed by Hedelius et al. [2016] that showed <0.05% errors arising from 10K temperature perturbations between the surface and 700 hPa for lower resolution FTS instruments, and thus the sensitivity should be smaller for the higher-resolution TCCON instrument. This is also smaller than the <1% diurnal variations in CH_4.
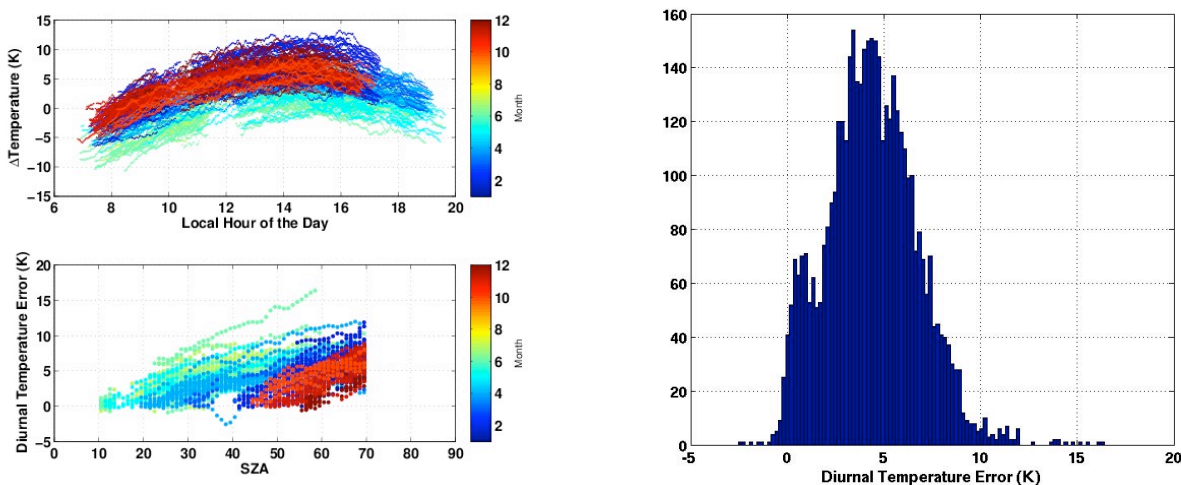
I have problems to understand that in Figure 2 the slope of the red dashed line differs between the top and bottom panels. If the slope is a function of time, why then is the slope in the upper panel so well defined (it encompasses data from several years, correct?).

The slopes of the red dashed lines in the quantile-quantile plots are not particularly important and the slope is not (necessarily) related to time. These plots were meant to show how we distinguished ambient SoCAB air from plumes. The bottom plot is for the time period when the Aliso Canyon leak was ongoing; the top panel is from the other time period. In this type of plot, data that are derived from a statistically similar set appear linear; that is, when the CO and CH_4 vary simultaneously, their quantile-quantile plot will be linear. When they do not co-vary (i.e. when there is a plume of CH_4),

their quantile-quantile plot will be nonlinear. We plotted the two times separately simply because the Aliso Canyon plumes will dominate the later data, and thus we wanted to be able to choose the filters that delineate between plumes and ambient air differently.

The error bars on the symbols in Figure 3 are difficult to read. It seems that while the 2013 and 2015 results from MkIV and the TCCON spectrometer agree nicely, the discrepancy in 2014 is much larger than the indicated error bars. Is this a sampling issue (dates of observations used?)

The Caltech annual mean slopes are calculated for September through August. If we calculate the mean 2014 monthly slope from the Caltech data for January-December, the slope increases from 2.1+/-0.4% (for September 2013 -August 2014) to 2.3+/-0.5%. (Closer to the 2.5+/-0.1% from MkIV.) We have also updated this figure to show the Caltech mean monthly slopes with the standard deviations as the uncertainties. (Previously we computed annual slopes and reported the slope errors, but we feel that was more complicated than necessary, and that the standard deviation of the monthly slopes provides a better estimate of the uncertainties.)

The scatter of the FTS deduced ethane to methane ratios in Figure 5 is large. The error bars on the individual data points are quite variable and especially in 2015, the scatter between the data points is much larger than the individual error bars. Why? Does this imply that the uncertainty budget is dominated by a sampling statistics issue?

Indeed, the variability in the FTS-deduced ethane to methane ratios is large. We would also point out that this is also true of the ratios in the delivered natural gas, which are very precise and accurate, and have very small error bars. This suggests to us that the delivered gas itself is quite a bit more variable than the reported withdrawn gas ratios from the Playa Del Rey storage facility, and our atmospheric measurements are able to detect that.

What is the level of significance for the derived slope value? Does the regression fit take into account a weighting of data points accoring to the individual error bars?

The slope of the ethane to methane ratios has an uncertainty of ~15%. The regression fit to the FTS data does take x and y errors into account, using the York et al. (2001) formulation.

York, D., N. M. Evensen, M. L. Martinez, and J. De Basabe Delgado (2004), Unified equations for the slope, intercept, and standard errors of the best straight line, Am. J. Phys., 72(3), 367, doi:10.1119/1.1632486.

The figure might suggest a superimposed peak of high ratio values in the mid of 2013.

We also noticed the mid-2013 peak, but given the large uncertainty, we're uncomfortable making any strong claims about that.

In Figure 7, the claimed steady rise of the slope during the observation period is hardly recognizable (due to the overlap of data points), perhaps a subdivision in several panels spanning fractions of the whole period would improve the readability.

We've updated the figure to show the slopes for each year overlaid.