PRELIMINARY Author’s reply to comments of reviewer #2 in AMTD

The authors would like to thank the constructive comments made by the reviewer and the time he/she took. Below please below find the responses. For the sake of interactivity this is a preliminary response. We would like to have a feedback of the reviewer, in particular, for 2.5 (consistency of measurements). After that a completed response will follow.

All changes made to the manuscript are marked in yellow.

Author’s replies are marked in green.

To summarize the main changes made to the manuscript, as a result of the reviewers comments we now think that it is not appropriate to a priori expect a good match between the instrumental CO2 concentrations, due to the potentially very different CO2 distributions sensed owing to their different alignment. Despite qualitative agreement between the DIAL and the OP-FTIR result the rather complex CO2 distribution along with the non-ideal measurement geometry and the early development stage of the DIAL prevent a robust quantitative comparison yet.

We finally propose to not publish the DIAL data as we found that the error introduced by not normalizing is too large to be able to have a useful interpretation of the data (see section 2.5 below). We think that this is a reasonable thing to do as this does not degrade the quality of the paper but on the contrary increases it. Some replies have become redundant but we have kept some of them nevertheless. Of course, we need to change the paper title accordingly.

The study is of considerable interest to the volcanology community because it discusses two rather new techniques for quantitative determination of CO2 concentrations and fluxes: the DIAL and the dispersion modelling. In principle, the combination of the two approaches might in the future be used to determine the CO2 flux emitted from a diffuse degassing region, an application that I believe the authors fail to mention but should be touched upon, as it lends additional importance to their study.

Reply 1:

Changed last phrase of manuscript “(...) an Eulerian dispersion model and optical remote sensing represent complementary techniques for monitoring non-uniform CO2 degassing.” to “(...) an Eulerian dispersion model and a differential absorption lidar represent complementary techniques for determining CO2 fluxes in regions with non-uniform degassing.

2.1. Difference in light paths. One problem arises from the different light paths chosen for the two open--path instruments. Though the paths are close to one another, the fact that the FTIR path runs at a constant 0.5 m above the ground while the DIAL path

starts 1 m above the ground and ends at the ground itself is unfortunate. The first meter of air above the ground in regions of diffuse CO2 degassing such as Caldara di Manziana is characterized by a very strong vertical gradient in the CO2 concentration. The CO2 mixing ratio can approach 100% a few mm above the ground (or water) surface, decrease rapidly and reach practically background concentrations at 1 meter. Since this gradient is caused by diffusion of CO2 into the background atmosphere, it is not expected to be constant, but rather exponential in first order approximation. Therefore, an instrument aimed at the ground will be extremely sensitive to degassing occurring near the point where the light path touches the ground. At the same time, the sensitivity to degassing occurring at the instrument’s location 1 m off the ground is negligible in comparison.

Given the different optical paths of the two instruments, it is unclear to me whether one should even expect consistent CO2 column amounts. It probably depends upon the spatial distribution of degassing sources along the light path. If a single significant source (in this case perhaps the water pool) were located in the center of the light path where both instruments’ views are approximately at the same height above the ground, then yes, a consistent measurement would be expected. However, if the main source is towards one end of the path or the other (as appears to be the case in your experiment), I would expect a significantly higher value for the instrument with the lower path at that location (in your case the DIAL). Given the spatial inhomogeneity of diffuse emissions in the crater, I expect this will be hard to work out. However, I believe that some sort of treatment of this issue is needed because if one simply assumes a spatially homogeneous soil emission along the path and an exponential decay of CO2 concentration with height above ground, the CO2 values obtained along the two different light paths would not be expected to be the same.

Reply 2:

First of all, we agree. We should not a priori expect a match. Also, I have to make a correction/clarification. Since for path #1 the DIAL light path ended (a few m) behind the IR source the region of overlap is in the second half of the path. This could explain why the measured CO2 concentrations somewhat agree, that is, their trends agree, apart from the strong small scale increase seen by the FTIR at places. This also could explain that the CO2 concentrations sensed by the DIAL are not systematically higher than those from FTIR.

We fully agree about the vertical gradient of CO2 mixing ratio and it should experience more consideration in the manuscript. We actually modeled the CO2 mixing ratio for heights different from 0.5 m . See reply for 2.2.

We also agree with your point that since line of sights cross not at the main CO2 vent and given the otherwise diffuse and inhomogeneous CO2 emission it is hard to work out a precise CO2 distribution, at least with the dispersion model we have.

2.2 Assessing the importance of a vertical CO2 gradient

One very simple approach to assessing the magnitude of the vertical CO2 gradient and its importance for the study is to look at measured concentrations as a function of height. The authors mention that at least one such measurement was performed above the water pool (page 4332, line 15). Please show the results of this measurement in the manuscript. What was the range of heights that were included? How did the CO2 concentration vary with height? Were any other such measurements performed in the area? Perhaps some general assumptions about vertical gradients could be gleaned from these?

Reply: This was written incorrectly in the paper. We did not measure the gradient. The corresponding phrase is changed.

Another approach to evaluating the importance of the vertical gradient in CO2 concentration might be to look at the model predictions for change in concentration with height above the ground. Unfortunately, though the authors state that a three-dimensional CO2 map is produced by the model, no vertically resolved modelling results are given or discussed. Somewhere, rather late in the manuscript, the author mention that the vertical resolution of the model is 0.5 m. My feeling is that this may not be enough. The authors are encouraged to, if possible, increase the vertical grid resolution in the lower-most meters above the ground to try to capture the vertical concentration gradient. At least this way the importance of the effect could be assessed. In principle, the DIAL light path might even be sampled out of the vertically resolved model data if multiple heights were available along the path.

Reply 3: This comment touches on the previous comment made by the reviewer, namely having different light paths at a given vertical gradient of CO2 concentration.

Knowing more about the vertical distribution might help to more explain small-scale mismatch in Fig. 4 more precisely.

For the second part of this point I suggest to add a new figure (Fig.5) where we can show some vertical gradients derived from the numerical code. In figure 4a should be added new labels at selected times (15:52, 16:00, 16:10, ….17:00, 10:02, 10:10, 10:20, see below) in order to refer to those in fig. 5.

Possible addition, somewhere in “Section 3.1 Modeling results”:

As expected, the predicted values of the CO2 air concentration tend to exponentially decrease with the altitude, following the general equation:

$y = a * e–bx$

where y is the height (in m), a (in m) the maximum altitude reached by the gas plume (i.e., the height where the plume density tends to match the air density), x is the air CO2 concentration (in ppm), and b (in ppm -1) indicates how rapidly the CO2 concentration decreases with height. Best fitting curves for the simulated layer (0–6m height) are reported in Figure 5, together with relative equations. The atmospheric layer affected by the gas dispersion ranges typically from 11 to ~70m (parameter a in the inset of Figure 5 referred to curves of the gray region), with the exception of conditions with high-speed winds blowing from W and SW, which raise the buoyant plume up to 200–400m (curves 1, 2, and 9)

This comment is very pertinent. First of all, we specified the vertical resolution of the model. The values of 0.5m is the typical maximum vertical resolution for this model, as a right compromise between calculation time and significance of results (Costa et al., 2005; Granieri et al., 2013, 2014).

We add this sentence at pg. 8 after “The computational domain is 1.0 km × 1.0 km with an x and y grid spacing of 10 m. Vertical grid dimensions range from Δz = 1.5 m in the uppermost layer (4.5–6.0 m) to Δz = 0.5 m in the lowermost layers (from ground level to 4.5 m) in order to better detect near-surface gas concentration gradients.

With reference to 0-1m gradient we also believe that it’s true, the concentration change is significant in the first metre since the curves decrease exponentially but this change is not dramatic because of the high concentration of the plume, whose top is at 10-60 m above the soil or more and not just at 1 m as the reviewer said somewhere and that the value assumed at 0.5 m is a reasonable compromise, obviously underestimated for the DIAL (measuring from 1m to ground) with respect to the FTIR (0.5m.

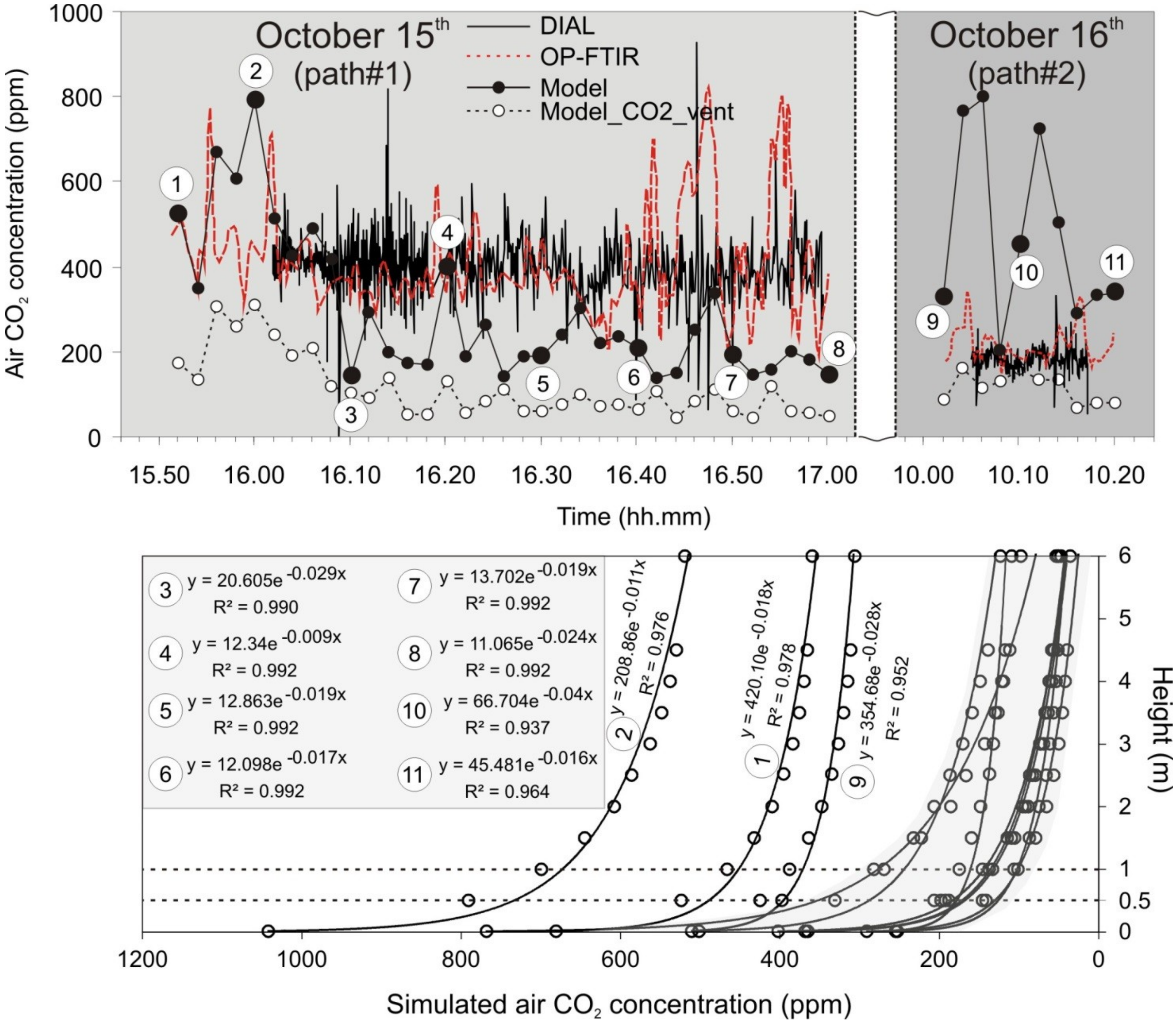


Figure 5. Just the graph below. In fig. 4a should be added new labels (as shown)

2.3. Assumptions made for the modeling study

It appears to me that the expected accuracy of the dispersion modeling is likely very dependent on the accuracy of the initialization of the model. It seems that quite a few assumptions must go into the initialization, a number of which are not currently mentioned in the manuscript. Parameters that are mentioned include the topography and soil flux (from the accumulation chamber survey). Also, the authors mention that the wind speed and direction are given by a meteorology station located within the model domain. However, this meteorology station only measures wind speed and direction (averaged over 2 minutes) at a single point in space. How is the rest of the model domain initialized? Surely not all points are initialized with the same values, or are they? Also, what about the influence of ground temperature? I can imagine that the temperature of the soil (and/or water) could significantly influence air flow at the small scales that are considered here. What other assumptions are made in the model study? Which effects are omitted and why?

Reply 4: to be added

2.4. Dealing with different time resolutions of the individual techniques

Clearly, the different techniques compared in this study operate at different time resolutions. The DIAL obtains a measurement every couple of seconds, the FTIR takes a few tens of seconds and the model only gets a result every 2 minutes because this is the time resolution of the meteorology station used to initialize it. This difference in time resolution is mentioned frequently throughout the manuscript and in places is called upon to explain differences in the results (e.g. page 4336 line 20). After reading the manuscript, it is not clear to me how this difference in sampling rate can cause significant differences in measurement/model results. If a range of wind speeds or wind directions are encountered during the 2 minutes integration time, wouldn’t the model results still approximately correspond to the measurement results averaged over 2 minutes? Could you please explain this in more detail?

Reply 5:

To be added

2.5. How consistent are the measurements really?

Another issue that I feel is not adequately addressed in the manuscript is the question of how well the results really compare to one another. For an paper titled ‘intercomparsion’, I find it odd that there is not a single figure plotting the results of one technique against those of another. Figure 4 is all the reader has to go on, but the time series plots leave something to be desired.

Reply 6:

Now obsolete. Maybe I misunderstand, but it would be redundant to have 3 plots (FTIR vs. model, DIAL vs. FTIR, DIAL vs model). Therefore it is all in one plot. For me this is a bit like comparing 3 pencils and pick the longest one. I would put them all next to each other and decide, instead of comparing two at one time. However, we now show DIAL vs. FTIR as you also ask for it further below.

And even though the residual (difference between FTIR and DIAL) is mentioned somewhere but it’s not plotted. In my opinion, there are some major issues with the comparison between FTIR and DIAL that are more or less glossed over by the authors. They state that the measured values tend to agree quite well when averaged over many tens of minutes or when looking only at long term trends. This may be true, but it's not that convincing of an argument because it could depend on some sort of absolute gain that is set by the user.

Reply 7:

I disagree. The trends, for instance, those we see over minutes do not change just by introducing an offset, which is what we do to "calibrate" the DIAL result with the FTIR result. Of course, this kind of data calibration is not what we should do or will do in the future, but this is one of the initial tests only.

Where any measurements performed outside the crater area? Were these used to in some way calibrate the DIAL?

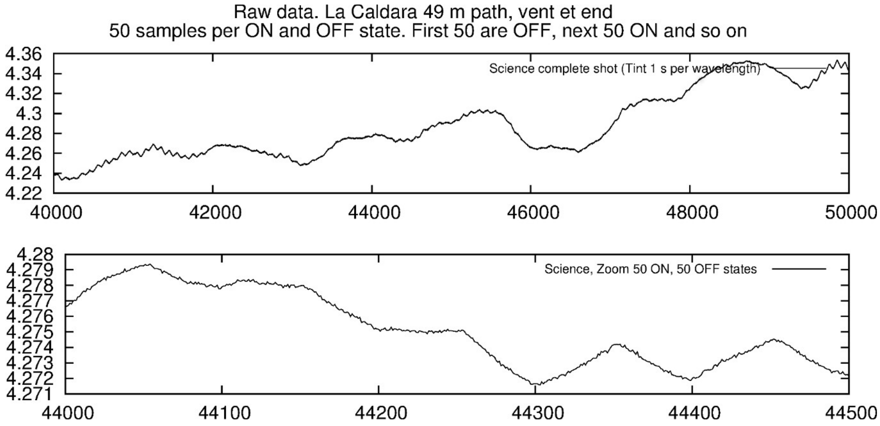
Since two different laser sources are used for the two different wavelengths (ON and OFF), don't you need to somehow normalize the relative signals from the two? In other words, you don't know the expected intensity ratio for the two wavelengths for a given a CO2 column. There must be some sort of calibration, and this is what is responsible for setting the long--term average value. How was this calibration performed for your measurements?

Reply 8: If I understand correctly what you refer to is the baseline intensity ratio, that is, finding the correct offset (instrumental offset mainly, i.e. Intensity ratio = 1+offset). Yes, we acquired DIAL data at the edge of La Caldara, but not together with the FTIR. We could use these to calibrate the DIAL result. In fact, I also evaluated these data assuming background CO2 levels to obtain the calibration offset. However, the area was still close to La Caldara, so CO2 levels were likely above background level. Instead, a data calibration was done by matching the first value of the time series (The mean would destroy systematic differences due to path mismatch etc. and using the first values is beneficial to test whether changes are similar). Of course, this may miss the true relative offset but we were more interested in trends. This was one of the first tests of the DIAL. The purpose was to see if the trends are similar (which, as we know now, we cannot a priori expect) not so much to compare absolute values. We see agreeing trends (see Replies 7 and 11). We agree with your argumentation that slight geometrical mismatches of the line of sights between DIAL and OP-FTIR cause severe differences in the measured CO2 concentrations due to the very strong CO2 concentration gradient. From this perspective it is interesting that we see a match between DIAL and FTIR at all. A key evidence of this study is that all three approaches (model, OP-FTIR and DIAL) indicate a decrease in CO2 concentration after 16:00, when the wind direction changed, despite the unfortunate alignment. The main message of the paper was to point this out. In other words, the bottom line is that despite the unfortunate geometry and a model not at all made for this kind of time and length scale and an unvalidated DIAL prototype we actually some sort of agreement at all! The degree of change is definitively quite different. I agree that this was not an ideal situation to test a new DIAL instrument due to uncertain CO2 distribution and uncertain DIAL performance at the same time. Please refer to Replies 9, 13 and 15 for further discussion of this.

Comparing shorter term trends (within the instruments' time resolution) seems more interesting to me. And here there appear to be some pretty serious discrepancies. For example, on 15 October between 16:45 and 16:50, the FTIR measured a column amount between 600 and 800 ppmm above background while the DIAL showed no increase whatsoever above the average value of 400 ppmm.

Reply 9: There is a peak at 16:47 which has not occurred before, coinciding with the CO2 concentration measured by the FTIR. Moreover, a second peak coincides with an increase in CO2 as measured by the FTIR at 16:55.

As for the short term fluctuations after 16:20, it is unlikely that the DIAL was not capable of measuring variations of this order since we have done corresponding sensitivity tests with CO2 before (see also Replies 13 and 14). The fact that positive changes are averaged out when using the full ensemble of 1e5 intensity ratios instead of a fraction of the data only means there were negative changes as well as positive, so that on average there was little change on scale of the DIAL integration time (2 s and 4 s respectively).



The figure above shows some raw science data from path #1. The first 50 samples correspond to OFF wavelength the next 50 to ON the next 50 to OFF aso. The oscillations are due to atmospheric turbulences. What becomes obvious when zoomed in (plot below shows 10 switch states) is the slope each state (ON or OFF) exhibits. For many states we can see this slope. If we compute the ratio between OFF and ON for each sample and average we get the flat time series with barely any small scale fluctuation. Now, before we compute an intensity ratio we can boxcar the samples, meaning we only chose to take the samples between sample 17 and 47, for instance. If we set the boxcar such that we chose the first few samples only we get a time series that follows the standard deviation in Fig. 4b., if we chose the last few samples only we get a time series that is almost anti-correlated to that one. We have no good explanation for this.

Figure 4bis below shows this result for taking samples between 44 and 45 (2 samples) [er ON and OFF states.

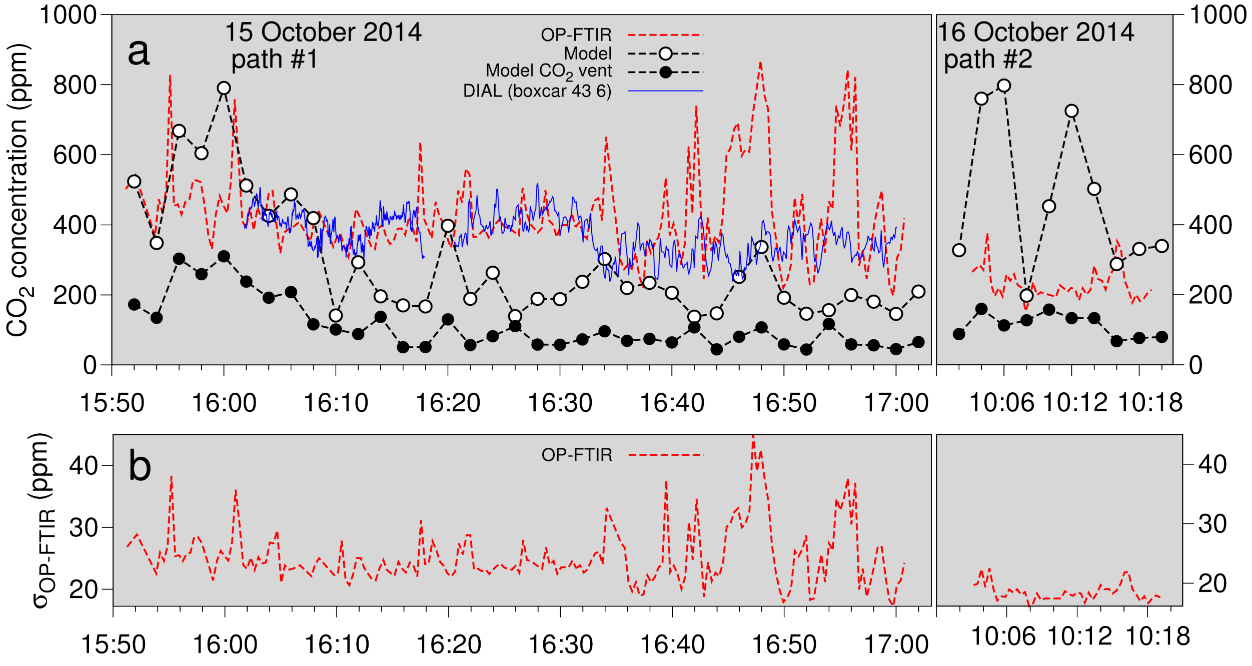


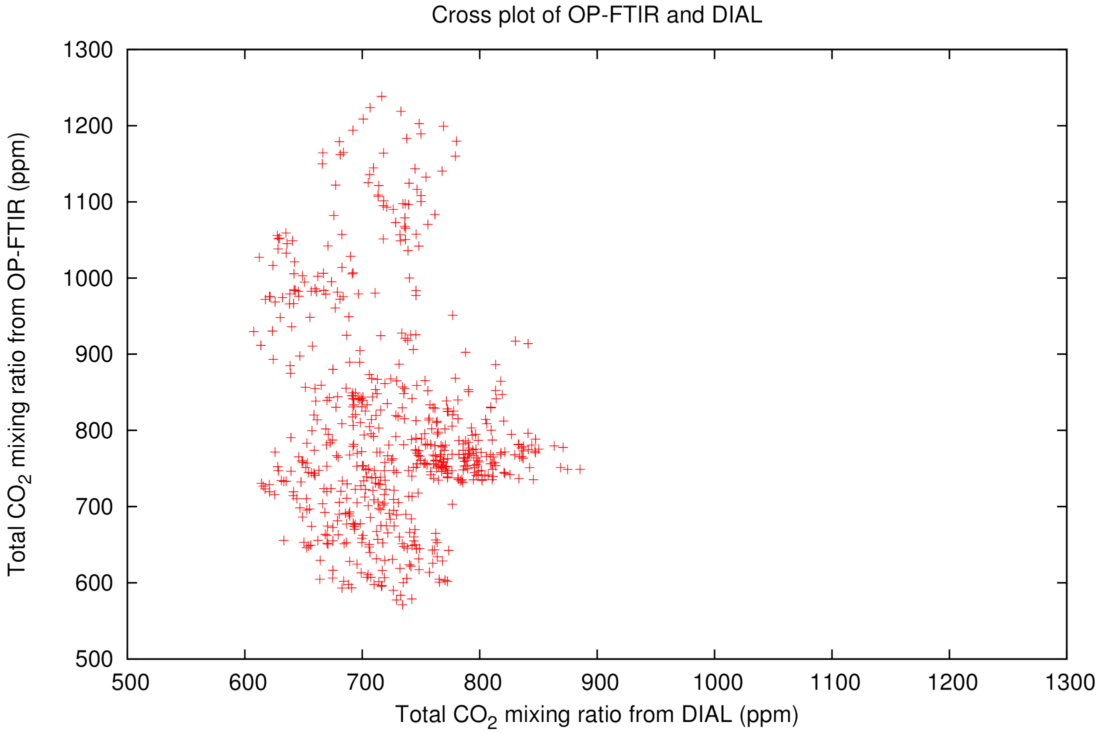
Figure 4bis

The DIAL follows somewhat the FTIR trend, features some small scale fluctuations that are not coinciding with the FTIR data though. This can be due to the vertical discrepancies in geometry. However, since we do not normalize the uncertainty of the time series is actually higher than I previously assumed, namely possibly 600 ppm total column amount. This is why I bellieve it is better to dismiss the DIAL from the paper.

This appears to be true in several other time periods as well. Could you please include a figure plotting the DIAL measurements (perhaps smoothed to the time resolution of the FTIR) vs the FTIR measurements?

Can you fit a straight line through those? What is the confidence of that fit? Is the correlation statistically relevant?

This figure shows a cross plot for the data of path 1 for the last 40 min of data using only data between sample 44 and 45 per ON and OFF state (Figure 4bis above). As expected there is an imbalance for the higher concentration from FTIR which are not matched by the DIAL.



Now of course one can argue that a correlation isn't necessarily expected due to the difference in light paths (see comments above).

Reply 11. Exactly. See also Reply 2. However, I still computed the Pearson cross-correlation coefficient, which is (for the data as shown in the discussion paper):

Path #1, 15 October: 0.07 (0.11 all pair, 0.03 with smoothed) (16:00 to 16:18), 0.09 (0.1 all pairs, 0.11 smoothed) (16:20 to 17:00)

Path #2, 16 October: 0.06 (0.14 with smoothed)

From comparing seismic well logs that have cm resolution with seismic surface data that has resolutions above 5 m I know that it is impossible to directly compare data that has been acquired on a very different scale (order of magnitude) in a robust manner. Moreover, a cross correlation coefficient alone is not always very meaningful. For instance, cross correlation is sensitive to phase differences. DIAL and FTIR time series sometimes have a very similar trend, but lag each other by tens of seconds, for instance around 16:07 (decrease by ca. 50 ppm) 16:08 (increase by ca. 50 ppm), 16:26 (increase by 150 ppm)). This could be simply because a dispersing CO2 pocket has reached the light path of one instrument, but not yet the light path of the other one. Moreover, for instance at 16:37 both DIAL and FTIR see an increase in CO2 concentration by ca. 100 ppm within ca. 4 min but at different rates and from different starting values. All this gives poor cross correlation, despite qualitative agreement.

However, the DIAL doesn't seem to show any change at all above the measurement noise during the measurements on October 15. This is surprising and to me indicates something probably isn't working properly.

Reply 12: Partly answered in Reply 9. Apart from the decrease in CO2 concentration after 16:00 (see Reply 7) the DIAL also indicates small scale variations by 200 ppm up to 500 (16:47)

Reply 13: We know from sensitivity tests done just before the La Caldara measurement that the DIAL instrument is capable of resolving variations in CO2 column mixing ratios by ca. 8000 ppm.m corresponding to a variation of at least 200 ppm at 49 m distance (length of path #1).

Given that the DIAL is aimed at the ground, it would expect it to be even more susceptible to variations in the gas flux (the 'pulsing' that you mention).

Reply 14: I agree that is should be like this. However, as written in Reply 2, the overlap actually was in the second half of that path and influenced by the CO2 vent. The area the beam was close to the ground (8 cm or less) was behind the CO2 vent though.

And the time resolution should be even better than that of the FTIR. So can you explain why you see practically no variations on timescales of around 5 minutes while the FTIR clearly sees these in the latter part of the measurement on October 15?

The bottom line is that the data seems to show that the chosen calibration of the DIAL gives similar mean CO2 columns as the FTIR. However, it also seems to show that it is incapable of measuring the variations on the order of 400 ppmm that the FTIR sees. This is worrisome.

Please see reply 9.

3. Minor comments and corrections

First of all, the manuscript would benefit a lot from English language proof reading. Parts of it are difficult to comprehend due to convoluted sentence structure and awkward vocabulary. It is beyond the scope of this review to correct all language related issues.

Reply:

The manuscript has been proof read.

4 P4328

L10 The instrument “is undergoing” validation studies. They are going on right now.

Corrected.

L15 – What are “Volcanic CO2 amounts”? please be more specific.

Changed to “CO2 concentrations at volcanoes”

L17 I fail to see how you would use a dispersion model to simulate CO2

concentrations in a

well mixed background atmosphere. I don’t think you would need a model for that, would you?

Changed ”In this case, using input data from point measurements to simulate column averaged CO2 amounts that are in line with measured CO2 column amounts can assumed to be straightforward” to “(…) dispersion modeling obsolete “

L21 – Non steady? I assume you mean variable in time?

Changed to: “variable in time”

L26 – “The FTIR used HERE is a MIDAC…”

Added.

P4329 L10 – “An interferogram is produced, which is…”

Changed accordingly.

L11 – “BY fitting …”

I am not a native English speaker. First I thought it is bad English, but I have seen a few papers where “through” is used instead of “b”y. Anyway, we have changed it to “by”.

L12 -- “path--averaged”

Changed.

L16 – “Many volcanoes exhibit quiescent degassing...”

Changed.

L19 – “Even IN good visibility,...”

Changed.

L28 – Corresponding to what? How about “slightly different wavelengths”

Changed accordingly.

P4330

L20 Here you talk about the normalization with the transmitted laser intensity. Later, you state that this was not done. If it’s not done, you don’t need to mention it here. Or say that it’s normally done, but couldn’t be done in this experiment.

I have moved the phrase that we did not normalize from below to here. Obsolete.

L25 - I did not understand the scheme used to mitigate the scintillation noise. Please explain it more clearly or just give a reference.

Obsolete now. Samples that are adjacent may correspond to the same statistical processes, e.g., an eddy. Averaging is more efficient if samples to be averaged are statistically independent. For this purpose we skip samples, taking every Nth to compute an average intensity ratio. Reference was already given.

P4331

L9 – “crossing each other IN HORIZONRAL DIRECTION at the IR source.”

Changed.

L10 “disperses INTO THE ATMOSPHERE”.

Changed.

P4333

L14 Instead of just giving the mean residual, showing it in a plot would be nice.

Obsolete as we do not show DIAL data anymore.

L15 – I think the main issue here is the height of the light path above the vent, not so much the FOV. See comments above.

I agree about the main issue, but we leave it as it still contributes.

5

L25 – Here and throughout the manuscript, all measurements are given in pmmm ABOVE THE BACKGROUND. This seems like a strange unit to work in. Both measurement techniques should measure the actual number of molecules in the light path, regardless of the background concentration. It is probably more of an issue related to the modelling approach – here perhaps you assume the background to be 0 and just work in ppmm above background. But that’s more of a technical issue, and to me it’s quite counterintuitive to work in ppmm above background. I would strongly recommend switching all values to absolute ppmm.

This is simply for display purposes. If you have a 100 ppm variation it becomes easier to see if you subtract the background.

P 4334

L2 – “data variability CAUSED BY atmospheric...”

Obsolete as we do not show DIAL data anymore.

L3 – “turbulence or water droplets emitted from the vent”

Obsolete as we do not show DIAL data anymore.

L5 – “it showed that, using the current...”

Obsolete as we do not show DIAL data anymore.

L11 – “the DIAL is currently being modified to allow for simultaneous sensing of the ON and OFF band signals.”

Obsolete as we do not show DIAL data anymor

L14 – I disagree with this statement. The DIAL has not yet been turned on at the beginning, and the FTIR is significantly lower (on average) than the model.

Changed to: The CO_2 concentrations from the OP-FTIR measurement decrease during the first ~20 min coinciding with the major change in wind direction (Fig. 4a). The modeled CO_2 concentrations from DISGAS do so as well (Model, Fig. 4a). This is remarkable as these values were not measured but simulated on a rather coarse scale, which we will discuss in detail further below.

L15 – you state that measured CO2 concentrations decrease during the first 15 minutes of the measurement, just like the model. But the measurements decrease much less than the model predicts! The model predicts a decrease to 20% of the initial value! Please be more quantitative! Also, it is unclear if the DIAL has a decrease at all.

We really hope that the revised paper will transmit the message better than this one obviously does: We are not comparing say optical power from a beam propagation in the lab with measured powers, but CO2 concentration indirectly retrieved from a dispersion model for utterly different scales. This comparison cannot be quantitative but is qualitative. At least for this experimental setup.

L22 – It is unclear to me why the 3m accuracy of the spatial location of the instruments is important if the model has a 10m spatial resolution. I assume the effect of this should be small...

I am quite sure that there are eddies with 10 cm scales influencing the CO2 distribute sensed by the instrument. So I think it may play a significant role. Added phrase: The path location in the model is based on GPS coordinates.

P4335

L27 – "... associated with October 16, where the atmospheric conditions..."

Changed accordingly

P4336

L3 – You mean 800 ppmm above background! That's about 1200 ppmm.

Added: above background level

L3 – "However, the mean modeled mixing ratio of 480 ppmm was still significantly higher than the measured value of 200ppmm..."

Changed to: The mean of the modeled mixing ratio of 480 ppm was still significantly higher than the measured value of 200 ppm, leading to a rather poor match between instrumental and model result.

P4337

L7 – “1. Discrete absorption measurement with DIAL”

obsolete as DIAL data removed from draft.

L8 – “2. Evaluation of absorption spectra...”

Evaluation is true, but inversion is more precise. Changed to evaluation anyway.

L14 – “... is currently being implemented...”

obsolete as DIAL part has been taken out.

6

P4342 – You might indicate the size of the modelling domain in figure 1.

We added a correponding note in the figure caption.

P4345 – The y axis label is misleading, as this is mixing ratio above background, not concentration. As mentioned before, I recommend switching all labels to absolute CO2 mixing ratio by adding the background value to all numbers.

In the caption it says above the background level of 370 ppm. In my experience, subtracting the background is common among volcanologists (I am not a volcanologist). It has advantages (see one of my replies above). We prefer to leave it that way.

Changes made by the authors

- introduced additional sub sections headers for better readability