Response to reviewer #1 Recent six-year atmospheric CO_2 concentration at the summit of Mt. Fuji observed by a battery-powered CO_2 measurement system

Shohei Nomura[a], Hitoshi Mukai[a], Yukio Terao[a], Toshinobu Machida[a] and Yukihiro Nojiri[a,b]

[a]Center for Global Environmental Research, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki, 305-8506, Japan.

[b]Hirosaki University, Bunkyo-1, Hirosaki, Aomori, 036-8560, Japan.

GENERAL COMMENTS

1. The paper present a new proof of concept and interesting application of a system for the execution of long-term continuous CO2 measurements to be used in a remote, unmanned measurement site characterized by the total absence (for several months) of electric power. The authors provided enough technical details on the proposed set-up to allow reproduction at other measurement sites. Moreover, new data were presented about CO2 variability at Mt. Fuji (Japan).

   >Thank you very much for your explanation.

2. My first general comment is that I'm not completely sure that this paper by Nomura et al., is completely fitting to AMT. As specified by AMT web site, AMT is dedicated to the" publication and discussion of advances in remote sensing, as well as in situ and laboratory measurement techniques for the constituents and properties of the Earth's atmosphere. The main subject areas comprise the development, intercomparison, and validation of measurement instruments and techniques of data processing and information retrieval for gases, aerosols, and clouds." The paper by Nomura et al. presented an automatic, low-consumption system (even if based on already well-known and used commercial system), for the execution of CO2 measurements at Mt. Fuji or at other remote, unmanned, adverse-weather measurement sites. Thus, for this aspect, it fits without any doubt to AMT. However, my feeling is that the paper is too unbalanced towards the analysis and exploitation of the CO2 data series provided by this system or in demonstrating the geographical representativeness of the measurement site (which is more fitting to ACP in my view), rather than (as should be done for AMT, I think) in presenting and commenting the experimental set-up or in demonstrating/assessing the quality and the strong/weak points of the presented system.

   >We re-arranged the manuscript for AMT. We made one section for explanation of system performance and added other explanations. We added our calibration information about stability of working standards, their usage duration and their scale differences from NOAA, which was the WMO standard scale. Also, we added the data about comparison between measurement results of a battery-powered measurement system and bottle sampling. In addition, we omitted a part of measurement results in atmospheric CO_2 concentration and shorten the measurement results part.

3. Another major point is that the authors did not provide enough information to allow a robust assessment of the system performance and, thus, possible application. I think that the scientific purposes of this set-up should be better addressed. As an instance, I'm wondering if this set-up (based on the NDIR LI-COR Li-840) can be able to meet the data compatibility goal (0.10 ppm) requested by WMO for long-term global CO_2 measurements. . .(I'm skeptical about this).

>The electrical power supply at Mt. Fuji is limited. No gridded power supply from September to following June. Therefore, Li-840 was better to be used in terms of saving energy consumption. We averaged the signal from Li-840 for 2 min in the case of standard gases, a linearity of the calibration and its reproducibility were very good. In the new section (3.1) which we added, we made explanation for checking the linearity and the differences between the measured values and the assigned values of the standard gases. The differences from linear fit were smaller than 0.05 ppm, suggesting that the uncertainty of measured values is within 0.1ppm.

4. Also the abstract should provide more details/information about the technical set-up and about the results of possible QA/QC test used to assess the quality of the data provided by the proposed set-up. Now, the abstract mostly provides a series of conventional atmospheric data analysis results, which (even if interesting) are not, I guess, the main topic for AMT.

   >We added the explanation about measurement system and QA/QC tests using two comparison works in the abstract. Also, we added the measurement performance for 6 years. According to your suggestion, we deleted some results of atmospheric data analysis.

5. I suggest that the authors can make more use of supplementary material (SM). Too many figures are presented in the manuscript up to now (17!). Some of them (e.g. Figure 2 and others – see my specific comments) can be moved to SM or skipped. The title clearly reflect the contents of the paper, but I would suggest recommend to modify keywords. I would suggest: carbon dioxide, unattended system, low-consumption system, Japan, long-term observations.

   > We decreased number of figures to six.
   Deleted: Table 2, Figure 2, 6, 14, 15 and 16,
   Merged: Figure 1 and 11, and Figure 3 and 4, and Figure 9 and 13,
   Created: one table and two figure, Move to supplementary material; Table 1).
   We changed the keywords based on the suggestion. But we would like to take "battery-powered system" instead of "unattended system", "low-energy consumption" instead of "low-consumption system", "Mt. Fuji" instead of "Japan", and "long-term monitoring" and. "Carbon dioxide"

6. By the way, several points should be addressed before publication (several of them are also related to assess the "performance" of this new set-up). Moreover, along the paper some inaccurate terms are used (e.g. "precision") and some analyses are too "qualitative" (i.e. lacking of robust statistic).

   >We changed the term "precision" to "repeatability" or "accuracy" depending on the cases. We tried to change qualitative explanation to more quantitative, using data and statistic.

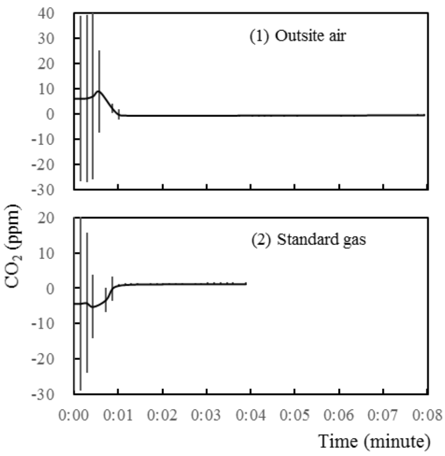
7. At some place in the paper, the authors should mention the exact period of operation of the measurements, i.e. "month 2009 – month 2015" (this info is only present in the abstract).

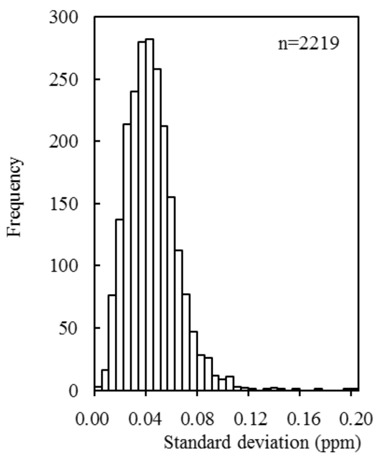
>We wrote the observation period (July 2009-December 2015) at the experimental section 2.4.

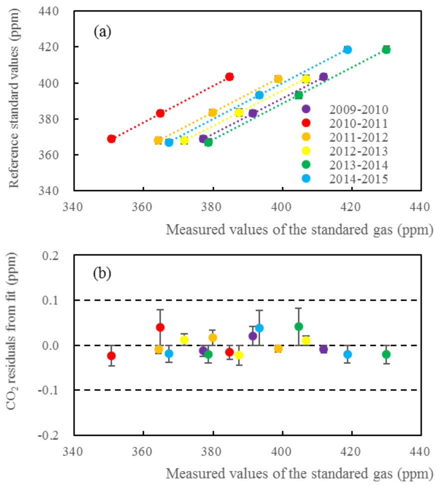
8. Thus, for summarizing, I think that the paper can be considered for AMT, but only after major revisions. Basically, I suggest to the authors to strength the analysis of the performance of their system, to clearly assess the typical application for which this system is suitable for (i.e. long-term monitoring – if yes which data quality objective can be reached- vs air quality vs event studies) and to consider the comparison with Mauna Loa time series as a “case study” or a “typical application” for the proposed system.

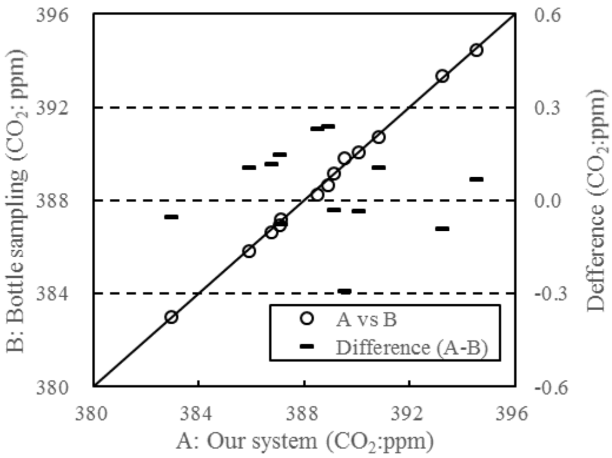
> We made major revision, as suggested. To strengthen the explanation about analytical performance, we added one section (3.1 analytical performances) which included calibration information on working standard gas and its stability. We added one table for the CO2 concentration stability in the standard gases cylinders. Linearity and deviation the data from the calibration curve between the measured values and the assigned values were evaluated. We also showed a results of comparison between CO_2 values of the system and the bottle sampling experiment.

	Before the cylinders installation				After the cylinders replacement				Change of concentration	
Cylinder No	Date of calibration	Calibrated value (ppm)	Pressure of Cylinder (MPa)	Date of installation	Date of replacement	Date of re-calibration	Re-calibrated value (ppm)	Pressure of Cylinder (MPa)	Change amount of the concentration (ppm)	Change rate (ppm year^{-1})
CPC-00449	17-Jun-2009	368.86	10.8	16-Jul-2009	24-Jul-2011	19-Aug-2011	368.82	3.4	-0.03	-0.02
CPC-00447	17-Jun-2009	383.10	11.0	16-Jul-2009	24-Jul-2011	19-Aug-2011	383.24	4.0	0.14	0.06
CPC-00448	17-Jun-2009	403.45	10.8	16-Jul-2009	24-Jul-2011	19-Aug-2011	403.54	3.2	0.09	0.04
CPC-00445	10-Jun-2011	367.94	11.4	25-Jul-2011	25-Jul-2013	6-Aug-2013	368.02	0.8	0.09	0.04
CPC-00450	10-Jun-2011	383.46	11.5	25-Jul-2011	25-Jul-2013	6-Aug-2013	383.42	3.6	-0.04	-0.02
CPC-00451	10-Jun-2011	402.29	11.5	25-Jul-2011	25-Jul-2013	6-Aug-2013	402.37	3.4	0.08	0.04
CPC-00043	23-Jun-2013	367.10	12.8	26-Jul-2013	1-Jul-2016	10-Jul-2016	367.12	2.5	0.02	0.01
CPC-00448	23-Jun-2013	393.17	12.6	26-Jul-2013	1-Jul-2016	10-Jul-2016	393.12	2.5	-0.05	-0.02
CPC-00449	23-Jun-2013	418.59	12.8	26-Jul-2013	1-Jul-2016	10-Jul-2016	418.44	2.5	-0.15	-0.05
CPC-00445	15-Jun-2016	389.18	13.2	2-Jul-2016						
CPC-00450	15-Jun-2016	409.15	13.2	2-Jul-2016						
CPC-00451	15-Jun-2016	429.16	13.2	2-Jul-2016						









SPECIFIC COMMENTS

9. Concerning the experimental set-up, I strongly recommend to change the PTEFE with stainless steel or Synflex 1300 tubing (PTFE is well known to be not optimal for CO2 sampling). By which frequency did you change the PTFE tubing? I would like to see more information about the working standards for routinely calibration of the system (i.e. flask material, mixture matrix, pressure regulators) and more discussion about the choice of your calibration strategy (are 4 min sufficient enough to obtain stable measurements?).

> We wanted to use stainless steel tube or Synflex for inlet tube. But Mt. Fuji including the station is categorized as the National park. When the equipment install outside the station, the equipment was required to be colorless and transparent to conserve the landscape. So we chose the PTFE tube for the inlet tube, as a second choice. We did not replace the PTFE tube after the first installation at the station. But we made leak check every summer. This detail was written in 2.2 measurement system.

We added table 1 which showed calibrated values, pressure and re-calibrated values for the standard gas cylinders. These standards were produced by Japan Fine Product Co. in aluminum 10L cylinder. CO2 gas was diluted by purified natural air as zero air. Special regulators for cold environment use which were provided by Nissan Tanaka Co. were used. This detail was written in 2.4 measurement sequence.

We added one figure (Fig. 3) to show the typical NDIR signal for standard gas and outside air measurement. As shown in the figure, signal became stable 2 min after starting measurement. So, we discard the data for first 2 min and averaged data for the rest of the time (in the case of standard data was averaged for 2 min, in the case of the air data was averaged for 6 min.) The measurement sequence (3 standards=> room air => outside air) was repeated for 4 times and 2^{nd}, 3^{rd}, 4^{th} data were averaged as represented data (1^{st} cycle data was discard, because we would like to most stable data). Tish detail was written in 3.1 Analytical performance.

We have already known the stability of the standard gas in 10 L aluminum cylinders from the many experiences in our other long-term observation. We added the data for the stability of the standard gases used for the Mt. Fuji.

10. Pag 3, line 16. You cannot use this comparison of data for evaluating the precision (i.e. the repeatability) of your measurements! I also have serious concern that you can provide an assessment of spatial representativeness by using this data-set. . .

>Yes. It was analytically incorrect to use CONTRAIL data for evaluation of repeatability. We added our comparison work using bottle sampling to assess the accuracy of our measurement system. The results also showed a good agreement within an analytical precision for bottle sampling and measurement system. This detail was written in 3.4 comparison with the bottle sampling data.

It may be difficult to compare our data with CONTRAIL data directly. But as a result, their good agreement in concentration suggested that data at Mt. Fuji could show the similar CO2 concentration level at the altitude over 3600m around Tokai-Kanto district.

11. I think that the authors should spent more work in presenting an assessment of the quality of the system performance. Along the manuscript, the authors provided a statistical analysis (Fig. 5 of the manuscript) of the standard deviation of the measured value of standard gases. This information can be useful to assess the "long-term repeatability" of the measurements (see e.g., Lebegue et al., AMT, 2016). However, to better exploit this point, I would like to see this analysis also split as a function of the different standard gases used and/or as a function of the different years. It should be interesting to see the absolute difference between the CO_2 measured values of the standard gases and the reference standard values. Moreover I would like to see information about the uncertainty or expanded uncertainty of the measurements achievable by this system, which is an important parameter missed in the paper.

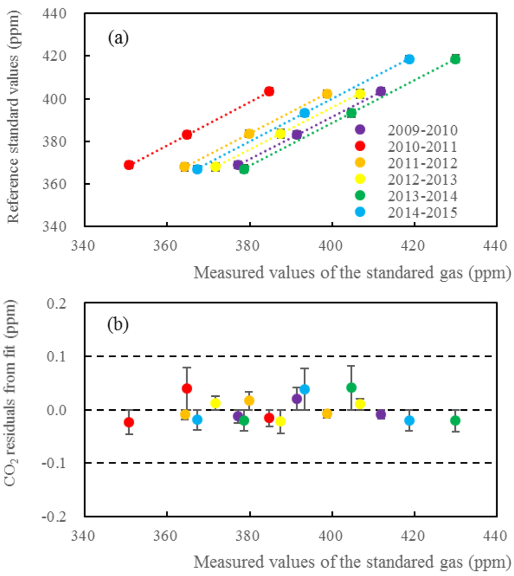
>We added Table 1 which showed the CO_2 first calibrated values of the working standard gas and the lastly calibrated values after use. The differences between them were about 0.1 ppm, means that they are very stable in the cylinders. Also, we can correct this kind of small deviation in the standard gas values by using weighted average with time (even so maybe small uncertainty (30% of 0.1 ppm = 0.03ppm = detection limit of NDIR in NIES lab) will be left). If we looked at the linearity of the calibration curve measured by the system, the deviation from the calibration curve was also small (at most 0.05ppm). Real signal (10 s average) had 0.05ppm on average as a standard deviation for repeating measurement. It is also said that NIES CO_2 standard series (NIES09 standard) matched with NOAA (WMO) standard within 0.09ppm.

Therefore, roughly we can estimate uncertainty for our measurement,

$$((0.03)^2 + (0.05)^2 + (0.05)^2 + (0.09)^2)^{0.5} = 0.12 \text{ ppm } (k=1)$$

If k=2 (expanded uncertainty), it should be 0.24 ppm.

If we remove the difference between NIES scale and NOAA scale, it is 0.08 ppm (k=1) or 0.16ppm (k=2)



12. Concerning the difference with MLO, which may be the role of using two calibration scales (NOAA vs NIES09)? Please quantify. . . How many often did you re-calibrate the working standard against your National reference (I suppose once per year, but this should be clearly stated in the paper)?

>We wrote the results of the 6th WMO/IAEA Round Robin inter-comparison whose results that the NIES09 scale was lower than the NOAA scale by 0.04-0.09 in a range of 376-404 ppm. We re-calibrate the working standard gas every two or three years. We created the table 1 which showed the information of standard gas (Days of calibration and re-calibration days of each cylinders).

	Before the cylinders installation				After the cylinders replacement				Change of concentration	
Cylinder No	Date of calibration	Calibrated value (ppm)	Pressure of Cylinder (MPa)	Date of installation	Date of replacement	Date of re-calibration	Re-calibrated value (ppm)	Pressure of Cylinder (MPa)	Change amount of the concentration (ppm)	Change rate (ppm $year^{-1}$)
CPC-00449	17-Jun-2009	368.86	10.8	16-Jul-2009	24-Jul-2011	19-Aug-2011	368.82	3.4	-0.03	-0.02
CPC-00447	17-Jun-2009	383.10	11.0	16-Jul-2009	24-Jul-2011	19-Aug-2011	383.24	4.0	0.14	0.06
CPC-00448	17-Jun-2009	403.45	10.8	16-Jul-2009	24-Jul-2011	19-Aug-2011	403.54	3.2	0.09	0.04
CPC-00445	10-Jun-2011	367.94	11.4	25-Jul-2011	25-Jul-2013	6-Aug-2013	368.02	0.8	0.09	0.04
CPC-00450	10-Jun-2011	383.46	11.5	25-Jul-2011	25-Jul-2013	6-Aug-2013	383.42	3.6	-0.04	-0.02
CPC-00451	10-Jun-2011	402.29	11.5	25-Jul-2011	25-Jul-2013	6-Aug-2013	402.37	3.4	0.08	0.04
CPC-00043	23-Jun-2013	367.10	12.8	26-Jul-2013	1-Jul-2016	10-Jul-2016	367.12	2.5	0.02	0.01
CPC-00448	23-Jun-2013	393.17	12.6	26-Jul-2013	1-Jul-2016	10-Jul-2016	393.12	2.5	-0.05	-0.02
CPC-00449	23-Jun-2013	418.59	12.8	26-Jul-2013	1-Jul-2016	10-Jul-2016	418.44	2.5	-0.15	-0.05
CPC-00445	15-Jun-2016	389.18	13.2	2-Jul-2016						
CPC-00450	15-Jun-2016	409.15	13.2	2-Jul-2016						
CPC-00451	15-Jun-2016	429.16	13.2	2-Jul-2016						

13. Another important information to provide is the data coverage (%) over the measurement period: I feel this is rather high, but this represent a parameter (you must mention it along the abstract!) to evaluate the success of the presented system.

>We got CO_2 daily data of 2219 days from July 2009 to December 2015 (for 2354 days), which covered 94% for the observation period. We wrote that on the abstract and results.

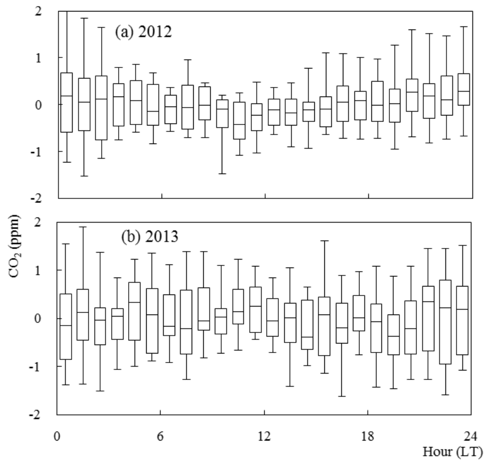
14. Pag 6, line 2: "The precision. . ..is calculated to be about 0.3ppm". As recommended by the WMO/GAW glossary, "precision" is a term to be used only in relative terms. How did you calculate this "precision"? If this is a standard

deviation, it is the "repeatability" or the "reproducibility". Any information about the "uncertainty" of the measurement?

>It was repeatability. We changed an expression from precision to repeatability.

15. Pag 7, line 24. For year 2012 it seems that the CO2 variability (error bars) experienced a diurnal behavior (higher variability during night-time): please comments. Does a diurnal cycle exist for internal temperature or other system parameters that can affect the measurements? Figure 8: you did not provide explanation in the Figure caption for the error bars (please, do it). To investigate if statistical significant diurnal variability exist, you should plot for each hour an average mean value with the 95% confidence interval (not simply the standard deviation).

   > In data of 2012, a fairly small lower tendency (lower than 0.4ppm) in concentration may be admitted before noon time. It often included vegetation effect. However, it was concluded that the timing of the air sampling over Mt Fuji did not appear to affect the monthly and yearly averages. We changed an expression of the figure 8. We plot each hour an average mean value with the 95% confidence interval.



16. Pag 8, line 6: ". . .a strong wind always blow". Too generic. Please provide data or references.

   >We add the sentence which is annual average wind speed is about 12 m^{-1}.

17. Pag 8, line 11: "it was evident that. . .often see in July". It is difficult from this plot to obtain information about monthly values. I would suggest to replace this plot with monthly mean values and related error bars (representing the associated 95% confidence interval). This would also allow to skip Table 1.

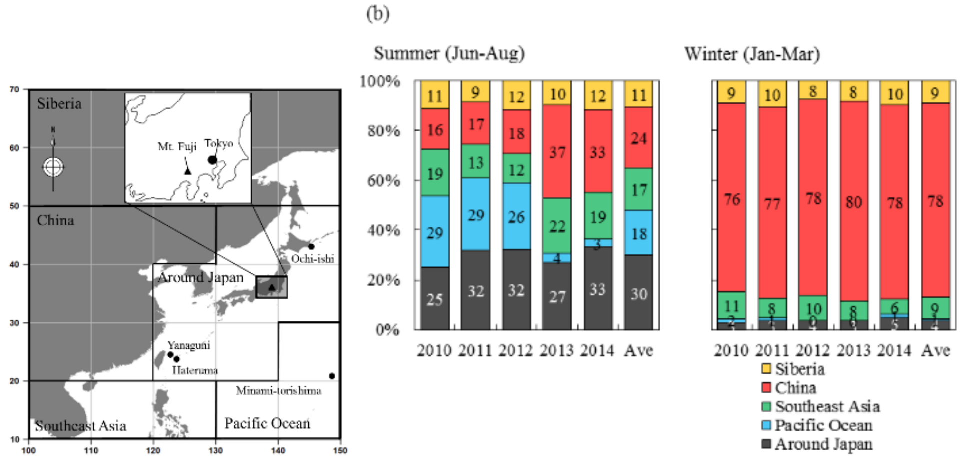
   >If the CO2 data of Mt. Fuji change from daily data to monthly data, extremely low concentration and high concentration of the CO2 data of Mt Fuji cannot be seen. These specific events are characteristic in the CO2 concentration of Mt. Fuji. We remained the figure of daily CO_2 data of Mt. Fuji. Table 1 as monthly average was transferred to supplementary material to skip it.

18. Pag 8, line 24: Since the CO2 is characterized by a long atmospheric lifetime, 72h back-trajectories can explain just a

fraction of its variability. Please comment in the paper. It is possible to plot the back-trajectories (as "number concentration field" or as centroids of the detected clusters) over the spatial domain in Figure 11? In any case, consider to move Figure 11 to SM. You should also provide the percentage of occurrence for the different air mass transport class. The discussion about results reported by Figure 12 is too qualitative. For the summer, it seems that a large amount of variability affect the data set. For each class of air mass transport please provide average values and 95% confidence interval. How did you define background air (Pag 8,line 27)?

>We added some explanation about meaning of 72 hr trajectory. To evaluate the variation of frequency of each sector, we added one Figure about that in the Fig 11. According to the variation of their frequencies we could see some reason for the variation of the concentration, especially in the case of summer. So, we rewrote this section and omitted a large part (e.g. long-term trend). This detail was written 3.5 effect of air mass origin on seasonal variation.

However, we left explanation about relative change in concentration depending on trajectory in each season.



19. Pag 9, line 1: In my opinion, a too low number of Pacific Ocean air-masses were observed during winter (at least basing on the visual inspection of Figure 12 since no information about their seasonal occurrences were provided) for provide any kind of comment. "air from China SOMETIMES. . .". Too qualitative, please be more specific! How many times (frequency) did you observed these spikes with air-mass from China? The sentence from line 9 to 12 is not clear. I think some words were missed somewhere. . .However, I do not think that this paragraph and figure 13 add really important information to the paper, thus they can be skipped.

> We showed the air mass origin in each season with figure that is showed the percentage of occurrence for the different air mass transport in winter. We added numbers in the explanation (e.g. frequency) if it is needed. However, we rewrote this section, based on the suggestion.

20. The comparison with CONTRAILS observations must be better commented (or even skipped). The authors claim for a good agreement but I do not completely agree: a rather large "average" bias (2 ppm) can be deduced by the linear correlation result. In general I do not think that you can use this comparison for "verify" the precision of the measurement system (the term "precision" is not correct at page 10, line 16)! Also concerning the assessment of the representativeness of the measurement site, I think that this analysis can provide only a "qualitative" information (also considering that CONTRAILS observations are not uniformly distributed on the geographical domain). As an instance,

the most part of the CONTRAILS observations are recorded not far from Tokyo. Would this introduce a bias on the contrails data-set due to the urban emissions? Can this (at least) partially explain the 2 ppm bias? Moreover, it is not clear what do you mean for "daily data". For CONTRAILS what kind of daily average did you consider? 24h mean values? Or did you select the hourly data on a time window in agreement with Fuji measurements (i.e. 14:100-17:28 and 21:00-00:28)?

>We changed explanation about this section. Bottle sampling data was included in the section for showing analytical performance. However, CONTRAIL data comparison was important to show the representativeness of measurement at Mt. Fuji.

The bias between Mt. Fuji and CONTRAIL is 0.05 ppm. Two ppm is standard deviation between Mt Fuji and CONTRAIL. The 2 ppm might be originated by the difference in measuring time (Mt. Fuji were taken at night and CONTRAIL's were taken throughout the day).and the measuring places (CONTRAIL data included many cases over Boso peninsula (Chiba prefecture), because Narita airport is located in Chiba prefecture. Some data originated from flights over Nagoya Chyubu airport). However, data of CONTRAIL at altitude of 3.6- 3.9 km was found to be unaffected by anthropogenic emission directory from Tokyo area (Shirai et al., 2012 at Tellus B), rather affected long-range transport or background air. In this case we are looking at the similarity in CO_2 concentration level over these areas.

21. Fig. 14 is almost unreadable. Mt. Fuji points completely overlap the "reference" time series of MLO. I'm wondering if it's really necessary to show the whole long term time series of MLO: for the most part of it, you do not have any data to overlap!!! I would suggest to start the x-axis since 2010 and make the MLO point more visible.

>We re- arranged the figure to expand it from 2010 to 2015.

22. Fig. 9, line 17: "The RANGE of the annual rates of increase during 2009-2015 was 1.5-2.7 ppm/yr". Actually, the variability in the growth rate can be also related to other factors, as an instance ENSO (as mentioned in the following by authors) or to change in fluxes between atmosphere and terrestrial biosphere.

>Exactly. In this case, we are checking general trend of the trend, comparing with MLO

23. Pag 10, discussion of Fig. 15. Some statistical analysis should be provided for the linear correlation (i.e. statistical significance). In general, I think that is dangerous to compare and comment (small) long-term CO_2 differences between two measurement sites without providing an assessment of the measurement compatibility. . .I'm wondering if these tendency in long-term differences are evident also at other measurement sites. In this case, this would reinforce your conclusions. I recommend to consider open-access data-sets available at the World Data Center for Greenhouse Gases by WMO/GAW and to repeat a similar analysis for other sites in the East Asia/Pacific regions vs MLO.

> We also agree with your suggestion for this case. The discussion for trend was omitted. We deleted Fig. 15.

TECHNICALS:

24. Pag 2 line 10: “(CDIAC, Boden et al., 2015)” line 15: “air mass transport” line 23: “Mt Fuji is positioned...of the Asian continent”. Please provide references or show back-trajectory analysis.

>We changed from CDIAC, Boden et al., 2015 to CDIAC 00001_V2016; Boden et al., 2015. We added the reference as (Igarashi et al., 2004)

25. Line 22: “Mt. Fuji is located. . .” Line 27: What do you mean for “irregular”? Please use a more specific terminology.

>We deleted “irregular”

26. Pag3, Line 28, point (v): it is difficult to obtain information on the annual rate of increase basing on so “sparse” measurement.

>Aircraft measurements which is showed the measurement results at the altitude of around Mt. Fuji (3776 m). We changed the sentence “(v) The observed annual rate of increase was comparable with the rate derived from aircraft measurements over Japan at the equivalent altitude the summit.”
In general, if even we have only data once a week, we can discuss trend roughly if the data shows regional background data.

27. Pag 4 Line 2: “shows a picture. . .” Line 3: “The system consists. . .” Line 16: “In addition, WHEN the temperature. . .”

> We changed this sentence as you indicated.

28. Pag 8 Line 27:” It was found,. . .IN SUMMER”

> We changed this sentence as you indicated.