Response to Referees #1-4

We thank the four anonymous reviewers for their valuable comments on our manuscript. The questions and requested revisions have been carefully considered and implemented to improve our manuscript. Our responses to the reviewers are listed below. We have noted the contribution from the four anonymous reviewers in our acknowledgement.

Anonymous Referee #1
General comments (GC):
(GC-1) (1)-I: Authors should prepare the manuscript carefully without necessary repetition and redundancy. Especially papers focused on the methodology should be prepared so that the readers can reconstruct the described method.
Reply: We have significantly revised the manuscript accordingly, providing much more detailed description of the methodology.

(GC-2) (1)-II: Please give manufacture's name and location (city, country) in parentheses for instruments used in this work.
Reply: In the revised manuscript, we have provided the information accordingly.

(GC-3) (1)-III: Author should standardize the term and wording throughout the manuscript (e.g., "concentration" and "mixing ratio", "flask sample analysis system" and "automated analysis system").
Reply: We have standardized the terms and wordings in the revision accordingly to avoid confusion in the reader.

(GC-4) (2): Authors made comparison between the flask and the surface continuous measurements to validate the flask sampling method. The authors stated that differences for each trace gas (CO2, CH4, CO) between these two measurements (0.2±0.3 ppm for CO2, -0.7±3.3 ppb for CH4, 2.4±4.1 ppb for CO) are not significant due to the uncertainty of sampling, analysis, and difference in the sampling height. However, the differences and corresponding uncertainties seem to be large for validation. In other word, these results cannot remove any potential bias within the uncertainty associated with the flask sampling method. Why did the authors choose other method to reduce these uncertainties? For example, comparison with in-situ measurements will remove the uncertainty due to the difference of sampling height. In addition, what about the validation for N2O?

Reply: We chose the field comparison method, because we viewed this approach as being the only and simplest way of validating the quality of the sampled air taken from the air-conditioning system during the actual flight of the C-130H aircraft. To achieve this purpose, the field inter-comparison using the aircraft was essential, although the experimental uncertainties in a field comparison are usually found to be larger than those found in a laboratory experiment, as pointed out by the reviewer. Comparison with in-situ measurements will indeed remove certain types of uncertainty, such as those due to the difference in sampling heights, as commented by the reviewer. However, such comparison of in-situ measurements with flask measurements on an aircraft will still have uncertainties resulting from such factors as atmospheric variability, flask-to-flask variability and possible flask sampling biases (Karion et al., AMTD 2012). For example, some previously reported uncertainty values are 0.14 ±0.11ppm for CO_2 and -1.10±2.12 ppb for CH_4 (Rella et al., AMTD, 2012), and -0.16±0.42 ppm for CO_2, 0.4±1.8 ppb for CH_4 and -0.6±2.8 ppb for CO (Karion et al., AMTD, 2012). Compared with these results however, our deviations are still larger.

Following the suggestion made by another reviewer, we re-evaluated the comparison results using both measurements at the nearest sampling time, instead of the hourly mean data as was done before. As a result of this re-evaluation, the differences and uncertainties have been largely reduced. The mean differences of 22 individual measurements (aircraft data - surface data) are now estimated to be +0.08 ±0.23 ppm for CO_2, +0.24 ±2.6 ppb for CH_4, and -0.17 ± 2.4 ppb for CO. These re-evaluated results are described in the revised manuscript, and the remaining uncertainties are discussed.

For N_2O, its mole fraction is expected not to change in the air-conditioning system, because this gas is very stable compared with other trace gases. Thus, we did not measure N_2O in this experiment.

(GC-5) (3): Authors stated that “we report some specific technical improvement of the laser-based instruments” in the last sentence of Introduction. I could not perceive such technical improvement. Please describe this improvement clearly.

Reply: We used the laser-based instruments for analyzing discrete air samples, and the instruments’ performances were evaluated. In the process we made some technical modifications to improve the measurement process. In response to the comment, we revised the manuscript to clearly identify those modifications.

Other comments (OC):

(OC-1) Section 1: Introduction is poorly written and constructed. Authors laid out research background with emphasis on Asian region only based on a small amount of references, dominated by CO2 research. Furthermore, none is mentioned for N2O. Authors should revise Introduction with a sufficient amount of references and mention the need for the observation of CO2, CH4, CO, and N2O. To my knowledge several aircraft observations have been performed including CARIBIC and MOZIC/IAGOS as well as CONTRAIL. These works should be referenced properly to explain the advantage of aircraft observation.

Reply: Introduction has been re-written to include additional information, including brief overviews of N_2O and other trace gases, as well as aircraft observations of CARIBIC, MOZAIC/IAGOS and CONTRAIL. Relevant papers have been referenced.

(OC-2) Section 2.1: P7070/L19-L23; These sentences describe the measurement system at MNM station, not for sampling flight. Please edit (the sentences may be moved to section 3.1.).

Reply: Following the suggestion, we moved the sentences to section 3.1.

(OC-3) Section 2.1: P7071/L5-L8; Sentences describe the flask air sampling. Please move to section 2.2.
Reply: Following the suggestion, we moved the sentences to section 2.2.

(OC-4) Section 3.1: Many description overlap with section 2.1 and 2.2. Please edit and shorten.
Reply: Following the comment, the overlapped descriptions were deleted and/or shortened.

(OC-5) Section 3.2: P7076/L2-L3; The conditioning method is already mentioned in the section 2.2.
Reply: Accordingly, the description of the conditioning method in section 3.2 has been deleted.

(OC-6) Section 3.5: This section should be shortened and be merged with section 3.6. Authors inferred the possible isotopic effects on the OA-ICOS measurements for CO and N2O based on the theory for WS-CRDS. Why the same theory can be applied to the OA-ICOS? At least, authors should explain the measurement principle of the OA-ICOS, and subsequently discuss the application of the isotopic correction.
Reply: The descriptions in this section have been significantly shortened by summarizing them in the Table 2, as well as merging them into section 3.6.

The same theory can be applied to the OA-ICOS, because the OA-ICOS analyzer measures the number of $^{12}C^{16}O$ and $^{14}N_2{}^{16}O$ molecules, and determines the total CO and N_2O mole fractions by dividing the fractional abundance of $^{12}C^{16}O$ and $^{14}N_2{}^{16}O$ in ambient air. If the fractional abundance of the standard gas used is different from that of the ambient air, the measurements of total mole fractions of CO and N_2O could be affected according to their isotopic ratios. Thus, the isotopic effects of the OA-ICOS system can be calculated by the differences in the fractional abundances between the standard gas and the ambient air, and could be estimated from the same theory. These explanations are now added in the revised text.

(OC-7) Section 3.6: What is main message to readers in this section? Estimation of the isotopic effects? or measurement comparison? If authors intended to do “to evaluate whole instrumental bias”, the comparison experiments should be done more carefully and described in great detail. For example calibration for each instrument should be described in detail. Please revise.
Reply: Our intent was to “evaluate whole instrumental bias” when we used the synthetic standard gases to calibrate the laser-based analyzers for the analysis of the ambient air samples. Following the reviewer’s suggestion, we largely revised this section to describe the calibration method and to provide more detailed discussions.

Specific comments (SC):
(S1) P7071/L11: What is the modification? Please specify it.
Reply: We modified a commercially available pump so that it can be operated by man-power instead of electrical power. The description has been revised to clarify this.

(S2) P7071/L25: What is the sample flow rate? can obtain constant flow?

Reply: The flow rate is more than 6-8L/min for flushing the sample flask. After the flushing for more than 5min, it takes about 1 min to pressurize the flask to about 0.4 MPa with sample air. Although the flow rate can not be controlled accurately because of the manual pumping, the flow rate is roughly monitored using a flow meter attached to the sample vent line. A brief description of the sample flow was added in the revised manuscript.

(S3) P7071/L25: The air sampling was made by manual pumping. Why? There is no available power source?

Reply: We can not use the electric power source on board the aircraft, because any emissions of electro-magnetic interference from the observation instruments are prohibited in the Japanese army aircraft. A brief description to this effect has been added in the Introduction section.

(S4) P7072/L3: Problems on the stability of the trace gases during storage should be explained briefly by citing previous reports especially for CO (e.g., Novelli et al., JGR 2003).

Reply: Following the comment, we added a brief explanation and related references.

(S5) P7072/L11: What is new? If the system is improved from the old one, please give an explanation of the old system and improvements.

Reply: The word “new” was used to denote construction for the purpose of use in the measurement program. Since this word is confusing, it was deleted.

(S6) P7072/L17: Is the NDIR and VURF installed permanently? If that so, why were these instruments installed while CO2 and CO are measured by WS-CRDS and OA-ICOS. Unnecessary measurements waste valuable sample. Please specify the reason.

Reply: As a permanent feature of our measurement program, we use both analyzers for cross-checking purposes, to ensure the analytical precision and as a precaution against any data loss due to equipment failure in one of the analyzers. We have added this information to the revised manuscript. As a footnote to this, we have, in order to utilize as much of the air in each flask as possible, started the analysis of O_2/N_2 and Ar/N_2 using sample air that remained after the mole fraction measurements.

(S7) P7072/L22: Each analytical instrument seems to have different sample flow length from the flask. Such configuration is usually unpreferred due to possible fractionation of the air sample. Why did you develop the system with such configuration? Please explain.

Reply: The difference in the sample flow length from the flask to each analytical instrument is small, because the corresponding air volume from this difference is quickly purged at a flow rate of a few hundred mL/min. In addition, the inner wall of the tubing is treated by electrochemical polishing to avoid any change in the mole fraction due to desorption/absorption on the inner wall. Thus, possible fractionation is negligible, as shown by the results in Table 1. This configuration is not special, because we use airflow tubing instead of a manifold.

(S8) P7072/L24: Please specify "a special operational program"
Reply: It refers to a custom-made control board connected to a personal computer (PC) with a custom-made operation software. In the revision, we have deleted the phrase, and instead added a brief explanation of the automated system.

(S9) P7072/L25: Please define "the main airflow line"
Reply: We have re-worded the phrase, to clearly indicate what it means in the revised manuscript.

(S10) P7072/L28: Please define "a vent line"
Reply: It is indicated in Figure 1 and a brief explanation is now added in the revised text to define it.

(S11) P7073/L3: What is the sample dew point?
Reply: It is roughly estimated to be less than -40 ~ -50ºC from the H_2O measurements of the WS-CRDS and ICOS analyzers, although the accurate dew point was not measured by other sensors. The H_2O content from the WS-CRDS is described in section 3.3.

(S12) P7073/L5ff: The subsample flows are substantially lower than those of default flows for the individual instruments. What is the reason for the flow regulation?
Reply: The flow regulation controls the consumption of sampled air, as well as working standard gases for the analysis. The default flow for the WS-CRDS is more than few hundreds cm^3 per minutes, which cannot be regulated precisely. The flow control is essential for the flask sample analysis because of the limited volume of sampled air available for analysis.

(S13) P7073/L10: Please give detailed description of the auto pressure controller.
Reply: In addition to deleting the word "auto" (because of possible confusion), we have added a more detailed description of the pressure controller in the revised text to make it clearer.

(S14) P7073/L16: Does the evacuation affect the absorption/desorption balance of CO2 and CO onto the inner wall of the line?
Reply: To minimize the effect of the absorption/desorption, the inner wall of all the tubing is treated by an electrochemical polishing. We can not quantitatively determine the effectiveness of this treatment, but no significant influences have been experimentally detected since we have a good agreement between the measured and assigned mole fractions, as shown in Table 1.

(S15) P7074/L1: Please specify the standard gases (balance gas, mixing ratios)

Reply: We have specified the balance gas and ranges of mole fractions of all the trace gases, including the standard gases, in the revised text. Approximate mole fraction for each standard gas is identified in Figure 4.

(S16) P7074/L21: Matsueda et al. (1996) is not found in the references list.
Reply: Thank you for catching this typo. It should have read “Matsueda and Inoue (1996)”, which is listed in the reference section.

(S17) P7075/L3: Stainless steel flask? The titanium flasks were not used.
Reply: The titanium flasks were not ready for use during the test flights, so we ended up using the available stainless steel flasks. The paper has been revised to more clearly point this out.

(S18) P7075/L5: What is the sample quality of N2O?
Reply: We were not able to evaluate the sample quality of N_2O, because the amount of sample air remaining after the analyses of other three species was insufficient for N_2O analysis.

(S19) P7075/L8-L9: Authors stated that the flask samples were also collected at the surface station for the comparison. But in next paragraph authors stated that the flask measurements were compared only with surface continuous measurements. What were the flask samples from the surface station? If the results from the flask measurements are not discussed in the manuscript, the description should be removed.
Reply: We agree, and the relevant sentence has been deleted.

(S20) P7076/L1: What was the water vapor concentration in the dry natural air?
Reply: The natural air was dried by a dryer tube (35cm long, 3/4 in) packed with magnesium perchlorate, the same drying procedure used for the aircraft flask sampling on board the aircraft. But we did not determine the water vapor concentration after the drying was completed. This is now pointed out in the revised.

(S21) P7076/L9: You did not make the storage test for N2O?
Reply: We did not do so because we had insufficient amount of sampled air to conduct N_2O analysis.

(S22) P7076/L26-L27: What does the sentence mean?
Reply: This sentence means that the frequency of the standard gas introduction is sufficient to precisely calibrate the sample measurements. Indeed, the instrumental drifts for all the analyzers used in this study were found to be negligible within this time interval. We modified the sentence to make it clearer.

(S23) P7080/L21: Nara et al. (2012) reported that the pressure-broadening effects due to the matrix gas variation can be significant impact on the CO2 and CH4 measurements by WS-CRDS. Your statement is inconsistent with that by Nara et al. (2012). Please revise.

Reply: Our description was insufficient. It has been revised to get rid of the inconsistency.

(S24) P7080/L25: Please describe the calibration gases used for the WS-CRDS and NDIR instruments.
Reply: We added a description of the calibration gases used in both instruments.

(S25) P7081/L2: Please notice the significant digits.
Reply: We have corrected this mistake by reducing the number of digits by one, thus consistent with the analytical precisions of our instruments.

(S26) P7083/L17-L21: These sentences should not appear in this section, they should rather be mentioned in Introduction. Please move to the section 1.
Reply: In accordance with the suggestion, we have moved these sentences to section 1.

(S27) Table 1, For CO2 difference, 0.09 ppm is significant, suggesting analytical bias between the JMA measurement system and "conventional methods". This should be considered carefully in this evaluation.
Reply: In response to this comment, we carefully checked the assigned value of CO_2 mole fraction and found that there was a typo. As a result of this correction, the CO_2 difference was reduced from 0.09 ppm to 0.06 ppm. Furthermore, we carefully checked all the 6 flask measurements, and found that one measurement out of 6 showed significant deviation due likely to insufficient filling of one of the flasks. When we excluded this data, the CO_2 difference was reduced to be 0.04 ppm. Thus, we revised the CO_2 data in Table 1. In addition, the mean data of the 5 flask measurements were also added to Table 1 and its explanation has now been included in the revised paper.

(S28) Table 1, footnote b: what are conventional methods? Please clarify.
Reply: The "conventional methods" refer to the use of the NDIR and gas chromatograph methods in the JMA standard gas calibration system. In response to the reviewer's comment we have added some additional description to clarify them in the footnote b of Table 1.

Anonymous Referee #2

General Comments (GC):

(GC-1) In my eyes, there is no completely new measurement development presented (Laser-based instruments in a similar setup are used elsewhere already), and the observed differences to the existing setup are scientifically not sufficiently interpreted. Because there is no data from the measurement campaigns presented here, the paper might rather be changed into a Technical Note to describe the current JMA calibration system.

Reply: We agree that there are less specific technical improvements in this study, but we first used the laser-based instruments for analyzing discrete air samples, and its performances evaluated in our measurement program. The results in our study could provide some useful information about the reliable techniques related to the use of laser-based instruments in an atmospheric observation platform like the one used in our program. Thus, following the reviewer's suggestion, we changed the tile of this paper, as described in the specific comment (SC-1).

(GC-2) However, the changes in the measurement procedures should be better motivated and investigated in greater detail; the manuscript needs major revisions. The main topics to work on are the correction of isotopic effects (between instruments and calibration gases), motivation of metal flask usage, and storage effects.

Reply: In accordance with these comments, we have largely revised the manuscript to clearly reflect and identify our study objectives. Furthermore, we significantly improved this section, incorporating the comments by other reviewers.

Specific Comments (SC):

(SC-1) P7069/L21: The title might be changed to "Technical Note: The JMA flask sampling and trace gas measurement system".

Reply: We have now changed the title to "Technical Note: The JMA flask sampling for a cargo aircraft C-130H and trace gas measurement system"

(SC-2) P7068/L20: The introduction gives an overview about the gas measurements in East Asia. Except the last paragraph, it does not give insights into the used measurement systems. However, for this paper, the location of the air sampling plays a minor role (it is not even used to argue about the observed differences at MNM station). It would be more interesting to give an overview about the flask sampling techniques that are currently used by different groups. At page 7069/22 I miss a discussion about the advantages and disadvantages of flasks over continuous measurements as a motivation for the whole paper. Why do you not use the instrumentation directly onboard the aircraft? What are the expected advantages and disadvantages of the new analyzing system?

Reply: We modified Introduction to identify more clearly as the main objective of the paper a detailed overview description of the flask sampling techniques and the measurement system introduced in our new aircraft observation program to sample trace gases at relatively low altitudes from Honshu (main Japanese island) to the JMA-operated MNM monitoring station in western Pacific.

We agree with the reviewer's comments that it is better to install the instrumentation directly onboard the aircraft. However, there are several logistical restrictions associated with conducting atmospheric measurements onboard a military aircraft, such as having no access to electric power and limited space. In particular, any emissions of electro-magnetic interference (EMI) from observational instruments are prohibited on any of the Japanese army aircraft. For these reasons, we were not able to install and operate a continuous measuring system on board the aircraft**.** Instead, we had to settle for the manual flask air sampling procedure without electric power supply, as described in the paper. We have added this information in the revised manuscript.

We have also made clearer identification of the improvements we have made in the new analyzing system we built, such as high precision, improved stability, low maintenance, and easier operation, compared with various conventional gas chromatograph methods. A further advantage is the simultaneous measurements of several trace gas species from one analyzer.

(SC-3) P7071/L11 etc: Please give the full name and location of company names throughout the manuscript: KNF, Swagelok, Picarro, Los Gatos, Licor, Aerolaser, Twinboard, JSP, … And please add the full model number (e.g. for Stiring collers and Valco valve (which rotator material is used?)).

Reply: Following the reviewer's suggestion, we have provided information regarding the full name and location of the companies, as well as the full model numbers for the instruments used in this study.

(SC-4) P7071/L12: The air is sampled through the air conditioning system. In the current news I have heard about problems with contaminated air in commercial passenger aircraft. Do you expect a similar effect in the cargo aircraft? Have you checked for potential pollution?

Reply: We have not heard about any problems with contamination of air through the air condition system on commercial passenger aircraft. Bases on our own experience [Matsueda and Inoue, Atmos. Environ. (1996) and Machida et al., JTEC (2008)] there has never been an issue with the air quality of the air-conditioning system on the JAL Boeing 747 aircraft. However, for the project described in this paper, we did examine the air-conditioning system used on the C-130H aircraft during the quality-assurance test under actual flight condition, as described in section 3.1. Our results indicate no significant contamination of the sampled air, although several experimental uncertainties remained in our results, as discussed in section 3.1.

(SC-5) P7071/L23: Do you see pressure effects e.g. on CO_2 by the magnesium perchlorate? How long is the inlet flushed before sampling?

Reply: We did not detect any change in CO_2 mole fraction due to pressure effects. We avoided such effects by having the dryer tube packed with the magnesium perchlorate flushed with dry air for more than 5 min before the sampling flight. We flushed the sample flask with a flow rate of about 6-8 L/min for more than 5 min, long enough to exchange the air in the flask before sampling.

(SC-6) P7071/L27: Why do you use Titanium flasks, no glass or other material?
Reply: Titanium flasks are solid and light for handling them safely in a tight space of the C-130H aircraft, whereas glass flasks are not.

(SC-7) P7072/L2: Please give a reference that proves the smaller drift. Silicon may have surface/storage effects on CO_2?
Reply: The silicon coating can prevent the corrosion of a metal surface, as reported by Smith et al. (2006), a study that is referred in the paper. We expect that the silicon coating is effective in reducing possible drift caused by corrosion over a relatively long time period, although Smith et al. (2006) did not conduct a storage test for trace gases. Thus, the storage tests described in section 3.2 for the flasks were performed in this study to examine its surface/storage effects. We found a small increase in CO_2, but we could not relate this drift to silicon coating of the inner wall of each flask. Our result indicates that an overall CO_2 drift in the flask is estimated to be less than +0.04±0.5 ppm over a 3-day period before analysis, as discussed in section 3.2 of the revised paper.

(SC-8) P7073/L11: “auto pressure controller”: How do they work? What was developed that has not been there before?
Reply: It refers to a pressure controller using a piezo valve system to maintain a constant pressure of the analyzer cell. We developed a custom-made pressure controller using a piezo valve (HORIBA STEC, Co., Ltd., Kyoto, Japan, model PV-1000/2000 series) to reduce the instability of the pressure controls caused by the relatively low flow rates. A description about the pressure controller has been added in the revised paper for clarification.

(SC-9) P7073/L15f: What happens to the ice in the cooler under vacuum conditions?
Reply: It is possible that the ice gets trapped in the cooler under vacuum conditions. We do not have any experimental evidence to indicate that the ice gets eliminated; however, it does not affect the analysis.

(SC-10) P7075/L8: The sentence implies that the ground station also has flask samples, although it is continuous [Wada, 2007].
Reply: It is true that the monitoring at the ground station at MNM is continuous. Since this sentence is not necessary, we deleted it in our revised manuscript.

(SC-11) P7075/15ff: Please avoid formulations like “nearly agree”, “correlated reasonably well”, “agree well”, “relatively constant”. Use quantitative measures instead and give numbers in comparison to previous published results.
Reply: We have revised the paper accordingly and used quantitative measures wherever possible. In particular, we state and discuss our results with respect to the numbers of the WMO-recommended inter-laboratory compatibility goal (e.g., 0.1 ppm for CO_2, 2 ppb for CH_4, 2 ppb for CO and 0.1 ppb for N_2O) (WMO, 2011).

(SC-12) P7075/18ff: Why are the differences and uncertainties so large? Please compare to the analytical precision of your and the ground based measurement. Why do you use hourly ground-based data only? What is the expected natural variability within >1000m distance between aircraft and station (you write: "slight difference in sampling height")? Do you try to consider the wind direction to decide at which time you compare with the ground based data? How does the comparison look like during the operational flight since February 2011?

Reply: In general, uncertainties associated with comparison of in-situ measurements with flask measurements in a field inter-comparison experiment are not completely eliminated. One still has many experimental biases resulting from atmospheric variability, flask-to-flask variability, and possible flask sampling biases (Karion et al., AMTD 2012). For example, some previously reported uncertainty values are 0.14 ±0.11ppm for CO_2 and -1.10±2.12 ppb for CH_4 (Rella et al., AMTD, 2012), and -0.16±0.42 ppm for CO_2, 0.4±1.8 ppb for CH_4 and -0.6±2.8 ppb for CO (Karion et al., AMTD, 2012). Compared with these results, our uncertainties are still larger.

In response to the reviewer's comment, we re-evaluated the comparison results using both measurements at the nearest sampling time, instead of the hourly mean data as was done before. As a result of this re-evaluation, the differences and uncertainties have been largely reduced. The mean differences of 22 individual measurements (aircraft data - surface data) are now estimated to be +0.08 ±0.23 ppm for CO_2, +0.24 ±2.6 ppb for CH_4, and -0.17 ± 2.4 ppb for CO. These re-evaluated results are described in the revised manuscript and the remaining uncertainties are discussed. Unfortunately, we did not have for analysis air samples from the operational low-level flight at about 1000 ft since February 2011 in time for this paper.

(SC-13) P7076/L1ff: Repetition. Shift to section 2.2.

Reply: Accordingly, we moved the repeated sentences to section 2.2.

(SC-14) P7076/L7: Did the pressure inside the flasks change during the storage?

Reply: We did not find any change in the pressure inside the flasks during the storage. However, we should note that the sample pressure used during the storage experiments was about half of the actual pressure inside the flask samples obtained from the C-130H aircraft; this is because of the consumption by the initial analysis before the storage of the flasks.

(SC-15) P7076/L9ff: The numbers are quite large and extremely noisy, and therefore insufficient to judge on storage effects. With the given numbers you cannot exclude a maximum drift of 0.012+0.017=0.029ppm CO2/day; then the CO2 drift exceeds 0.1 ppm change already after 3 days! It is similar for the other species. Given the targeted accuracy of the measurement the storage test does not give valuable information. Moreover, it would be very interesting to know the storage effect on longer time scales of weeks to months. That would allow the usage of the data also after a longer interruption due to transportation delays or instrumental problems.

Reply: Flask-to-flask variability is large in our experiments, but we estimate the total drift after 3 days for each gas species to be about +0.04±0.5 ppm for CO_2, +0.2±0.8 ppb for CH_4, and +0.6±1.3 ppb for CO. These estimated drifts are likely overestimated by a factor of two, since the pressure in the flask samples used for the storage experiments was about half of the actual pressure of the flask sampling used on the C-130H aircraft. The mean drift of each gas after 3 days is smaller than the WMO-recommended limits (e.g., 0.1 ppm for CO_2, 2 ppb for CH_4, and 2 ppb for CO) (WMO, 2011), although a significant scatter around each mean drift remains. We agree with the reviewer that more experiments by longer time storage tests will be necessary for precisely confirming the stability of the sample flasks. These discussions are now included in the revised text.

(SC-16) P7076/L25: Is there any reason why you use so many cylinders? You stated before the excellent linearity of the systems; thus, 1 or 2 cylinders might be sufficient. Are there Allan Variance tests that prove how often the system has to be calibrated? Please scientifically motivate and explain the sentence page 7076/26f.

Reply: The initial purpose of using 5 cylinders was to evaluate the linearity of the analyzers and the consistency of the WMO mole fraction scales; however, we have continued to use all these 5 cylinders because we felt that it would be useful to keep on evaluating the stability of the analyzers at this level for the sample measurements. Also it would reduce the risk of calibration errors when one of the standard gas measurements fails.

We analyzed the measured data by the successive analysis over a 10-hour period, to examine how often the system would require calibration. It was found that the instrumental drifts of all the analyzers were relatively small over a 30-min period, and the analytical precisions were not influenced. These discussions have been added in the revision.

(SC-17) P7077/L12: How did you measure the H_2O? How has the instrument been calibrated for the low water vapour amount? 0.01% H_2O is still 100 ppm H_2O, it does not fully cancel the dilution + broadening effect on CO_2 of about 0.05 ppm.

Reply: We measured the H_2O amount by using the WS-CRDS instrument. Although an accurate instrumental calibration for a low vapor amount were not carried out in this study, any differences in the H_2O measurements between dried air samples and standard gases with a dew point of less than -70°C were not seen, as shown in Figure 3f.

A level of 0.01% H_2O was found only in very wet samples in the preliminary experiments. But since the H_2O content in the aircraft samples was always reduced to less than 0.005% through drying (as mentioned in the text), the estimated effect was shown to be less than 0.025ppm for CO_2, which is smaller than our analytical precision. Since the discussion related to wet samples is not relevant to the main presentation, it was deleted in the revision.

(SC-18) P7077/L20: The values do not really give the “reproducibility”. They are an estimate of the analytical precision, not even the accuracy of the in the instrument. The requirements of the WMO include the uncertainties to the calibration scale + cylinders, the errors of the instrumentation, etc.

Reply: Following the suggestion, we changed “reproducibility” to “analytical precision”.

(SC-19) P7079/L1ff: Would it be possible to use calibration gases with natural isotopic composition?

Reply: It is possible to make synthetic calibration gases with natural isotopic composition, but they are expensive in Japan.

(SC-20) P7079/L12f: Do you correct for this isotopic error? I suggest an isotopic measurement of the used calibration gases. Given the high measurement quality standards of WMO, it is also difficult to follow your discussion at page 7081/6f, where biases of up to 0.16 ppm in CO_2 are not further commented.
Reply: We agree with the reviewer that we should correct for the isotopic error when the data from on analyzer are compared with the data obtained by another analyzer with different isotopic effects. As suggested by the reviewer, the isotopic compositions of CO in all the calibration gases were recently measured; the isotopic compositions of other trace gases will be measured in the near future. In further response to the reviewer, we have modified the discussions at page 7081/6f to make the explanation more understandable.

(SC-21) P7083/L15ff: Before widely using the instrumentation, some of the observed biases might need further discussion. A bias of 0.1 ppm CO_2 in a global network changes the net carbon amount by 0.2 PgC in the atmosphere.
Reply: We agree that for a baseline monitoring, a bias of less than 0.1 ppm CO_2 is needed, as recommended by the WMO (2011). As shown in Figure 5a, the bias of the WS-CRDS instrument relative to the NDIR is 0.03±0.07 ppm for CO_2, although large uncertainties still remain. Thus, in the revised paper we have added the comment to the effect that more precise experiments will be needed.

(SC-22) It is not clear to me, what A and B stand for. From my understanding, the conventional method uses NDIR for CO_2; thus, the values in the first columns should be the same, shouldn't they? The difference between A and B in the CO_2-WS-CRDS column are quite high.
Reply: Indeed, the values in the first columns for CO_2 and CO are mistyped. After the correction, the difference between A and B in the CO_2-WS-CRDS column is reduced from 0.09 ppm to 0.06 ppm. Furthermore, we carefully checked all the 6 flask measurements, and found that one measurement out of 6 showed significant deviation due likely to insufficient filling of one of the flasks. When we excluded this data, the CO_2 difference was reduced to be 0.04 ppm. Thus, we revised the CO_2 data in Table 1. In addition, the mean data of the 5 flask measurements were also added to Table 1 and its explanation has now been included in the revised paper.

(SC-23) Figure 1: I am puzzled by the cooler/heater setup: How is it possible that the sample air can flow through the cooler first and then be heated up again in the shown circle? How do you remove the frozen ice? If you use setup alternately, do you see differences between the setup?
Reply: We used the heater to warm the air samples, in order to bring the sample air temperature close to the analyzer cell temperature. After the analysis, the frozen ice was removed by a vacuum pump as the temperature increased gradually after the cooler was switched off. We did not see the difference between the setup.

(SC-24) Figure 3: Units for the gas components are missing. Maybe it would be more informative about the instruments responses to show calibrated data rather than raw data (that should be also clarified in the Figure description).
Reply: We thank the reviewer for pointing out this omission. Units for the y-axis were added and the instruments responses were shown by the calibrated data. In addition, the figure descriptions were also revised.

(SC-25) Figure 5: Especially CO and CH4 show obviously skewed distributions. What are the reasons for these; isotopic composition?
Reply: It is not from the isotopic composition. We believe that the skewed distributions are influenced partly by several outlier data scattered around the average.

(SC-26) P7071/L19: You mention the diameter ¼ inch 2 times. Skip one.
Reply: We deleted one. Thanks.

(SC-27) P7072/L8: 24 flask "are divided" into 4 packages sounds strange. "are distributed" might be better.
Reply: We revised it accordingly.

(SC-28) P7072/L24: "by a special operational program in PC" – What does it mean?
Reply: This is reference to a custom-made operation software. We deleted this phrase and added a brief explanation instead.

(SC-29) P7073/L14: "." Missing
Reply: Inserted.

(SC-30) P7074/L2: "maintained in a JMA calibration system" -?- Please rewrite.
Reply: This sentence was rewritten to make it clearer.

(SC-31) P7077/L26ff: references missing
Reply: The references have been added.

(SC-32) P7077/L28: output signals > raw signals
Reply: We have revised it accordingly.

(SC-33) P7078/L4: What does "re-calibration" mean?
Reply: "Re-calibration" means "calibration by the VURF" after the first calibration by the GC/HgO. Additional information has been added in the revised paper to make it clearer.

Anonymous Referee #3

General Remarks (GR):

(GR-1) The presentation quality of the paper (especially English language, redundancies) needs to be significantly improved before publication in AMT can be considered. Over large parts, it is very descriptive, and sentences are sometimes not clear.

Reply: The same comments by Referee #3 are also pointed out by Referee #1. The English language of our revised manuscript was improved by a native speaker.

(GR-2) As referee #1 already mentioned, the introduction is focused on Asia and does not properly account for other aircraft measurements that have been made. The introduction needs to be revised.

Reply: Taking the comments of Referees #3 and #1 into account, Introduction has been re-written to include additional information, including brief overviews of N_2O and other trace gases, as well as aircraft observations of CARIBIC, MOZAIC/IAGOS and CONTRAIL. Relevant papers have been referenced.

(GR-3) Concentration should not be used; use mole fraction instead.

Reply: We changed the word "concentration" throughout the text to "mole fraction".

(GR-4) A general concern is that the presented work used flasks that were only stored up to a few days before they were analyzed. Often the time between sampling and analysis will be longer, and more information about long term stability is needed.

Reply: We agree with the comment that more experiments by longer time storage tests will be necessary for a more precise determination of the stability of the sample flasks. From our shorter storage experiments, we estimate the total drift after 3 days for each gas species to be about +0.04±0.5 ppm for CO_2, +0.2±0.8 ppb for CH_4, and +0.6±1.3 ppb for CO. The mean drift of each gas after 3 days is smaller than the WMO-recommended limits (e.g., 0.1 ppm for CO_2, 2 ppb for CH_4, and 2 ppb for CO) (WMO, 2011), although a significant scatter around each mean drift remains. We agree with the reviewer that more experiments by longer time storage tests will be necessary for precisely confirming the stability of the sample flasks. These discussions are now included in the revised text.

(GR-5) It is also not very clear what the additional value of airborne flask samples is. These new analytical techniques would allow highly time resolved measurements during flights if the instruments are directly deployed in the aircraft (as it is done in other programs). Such measurements would provide altitude dependent mole fraction profiles, which could be especially valuable in remote sites. Are airborne flask samples really a contribution that is needed?

Reply: We agree with the reviewer's comments that it is better to install the instrumentation directly onboard the aircraft. However, there are several logistical restrictions associated with conducting atmospheric measurements onboard a military aircraft, such as having no access to electric power and limited space. In particular, any emissions of electro-magnetic interference (EMI) from observational instruments are prohibited on any off the Japanese army aircraft. For these reasons, we were not able to install and operate a continuous measuring system on board the aircraft. Instead, we had to settle for the manual flask air sampling procedure without electric power supply, as described in the paper. We have added this information in the revised manuscript.

One of the advantages of a flask sampling program is that we can make simultaneous measurements of various trace gas species although the profile resolution is lower, as pointed out by the reviewer. As stated in the revised Introduction, regular and long-term measurements of various trace gases by flask sampling are still needed to better understand the widespread dispersion of trace gases from various Asian emissions, because there are still limited observations in the middle troposphere over the western North Pacific.

Specific comments (SC):

(SC-1) Figure 2: It should be considered to use a different color for ambient data and flask results.

Reply: Accordingly, we revised Figure 2 using a different color.

(SC-2) Section 2.2.: The use of the manual pump (modification? What do you mean with manual?) is not well enough explained.

Reply: We have revised the description to clearly denote the operation of the pump by human power instead of electrical power.

(SC-3) It is mentioned that the new flasks are made from titanium, but low level flight flasks over MNM (p 7075, line 3) were collected in SS flasks. Is this a mistake, or were different types of flasks collected?

Reply: We used stainless steel (SS) flasks for the low-level flight sampling in 2010, because the titanium flasks used for the operational flights after 2011 were not ready for deployment at that time. We have revised the paper to make this clear.

(SC-4) Section 3.2.: Why were the flasks only stored for 2-10 days between measurements? To properly evaluate the drift, longer storage times would allow better quantification of the drift and reduce the corresponding uncertainties. If such data are available, they should be included in a revised version.

Reply: We agree with the comment that longer storage times would allow better quantification of the drift and reduce the corresponding uncertainties. We only conducted a 2-3 day storage drift test based on the assumption that it would be a typical time lapse between the flask sampling and the arrival of the flasks at JMA for analysis. We did not refer in this paper to the results we obtained from a few longer storage tests since these results are very preliminary and have not yet gone through rigorous analysis. We do however mention in the revised manuscript the value of conducting longer storage tests.

(SC-5) Table 1: It is unclear what you mean with ‘conventional method’.

Reply: The ‘conventional method’ refers to the analytical methods of the JMA standard gas calibration system using a NDIR for CO_2 and gas chromatographs for CH_4, CO, and N_2O. The footnote in Table 1 has been rewritten to make it clearer.

(SC-6) Figure 4: The uncertainties are mentioned in the text, but never quantitatively. It would be good to show uncertainties in the figure (error bars) to prove that the instruments are linear within the uncertainties of the analytical methods.
Reply: Error bars have been included in the revised Figure 4.

(SC-7) Section 3.6.: Discussion of CH4 and N2O isotope effects is somewhat useless considering the much larger uncertainties of the GC methods. This should clearly be pointed out in a revised manuscript, in which a more detailed discussion of the different contributions to the overall uncertainties should be given.
Reply: In accordance with this helpful comment, we revised the paper and added more detailed discussions, including a comparison of our results with the results of other previous studies.

(SC-8) The consequences of the isotopic effect of the CO2 analysis should be more clearly described. There are several possible approaches to solve this issue. E.g. a correction could be applied if the isotopic composition of the standards is known.
Reply: We have added a statement in the revised text to this effect.

(SC-9) The improvements of the system (lower flow rates but still sufficiently fast response times?) are several times mentioned in the manuscript, but never described in detail. More information is need before publication can be considered.
Reply: In agreement with the reviewer, we recognize the relationship between the flow rate and response time to be one of the key issues in improving a measurement system. However, such improvement was not the main focus of our presentation (relative to other improvements we made in the system). We mentioned it to highlight the fact that it is an important improvement we need to make in the future (see Conclusion in the revised paper).

Anonymous Referee #4

Specific Comments (SC):

(SC-1) P7071 Line 5: How long does it take fill to a sample and what distance does the aircraft cover during filling?

Reply: It takes about 1 min to pressurize a flask with a sample, so that the corresponding distance traveled by the aircraft during filling is estimated to be about 10 km. This information has been added in the revised paper.

(SC-2) Page 7071 Line 10: The sample collection system is run manually, requiring two attendants. Was an automated collection system considered?

Reply: We agree with the reviewer's comments that it is better to install an automated collection system. However, there are several logistical restrictions associated with conducting atmospheric measurements onboard a military aircraft, such as having no access to electric power and limited space. In particular, any emissions of electro-magnetic interference from observational instruments are prohibited on any of the Japanese army aircraft. For these reasons, we were not able to install and operate an automated collection system on board the aircraft. Instead, we had to settle for the manual flask air sampling procedure without electric power supply, as described in the paper. We have added this information in the revised manuscript.

(SC-3) Page 7071 Line 20: What kind of pump is used?

Reply: We use a diaphragm pump (Model N022, KNF) modified to be operate by hand without the use of electric power supply. This is described in the previous sentences in this section.

(SC-4) Page 7071 Line 25: Describe how the sample is collected using a manual pump.

Reply: The same diaphragm pump is used, as described in this paragraph. Thus, we revised it to make it clearer.

(SC-5) Page 7072 Line 12: Where is the analysis system located?

Reply: The system is located at JMA. Thus, it is described in the revised text to make it clearer.

(SC-6) Page 7072 Line 23, Figure 1: The figure suggests two separate cooling lines, Why is this? Where are heaters located before cold traps?

Reply: One cooling line is a backup system. When the flow is stopped by trapped ice in the cooler, the other cooling line can be used. The heater located before the cold traps is used to keep the temperature of the sampled air relatively constant so that the variability of cold-trapping efficiency may be reduced. We added a brief explanation for the use of the two separate cooling lines in the revised text.

(SC-7) Page 7073 Line 6: What is meant by `subsample flow`? Is this the purge gas?

Reply: The phrase ‘subsample flow” is used to indicate ‘branched flow” from the airflow line to each analyzer. Since the phrase does not appear to be clear, we modified it to make it clearer.

(SC-8) Page 7073 Lines 13-15: I do not understand what is meant by ‘….. subsample flow maintained for 10 minutes … last 3 are used to calculated results’. Please clarify
Reply: This has been modified to make the sentence clearer in the revised paper.

(SC-9) Page 7074 Line 2: How are stabilities of primary reference gases determined?
Reply: According to the WMO recommendation, the JMA primary reference gases are calibrated at NOAA every 2 years to evaluate their stabilities. In addition, they are compared with the reference gases at MRI twice a year to check for any possible drifts.

(SC-10) Page 7075 Line 4-5: Why were the sample measured at MRI? Is this where the JMA analysis system is operated?
Reply: Since the JMA trace gas measurement system was not ready at the time of the test flights in 2010, we analyzed the flask samples at MRI. This explanation is added in the revised text to make it clearer.

(SC-11) Page 7075 Lines 11-25: The differences between aircraft and surface samples are quite large. However this is always the possibility that on any given day the two altitudes sample different air parcels. Do air samples collected during descent to Minamitorishima show a vertical gradient?
Reply: As pointed by the comment, the vertical gradients during the descent flight to Minamitorishima were observed to be variable.

(SC-12) Page 7075 Lines 18-20: The differences between in situ surface measurements and the aircraft flask results are large enough for concern. But I also agree when the authors indicate the differences hard difficult to interpret. The correct experiment is to equip the aircraft with the lab instruments and compare the in situ versus flask results.
Reply: We agree with the comment that the correct experiment is to equip the aircraft with the lab instruments, but this was difficult to do in this study because of various restrictions related to the use of the electric power supply and available space on the plane.

As pointed out by the comment, it is difficult to properly interpret the differences between in situ surface measurements and the aircraft flask results for the field inter-comparison experiments. In fact, the uncertainties for comparison of in situ with flask measurements in the aircraft campaigns are reported to be not completely eliminated, because Karion et al., AMTD (2012) pointed out the many experimental biases such as atmospheric variability, flask-to-flask variability, and possible flask sampling biases.

Following the suggestion made by another reviewer, we re-evaluated the comparison results using both measurements at the nearest sampling time, instead of the hourly mean data as was done before. As a result of this re-evaluation, the differences and uncertainties have been largely reduced. The mean differences of 22 individual measurements (aircraft data - surface data) are now estimated to be +0.08 ±0.23 ppm for CO_2, +0.24 ±2.6 ppb for

CH_4, and -0.17 ± 2.4 ppb for CO. These re-evaluated results are described in the revised manuscript and the remaining uncertainties are discussed.

(SC-13) Page 7076 Lines 1-20: The mean growth rates and large standard deviations of the means suggests the may have been significant change in the samples. Different flasks in a batch may behave differently. The statement that there was no significant drift in the mean of 28 samples is deceptive: it does not mean there was no drift in the flasks.
Reply: We agree with the reviewer and deleted the relevant statement and added a more detailed discussion, properly evaluating the results considering the large standard deviations due to flask-to-flask variability.

(SC-14) Page 7076 Line 23: Figure 3: I do not see the values in this figure, it should be removed. The key results are in Table 1.
Reply: While we agree with the comment that the key results are in Table 1, other necessary information such as analytical sequence, sample pressure, and H_2O content are shown in Figure 3, as described in the first paragraph of section 3.3. Since we believe that this figure is helpful to the reader in understanding the outline of the actual analysis procedure, we have decided to retain this figure.

(SC-15) Page 7077 Lines 20-25: What is meant by 'comparable'? Please be more quantitative.
Reply: The analytical precisions for the conventional methods were added for quantitative comparison.

(SC-16) Page 7078-7080: The isotopic signature of the working standards should be measured not estimated.
Reply: According to the comment, we directly measured the isotopic compositions of CO in our working standard gases used in this study by using a continuous-flow isotope ratio mass spectrometry (Tsunogai et al., 2002). Although the direct measurements of isotopic ratios for CH_4 and N_2O are not available, the sensitivity tests strongly suggest that the isotope effects for CH_4 and N_2O are not significantly changed when we use a wider range of isotopic ratios in standard gases. On the other hand, the isotopic effects for CO_2 (~0.05 ppm) depend slightly on the estimated isotopic signature, as discussed in pages 7078-7080. Direct measurements for CO_2 isotopic signature is the subject of a future study.

(SC-17) Page 7081 Lines 8-10: A brief description of this experiment is needed.
Reply: We added more description of this experiment in the revised text.

(SC-18) Page 7081 Lines19-21: This statement is not sound and should be removed.
Reply: We removed it accordingly.

(SC-19) Page 7081 Lines 22-25: Note that C. Zellweger has a new paper in ATM which examined Picarro and Los Gatos instruments.

Reply: We thank the reviewer for alerting us to this paper. We have now quoted and referenced the new paper of C. Zellweger et al., ATM (2012).

(SC-20) Page 7081: This paper needs to quantitatively examine the total measurement errors a combination of the error in step of the sample analysis. Simply looking at the errors discussed in this manuscript; that related to the flask sampling (the difference between surface and aircraft samples) is the largest. This could easily be studied more rigorously.
Reply: We revised the paper accordingly to discuss total measurement errors whenever possible.