Response to reviews of manuscript: "Inter-comparison of two cavity ring-down spectroscopy analyzers for atmospheric 13CO2/12CO2 measurement" (amt-2016-95-RC1)

Dear Editor,

We deeply appreciate you for giving us an opportunity to improve our manuscript. would like to thank all of you and the two reviewers for the valuable suggestions. are the point-to-point responses (responses in upright Roman) to the comments *(original queries in Italic)*.

Response to Reviewer #1

Major comments:

1) *The authors present valuable data on new technology, but they stop short of making their findings transferable beyond these 2 specific analyzers tested. They should make suggestions for calibration 'best practices' and instrument performance diagnoses that other researchers can apply to their analyzers.*

Response: Thank you for your suggestions. We have made suggestions for calibration practice and instruments performance diagnoses in the discussion section as following: "In general, all of the IRIS instruments aim to maintain high enough precision and accuracy such that the data is traceable to international scales. However, sensitivity to changing environmental conditions (e.g., temperature dependence) and dependence of $\delta^{13}C$ on the CO_2 concentration affect the performance of IRIS measurements (Wada et al., 2011; Guillon et al., 2012; Wen et al., 2013). Reliable and accurate measurements similar to that of IRMS can be obtained with proper calibration (Bowling et al., 2005; Guillon et al., 2012; Hammer et al., 2013; Vogel et al., 2013; Wen et al., 2013). In theory, both issues of delta scale stretching and the concentration dependence should be corrected by generating multiple delta values over a range of mixing ratios under ambient conditions. In practice, ignoring the effect of the delta scale stretching, the two-point mixing ratio gain and offset calibration method was successfully applied to calibrate the mixing ratios of $^{12}CO_2$ and $^{13}CO_2$ separately (Bowling et al., 2003; Wen et al., 2013). For the instrument performance diagnoses, it was suggested that another reference gas be measured to monitor the long-term precision and accuracy. Three–point or higher calibration schemes with CO_2 concentration signals spanning the range of ambient concentrations were suggested to ensure the linearity of the analyzer and diagnose the instrument performance. With proper calibration frequency, the instrument drifts would be eliminated. Calibration frequency and sampling interval are instrument-specific characteristics. Note that considering the $\delta^{13}C$ dependence on H_2O, researchers should consider drying moist sample air when H_2O is above 2.4% as is factory recommended, even though the water correction works sufficiently well (Fig. 5)." (see Page 19 lines 379-396)

2) *Please provide more context for the precision requirements for this field of*

research. GWA target (WMO, 2011) is 0.01 permil, but that is hard to achieve, even with mass spectrometery. What implications would a difference in KP intercepts of 1.24 or 0.36 permil have? Maybe have a look at Pataki et al. (2003) and Zobitz et al. (2006) and see if they present some precision targets for Keeling Plot applications in ecology.

Response: Thank you for your suggestion. We have rephrased and clarified as following: "With proper calibration, high enough precision ($\pm$0.1‰) for δ^{13}C research, similar to that of IRMS, should be obtainable by all of the IRIS instruments. It is difficult, however, to achieve $\pm$0.01‰ precision, as recommended by the The Global Atmosphere Watch Programme of the World Meteorological Organization (WMO-GAW; WMO, 2011)" (see Page 22 lines 456-460)

"In this study, the Keeling plot intercepts of G1101-i and G2201-i measurements should be identical because of the common air samples. Differences in KP intercepts of 1.24 or 0.36‰ were caused by a systematic error between G1101-i (before and after upgrade) and G2201-i. Note that the uncertainty of the Keeling plot intercept was related to its underlying assumption, CO_2 range, and uncertainty in the CO_2 and isotopic measurements. Generally speaking, the standard error of the Keeling plot intercept should be less than 1‰ (Pataki et al., 2003; Zobitz et al., 2006)."(see Page 22 lines 443-449)

3) I'd like to see the authors present some recommendations for 'best practices' for proper calibration strategy from their experience with these analyzers. What is the frequency of calibration required to obtain quality data? How long of sampling intervals are needed, etc. If the d13C dependence on H2O is above what threshold, then researchers should consider drying moist sample air.

Response: Thank you for your valuable suggestion. Please also see the response of major comments #1 by Reviewer #1. (see Page 19 lines 379-396)

4) Essentially, the authors tested 3 analyzers: the G1101-i before upgrading, the G1101-I after upgrading and the G2201-i. It is sometimes hard to determine whether the G1101-i discussion is considering data collected before or after the upgrade. I recommend giving these analyzer unique identifiers (e.g. G1101-i-original and G1101-i-upgraded, or something shorter).

Response: Thank you for your suggestion. We have given the G1101-i before upgrading and after upgrading with unique identifiers with "G1101-i-original" and "G1101-i-upgraded" in our manuscript.

5) English grammar is good, but there are mistakes in singular/plural and article use that are common to non-native English speakers. This reviewer recommends the journal provide grammatical editing for the manuscript.

Response: Thank you for your suggestion. In response, we carefully revise the grammatically style and word use of this manuscript by ourselves, and also asked for an edit on our revised manuscript by the Elsevier Language Editing Services.

Specific comments:

1) ln27: Mention that the G2201 is newer model version than the G1101.
Response: Change has been made. (see Page 2 line 27)

2) ln 27-30: Should break these up into separate sentences.
Response: Thank you for this suggestion. Change has been made. (see Page 2 lines 28-31)

3) ln 59-61: Vague. What type of studies would comparability among analyzers enable?
Response: Thank you for the comment. Clarified as: "To assess the data comparability of different experiments, it is important to conduct an inter-comparison of different IRIS instruments to ensure their compatibility (Flowers et al., 2012; Griffith et al., 2012; Wen et al., 2013)." (see Page 4 lines 63-65)

4) ln 69: "It is important to get precise and accurate measurements… by constructing proper calibration strategy."
Response: Change has been made. (see Page 5 lines 72-74)

5) ln 84: consistent d13C offset
Response: Change has been made. (see Page 5 lines 86)

6) ln 88: Can you expand on what the 'non-linear absorption effect' is?
Response: Thank you for your suggestion. Clarified as "Tuzson et al. (2008) found a difference between QCLAS and flask-IRMS measurement of 0.28~2‰ that was probably caused by non-linearity of the QCL-instrument at elevated CO_2 concentrations and laser intensity variation. Note that an ideal IRIS instrument should be free of non-linear absorption or concentration dependence effects, meaning that its measurements should not change with the changing CO_2 concentrations at a constant isotopic composition." (see Page 6 lines 91-95)

7) ln 99: Was the concentration dependence not accounted for in that study?
Response: Thank you for your comment. Clarified as "which had excellent agreement over a 7-day atmospheric measurement period with a difference of only -0.02 ± 0.18‰ after proper calibration. However, there was still a slight correlation of the difference between the two analyzers with concentration. This slight concentration dependence resulted in a much larger difference (2.44‰) for the Keeling intercept by propagating through the Keeling analysis." (see Page 6 lines 103-107)

8) ln 118: What is the turnover time of sample air in the analyzer? Is it flowing fast enough for 1 Hz data to be meaningful?
Response: Thank you for your comment. Clarified as "A diaphragm pump was used to pump the sample air and calibration gas continuously to the cavity (volume of 35

mL) at a flow rate of 0.03 L min^{-1} at standard temperature and pressure (STP); measurement frequencies were approximately 0.3 and 1 Hz for G1101-i and G2201-i, respectively. Note that the turnover time of sample air in the analyzer is not fast enough for 0.3 Hz and 1 Hz data to be meaningful. In this study, the data reported were block-averaged to average time intervals after deleting the data collected during transitional periods in response to valve switching between the two sample intakes." (see Page 7 lines 128-134)

9) *ln 120: Mention which experiments were repeated after the upgrade.*
Response: Thank you for your suggestion. Clarified as "In the following laboratory and atmospheric measurements, the water vapor sensitivity test and atmospheric measurements were done before the upgrade of G1101-i (G1101-i-original) and after the upgrade of G1101-i (G1101-i-upgraded) in August 2014." see Page 8 lines 136-138)

10) *ln 127: The period of transition depends on the turnover time of sample in the analyzer. I don't see how citing other studies is relevant when each set up could have a different pumping speed. This would be a good place to provide some statistics on the 3 min transition time.*
Response: Thank you for your comment. Rephrased as "Data from the transitional periods, i.e., the first 180 s of each sample measurement cycle after valve switching, were discarded. The transitional periods in response to valve switching between two air sample intakes were about 120 s." (see Page 8 lines 143-145)

11) *ln152: Was the standard gas bubbled through the DPG, or was CO2-free wet air from the DPG mixed into the dry CO2 standard stream. Did this change the 'true' 13C signal at all because of carbonate chemistry in solution and temperature changes?*
Response: Thank you for your comment. Rephrased as "The reference gas (Ref1) bubbled through the reservoir of the dew point generator to produce different humidity levels by setting different dew point temperatures. The dissolution of CO_2 in the reservoir (25-30 mL) of the dew point generator moved quickly into a dynamic equilibrium state because of the carbonate chemistry in solution at different dew point temperatures, which did not change the true $\delta^{13}C$ signal because of lasting bubbled processes." (see Page 9 lines 171-176)

12) *ln 167: Std2?*
Response: Change has been made. (see Page 10 line 192)

13) *ln 183: Be consistent with 13CO2/12CO2 or 13C/12C.*
Response: Change has been made. (see Page 11 line 208)

14) *ln 188: Cite the statistical method used.*
Response: Thank you for your suggestion. Cited as "Figure 2 shows the Allan

variance (Allan, 1987) as a function of the average time of the $\delta^{13}C$ measurements for Ref1, Ref2, and Ref3 measured by G1101-i-original and G2201-i." (see Page 11 line 213-214)

Reference:
Allan D W.: Should the classical variance be used as a basic measure in standards metrology?, IEEE T. Instrum. Meas., 1001(2), 646-654, 1987.

15) *ln 191: Why do you think you got different optimal averaging times? Was it a function of the analyzer or the gas source?*
Response: Thank you for your comments. Rephrased as "If the Allan variance is dominated by the random white (Gaussian) noise, the Allan variance should decrease linearly with average time, and the precision should increase with the average time. However, for longer average times, the precision worsens because of instrumental drift. In addition, the precision should increase with increasing CO_2 concentrations because of high signal-to-noise ratio." (see Page 11 lines 214-218)

16) *ln 195: Mention the manufacturer and spell out abbreviated instruments. TDLAS = Campbell, QCLAS = Aerodyne, etc.*
Response: Thank you for your suggestion. The manufacturers and abbreviated instruments had been showed in the introduction section as "To date, various IRIS techniques are commercially available for measuring stable carbon isotopes, including lead-salt tunable diode laser absorption spectrometry (TDLAS, Campbell Scientific Inc.), wavelength-scanned cavity ring down spectroscopy (WS-CRDS, Picarro Inc.), off-axis cavity output spectroscopy (OA-ICOS, Los Gatos Research), quantum cascade laser absorption spectrometry (QCLAS, Aerodyne research), and difference frequency generation laser spectroscopy (DFG, Thermo Scientific; Griffis, 2013; Wen et al., 2012, 2013)." (see Page 4 lines 55-60)

17) *ln 228-229: What was the temporal frequency of calibration?*
Response: Thank you for your comment. Rephrased as "The two-point mixing ratio gain and offset calibration method (Bowling et al., 2003) was used to calibrate the measured $\delta^{13}C$ value for each 120 min measurement cycle." (see Page 13 lines 259-261)

18) *ln 231: Explain how the temporal drift is calculated. Is this assumed linear, unidirectional, or low-pass variability? Any idea what causes the drift? Analyzer temperature or pressure variaitions?*
Response: Thank you for your comment. Rephrased as "The instrument temporal drift was calculated as the maximum variability during the measurement period, which mainly resulted from the sensitivity to the changing environmental conditions (e.g., temperature dependence)." (see Page 13 lines 261-263)

19) *ln 237-238: It's not clear what you mean by interpolation versus extrapolation.*

Response: Thank you for your comment. Rephrased as "Much improved accuracy was obtained when the calibration was interpolated for Ref2 with Ref1 and Ref3 rather than extrapolated for Ref1 with Ref2 and Ref3 or Ref3 with Ref1 and Ref1." (see Page 14 lines 271-273)

20) ln 257: What is CEAS?
Response: Thank you for your comment. Clarified as "optical feedback cavity enhanced absorption spectroscopy (OF-CEAS)" (see Page 15 line 293)

21) ln 263: Again, what was the calibration frequency? What is 'proper calibration' for this analyzer?
Response: Thank you for your comment. Please see the response of specific comments #17 by Reviewer #1. (see Page 15 lines 298-300)

22) ln 270: What are peak-to-peak amplitudes? Of what?
Response: Thank you for your comment. Clarified as "and the peak-to-peak amplitudes of δ13C under different water vapor mixing ratios were 1.96‰ and 0.45‰, respectively." (see Page 15 lines 308-309)

23) ln 298: Specify what range of H2O this variability was calculated for.
Response: Thank you for your comments. Clarified as "In this study, the standard deviations of $\delta^{13}C$ measured by G2201-i (0.07‰ and 0.08‰) under a dew point in the range 5~20 ℃ are better than the precision given by manufacturer (0.15‰)," (see Page 17 lines 339-340)

24) ln 335: I assume total CO2 was calculated from the Picarro data? What was the range of CO2 concentrations observed. What was the uncertainty of CO2 and d13C used in the error propagation?
Response: Thank you for your comment. Clarified as "The total CO2 was calculated using the 12CO2 and 13CO2 from the Picarro data. During the first atmospheric measurement period, the δ13CS values were -24.80 ± 0.39‰ and -23.98 ± 0.30‰, with a mean difference of 0.82‰, respectively, for G1101-i-orginal with a range of CO2 concentrations from 390.92 to 630.92 ppm and G2201-i with a range of CO2 concentrations from 391.76 to 631.29 ppm. Note that the uncertainties are the standard error of the intercept from the fitting algorithm. If we assumed that the atmospheric δ13C is a linear function of 1/CO2 with a small concentration dependent error d (Eq. 18; Wen et al., 2013), then error propagation through the concentration dependence would be a function of ε with respect to the intercept of the Keeling plot. When ε=0.05‰，this error would propagate through the Keeling Plot and cause a difference of 0.99‰. This result is close to the actual difference of 0.82‰ between G1101-i-orignal and G2201-i." (see Page 20 lines 415-424)
"During the second atmospheric measurement period, the $\delta^{13}C_S$ values were -25.90 ± 0.28‰ and -25.97 ± 0.12‰, with a mean difference of 0.07‰, for G1101-i-upgraded with a range of CO_2 concentrations from 398.51 to 552.66 ppm and G2201-i with a

range of CO_2 concentration from 399.92 to 555.90 ppm, respectively." (see Page 21 lines 431-434)

25) *ln 351: Can you relate these differences to something in the calibration procedure or precision differences among the analyzers? It seems that if both analyzers are calibrated using common standards, that should eliminate bias between the measurements and force the KP intercepts to converge. The only differences left I can think of are un-corrected dependencies on CO2 or H2O concentrations, or larger standard deviation propagating into intercept uncertainty. Or there were different ambient ranges of CO2 and H2O affecting the analytical and KP precision.*

Response: Thank you for your comments. Rephrased as "The results confirm that we should pay attention to the measurement difference resulting from the uncorrected dependencies on CO_2 or H_2O concentrations among different IRIS instruments and that this difference will result in error propagation through Keeling plot analysis (Wen et al., 2013)." (see Page 21 lines 437-440)

26) *ln 354-364: I feel like this information should come before the KP discussion. It's relevant to the intercept differences found.*

Response: Thank you for your suggestion. Change has been made. (see Page 20 lines 399-410)

27) *ln 368: What is the precision for a practical measurement duration. I doubt anyone is going to do KP calculations with 2 hour averaging intervals. What would 5 min averaging get you?*

Response: Thank you for your comments. Rephrased as "The Allan variation test indicates that the best precision was 0.08~0.15‰ and 0.01~0.04‰, measured respectively by G1101-i-orginal and G2201-i with a CO_2 range from 368.1 to 550.1 ppm; the five-minute precision was 0.24-0.34‰ and 0.08-0.12‰, respectively." (see Page 22 lines 454-456 and Page 29 lines 609-611)

Table 1. The $\delta^{13}C$ precision with averaging time derived from the Allan variance analysis.

species		G2201-i			G1101-i-original		
		1min	5min	optimum	1min	5min	optimum
$^{12}CO_2$	Ref1	0.03	0.02	0.01	0.07	0.03	0.01
	Ref2	0.04	0.02	0.01	0.07	0.04	0.02
	Ref3	0.04	0.02	0.01	0.08	0.04	0.01
	mean	0.04	0.02	0.01	0.07	0.04	0.02
$^{13}CO_2$	Ref1	0.0009	0.0005	0.0001	0.0025	0.0014	0.0006
	Ref2	0.0009	0.0004	0.0003	0.0025	0.0013	0.0008
	Ref3	0.0009	0.0004	0.0001	0.0032	0.0016	0.0007

	mean	0.0009	0.0004	0.0002	0.0027	0.0014	0.0007
	Ref1	0.22	0.12	0.03	0.63	0.34	0.08
$\delta^{13}C$	Ref2	0.17	0.09	0.04	0.51	0.26	0.15
	Ref3	0.16	0.08	0.01	0.44	0.24	0.10
	mean	0.18	0.09	0.03	0.53	0.28	0.13

28) ln 371: Is this gradient a vertical gradient of ambient air?

  Response: Thank you for your comment. Rephrased as "For the gradient switching test lasting 48 h among Ref1, Ref2, and Ref3, the dependence of $\delta^{13}C$ on the CO_2 concentration was 0.46‰ per 100 ppm for G1101-i-orginal and 0.09‰ per 100 ppm for G2201-i in the range of 368.1~550.1 ppm, and the drift of the instruments ranged from 0.92 to 1.09‰ and 0.19 to 0.37‰, respectively." (see Page 22 lines 461-464)

29) ln 374-375: I assume you are comparing values for Std2 here? and the 2 values given are the 2 different analyzers? Before/after upgrade?

Response: Thank you for your comment. Clarified as "the average $\delta^{13}C$ values of Ref1, Ref2, and Ref3 were -20.34 ± 0.07‰ by G1101-i-orginal and -20.45 ± 0.09‰ by G2201-i, similar to the actual values measured by IRMS (-20.38 ± 0.06‰)." (see Page 23 lines 465-467)

30) ln 377-378: For those readers who don't read the entire paper, let them know that the G1101-I analyzer was upgraded before the second test.

Response: Thank you for your suggestion. Change has been made as "With dew point temperatures in the range of 5~20 ℃, the sensitivity of $\delta^{13}C$ to the water vapor mixing ratio was 1.01‰/% H_2O and 0.09‰/% H_2O by G1101-i-original and G2201-i, respectively, during the first test (before the upgrade of G1101-i) and 0.15‰/% H_2O and 0.13‰/% H_2O by G1101-i-upgraded and G2201-i, respectively, during the second test (after the upgrade of G1101-i)." (see Page 23 lines 468-471)

31) ln 382: Define 'rapidly changing'. On what time scale?

Response: Thank you for your comments. Clarify as "Atmospheric $\delta^{13}C$ measured by G1101-i and G2201-i captured the rapid changes in atmospheric $\delta^{13}C$ on hourly to diurnal cycle scales." (see Page 23 lines 475-476)

32) ln 383-384: The range and mean bias between the two analyzers wasn't improved all that much by the upgrade. Is this the difference in hourly averages?

Response: Thank you for your comments. Clarified as "Before G1101-i was upgraded (DOY164 - 174), the difference of hourly $\delta^{13}C$ averages measured by G1101-i-original and G2202-i analyzers ranged from -0.62‰ to 0.76‰, with an average value of 0.07 ± 0.24‰. After G1101-i was upgraded (DOY348-356), the difference in hourly $\delta^{13}C$ averages measured by G1101-i-upgraded and G2201-i analyzers ranged from -0.57‰ to 0.85‰, with an average value of 0.05 ± 0.30‰. This difference exhibits a Gaussian distribution. Before the upgrade of G1101-i, a significant linear correlation was observed between the $\delta^{13}C$ difference and water

vapor concentration (P<0.01), but there is no significant correlation (P>0.05) after the upgrade of G1101-i." (see Page 23 lines 476-483)

33) ln 390: Has Picarro released information on what changes were made during the upgrade? Software, hardware or calibration?

Response: Thank you for your comment. Clarified as "This is mainly due to the improvement of the interference of water vapor to the $\delta^{13}C$ measurement by the upgraded algorithm of the G1101-i software.". (see Page 23 lines 483-484)

34) Figure 1: Are these averages plotted at mid-point or end-points? Did the Allen variation change for the G1101-I after the upgrade? Marking some kind of common threshold would aid the eye in seeing the improved performance of the G2202-i.

Response: Thank you for your comment. Change has been made as "Figure 2: Allan deviation of the $\delta^{13}C$ at end-point for the (a) G1101-i before upgradation (G1101-i-original) and (b) G2201-i analyzers with 3 different CO_2 concentration with same $\delta^{13}C$ standard gases." In this study, the Allan variance test was conducted only before upgradation of G1101-i. However, the upgradation was focus on the algorithm of cross interference of water vapor mixing ratio, we think this upgrade would not change the precision of analyzer. In addition, and the threshold of Allan deviation at 0.1‰ was plotted in Figure 2 (see Page 32 lines 619-621)

Figure 2
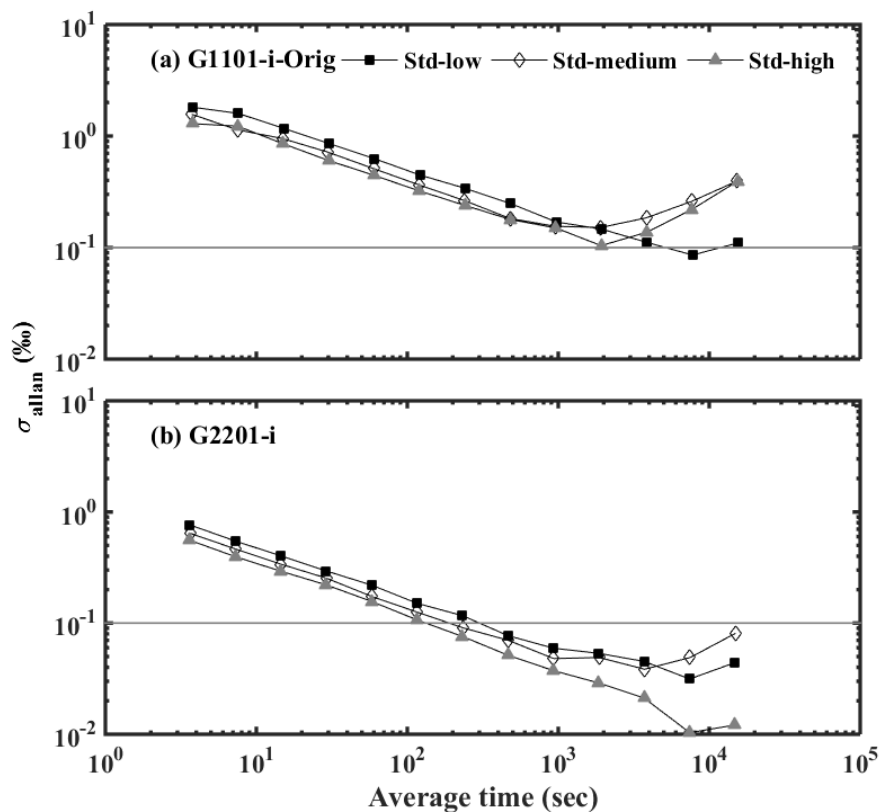
35) Figure 2: Is this the G1101-I before or after the upgrade?

Response: Thank you for your comment. Clarified as "Figure 3:   Dependency of the measured $\delta^{13}C$ of G1101-i before upgradation (G1101-i-original) and G2201-i analyzers on the measured CO_2 concentration with 3 different CO_2 concentrations with same $\delta^{13}C$ reference gases." (see Page 33 lines 627-629)

36) Figure 3: Would this be better as a data table with statistical metrics added?

Response: Thank you for your suggestion. Change has been made "Table 2. The stability and accuracy of $\delta^{13}C$ values of the 3 reference gases (Ref1, Ref2 and Ref3) with same $\delta^{13}C$ measured by G1101-i before upgradation (G1101-i-original) and G2201-i analyzers." (see Page 30 lines 613-614)

Table 2

|	G1101-i		G2201-i		IRMS
	Measured	**Corrected**	**Measured**	**Corrected**	
Ref1	-23.46±0.26	-20.29±0.34	-21.65±0.07	-20.51±0.21	-20.38 ±0.06
Ref2	-22.99±0.28	-20.42±0.20	-21.51±0.08	-20.35±0.08	-20.38 ±0.06
Ref3	-22.62±0.27	-20.32±0.21	-21.49±0.05	-20.48±0.14	-20.38 ±0.06

37) Figure 4: Why are the error bars larger at low H2O?

Response: Thank you for your comment. For this water vapor sensitivity experiment, we have test the reference gas at each humidity level was measured for 20 minutes, and a total of three times. A relatively enriched $\delta^{13}C$ value of -21.7021was reported for the humidity level of 0.65 during the first test, which caused the much larger error at this low humidity level. This is probably due to the unstable situation of the analyzer. We have deleted this data and recalculated it, $\delta^{13}C$ value for the humidity level of 0.65 was 22.11±0.11‰.

$H_2O\%$	$\delta^{13}C$ of G2201-i			average	std
0.65	-21.70	-22.10	-22.12	-21.97	0.23
0.87	-22.42	-22.41	-22.45	-22.43	0.03
1.22	-22.27	-22.38	-22.45	-22.36	0.09
1.69	-22.34	-22.31	-22.31	-22.32	0.02
2.32	-22.24	-22.32	-22.18	-22.25	0.07

Change has been made as "Figure 5: Sensitivity of the measured $\delta^{13}C$ by G1101-i and G2201-i on water vapor mixing ratio. (a) measured before G1101-i upgraded (G1101-i-original and G2201-i) and (b) measured after G1101-i upgraded (G1101-i-upgraded and G2201-i)." (see Page 35 lines 636-639)

Figure 5
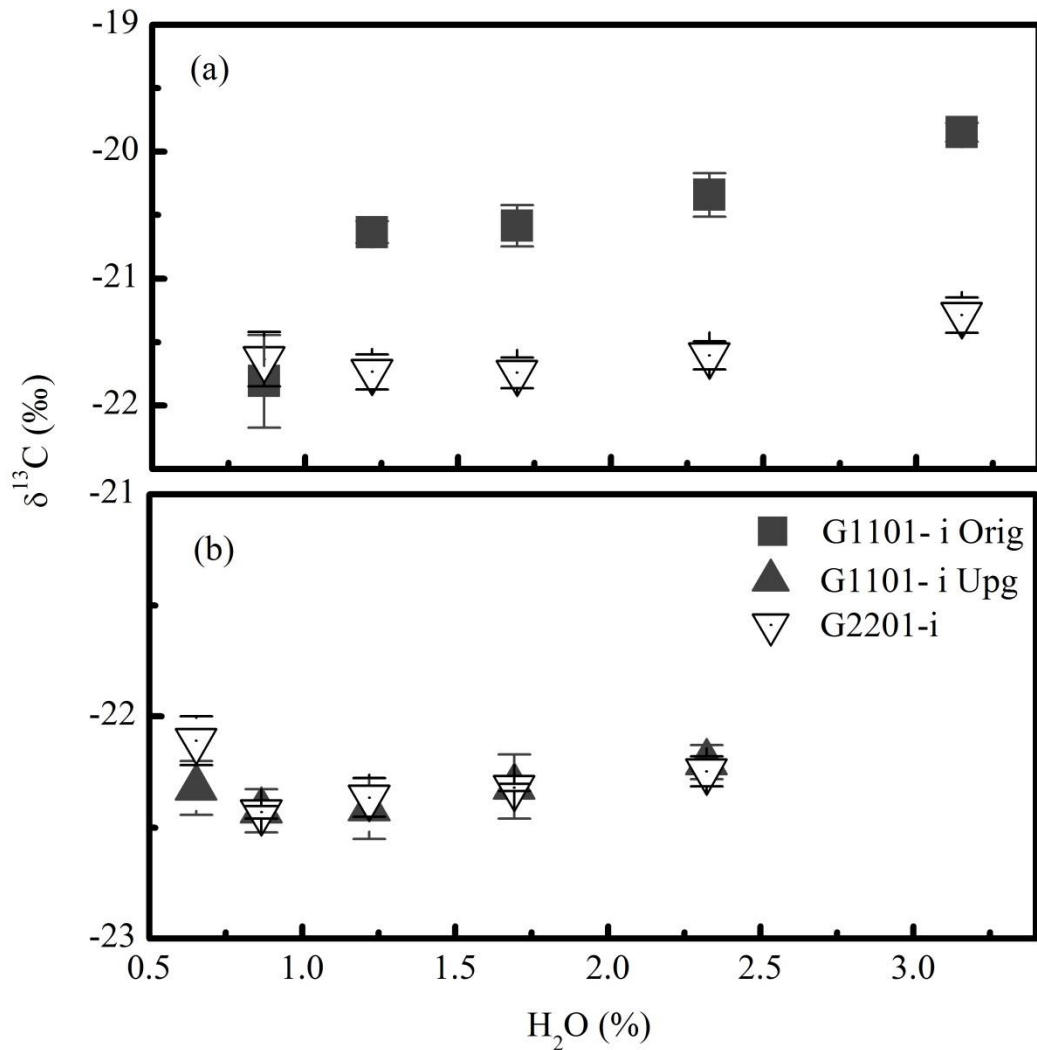
Response to Reviewer #2

General comments:

1) *The conclusion section should to be more instructive, e.g. the authors find that the water correction works sufficiently well. However, it is never established – sufficiently well to do what? Unfortunately, the authors do not make wider recommendations for the IRIS user community based on their results or the need for future IRIS comparison studies. After addressing these minor issues (and the specific comments) this study seems suitable for publication in AMT.*

Response: Thank you for your comment. We have made suggestions for calibration practice and instruments performance diagnoses in the discussion section. Please also see the response of major comments #1 by Reviewer #1. (see Page 19 lines 379-396)
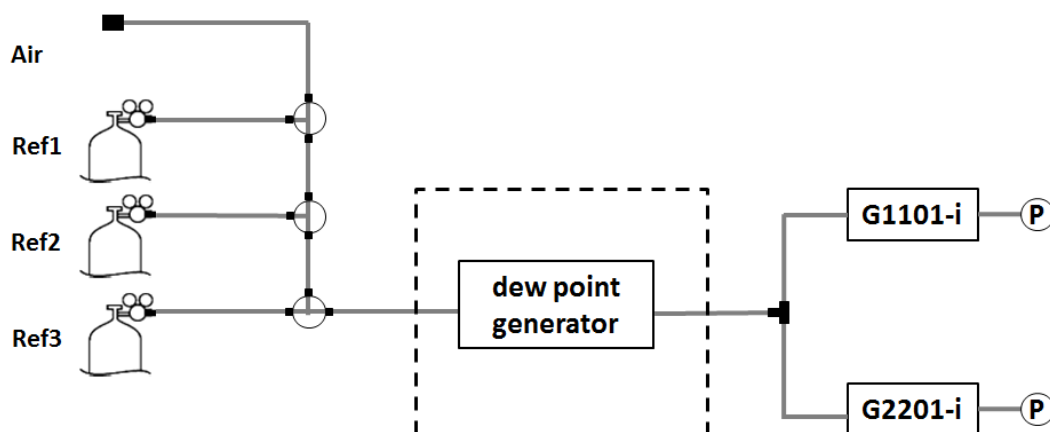
Specific comments:

1) *Line 113f – Including a graphic of the test setup would be most helpful/necessary*

here.

Response: Thank you for your suggestion. Change has been made as "Figure 1: Schematic setup of the laboratory and ambient measurements of two Picarro CO_2 $\delta^{13}C$ analyzers." (see Page 31 lines 617)

Figure 1

2) *Line 116f – Please add the information about the observed stability of the set points of 45C, 140Torr and 148Torr. Both for G1101i and G2201i instruments those parameters change within a small range and it would be beneficial for the reader to see if the variations could be a relevant sources of error.*

Response: Thank you for your suggestion. Change has been made as "The built-in pressure and temperature monitoring systems of G1101-i and G2201-i maintained the cavity temperature of both systems at 45 ℃ and the cavity pressures at 140 Torr and 148 Torr, respectively. The observed stability of temperature over 24 h was 45.0 ± 0.0024 ℃ and 45.0 ±0.0005 ℃, and the observed stability of pressure was 140.0429 ±0.0580 Torr and 147.9990 ±0.0165 Torr for G1101-i and G2201-i, respectively. No relationship between temperature and pressure variation and the $\delta^{13}C$ difference of either instrument was found." (see Page 7 lines 122-128)

3) *Line 130 – Please clarify your use of the term "standard gases" here. I suggest using "reference gas" instead as the gas was not provided by a NMI or a central calibration laboratory, but a private company. I would strongly suggest to follow the nomenclature of WMO/GAW-VIM throughout the paper to avoid confusion. (http://www.wmo.int/pages/prog/arep/gaw/documents/Final_GAW_213_web.pdf)*

Response: Thanks for your comments. Change has been made. (see Page 8 line 148)

4) *Line 138 – Please cite original papers or sources closer to the original studies see e.g. Allan 1987 (IEEE Transaction on instrumentation and measurements, IM36-2, 1987)*

Response: Thanks for your comment. Change has been made (see Page 8 line 156)

5) *Line 163ff – Please consider adding a signifier for the upgraded G1101-i e.g. G1101-i* or G1101-iup to clearly and efficiently distinguish the instruments in the plots and following text.*

Response: Thank you for your suggestion. We have given the G1101-i before upgrading and after upgrading with unique identifiers with "G1101-i-original" and "G1101-i-upgraded" in our manuscript. Please also see the response of major comments #4 by Reviewer #1. (see Page 10 lines 186)

6) *Line 187ff – I would suggest to summarize the key results presented in this section in a table, which could make the text easier to read.*

Response: Thanks for your comment. Change has been made. Please also see the response of specific comments #27 by Reviewer #1. (see Page 22 lines 454-456 and Page 29 lines 609-611)

Table 1. The $\delta^{13}C$ precision with averaging time derived from the Allan variance analysis.

species		averaging time					
		G2201-i			G1101-i		
		1min	5min	optimum	1min	5min	optimum
$^{12}CO_2$	Ref1	0.03	0.02	0.01	0.07	0.03	0.01
	Ref2	0.04	0.02	0.01	0.07	0.04	0.02
	Ref3	0.04	0.02	0.01	0.08	0.04	0.01
	mean	0.04	0.02	0.01	0.07	0.04	0.02
$^{13}CO_2$	Ref1	0.0009	0.0005	0.0001	0.0025	0.0014	0.0006
	Ref2	0.0009	0.0004	0.0003	0.0025	0.0013	0.0008
	Ref3	0.0009	0.0004	0.0001	0.0032	0.0016	0.0007
	mean	0.0009	0.0004	0.0002	0.0027	0.0014	0.0007
$\delta^{13}C$	Ref1	0.22	0.12	0.03	0.63	0.34	0.08
	Ref2	0.17	0.09	0.04	0.51	0.26	0.15
	Ref3	0.16	0.08	0.01	0.44	0.24	0.10
	mean	0.18	0.09	0.03	0.53	0.28	0.13

7) *Line 293ff – Similar work has been conducted for CRDS instruments for GHG concentrations, e.g. Yver-Kwok et al. 2015 (AMT-8-3867-2015) investigated the cross-sensitivities (e.g. for H2O) of over 60 instruments. This could be a useful reference.*

Response: Thanks for your comment. Cited as "The incompatibility of these results indicates the need for more precise experiments to evaluate the transferability of water correction functions (Kwok et al., 2015)." (see Page 17 lines 334-335)

Reference:
Kwok, C. Y., O. Laurent, A. Guemri, C. Philippon, B. Wastine, C. W. Rella, C. Vuillemin, F. Truong, M. Delmotte, V. Kazan, M. Darding, B. Lebegue, C. Kaiser, I.

Xueref-Remy and M. Ramonet: Comprehensive laboratory and field testing of cavity ring-down spectroscopy analyzers measuring H_2O, CO_2, CH_4 and CO, Atmos Meas Tech, 8, 3867-3892, doi:10.5194/amt-8-3867-2015, 2015.

8) Line 337 – Please cite the appropriate publication by C.D. Keeling here
Response: Thank you for your suggestion. Change has been made as: "In this study, $\delta^{13}C_S$ was calculated using the calibration dataset of $\delta^{13}C$ and $1/CO_2$ by the Keeling plot intercept method (Keeling, 1958; Fig. 9)." (see Page 20 lines 413-415)

Reference:
Keeling, C. D.: The concentration and isotopic abundances of atmospheric carbon dioxide in rural areas, Geochim. Cosmochim. Ac., 13(4), 322-334, 1958.

9) Line 338f – Please expand if the uncertainties given include the full error budget or are the uncertainty of the fit (fitting algorithm?).
Response: Thanks for your comment. Clarified as "Note that the uncertainties are the standard error of the intercept from the fitting algorithm." (see Page 21 lines 419-420)

10) Line 381ff – see general comments
Response: Thanks for your comments. We have made suggestions for calibration practice and instruments performance diagnoses in the discussion section. Please also see the response of major comments #1 by Reviewer #1. (see Page 19 lines 379-396)

11) Line 392ff – the importance of compatibility between difference instruments can be easily agreed upon. However, the potential problems caused by non-compatibility could be easily illustrated here (e.g. Levin et al. 2012, Nature, doi:10.1038/nature11175). The difference in the keeling intercept decreases from 1.24permil to 0.36permil, but what are the scientific implications of this? Why is 0.36permil good enough and 1.24permil not? Please given an example of an application or demonstrated how/if the higher bias of the d13Csource would lead to different scientific interpretations of the observations.
Response: Thanks for your comments. Change has been made "The potential problems caused by non-compatibility include the integrity of an internal calibration scale and modifications to analytical procedures in decade-long records (Levin et al. 2012). In this study, the Keeling plot intercepts of G1101-i and G2201-i measurements should be identical because of the common air samples. Differences in KP intercepts of 1.24 or 0.36‰ were caused by a systematic error between G1101-i (before and after upgrade) and G2201-i. Note that the uncertainty of the Keeling plot intercept was related to its underlying assumption, CO_2 range, and uncertainty in the CO_2 and isotopic measurements. Generally speaking, the standard error of the Keeling plot intercept should be less than 1‰ (Pataki et al., 2003; Zobitz et al., 2006)" Please also see the response of major comments #2 by Reviewer #1. (see Page 22 lines

442-449)

Reference:

Levin, I., C. Veidt, B. Vaughn, G. Brailsford, T. Bromley, R. Heinz, D. Lowe, J. Miller, C. Poß and J. White: No inter-hemispheric $\delta\ ^{13}CH_4$ trend observed, Nature, 486(7404), E3-E4, 2012.