Review of: “Inter-comparison of two cavity ring-down spectroscopy analyzers for atmospheric 13CO2/12CO2 measurement”
By: Pang et al.
Submitted to: Atmos. Meas. Tech. Discuss.

Summary:

The authors present an analysis of two different models for Picarro 13C-CO2 analyzers: performance characteristics and bias between the analyzers. They show 2 example applications of Keeling Plot intercept calculations and find that the analyzers get different estimates of the source d13C.

Major comments:

This paper is presented logically and is well written with minor grammatical errors. The subject matter falls within the scope of AMT. 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.

1. 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.
2. 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.
3. 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).
4. 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.

Specific comments:

ln27: Mention that the G2201 is newer model version than the G1101.
ln 27-30: Should break these up into separate sentences.
ln 59-61: Vague. What type of studies would comparability among analyzers enable?
ln 69: "It is important to get precise and accurate measurements… by constructing proper calibration strategy."
ln 84: consistent d13C offset
ln 88: Can you expand on what the 'non-linear absorption effect' is?

ln 99: Was the concentration dependence not accounted for in that study?
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?
ln 120: Mention which experiments were repeated after the upgrade.
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.
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?
ln 167: Std2?
ln 183: Be consistent with 13CO2/12CO2 or 13C/12C.
ln 188: Cite the statistical method used.
ln 191: Why do you think you got different optimal averaging times? Was it a function of the analyzer or the gas source?
ln 195: Mention the manufacturer and spell out abbreviated instruments. TDLAS = Campbell, QCLAS = Aerodyne, etc.
ln 228-229: What was the temporal frequency of calibration?
ln 231: Explain how the temporal drift is calculated. Is this assumed linear, uni-directional, or low-pass variability? Any idea what causes the drift? Analyzer temperature or pressure variaitions?
ln 237-238: It's not clear what you mean by interpolation versus extrapolation.
ln 257: What is CEAS?
ln 263: Again, what was the calibration frequency? What is 'proper calibration' for this analyzer?
ln 270: What are peak-to-peak amplitudes? Of what?
ln 298: Specify what range of H2O this variability was calculated for.
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?
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.
ln 354-364: I feel like this information should come before the KP discussion. It's relevant to the intercept differences found.
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?
ln 371: Is this gradient a vertical gradient of ambient air?
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?

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.
ln 382: Define 'rapidly changing'. On what time scale?
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?
ln 390: Has Picarro released information on what changes were made during the upgrade? Software, hardware or calibration?

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.
Figure 2: Is this the G1101-I before or after the upgrade?
Figure 3: Would this be better as a data table with statistical metrics added?
Figure 4: Why are the error bars larger at low H2O?

References:

Pataki, D. E.: The application and interpretation of Keeling plots in terrestrial carbon cycle research, Global Biogeochemical Cycles, 17(1), 1022, doi:10.1029/2001GB001850, 2003.

WMO: 15th WMO/IAEA meeting of experts on carbon dioxide, other greenhouse gases and related tracers measurement techniques, wmo.int [online] Available from: http://www.wmo.int/pages/prog/arep/gaw/documents/ (Accessed 24 May 2012), 2011.

Zobitz, J. M., Keener, J. P., Schnyder, H. and Bowling, D. R.: Sensitivity analysis and quantification of uncertainty for isotopic mixing relationships in carbon cycle research, Agricultural and Forest Meteorology, 136(1-2), 56–75, doi:10.1016/j.agrformet.2006.01.003, 2006.