Response to Anonymous Referee #1

Responses in blue

The presented article address the application of low-cost NDIR CO2 sensors for urban measurement networks aimed at assessment of CO2 fluxes over an urban areas using inverse modelling technique. While such sensors has not enough precision in case of background studies, their application in urban areas where amplitude of atmospheric CO2 mixing ratios is order of magnitude higher is possible. These circumstances make the presented study an important contribution to the construction of such measurement networks. Authors focused on evaluation of several copies of SenseAir K30 NDIR CO2 sensor. Authors demonstrated that all the sensors fulfill the technical specification of manufacturer however this specification is not enough in above application. A series of long term measurements showed that application of correction factors determined by two statistical approaches, subsequent univariate and multivariate linear regression analysis significantly improve the sensor performance.

Thank you for your time in reviewing our manuscript, we appreciate your acknowledgement that this work has the potential to help improve the constraint of CO_2 emissions from urban areas.

Detailed comments:

p.1 l.15: there is no info on RMS of research-grade analyzer used by authors which is compared to low-cost

The LGR analyzer was calibrated as described in section 3 with two NIST-traceable gas standards as well as evaluated for drift over the experiment using a tank of breathing air. The Allan variance was computed for the LGR and determined that ~100 seconds was the optimum averaging time for noise. Additionally, after correcting for the drift, the noise in the LGR for the breathing air tank was ±0.3 ppm for 2-second data, which is within the manufacturer's specifications. This 1-sigma standard deviation has been added to the manuscript.

p.4 l.3 Why authors decided to use such narrow range of CO2 mixing ratios. In real urban environment values close to 500 ppm or more are frequently observed.

The ambient experiment was conducted in College Park, Maryland, approximately 12 km from the central business district of Washington, DC, and a couple of the nights the concentration did exceed 50 0ppm, but was usually below 450 ppm. These two tanks were used for calibration because they were readily available from other experiments at the University of Maryland, where the application is to calibrate boundary layer observations on aircraft, where the concentrations are generally lower. In addition, three breath air cylinders with higher CO_2 mole fractions of

449.73, 486.53, and 516.41ppm (all NIST-traceable) were also used to calibrate the LGR.

p.6. L23 it is not clear why the calibration strategy has been changed during the experiment. Why some standards were flushed for 10 min and other for one hour?

We were initially unsure how long the LGR needed to equilibrate, and wanted an idea of the variability/stability at the fixed concentration, which is why we initially ran the calibration gas for 60 minutes. Looking at the raw data, it takes somewhere on the order of 90-120 seconds to fully equilibrate. It was switched later to conserve the breathing air tank.

p.11 l.19 Authors decided to use natural synoptic variability to perform a regression analysis aimed at determination of correction factors taking into account the influence of temperature, humidity and pressure variability of NDIR sensors on CO2 measurements. Such procedure requires long time and is depended on existing natural variability. To standardize and shorten the procedure maybe a construction of special environmentally controlled chamber should be taken into account?

In an ideal situation, this is what should be done, but is impractical for a large number of sensors. To cycle temperature, pressure, and relative humidity throughout typical ambient ranges is not difficult for one instrument, but for several requires a large enough chamber, and needs to be set up to be autonomous, otherwise if someone is manually controlling these parameters for days/weeks, the low-cost aspect of these instruments becomes much more labor intensive. An additional point that was added to the revised manuscript as suggested by another reviewer, is to see if each sensor has to be individually evaluated. Since it appears that a uniform set of regression coefficients does not work, each sensor requires a 15-30 day evaluation period and a chamber would need to be large enough to contain at least 6, but preferably 10 or more to be viable. Using the natural variability to calibrate the sensors allowed the experiment to be conducted without supervision once it was set up.

Response to Anonymous Referee #2

General comments

This study characterizes the performance (accuracy, precision, drift) of one type of low-cost CO_2 sensor for ambient air measurements. The paper describes an experiment where duplicate sensors are used to make continuous ambient measurements in an environment with conditions that change slowly over time. Ambient pressure, temperature, and humidity are monitored simultaneously and used to derive empirical corrections for individual sensors, which significantly improve the accuracy of the final datasets. The overall experiment is well-designed and the analysis of the resulting data is sound.

The paper topic is highly relevant and potentially useful to the broader atmospheric measurement community, but currently falls short of that potential. There are several additional experiments and analyses one could imagine that would fit in the same paper and would improve the scope and significance, such as: (i) test whether a unique correction is needed for each unit and what the uncertainty would be if a generalized correction were applied, (ii) demonstrate an experiment which would allow correction factors to be rapidly derived in the lab, (iii) test the K30 in a real-world environment, (iv) test the K30 for long-term drift. At a minimum, it appears that the authors can use the existing dataset to address point (i).

First off, thank you very much for your thorough review of our paper and for the helpful comments and suggestions. For your general comments:

(i)   This additional analysis has been added to Section 6 as was suggested, but to summarize, yes a unique correction is required. By taking the average coefficients and intercepts computed by the multivariate regression of the five best performing K30s, all six sensors had higher RMSEs than with independent coefficients, which is expected. But at a minimum, the RMSE doubled with in some cases actually became worse than with no correction at all. Thus, instead of 1.5-3 ppm RMSE after correction, it can range from 3 to 24 ppm for 1 minute data.

(ii)   The main reason why we did an ambient calibration, and the main concern with doing a laboratory correction is the labor and equipment required to do this. To cycle temperature, pressure, and relative humidity throughout typical ambient ranges is not difficult for one instrument, but for several requires a large enough chamber, and needs to be set up to be autonomous, otherwise if someone is manually controlling these parameters for days/weeks, the low-cost aspect of these instruments becomes

much more labor intensive. At this time, we did not have the resources or an environmental chamber available to conduct this experiment. This is something we hope to do with future work, as mentioned in the conclusions.

(iii) The experiment described in the manuscript is a quasi-real-world environment, as we used ambient air and ambient variations in the environmental variables, but the main difference was the sensors were not outside in direct sunlight or exposed to weather. This is something we hope to do in the future, but we wanted to evaluate them without these engineering concerns initially.

(iv) This is also something we hope to achieve with future work (as mentioned in the conclusion portion of the paper). While it would have been nice to include long-term drift in this manuscript, that would require 6-12 months of data with either a continuous gas analyzer as reference or calibration gas introduced periodically. We believe it is sufficient to publish these initial results characterizing month-long drift, as the results will be useful to others working with similar sensors. We will include an evaluation of long-term drift of this sensor and others in a future manuscript.

Specific comments

Los Gatos Instrument

You use the LGR dataset as a control, but I am concerned that there could be large uncertainty associated with the LGR water correction. Do you know the accuracy of the LGR water correction? I have never seen an assessment of it.

Yang et al. 2016 describes a comparison between a Picarro dried with Nafion and a LGR for flux measurements and found an R^2 of 0.99 for CO_2. While the correction is not perfect, it seems sufficient for our purposes, particularly that we are attempting to determine the accuracy and precision of the NDIR sensors relative to the LGR, not to the absolute concentration.

Additionally, previously this instrument's water vapor correction was evaluated, but was focused on CH_4 by adding different moisture into the sample line with a calibration gas flowing. The measured [CH_4_dry] is constant at different RH values. This experiment was not designed for CO_2 assessment because the calibration gas went through a bubbler at different flow rates and some CO_2 dissolved into the water.

You calibrate the instrument with two points by assuming a linear fit. How do you know the true instrument response is linear?

A previous calibration for this instrument was done with five NIST-traceable standards ranging from 369.19 to 516.41 ppm which showed its linearity. This has been clarified in the text.

How did you decide to measure the tank at 23 and 47 hour intervals? How do you know that significant drift did not occur over shorter intervals?

We wanted calibrations to occur at different times throughout the day, thus the 23 and 47 hour intervals, and was not performed more often in an attempt to have enough calibration gas for the entire period. As described below, we were unsure how long we would need to equilibrate/get an idea of the repeatability of the LGR instrument.

Why do you need to measure the tank for such long time periods (10-60 minutes)? Does it take that long for the measurement to equilibrate? If you are using the proper materials in your plumbing, the measurement should equilibrate in a matter of seconds to minutes. If it is taking a long time for the CO_2 signal to equilibrate, that suggests that CO_2 may be absorbing/desorbing onto the walls in your plumbing.

The tank is located downstairs indoors about 5 meters or so from the LGR, so there is some time required for the tubing as well as the LGR cavity to flush. Looking at the raw data, it takes somewhere on the order of 90-120 seconds to fully equilibrate. We were initially unsure how long the LGR needed to equilibrate, and wanted an idea of the variability/stability at the fixed concentration, which is why we initially ran the calibration gas for 60 minutes. It was switched later to conserve the breathing air tank.

What is the purpose of the Dasibi calibrator? Did you have to dilute the tank air to get ambient values?

The Dasibi calibrator is purely used as the scheduler for the tank used in the LGR stability evaluation. It contains a clock that turned the calibration gas on/off at the specified intervals. No dilution was performed, as the tank of breathing air provided a concentration within the normal range observed.

Figure 3: If you take the linear trend out, are the remaining variations related to a physical parameter such as temperature (ambient or cell)?

Residual Means vs Cavity T/P
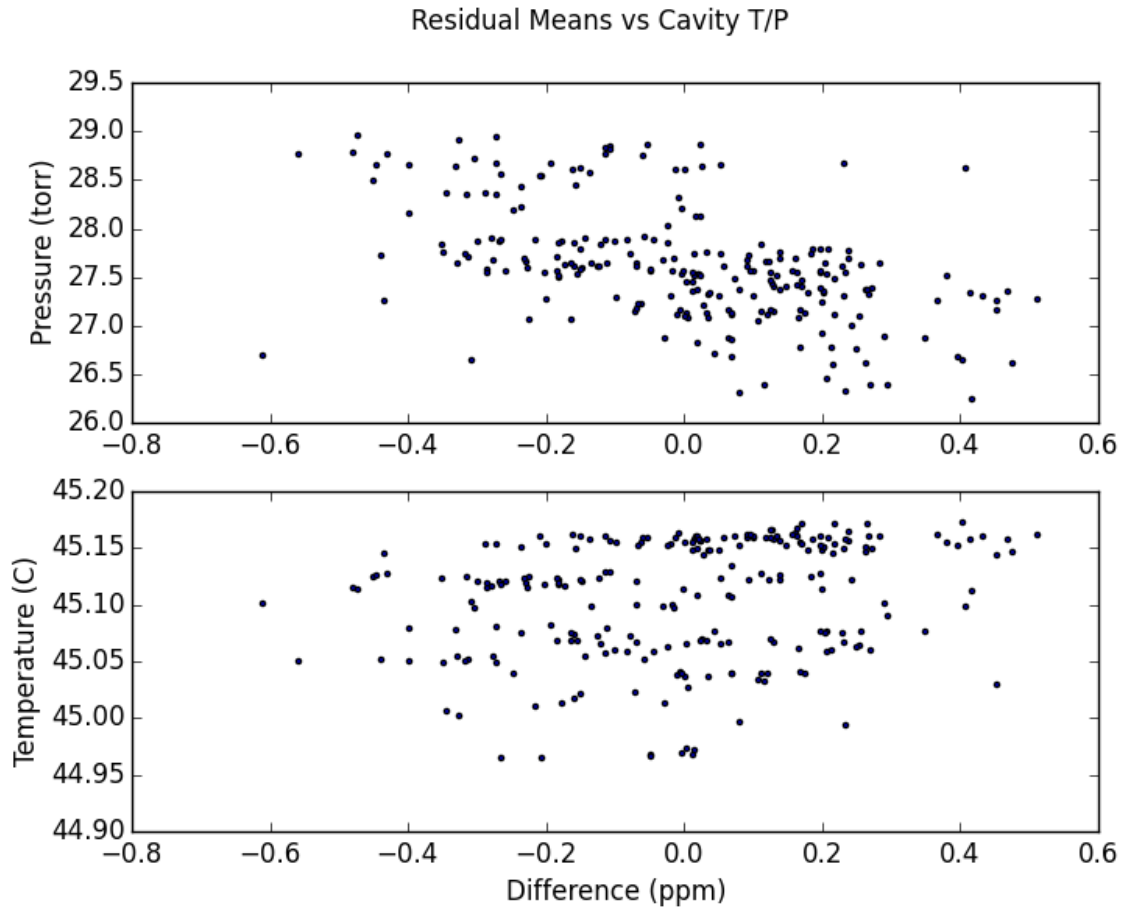
The mean residuals from the linear trend plotted against the mean cavity temperature and pressure during the calibration periods show some correlation with pressure (R^2 of 0.27) and virtually no correlation with temperature (R^2 of 0.06)

Your drift correction technique of fitting a line to each subsequent pair of calibration points will introduce discontinuities into the corrected dataset that do not represent the real-world. It would be better to fit a smoothed curve (captures short-term drift) or a single linear fit (captures the long term drift).

Thank you for the excellent suggestion, using the single linear fit actually improves the RMSE slightly across all the sensors. The analysis throughout the paper (figures, table, numbers) will be update to reflect the results with the linear correction of the LGR drift.

When describing the differences between the K30 and LGR, you say that the LGR cavity temperature and pressure are relatively well controlled. Please give some numbers to give us a sense of how well controlled they are.

Over the entire evaluation period, the standard deviation of the 2-second data is 0.44 torr for cavity pressure and 0.06 °C for cavity temperature. This has been added to the manuscript.

You average the datasets into 1-minute bins based partially on the Allan variance results for the K30. Did you also do an Allan variance for the LGR?

Doing an Allan variance on the 0.5Hz LGR data using the breathing tank reveals that the noise is also Gaussian and that the optimum averaging interval is ~100 seconds, so 1 minute is also appropriate for the LGR data.

K30 Sensor Performance and Evaluation

Sect 2.1 – Did you compute an Allan variance for more than one sensor? Do they all perform similarly?

Yes, they all perform similarly. This has been added to the paper.

Figure 4 – CO2 traces show periods of higher noise on some sensors (e.g. K30- 3 during the second half of the time period shown). In particular, I am wondering about the smattering of points that appear as outliers. In these cases, are the sensors still meeting the manufacturer's specification of +/- 30 ppm? Is there evidence that a sensor's precision can diminish over time?

Yes, there are some outliers of the two poor performing sensors that are outside of the ±30 ppm range from the original dataset, but after accounting for the zero offset, only K30-3 has any values outside of this range. This particular sensor has much more noise compared to the others, but for 1-sigma, it is within the specifications.

It's too difficult to say from this dataset if there is any evidence of precision diminishing over time. There is definitely a possibility as dust that could collect on the internal mirrors could change the absorption/path length of the IR light, and the light source could also potentially change with use. This would be something we hope to investigate in future work by performing a long-term (6-12 month) evaluation.

At the beginning of section 5, you state that CO2 measurement differences are correlated with environmental variables, but you have not demonstrated the correlation. Can you show some scatter plots?

The correlations are not perfect by any means, but shown below are scatter plots for K30-1 for before the four stages of the regression:

For each plot the x-axis is the original K30 data – LGR. Top left: the difference versus the LGR

values, top right: the difference versus atmospheric pressure, bottom left: difference versus temperature, and bottom right: difference versus water vapor. Temperature is the least correlated (the K30 includes a crude first order temperature correction) and is the variable that has little effect on the RMSE (at least in this range of observed values), and pressure is the highest. Note the density of points are not well resolved in the figures and will skew the fit lines, this is shown below with the second plot (larger figure).

K30-1 Regression Residual Correlation
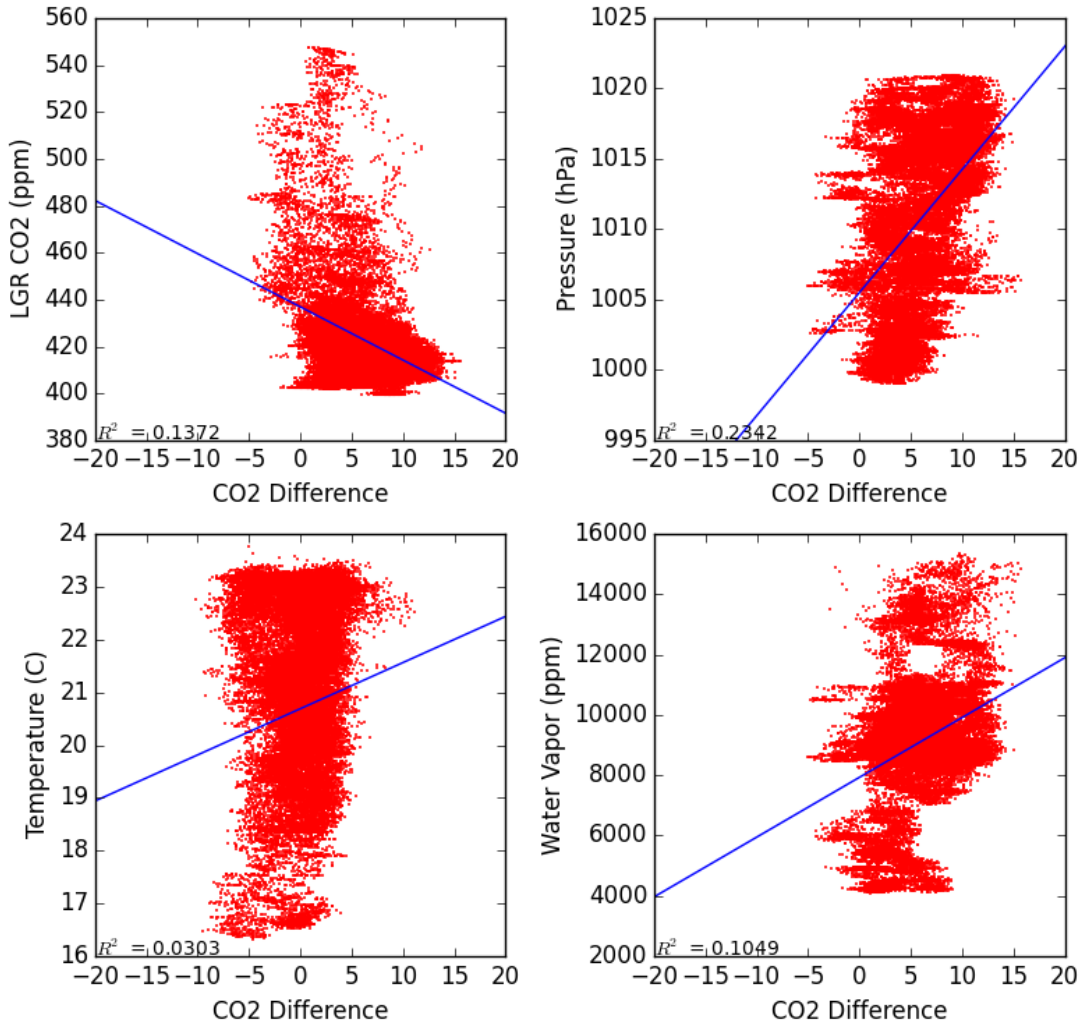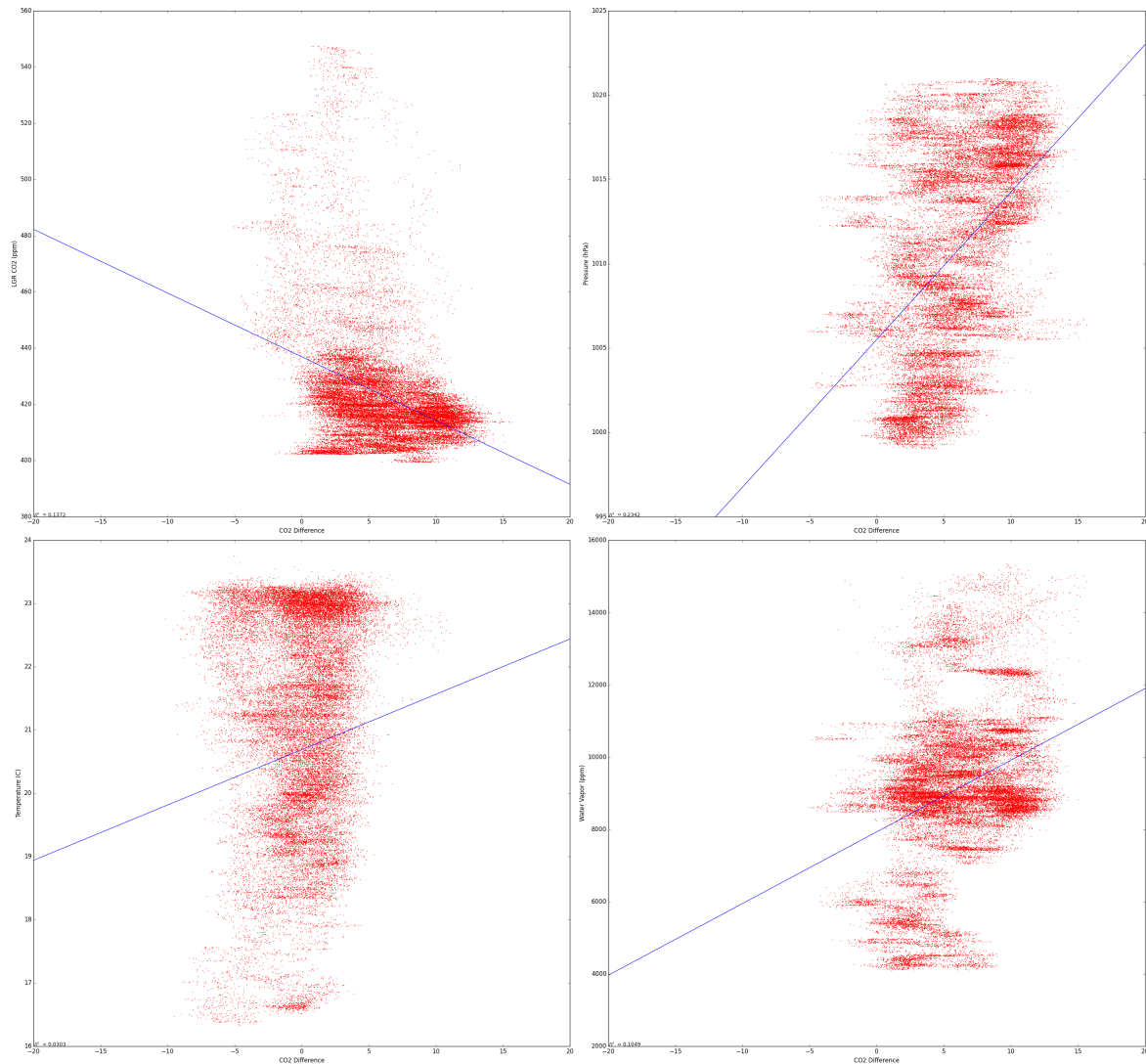
Before doing an empirical fit to the environmental parameters, it would seem sensible to account for the dilution of the CO2 mixing ratio in humid air. See section 2 of Shusterman et al. 2016 for an example.

The multivariate regression takes into account the water vapor mixing ratio, as well as temperature and atmospheric pressure, so this should be accounted for when regressing against the CO_2_dry output from the LGR. The correction described in Sect. 2 of Shusterman et al. 2016 is essentially a simplified version of the multivariate regression where they correct for varying T,P,q.

Table 1 – Are all of the regression coefficients significant? Which parameter leads to the biggest improvement and which leads to the smallest improvement?

The most significant correction comes from the simple regression against the LGR reported CO2 values. Otherwise, because the sensor uses the absorption of infrared light, from the ideal gas law, it relates the concentration relative to a reference pressure, so atmospheric pressure has the largest correction. If you change the order, the final result is still the same, but since T/P/q are all correlated from weather and diurnal variations, the first one can often have the most significant impact.

Section 6.1 – Do you find that the K30 sensors that were closer to the LGR inlet have shorter lag times relative to the LGR response? You should try computing cross- correlation functions for each K30 against the LGR to improve the time-matching of the different time series.

Unfortunately, we do not have an exact record of the location or distance from each K30 to the LGR inlet. The cross-correlation functions would be difficult on this dataset because the lag can vary depending on things like time of day and weather, thus the lag correction may change throughout the time series. In a real-world setting, there could be occasional dramatic shifts in concentrations from plumes, boundary layer dynamics, or other reasons, so this demonstrates that there is a need to wait several minutes until the signal equilibrates somewhat.

Significant figures - Most of the performance metrics stated for the K30 sensors in units of ppm CO2 are given with two decimal places, yet you state the K30 measurement resolution is only 1 ppm.

Yes, the output is 1 ppm resolution, but for 2-second data, when using the 1-minute averages, the effective resolution is higher. We have now changed these units to use only one decimal place, which is more consistent with the precision of the 1-minute averages but still shows the impact the regressions have on the RMSE.

Figures 5,6,7,8,10 are all shown for K30 #1, which, from Figure 4, appears to be one of the best performing sensors. I would be curious to see a residual plot for sensor #3, 5 or 6.

Here is Figure 8, but rather for K30 #5 than K30 #1.

K30-5 vs Los Gatos Multivariate Regression Results

Averaging Length = 1 minute(s)
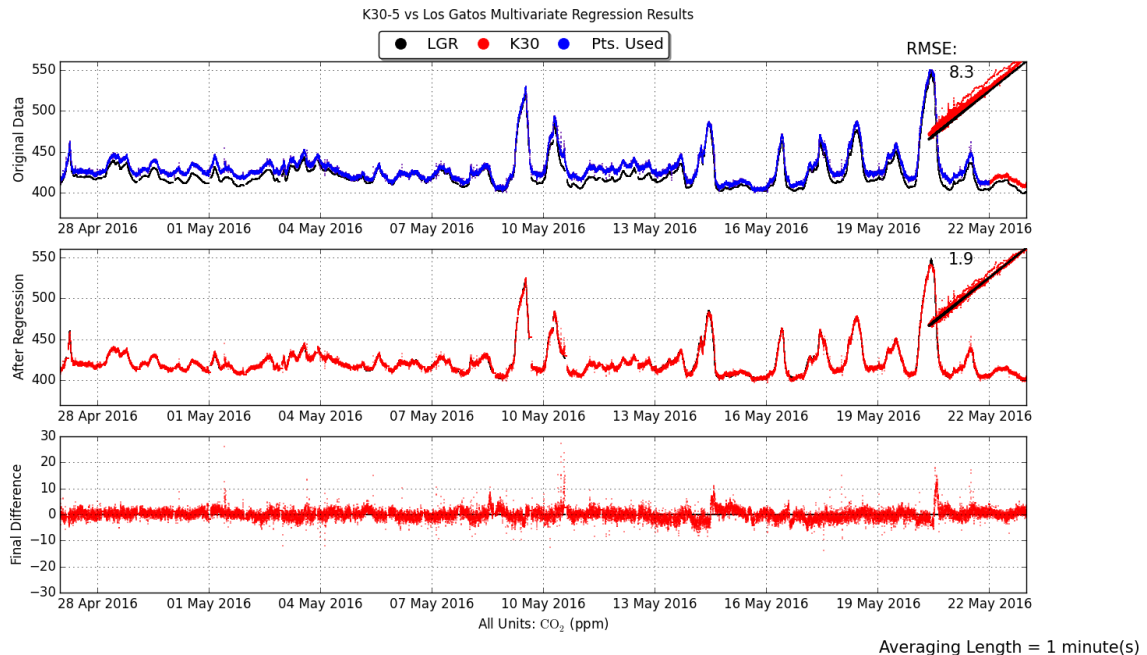
You state that one goal of this work is to understand how correction coefficients can be derived quickly. Wouldn't it be more efficient to design a controlled experiment where controlling variables are deliberately varied across the full range of operating conditions?

We were looking for a way to derive the coefficients for a large group of sensors with minimum human labor. By artificially controlling air temperature, pressure, and moisture content, the cost of the evaluation both in time and money would increase, negating some of the benefits of the price of these sensors.

Section 7 – In future work, you aspire to characterize the sensors' maximum performance in a controlled environment. Yet, if the big-picture goal is to use these sensors is to generate science quality ambient air measurements, I believe a more worthy goal would be to characterize their minimum performance in an uncontrolled environment.

This is another area we hope to pursue with future work, but would need to devise a way that would be uncontrolled but also meaningful enough for publishable data. Ideally we would like some installed outdoors next to an inlet for a gas analyzer as the reference, but would need to ensure that the sensors are in an enclosure that provides adequate ventilation but also protection from weather.

Technical comments

Title – I don't think "enhancement" is the right word. How about something like "Evaluation and correction of CO2 measurement in ambient air from a low-cost sensor"

Thanks for the suggestion, we changed it to "Evaluation and environmental correction of ambient CO_2 measurements from a low-cost NDIR sensor"

Abstract – The quantities reported have different numbers of significant figures. These should be uniform and reflect the precision of the measurement.

As mentioned above, we now use one decimal place for the RMSE of the K30 sensors.

Pg 1, Ln 25 – "dry air" is used twice in this sentence.

Fixed.

Pg 1, Ln 28 – The WMO compatibility goal is a goal, but is not always achieved, and certainly not for historical measurements.

We agree with the reviewer. We have changed this to state that this is the WMO compatibility goal.

Pg 1, Ln 29 – Suggest: '. . .to collect samples, which are subsequently transported'

Thanks, changed it to reflect this suggestion.

Pg 1, Ln 30 – You mention two expensive types of measurements – flasks and Picarros, but you do not mention moderately-priced analyzers from LiCor and Los Gatos, which are used at many research-grade monitoring sites.

This is true, but the added costs for calibration and maintenance can make a LiCor (and Los Gatos as you've seen from our analysis) comparable in total cost to a Picarro. The models are not explicitly stated and this is merely to show that either flasks or continuous observations are prohibitively expensive to do at high spatial resolution. We have added a phrase to indicate that this cost includes labor and calibration costs.

Pg 2, Ln 1 – Be consistent about whether you spell out carbon dioxide or use the abbreviation.

This has been corrected throughout the manuscript. Thanks.

Pg 2, Ln 2 – Suggest paragraph break at "Recent research"

Done.

Pg 2, Ln 8 – Is 8-12 sites typical? There are ~5 sites in Boston, SLC, and Paris.

For the LA Megacities project there are 14 sites, 11 current sites in Indianapolis, and a planned 14 sites for DC/Baltimore, 8-12 was used as a rough average of these 6 cities. The text has been changed to 3-12 to reflect the inclusion of the smaller networks listed. Additional references have been added here to show a sample of these networks both in size and geographic location.

Pg 2, Ln 9 – You say that more dense observations, even with larger uncertainties, yield better inversion constraints, but this is all relative and depends on the inversion setup/goals. See Turner et al., 2016, ACP for an exploration of the tradeoffs.

The text has been updated to state that this depends on the methodology used, and a citation to Turner et al., 2016 has been added.

Pg 2, Ln 14 – Suggest deleting "however"

Done.

Pg 2, Ln 16 – Suggest changing the phrasing to: "Recent evaluations and implementations of new low-cost sensors demonstrate their promise for ambient air monitoring."

Changed to "Evaluation and implementation of some of these new low-cost sensors demonstrate their promise for ambient air monitoring."

Pg 2, Ln 26 – Can you give some numbers to scope what you mean by "reasonably accurate"?

Based on the cited texts, ±3-5ppm, has been added to the manuscript.

Pg 3, Ln 7 – Suggest: "The K30 sensor module from SenseAir (Sweden) is the low-cost NDIR CO2 sensor that was tested for this study".

Changed to "The K30 sensor module (K30) from SenseAir (Sweden), is the low-cost NDIR CO_2 observing instrument used in this study[1]."

Pg 3, Ln 10 – Suggest deleting "given as"

Thanks. Changed.

Pg 3, Ln 13 – Suggest: "The K30 was chosen not only because it has the highest manufacturer-specified accuracy, but also because initial testing showed reliability and consistency with higher-quality observations."

Changed to close to your suggestion: "The K30 was chosen not only because of it has the highest manufacturer-specified accuracy, but also because initial testing showed reliability and

consistency when compared to higher-quality observations."

Pg 3, Ln 17 – You should give the units (relative humidity) for the 3% and 0.008% quantities.

Thanks for the suggestion. We have revised the sentence as "…has an average absolute accuracy of ±1 ℃, ±3 %, and ±1 hPa, and an output resolution of 0.1 ℃, 0.008 % and 0.01 hPa for temperature, relative humidity, and pressure, respectively"

Pg 3, Ln 24 – "less than one percent" "< 1%"

Fixed.

Pg 4, Ln 5 – Another difference between the two analyzers could be their sensitivity to the isotopes of CO2.

This could be true, but without knowing for sure, we prefer to not add this point to the paper. The LGR is only sensitive to ^{12}C, but the standards used to calibrate the LGR account for all isotopes of CO_2. Additionally, the component of ^{13}C is around 1% relative to ^{12}C, and thus the difference would be small. This is now briefly addressed in the text.

Pg 4, Ln 29 – Can you briefly describe what you mean by "various complications"?

The Raspberry Pi runs a full Linux OS, so because of the complexity of the OS, sometimes there is a delay in when certain tasks execute, which may compound into some sensors collecting (at times) observations out of phase from others. The LGR 0.5Hz data starts whenever the system initializes. Thus perfect synchronization is difficult, but all have recorded time stamps and can be averaged / regularized for comparison. A brief explanation of this has been added to Section 2.

Pg 5, Ln 1 – My understanding is that you merged datasets by their timestamps. Did you have to do something to keep the clocks synchronized?

All of the Raspberry Pi data loggers use an internet server to synchronize their time, and the LGR uses an internal clock with battery that was set to the same time as the Pis at the beginning of the experiment. This has been added to Section 2.

Pg 5, Ln 13 – "longer averaging times do not reduce the noise"

Fixed. Thank you for noticing this.

Pg 7, Ln 21 – Suggest deleting "However".

Done.

Pg 7, Ln 22 – Is the statement about each K30 meeting the manufacturer's uncertainty specification in regards to the raw (2-second) data or 1-minute averages? Please clarify in the text.

The manufacturer specifies the range for the raw data but our analysis is for the 1-minute averages. Text is updated to reflect this, both in section 2 that the datasheet is for 2-sec and in section 4 that our analysis is for 1-min.

Pg 10, Ln 21 – suggest: ". . . and 1.48 ppm, for 1-minute, 10-minute, and hourly averages, respectively.

Text is changed to this suggestion.

Pg 10, Ln 26 – suggest: "One goal of this work is to develop a methodology to evaluate individual sensors quickly. . ."

Changed to "One goal of this work is to develop a methodology to evaluate individual sensors quickly so that they can be used in scientific applications."

Pg 11, Ln 32 – "less than five parts per million" "< 5 ppm"

Fixed here as well.

Figure 1 – A ballpoint pen is included in the picture for size reference. A ruler instead of a pen would be more useful.

We liked this idea, and Figure 1 now includes a ruler instead of a ballpoint pen.

Figure 2 – What was the CO2 concentration of the tank used?

This is the breathing air tank used for the LGR drift, so estimated to be 463.7 ppm after calibrating the LGR with NIST standards, as noted in the text at the end of Section 3.

Figure 4 – State the time interval of the data shown. I can't tell if this is raw 2-second data or 1-minute averages.

1-minute averages, all figure captions have been updated to clarify this.

Figure 8 – I don't understand the difference between the red and blue points.

The blue data points are used in the regression, and the red is the complete dataset. This is done for consistency with internal plots that show the time series for regression periods of varying

length. Captions for Figs. 8 and 10 have been updated to clarify this.

Figure 9 – Can you put error bars on each point for the y-variable?

We decided to instead show a box plot as well as all six sensors' values, as this gives a better picture of the variability than just error bars. Please see updated Figure 9.

Response specific citations:

Yang, M., Prytherch, J., Kozlova, E., Yelland, M. J., Parenkat Mony, D., and Bell, T. G.: Comparison of two closed-path cavity-based spectrometers for measuring air–water CO_2 and CH_4 fluxes by eddy covariance, Atmos. Meas. Tech., 9, 5509-5522, doi:10.5194/amt-9-5509-2016, 2016.

Evaluation and environmental correction of ambient CO_2 measurements from a low-cost NDIR sensor

Cory R. Martin[1], Ning Zeng[1,2], Anna Karion[3], Russell R. Dickerson[1,2], Xinrong Ren[1,4], Bari N. Turpie[1], Kristy J. Weber[1,5]

[1]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742, USA
[2]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20742, USA
[3]National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
[4]Air Resources Laboratory, National Oceanic and Atmospheric Administration, College Park, MD 20740, USA
[5]Now at: Department of Geography, University of Colorado at Boulder, Boulder, CO 80309, USA

Correspondence to: Ning Zeng (zeng@umd.edu)

Abstract. Non-dispersive infrared (NDIR) sensors are a low-cost way to observe carbon dioxide concentrations in air, but their specified accuracy and precision are not sufficient for some scientific applications. An initial evaluation of six SenseAir K30 carbon dioxide NDIR sensors in a lab setting showed that without any calibration or correction, the sensors have an individual root mean square error (RMSE) between ~5 to 21 parts per million (ppm) compared to a research-grade greenhouse gas analyzer using cavity enhanced laser absorption spectroscopy. Through further evaluation, after correcting for environmental variables with coefficients determined through a multivariate linear regression analysis, the calculated difference between the each of six individual K30 NDIR sensors and the higher-precision instrument had an RMSE of between 1.7 ppm and 4.3 ppm for one minute data. The median RMSE improved from 9.6 for off the shelf sensors to 1.9 ppm after correction and calibration, demonstrating the potential to provide useful information for ambient air monitoring.

1 Introduction

Carbon dioxide (CO_2) is a major greenhouse gas, with fundamental importance to Earth's climate. Since measurements started at the Mauna Loa Observatory in the 1950s (Keeling et al., 2005), the global mean concentration of CO_2 has steadily risen from the preindustrial mole fraction of approximately 280 μmol mol^{-1} of dry air (parts per million, or ppm) to today's level exceeding 400 ppm. These observations, both from flask samples and state-of-the-art continuous measurement instruments, have a typical compatibility goal of ~0.1 ppm, recommended for observations at background global network sites (World Meteorological Organization, 2013). Flask-based measurements require observers to collect samples, which are subsequently transported to a lab for analysis, at significant cost. Continuous in-situ CO_2 analyzers located at towers do not suffer from these regular costs, but these high-precision analyzers can cost

1

upwards of $100,000 per site (especially once labor and calibration gases are included), plus any additional installation costs including adding inlets to towers or rooftops. High-accuracy CO_2 observations are thus relatively sparse compared to other climatological variables such as temperature and precipitation.

Deleted: carbon dioxide

5   Recent research efforts have focused more locally and on the use of networks of observing sites that use instrumented towers similar to what is used for global monitoring, but applied to the urban environment (Pataki et al., 2003; Briber et al., 2013; Kort et al., 2013; McKain et al., 2012; Turnbull et al., 2015). High-accuracy observations from these tower sites are then used to create inversions to estimate the total greenhouse gas flux from the urban area in question (McKain et al., 2012; Bréon et al., 2015; Lauvaux et

10   al., 2016). However, due to the cost of these networks being comparable to ones at the global scale, the observation towers are still sited at a relatively low density of typically 3 to 12 sites in a single metropolitan area (McKain et al., 2012; Kort et al., 2013; Turnbull et al., 2015; Bréon et al., 2015). Depending on the methodology used, a higher spatial density of observations in these urban regions has been shown to better constrain the inversion estimates, even if the absolute uncertainty of the observations is higher (Turner et

15   al., 2016; Wu et al., 2016; Lopez-Coto et al., under review), but a trade-off between total network cost and inversion constraint must be balanced.

Deleted:

Deleted: 8

Deleted: Turnbull et al., 2015;

Deleted: A

Recently, a wave of small, low-cost sensors, some of which measure trace gases or particulate matter, in addition to traditional meteorological variables, using various technologies have become commercially

20   available. Evaluation and implementation of some of these new low-cost sensors demonstrate their promise for ambient air monitoring. (Eugster and Kling, 2012; Holstius et al., 2014; Piedrahita et al., 2014; Young et al, 2014; Wang et al., 2015; Shusterman et al., 2016). Many of these instruments are based on electrochemical reactions to measure the concentrations of trace gases. With the advent of widely available and low cost mid-IR light sources and detectors, a small group of non-dispersive infrared (NDIR) CO_2

25   sensors have also become commercially available. They are designed for use in a number of applications including ventilation control, agricultural and industrial applications, and inclusion in stand-alone commercial products. Additionally, with the high volume of possible applications, these small NDIR CO_2 sensors are affordably priced on the order of $100 to $200 per sensor. Previous studies have compared some of these NDIR CO_2 devices and concluded that after application of some type of calibration

30   procedure, some of these devices can provide reasonably accurate measurements ($\pm$3-5ppm) of ambient CO_2 concentrations (Hurst et al., 2011; Yasuda et al., 2012; Shusterman et al., 2016).

Deleted: , however

Deleted: these sensors is quite promising

In this paper, one of these small NDIR CO_2 devices is assessed by determining its precision with and without environmental corrections. Section 2 describes the CO_2 sensor and its Allan variance, the other

2

instruments included in the system, and the data collection and processing methodology. Section 3 describes the calibration and shows the stability of the reference high-precision gas analyzer, and the initial results from the NDIR sensor are shown in Sect. 4. In Section 5, two methods are described to determine functional relationships and coefficient values to correct the observed values of the instrument for

5 environmental variables and Sect. 6 discusses the potential utility of observations from this sensor after correction and temporal averaging.

2 Instruments and methods

To test the validity of using low-cost sensors for scientific applications, a sensor package was implemented consisting of various off-the-shelf components. The K30 sensor module (K30) from SenseAir (Sweden), is

10 the low-cost NDIR CO_2 observing instrument used in this study[1]. The K30 is a microprocessor-controlled device with on-board signal averaging, has a measurement range of 0 to 10,000 ppm, observation frequency of 0.5 Hz, and resolution of 1 ppm. The manufacturer's stated accuracy of the K30 sensor is ±30 ppm ±3 % of reading (SenseAir, 2007) for the 0.5Hz raw output. Additional NDIR sensors were initially evaluated before selecting the K30, including the COZIR ambient sensor and Telaire T6615, having

15 accuracies described as being ±50 ppm ±3 % and ±75 ppm respectively (Gas Sensing Solutions, 2014; General Electric, 2011). The K30 was chosen not only because it has the highest manufacturer-specified accuracy, but also because initial testing showed reliability and consistency when compared to higher-quality observations. In addition to CO_2, temperature, relative humidity, and pressure readings are recorded using a breakout board purchased from Adafruit. This board features a Bosch Sensortec BME280, which

20 according to the manufacturer's datasheet has an average absolute accuracy of ±1 ºC, ±3 %, and ±1 hPa, and an output resolution of 0.1 ºC, 0.008 %, and 0.01 hPa for temperature, relative humidity, and pressure, respectively (Bosch Sensortec, 2015).

To compare the performance of the K30 to better-performing research instrumentation, a greenhouse gas

25 analyzer based on cavity enhanced absorption spectrometry (CEAS) was used as the control. The LGR-24A-FGGA fast greenhouse gas analyzer (FGGA) from Los Gatos Research (LGR, San Jose, CA) provides

---

[1]Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental

30 procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

3

Deleted:), manufactured by the Swedish company

Deleted: ,

Deleted: [1]

Deleted: . ... [1]

Deleted: given as

Deleted: of

Deleted: its

Deleted: 's

Deleted: of observed reliability and consistency with observations

CO$_2$, CH$_4$, as well as water vapor mixing ratios at a frequency of 0.5 Hz and has an un-calibrated uncertainty of < 1 % (Los Gatos Research, 2013). The FGGA was connected to a tee connection, to allow either ambient air or a calibration source (during calibrations) to be sampled continuously by the analyzer at a flow rate of 400 standard mL min^{-1}. Calibrations for CH$_4$ and CO$_2$ were conducted using several NIST-

5 certified standard mixtures every 23 to 47 hours for a period of one month with molar mixing ratios ranging from 1869.6 parts per billion (ppb) to 2159.4 ppb for CH$_4$ and from 369.19 ppm to 429.68 ppm for CO$_2$. See Sect. 3 for details and results of this calibration period.

It is important to note that there are differences in how CEAS works compared to NDIR, most notably that

10 the FGGA and other CEAS instruments have a controlled cavity where pressure and temperature are kept nearly constant (with a standard deviation of under 0.5 torr and 0.1 ºC for 2-second data), removing potential environmental interference and the need for corrections, whereas the NDIR K30 works in the ambient environment without any mechanism for keeping temperature or pressure constant. Additionally, the FGGA implements a water vapor correction on its greenhouse gas concentrations to estimate the dry

15 gas mixing ratio, while the K30 makes no water vapor corrections. A difference between the two analyzers with regard to their sensitivity to the isotopes of CO$_2$ is expected to be small because the standards used to calibrate the FGGA account for all CO$_2$ isotopes. To increase the effective path length, both the K30 and FGGA use mirrors, but the FGGA system uses highly reflective mirrors that allow for an effective path length that is many times longer than that of the K30. Additionally, the CEAS instrument determines the

20 concentration of a gas by how long it takes for the signal to degrade inside the cavity (the e-folding time), whereas an NDIR sensor merely measures the intensity of the signal received relative to the total intensity emitted.

For data collection, a Raspberry Pi (RPi) computer is used (Raspberry Pi Foundation, 2015). The RPi is a

25 credit card sized (approximately 6 x 9 cm) computer running a full Linux distribution, allowing for easy customization and usability, that is priced at around $25. The K30 is connected to the RPi over Universal Asynchronous Receiver/Transmitter (UART) Serial, and the BME280 over Inter-Integrated Circuit (I^{2}C) serial. An image of the complete sensor package is available in Fig. 1. Data is archived on the RPi and uploaded to a centralized data storage and processing server. The FGGA collects and archives its own data,

30 but an RPi is used here as well to collect the data from the FGGA over a local area network and transfer it to the same centralized server. The added computational power of a Raspberry Pi over traditional data loggers allows for the ability to archive two levels of data: the raw data collected every two seconds, and one-minute averages.

4

Deleted: less than one percent

Archiving and comparing multiple datasets proved to be challenging, so steps are taken to ensure that each compared value is at the same observed time. All of the RPis use an internet server to synchronize their time, and the LGR uses an internal clock with battery that was set to the same time as the RPis at the beginning of the experiment. Because of various complications including the exact LGR start time and the

5   potential for delays in the RPi's Linux operating system, the data collection times of each K30 sensor package and the FGGA are asynchronous. Additionally, power issues can corrupt parts of the plain text data files stored on the RPi's SD card with random characters. Thus, a post-processing procedure has been developed that filters extraneous characters, and then each dataset is synchronized based on recorded time stamps and averaged over selected time periods. These new datasets can then be directly compared without

10   missing or out of phase data points.

2.1 K30 Allan variance

Allan variance (Allan, 1966) is a measure of the time-averaged stability between consecutive

15   measurements or observations, often applied to clocks and oscillators. In addition, an Allan variance analysis can be used to determine the optimum averaging interval for a dataset to minimize noise without sacrificing signal. Figure 2 shows the Allan deviation (the square root of the variance) for one K30's raw two-second data when exposed to a known reference gas. The original two-second data shows the maximum noise, with a standard deviation comparable to the manufacturer's specifications of ± 30 ppm, but

20   averaging for even ten seconds drops the variance significantly. According to this analysis, the optimum averaging time, when the Allan variance is at a minimum (Langridge et al., 2008), is approximately three minutes; longer averaging times do not reduce the noise. The other sensors were found to perform similarly. For the subsequent analysis, an averaging time of one minute is used, as the Allan variance is only slightly higher than for three minutes, and one minute observations allow for resolution of

25   atmospheric variability at shorter time scales.

2.2 Experiment

The need to quickly and effectively evaluate a relatively large number of sensors under conditions with relatively stable CO_2 led to the use of a rooftop observation room on the University of Maryland campus in College Park, Maryland. Because this rooftop room had limited access, and it was not part of the building's

30   HVAC system, it served as an ambient evaluation chamber with minimal influence from human respiration. The room was slightly ventilated for the entire evaluation period to allow outside air to slowly diffuse into the room, with a small household box fan also in the room to ensure that the air was well mixed. The room

5

Deleted:

Deleted: to now

Deleted: D

also features a small, independent heating and cooling unit, but it was only used to keep the room from exceeding a certain temperature, thus the room was not fully temperature controlled. Even with this control, the diurnal fluctuations of temperature in the room were similar to that of the outdoor environment. This ventilation strategy was intentional so that the room then mimicked the ambient CO_2 concentration of the

5   surrounding atmosphere, and approximated the outdoor temperature and humidity, while protecting instruments from direct sunlight, extreme temperatures, and inclement weather. This provided an advantage over controlled tests in a laboratory setting in that rather than just a multi-point calibration, comparing datasets over ambient concentrations and environmental conditions allowed for a realistic evaluation of these instruments in more real world scenarios.

10

For a continuous period of approximately four weeks in spring 2016, six K30 sensor packages as described in Sect. 2 were deployed alongside the LGR FGGA in the rooftop room, all sampling room air. The FGGA was also connected to a mass flow controller and standard tank to periodically provide a reference for stability (details in Sect. 3). For the reference dataset, the dry CO_2 ($CO_{2\,dry}$) output calculated by the FGGA

15   was used. This output includes an applied correction to the mole fraction of CO_2 to give the dry air mole fraction in ppm. The raw CO_2 values were recorded from each K30, temperature and pressure were recorded from each BME280 sensor, and water vapor mole fraction was also recorded by the FGGA. All of the observations were recorded every two seconds, and averaged into one minute values. The next two sections describe the stability of the FGGA as well as the initial comparison between the K30 and FGGA

20   observations.

3 Los Gatos evaluation and correction

To evaluate the K30 NDIR sensor performance compared to a research-grade analyzer, first the control dataset needs to be calibrated and corrected for drift. To calibrate the FGGA, after the experiment concluded the dataset was corrected using a two-point calibration curve derived from using two NIST-

25   traceable gas standards, one with a CO_2 mole fraction of 369.19 ppm, and the other with a mole fraction of 429.68 ppm. A linear fit was then assumed between the two calibration points, with the recorded values as the dependent variable and the NIST-assigned tank values as the independent variable. In addition, three cylinders of breathing air with higher CO_2 mole fractions of 449.73, 486.53, and 516.41ppm (that are NIST-traceable) were also previously used to calibrate the FGGA and showed its linearity. Once the

30   coefficients were determined, the entire FGGA dataset was then corrected for further analysis.

Deleted: carbon dioxide

In addition to the calibration described above, there was a need to quantify any drift in the FGGA analyzer. During the experiment period, the FGGA was attached to a tee connector, which pulled ambient air from the aforementioned evaluation chamber using its included pump most of the time, but received periodic calibration every 23 to 47 hours for a period of initially one hour and later, ten minutes, to conserve the tank, using a reference tank of breathing air connected to a Dasibi Model 5008 calibrator. This breathing air tank is assumed to have a fixed CO_2 mole fraction, which was estimated by using the FGGA to be 463.7 ppm and was used to quantify and subtract the drift of the FGGA over the comparison period.

In Fig. 3, the ambient data from the FGGA has been filtered out to show only each calibration period performed during the month long experiment. The data during each calibration period was averaged (either a total of 10 minutes or one hour depending on the calibration period) and the averages are plotted on Fig. 3. While there is some small variation in the mean mole fraction observed during each calibration from day-to-day, there was an upward trend in the recorded value, by over 1.2 ppm over a 30-day period. This observed drift, while not insignificant, is well within the manufacturer's specifications for this analyzer. However, the observed standard deviation of the two-second points used in each average (the error bars on Fig. 3) remained relatively constant throughout the period with a mean standard deviation of ±0.3 ppm, which is the manufacturer's specified repeatability for 2-second data. This high-frequency noise is not a problem for the analysis with the K30 sensor because both datasets are averaged to one minute values, which removes most, if not all, of this noise. For comparisons between the K30s in the remainder of this paper, the FGGA drift is corrected by first computing a linear fit to the calibration points in time (red line, Fig. 3) and then subtracting from the FGGA dataset the difference of this fit line from the tank's assigned value of 463.7 ppm. After this linear correction, the means of each calibration had an RMSE of 0.2 ppm from the fit line.

4 Initial K30 results

Figure 4 shows the original time series of data recorded during the evaluation experiment described in Sect. 2.2. The top panel shows raw CO_2 mole fractions reported by six K30 sensors as well as the LGR FGGA analyzer, each of which is located in the same rooftop evaluation chamber. The middle panels show the reported atmospheric pressure and temperature values from one BME280 sensor, and the water vapor mole fraction from the FGGA. Then, the bottom panel is the difference between the original recorded K30 value and the corrected FGGA recorded CO_2 mole fraction with the calibration periods removed.

Deleted: Additionally

Deleted: as well as to correct the FGGA reported values based on known gas concentrations.

Deleted: would also receive

Deleted: either

Deleted: 10

Deleted: minutes or

Deleted: one hour

Deleted: at around

Deleted: 4

Deleted: first linearly interpolating between each

Over this four-week period, the FGGA observed an ambient variation of CO_2 with an average value of just over 423 ppm, and a standard deviation of just under 21 ppm. There is distinct synoptic variation in the diurnal cycle observed, with the magnitude varying from as little as 10 ppm over 24 hours to more than 100 ppm. Each of the K30s was successfully able to resolve the ambient variations in CO_2 over this evaluation

5   period, although none of the K30s matched the FGGA perfectly in both absolute concentration and relative change. However, without any correction or calibration, each K30 was well within the manufacturer's stated uncertainty of ±30 ppm ±3 % of the reading for 1-minute values.

From the difference plot (Fig. 4, bottom panel), there are some important things to note. First and foremost,

10   each individual K30 sensor has a distinct zero offset. A few of the sensors are approximately the same as the FGGA, but many can have an offset that is as much as 5 % (20 ppm) from the LGR FGGA. The differences between each K30 and the FGGA all have standard deviations between 4 ppm to 6 ppm and root mean square errors (RMSE) between 5 ppm to 21 ppm. This means that after accounting for the offset of each individual K30, the practical accuracy of the K30 CO_2 sensor can be within 1 % of the observed

15   concentration. Secondly, each K30 difference time series appears to feature two wave patterns, one with a period of around one week, and another with a period of approximately one day. Given that the cycles seem fairly consistent and are present in each K30, this suggests that the difference between the recorded values from the FGGA and each K30 is not random, but instead that there are external factors that can be assessed for potential compensation in the K30 response.

20   **5 Environmental correction**

In Fig. 4, the difference between the FGGA and each K30 is shown in the bottom panel below time series of environmental data from the evaluation chamber. Just like in the difference plot, each of the environmental variables features two distinct time scales of variability. There is a diurnal cycle of each variable, as well as synoptic-scale variability attributed to weather systems that occurs on the order of one

25   week. Because the observed CO_2 differences and the environmental variables are correlated on both short and long time scales, statistical regression methods were used to correct the observed concentration of CO_2 from the K30 sensor to a value approximately that of the concentration determined from the calibration-corrected FGGA measurements. Generally, a multivariate linear regression is of the form shown in Eq. (1):

30   $$y = a_1 x_1 + a_2 x_2 + \cdots a_n x_n + \varepsilon_n \quad (1)$$

In this case, the measured value y is influenced by: the 'true' CO_2 value (taken as the value from the LGR FGGA instrument), pressure, and other environmental variables as the dependent variables x_1, x_2, x_n, respectively. A multivariate regression analysis can then be used to find the corresponding coefficients. In addition, in order to better identify the contribution from each individual factor, the data was also analyzed in a successive regression analysis, as described below.

5.1 Successive regression method

Each individual K30 sensor's original observed CO_2 dataset is first regressed to the LGR FGGA dry CO_2 dataset. This regression accounts for the traditional zero and span corrections made during an instrument calibration. The calibration curve of one K30 for just zero and span is shown in Fig. 5. But to include biases due to environmental factors, then the residual, epsilon (ε), is calculated in Eq. (2) as:

$$\varepsilon = y - ax - b \tag{2}$$

where in this instance x, the independent variable, is the FGGA dataset and y, the dependent variable, is the K30 dataset.

This process is repeated for each environmental variable pressure (P), temperature (T), and water vapor (q), where (P,T,q) is the independent variable, x, and the ε from the previous step is the dependent variable, y. This linear regression method leads to eight correction coefficients, of the form a_n and b_n, where n is from 0 to 3 representing each of the independent variables included in the regression. These coefficients can then be used in Eq. (3) along with the environmental variables to correct K30 CO_2 observations for environmental influences.

$$y_{corrected} = \frac{y - b_0 - (a_1 x_1 + b_1) - (a_2 x_2 + b_2) - \cdots - (a_n x_n + b_n)}{a_0} \tag{3}$$

For one typical K30, the initial standard deviation of the difference between the K30 and FGGA, the RMSE of the data was 6.9 ppm. Using the cumulative univariate regression method described above for the entire evaluation period, the RMSE decreased after each step. After the span and offset regression, it dropped significantly to 3.3 ppm. Then after correcting for atmospheric pressure, the RMSE dropped even lower to 2.7 ppm. Furthermore, including air temperature and water vapor mixing ratio resulted in a RMSE of 2.7 ppm and 2.1 ppm respectively. It is important to note that the temperature regression did slightly reduce the RMSE, but not significantly enough to be resolved with only two significant figures. Therefore, using the successive regression method, the RMSE of the observed difference dropped from 6.9 ppm to 2.1 ppm, a

reduction of the error by over a factor of three. Fig. 6 shows the results and scatter plots for each step of the correction for this K30; Fig. 7 shows a difference plot at each step for this same K30 unit. Similar results were observed for each K30 sensor evaluated and a summary can be found in Table 1.

5.2 Multivariate linear regression method

5 Alternatively, a multivariate linear regression statistical method can be used to calculate the regression coefficients for each K30 sensor. This results in five correction coefficients a_n and b where n represents each independent variable, the dry CO_2 from the LGR FGGA, pressure P, temperature T, and water vapor mixing ratio q. Like the successive method above, these coefficients can be used in Eq. (4) along with the original K30 data, y, and the environmental variables to predict the true CO_2 concentration observed.

10

$$y_{corrected} = \frac{y - b - (a_1 x_1) - (a_2 x_2) - \cdots - (a_n x_n)}{a_0} \tag{4}$$

Using the multivariate regression function provided by Python-SciPy-Stats (Jones et al., 2001), differences from the FGGA of the same K30 described in Sect. 5.1 were reduced to an RMSE of 2.1 ppm, slightly
15 better than the iterative method. This consistently better performance from the multivariate method is shown in the other K30 sensors evaluated. Figure 8 shows the final results of the multivariate regression for the same K30 as in Fig. 6 and Fig. 7, as well as the difference between the corrected K30 dataset and the FGGA. As with the univariate method, similar results were observed from each K30 sensor evaluated and a summary can also be found in Table 1.

20

6 Discussion

6.1 Time averaging

There are two observations to note based on the evaluation and analysis. First, both before and after the multivariate regressions, there are frequent shifts in the sign of the difference between each K30 and the
25 FGGA; these sudden changes occur at or around sunrise most days. Because of the rapid change in atmospheric CO_2 concentration at this time, the ambient calibration chamber may not be well mixed during this time period. Each K30 is located in a slightly different location in the ambient calibration chamber, and are all approximately 1 to 2 meters away from the FGGA inlet. This effect, combined with the different response time of the K30s compared to the FGGA, can lead to dramatic differences between what each
30 K30 observes and what the FGGA observes at the same timestamp for a short period of time each day.

10

Atmospheric inversion methods often use hourly averaged data from tower observations (McKain et al., 2012; Bréon et al., 2015; Lauvaux et al., 2016), so after the multivariate regression was applied, the K30 and FGGA datasets were further averaged to 10 minute and hourly datasets. The average RMSE for the six K30s with the one-minute data is 2.3 ppm, 2.0 ppm for 10-minute averages, and 1.8 ppm for hourly-averaged data. Throughout this analysis period, one of the six K30s evaluated performed consistently worse than the others, and after removing it from the averages, the RMSE values dropped to 1.9 ppm, 1.6 ppm, and 1.5 ppm, for 1-minute, 10-minute, and hourly averages, respectively. Thus, by using hourly averages and discarding underperforming sensors, the average RMSE of the difference between the LGR FGGA and a K30 NDIR sensor can be reduced to approximately 1.5 ppm.

6.2 Regression period

The RMSE described above and in Table 1 are for regressions calculated over the entire experiment period of approximately four weeks. One goal of this work is to develop a methodology to evaluate individual sensors quickly so that they can be used in scientific applications. In Fig. 9 the average RMSE calculated over the entire month of all six K30s is plotted with respect to the number of days used in the multivariate regression from Sect. 5.2. While the RMSE is generally minimized with increasing regression length, after a regression period of just a few days, the RMSE drops significantly from its initial values. Once a few diurnal cycles of varying amplitude have been incorporated, as well as the synoptic scale variations in the atmosphere (with a time scale of around one week), the regression stabilizes. Thus, a regression length of around two weeks is recommended to maximize correction while minimizing the required amount of time the sensor needs to run concurrently with the FGGA.

In Fig. 10, a multivariate regression is applied to the same K30 as described in the aforementioned sections and shown in Figs. 6, 7 and 8, but the coefficients are only calculated for the first 15 days. The change in the RMSE between the two regressions is 0.1 ppm, going from 1.8 ppm when using all data points to 1.9 ppm when using only approximately the first half. This small, but not insignificant change is most likely attributed to the fact that during the first half of the evaluation period, the ambient CO_2 concentrations do not vary significantly, especially relative to the second half, where both the minimum and maximum values occur. In fact, when instead regressing for the last 15 days of the period, the RMSE is 1.8 ppm, a difference not distinguishable with only one decimal place. So as stated above, the diurnal cycles act as a range of calibration points, but values above and below what is included in the regression period may cause the corrected data to still have large errors during these periods, increasing the RMSE for the entire evaluation cycle. Based on these results, it is reasonable to assume that there is either no noticeable baseline drift or that it is assumed to be linear and removed by the multivariate regression in the sensors observed on the

11

Deleted: 2
Deleted: 4
Deleted: 87
Deleted: 0
Deleted: 66
Deleted: 48

Deleted: these sensors individually

Deleted: 25
Deleted: 89
Deleted: 2.14

Deleted: 94
Deleted: of only 0.05 ppm

weekly to monthly timescales. The longer-term drift of the sensors for periods greater than one month is not known at this time, however, and would require a longer evaluation period of at least six months.

6.3 Uniform Regression Coefficients

All of the final RMSEs calculated in this analysis are from using individual regression coefficients for each K30 sensor. However, it would be beneficial to determine if a uniform set of regression coefficients could be applied to any K30 sensor, and what the RMSE over the evaluation period would be. To calculate the uniform coefficients, the four slopes for each variable as well as the intercepts for each of the five remaining sensors were averaged together, K30-3 was omitted due to the fact that it was the poorest performing sensor, and that its coefficients were significantly different from the other five. After correction using the same set of coefficients, the RMSEs of the six sensors ranged from 3.1 ppm to as high as 23.9 ppm. The final RMSEs in some cases were higher than with the original, uncorrected data. Thus, it appears that for each K30 sensor, an independent evaluation must be completed to provide observations with a sufficient level of quality.

7. Conclusions and future work

The K30 is a small, low-cost NDIR CO_2 sensor designed for industrial OEM applications. Each of the sensors tested falls within the manufacturer's stated accuracy range of ±30 ppm ±3 % of the reading when compared to a high-precision CEAS analyzer, but these ranges are not particularly useful for scientific applications aimed at measuring ambient atmospheric CO_2. If these sensors are individually calibrated, selected for stability, and corrected for sensitivity to temperature, pressure, and RH, the practical error of these sensors is < 5 ppm, or approximately 1 % of the observed value. The final RMSE of the six K30 ranged between 1.7 ppm and 4.3 ppm for 60 s averaging times. Averaging for 200 s further reduces the noise by about 30 %, but longer times did not further improve precision. With errors in this range, these instruments could be used in a variety of scientific applications, including observations at high spatial density to better represent the range and distribution of an urban or natural region's CO_2 concentration.

In the future, further analysis will be performed evaluating the K30 as well as other low-cost CO_2 sensors in a laboratory setting with controlled temperature, pressure and relative humidity. A Picarro cavity ring-down spectroscopic greenhouse gas analyzer will be used as a high-precision control and the various instruments will be subjected to ambient air as well as periodic reference gases. From this lab analysis, we hope to determine the theoretical maximum performance of these sensors in a controlled environment. This

12

subsequent study will additionally attempt to quantify any long-term drift over the course of multiple months.

5

10

Deleted: The K30 is a small, low-cost NDIR CO_2 sensor designed for industrial OEM applications. Each of the sensors tested falls within the manufacturer's stated accuracy range of ± 30 ppm ± 3 % of the reading when compared to a high-precision CEAS analyzer, but these ranges are not particularly useful for scientific applications aimed at measuring ambient atmospheric CO_2. If these sensors are individually calibrated, selected for stability, and corrected for sensitivity to temperature, pressure, and RH, the practical error of these sensors is less than five parts per million, or approximately 1 % of the observed value. The final RMSE of the six K30 ranged between 1.6 ppm and 4.4 ppm for 60 s averaging times. Averaging for 200 s further reduces the noise by about 30 %, but longer times produce no lower noise. With errors in this range, these instruments could be used in a variety of scientific applications, including observations at high spatial density to better represent the range and distribution of an urban or natural region's CO_2 concentration.

13

References

Allan, D. W.: STATISTICS OF ATOMIC FREQUENCY STANDARDS, Proceedings of the Institute of Electrical and Electronics Engineers, 54, 221-&, 10.1109/proc.1966.4634, 1966.

5 Bosch Sensortec: BME280 Digital Pressure Sensor Datasheet, available at: https://cdn-shop.adafruit.com/datasheets/BST-BME280_DS001-10.pdf (last access: 7 June 2016), 2015.

Bréon, F. M., Broquet, G., Puygrenier, V., Chevallier, F., Xueref-Remy, I., Ramonet, M., Dieudonne, E., Lopez, M., Schmidt, M., Perrussel, O., and Ciais, P.: An attempt at estimating Paris area CO_2 emissions from atmospheric concentration measurements, Atmospheric Chemistry and Physics, 15, 1707-1724, doi:(?)

10 10.5194/acp-15-1707-2015, 2015.

Briber, B., Hutyra, L., Dunn, A., Raciti, S., and Munger, J.: Variations in Atmospheric CO_2 Mixing Ratios across a Boston, MA Urban to Rural Gradient, Land, 2, 304, 2013.

Eugster, W., and Kling, G. W.: Performance of a low-cost methane sensor for ambient concentration measurements in preliminary studies, Atmospheric Measurement Techniques, 5, 1925-1934, 10.5194/amt-5-

15 1925-2012, 2012.

Gas Sensing Solutions: COZIR Ultra Low Power Carbon Dioxide Sensor, available at: http://www.gassensing.co.uk/media/1050/cozir_ambient_datasheet_gss.pdf (last access: 29 December 2015), 2014.

General Electric: Telaire T6615 Sensor Dual Channel Module, available at: http://www.avnet-

20 abacus.eu/fileadmin/user_upload/Products_Menu/Amphenol/AmphenolAdvancedSensors_CO2_double_channel_module.pdf (last access: 29 December 2015), 2011.

Holstius, D. M., Pillarisetti, A., Smith, K. R., and Seto, E.: Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California, Atmospheric Measurement Techniques, 7, 1121-1131, doi:(?) 10.5194/amt-7-1121-2014, 2014.

25 Hurst, S., Durant, A. J., and Jones, R. L.: A low cost, disposable instrument for vertical profile measurements of atmospheric CO_2, Centre for Atmospheric Science, Department of Chemistry, University of Cambridge, 2011.

Jones E, Oliphant E, Peterson P, et al.: SciPy: Open Source Scientific Tools for Python, http://www.scipy.org/ (last access: 10 October 2016), 2001.

Keeling, C. D., Piper, S. C., Bacastow, R. B., Wahlen, M., Whorf, T. P., Heimann, M., and Meijer, H. A.:

30 Atmospheric CO_2 and $^{13}CO_2$ exchange with the terrestrial biosphere and oceans from 1978 to 2000: Observations and carbon cycle implications, History of Atmospheric CO_2 and Its Effects on Plants, Animals, and Ecosystems, 177, 83-113, 2005.

14

Kort, E. A., Angevine, W. M., Duren, R., and Miller, C. E.: Surface observations for monitoring urban fossil fuel CO_2 emissions: Minimum site location requirements for the Los Angeles megacity, Journal of Geophysical Research-Atmospheres, 118, 1-8, 10.1002/jgrd.50135, 2013.

Lopez-Coto, I., Ghosh, S., Prasad, K., and Whetstone, J.: Tower-Based Greenhouse Gas Measurement Network Design – the NIST North East Corridor Testbed, Advances in Atmospheric Science, under review, 2016.

Los Gatos Research: Fast Greenhouse Gas Analyzer (Enhanced Performance Model) Datasheet, available at: http://www.lgrinc.com/documents/LGR_FGGA_Datasheet.pdf (last access: 29 December 2015), 2013.

Langridge, J. M., Ball, S. M., Shillings, A. J. L., and Jones, R. L.: A broadband absorption spectrometer using light emitting diodes for ultrasensitive, in situ trace gas detection, Review of Scientific Instruments, 79, 14, 10.1063/1.3046282, 2008.

Lauvaux, T., Miles, N. L., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., Gaudet, B., Gurney, K. R., Huang, J., and O'Keefe, D.: High-resolution atmospheric inversion of urban CO_2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX), Journal of Geophysical Research: Atmospheres, 2016.

McKain, K., Wofsy, S. C., Nehrkorn, T., Eluszkiewicz, J., Ehleringer, J. R., and Stephens, B. B.: Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region, Proceedings of the National Academy of Sciences of the United States of America, 109, 8423-8428, 10.1073/pnas.1116645109, 2012.

Pataki, D. E., Bowling, D. R., and Ehleringer, J. R.: Seasonal cycle of carbon dioxide and its isotopic composition in an urban atmosphere: Anthropogenic and biogenic effects, Journal of Geophysical Research-Atmospheres, 108, 8, 10.1029/2003jd003865, 2003.

Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., Li, K., Dick, R. P., Lv, Q., Hannigan, M., and Shang, L.: The next generation of low-cost personal air quality sensors for quantitative exposure monitoring, Atmospheric Measurement Techniques, 7, 3325-3336, 10.5194/amt-7-3325-2014, 2014.

Raspberry Pi Foundation: Raspberry Pi Hardware Documentation, available at: https://www.raspberrypi.org/documentation/hardware/raspberrypi/ (last access: 29 December 2015), 2015.

SenseAir : CO_2 Engine K30 Specification, available at: http://www.senseair.com/wp-content/uploads/2015/03/CO2-Engine-K30_PSP110-R7.pdf (last access: 29 December 2015), 2007

Shusterman, A. A., Teige, V., Turner, A. J., Newman, C., Kim, J., and Cohen, R. C.: The BErkeley Atmospheric CO_2 Observation Network: initial evaluation, Atmos. Chem. Phys. Discuss., 2016, 1-23, 10.5194/acp-2016-530, 2016.

Turnbull, J. C., Sweeney, C., Karion, A., Newberger, T., Lehman, S. J., Tans, P. P., Davis, K. J., Lauvaux, T., Miles, N. L., Richardson, S. J., Cambaliza, M. O., Shepson, P. B., Gurney, K., Patarasuk, R., and Razlivanov,

15

I.: Toward quantification and source sector identification of fossil fuel CO_2 emissions from an urban area: Results from the INFLUX experiment, Journal of Geophysical Research-Atmospheres, 120, 292-312, 10.1002/2014jd022555, 2015.

5   Turner, A. J., Shusterman, A. A., McDonald, B. C., Teige, V., Harley, R. A., and Cohen, R. C.: Network design for quantifying urban CO_2 emissions: assessing trade-offs between precision and network density, Atmos. Chem. Phys., 16, 13465-13475, 10.5194/acp-16-13465-2016, 2016.

Wang, Y., Li, J. Y., Jing, H., Zhang, Q., Jiang, J. K., and Biswas, P.: Laboratory Evaluation and Calibration of Three Low- Cost Particle Sensors for Particulate Matter Measurement, Aerosol Science and Technology, 49, 1063-1077, 10.1080/02786826.2015.1100710, 2015.

10   World Meteorological Organization: GAW Report No. 213: 17th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques (GGMT-2013), http://www.wmo.int/pages/prog/arep/gaw/documents/Final_GAW_213_web.pdf (last access: July 11, 2016), 2013.

Wu, L., Broquet, G., Ciais, P., Bellassen, V., Vogel, F., Chevallier, F., Xueref-Remy, I., and Wang, Y.: What

15   would dense atmospheric observation networks bring to the quantification of city CO_2 emissions?, Atmos. Chem. Phys., 16, 7743-7771, 10.5194/acp-16-7743-2016, 2016.

Yasuda, T., Yonemura, S., and Tani, A.: Comparison of the Characteristics of Small Commercial NDIR CO_2 Sensor Models and Development of a Portable CO_2 Measurement Device, Sensors, 12, 3641-3655, 10.3390/s120303641, 2012.

20   Young, D. T., Chapman, L., Muller, C. L., Cai, X. M., and Grimmond, C. S. B.: A Low-Cost Wireless Temperature Sensor: Evaluation for Use in Environmental Monitoring Applications, Journal of Atmospheric and Oceanic Technology, 31, 938-944, 10.1175/jtech-d-13-00217.1, 2014.

Formatted: Subscript

Table 1. Root mean square error in ppm between the CEAS LGR FGGA and each K30 NDIR sensor's one-minute averaged data for: the original dataset before correction, at each step of the successive regression correction (correcting for 1. zero/span, 2. atmospheric pressure, 3. temperature, and 4. water vapor mixing ratio), and after the multivariate regression correction. Each value shown is for a regression calculated using data from the entire evaluation period.

	Original	Zero/Span	Pressure	Temp	q (final)	Multivariate
K30 # 1	6.9	3.3	2.7	2.7	2.1	1.8
K30 # 2	5.4	3.5	2.2	2.2	1.9	1.7
K30 # 3	10.9	6.0	5.0	4.9	4.5	4.3
K30 # 4	20.8	3.7	2.5	2.4	1.9	1.7
K30 # 5	8.3	3.7	2.6	2.6	2.2	2.0
K30 # 6	15.2	4.9	3.6	3.5	2.7	2.2

17

Figure 1. Photograph of a Raspberry Pi computer (top), a SenseAir K30 (NDIR) CO$_2$ sensor (bottom center), a Bosch BME280 temperature and pressure sensor (bottom left), and a ruler for size reference.

5
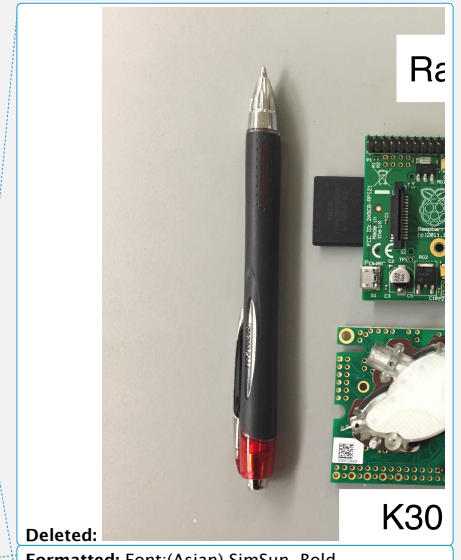
Deleted:

Formatted: Font:(Asian) SimSun, Bold

Formatted: Centered

Deleted: right

Deleted: ballpoint pen

K30-4 Allan Deviation Analysis
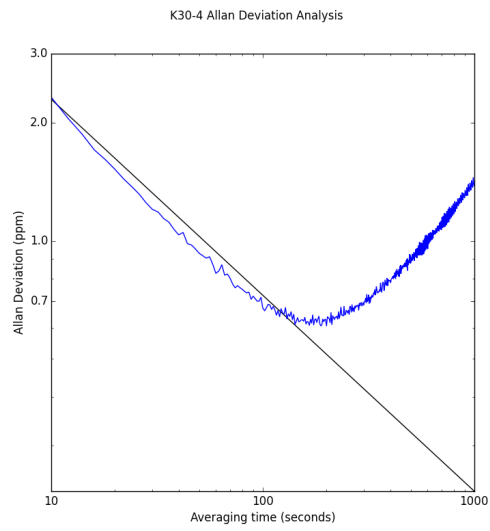
Figure 2. Allan variance analysis for an NDIR (K30) CO_2 sensor when introduced to breathing air from a high-pressure cylinder of a constant and known CO_2 concentration. Averaging times between 10 and 1,000 seconds are shown. The black line (slope -0.5) shows where the noise is white or Gaussian. Averaging times greater than about 200 s produce no improvement.

19
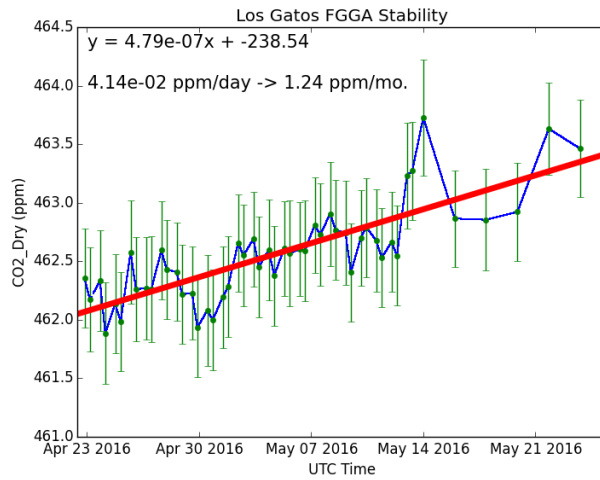
Figure 3. Stability of the Los Gatos Fast Greenhouse Gas Analyzer shown over a 30-day period. Excess
breathing air with a fixed CO_2 concentration was introduced periodically using a mass flow controller. The
mean of each calibration period is plotted in green with the standard deviation as error bars. The blue line is
the linear interpolation between each calibration point, and the red line is a linear fit of each calibration
point over the entire time series. The red line is subtracted from the dataset to account for the drift of the
analyzer over this period.

20
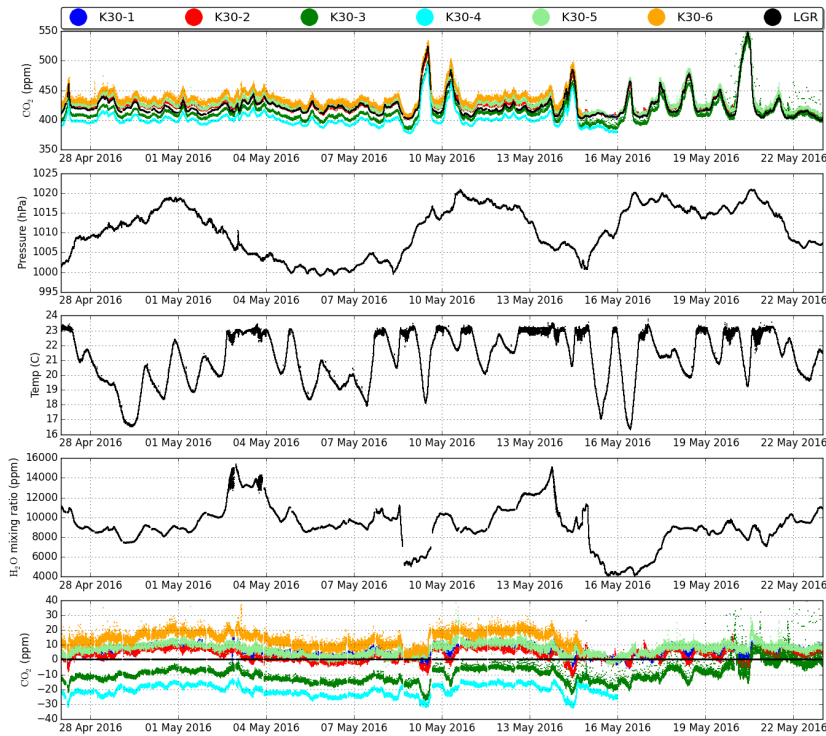
Figure 4. Continuous 1-minute time series data during the evaluation experiment. Top panel: CO_2 observed by six K30 sensors as well as the Los Gatos Research Fast Greenhouse Gas Analyzer. Middle panels: observed atmospheric pressure, temperature, and water vapor mixing ratio, respectively. Bottom panel: difference of each K30 from the Los Gatos instrument.
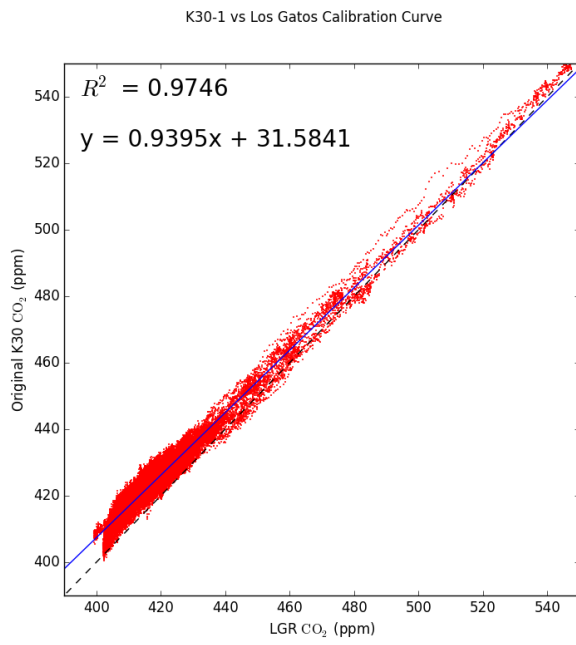
21

Figure 5. Calibration curve of K30-1 vs LGR FGGA for 1-minute averages without any environmental correction, only span and zero offset are corrected. Solid line is the best fit; dashes represent the 1:1 line

5
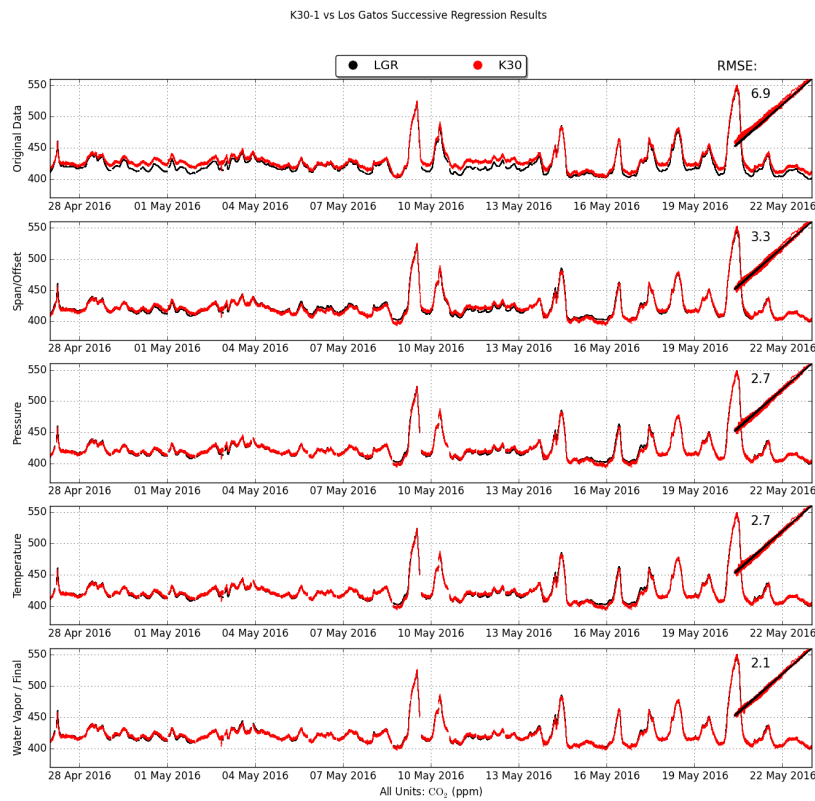
Figure 6. A continuous time series of 1-minute averages as well as scatter plots for K30 #1 compared to the LGR instrument during each step of the successive regression described in Sect. 5.1. Cumulative, in order from top to bottom: the original dataset, after correcting for span and offset, after correcting for pressure, after correcting for temperature, and finally, after correcting for water vapor. The root mean square error (RMSE) of the K30 data compared to the LGR FGGA at each step is annotated to the upper left of the scatter plot. This regression contains all data points observed in the evaluation period.

23
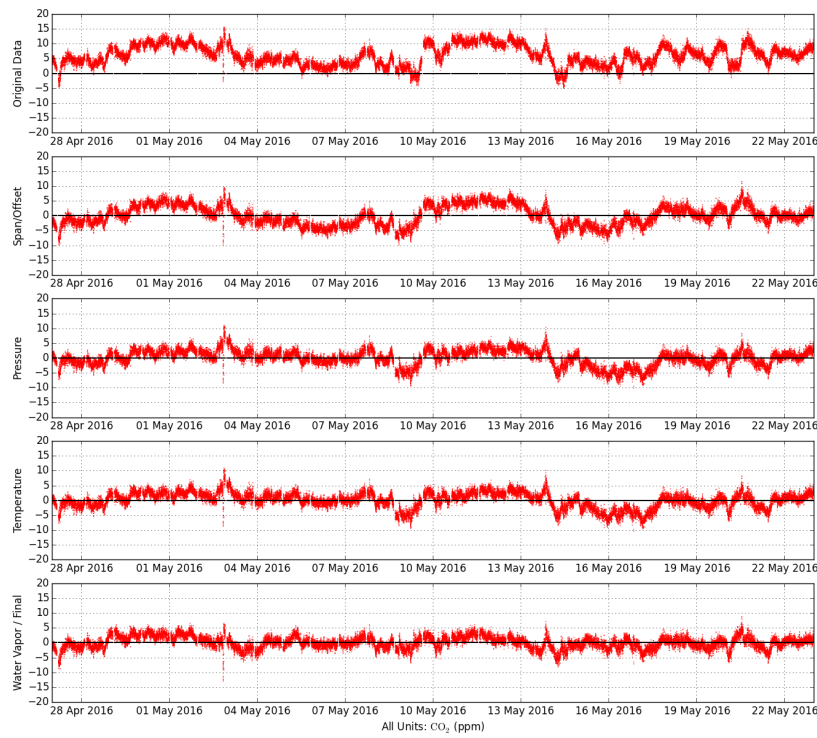
Figure 7. Difference plots for K30 #1 compared to the LGR FGGA during each step of the successive regression described in Sect. 5.1 and shown in Fig. 6 for 1-minute averages. Cumulative, in order from top to bottom: the original dataset, after correcting for span and offset, after correcting for pressure, after correcting for temperature, and finally, after correcting for water vapor.
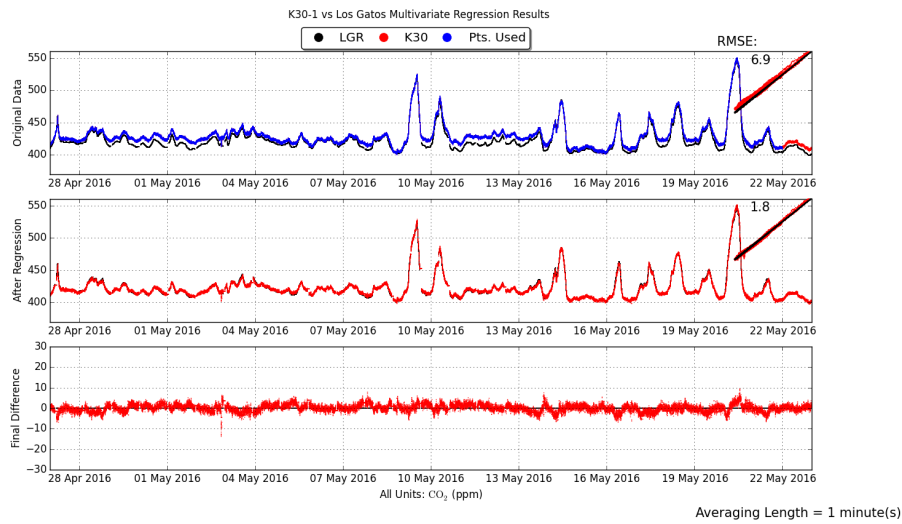
24

Figure 8. A continuous time series of 1-minute averages as well as scatter plots for K30 #1 compared to the LGR FGGA for the multivariate regression described in Sect. 5.2. Top panel: the original data, middle panel: final time series after correction, and the bottom panel: difference plot between the corrected K30 dataset and the original FGGA dataset. The root mean square error (RMSE) of the K30 data compared to the FGGA before and after the regression is annotated to the upper left of the scatter plot. The blue data points are used in the regression, which cover the first 25 days of the evaluation period, whereas the red points represent the entire dataset.

25

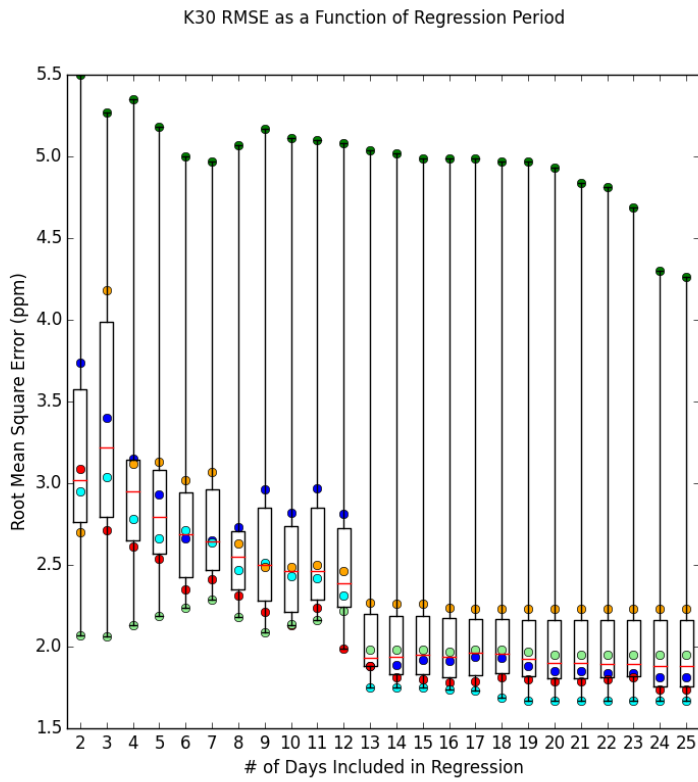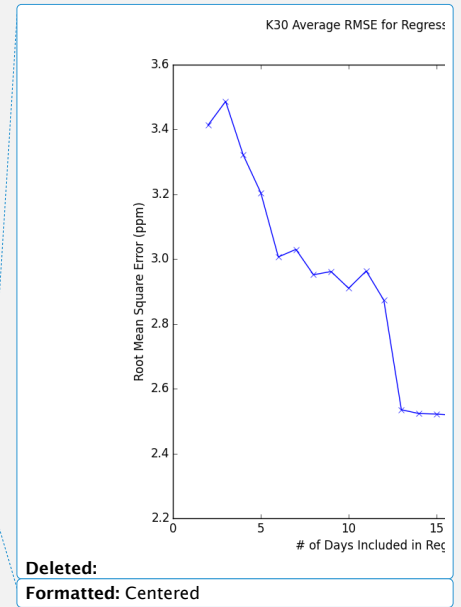**K30 RMSE as a Function of Regression Period**

Figure 9. The RMSE of all six K30 NDIR sensors when compared to the LGR FGGA over the entire experiment as a function of how many days the regression analysis was performed. The colored dots represent each K30's RMSE, and the box plot shows the median in red, the first and third quartiles within the box, and the min and max values on the whiskers.

K30 Average RMSE for Regress

Deleted:

Formatted: Centered

Deleted: average RMSE of all six K30 NDIR sensors when compared to the LGR FGGA over the entire experiment as a function of how many days the regression analysis was performed.
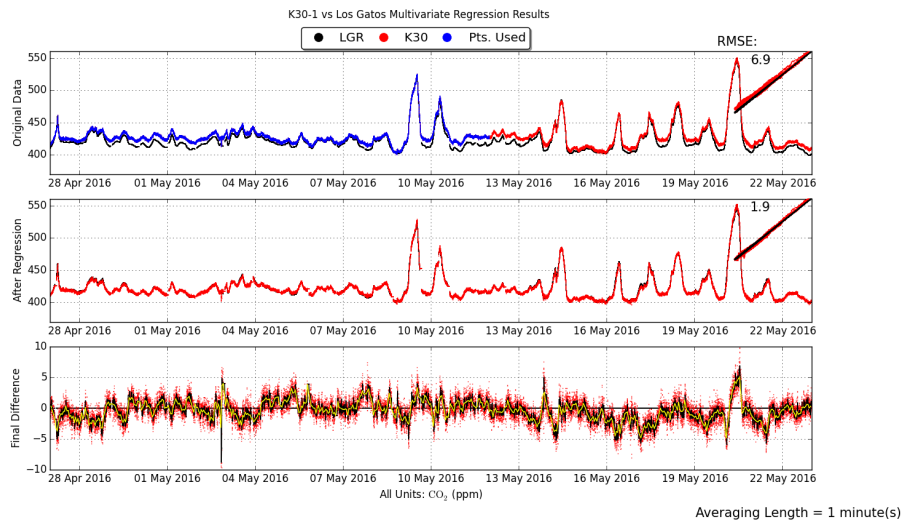
Figure 10. As depicted in Fig. 8, a continuous time series as well as scatter plots for K30 #1 compared to
5   the LGR FGGA for the multivariate regression described in Sect. 5.2. Top panel: the original data, middle
panel: final time series after correction, and the bottom panel: difference plot between the corrected K30
dataset and the original FGGA dataset. However, this regression only includes the first 15 days of data
(regression training data in blue, the entire dataset in red) to compute the correction coefficients. The
difference plot (bottom) also shows running means for 10 minute (black) and hourly (yellow) averages.
10

[1]Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.