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.