We appreciate the referee's time and feedback, which have resulted in significant improvements to our manuscript, as detailed below:

The manuscript describes a novel CO2 monitoring network consisting of low-cost sensors. First results from a measurement period are presented. The manuscript is well written, with mostly well-prepared figures and a clear structure. I recommend publication after the following comments have been addressed.

1. The procedure used for bias correction is somewhat unclear the way it is described. What I understand: a CO2 background determined as the weekly minimum at a reference site is subtracted from the CO2 time series at all sites, then each timeseries is deseasonalized, and the weekly minimum of the resulting timeseries is fitted as a piecewise linear function of time to derive the time-dependent part of bias (B_temporal). After removing this time-dependent part of the bias, the mean of the weekly minima at each site are taken as the constant bias term and subtracted from the data. Overall it looks like a high-pass filtering of the data (after de-seasonalizing), as slowly varying or constant contributions are subtracted. The question is if any contribution of constant gradients between the different stations (as expected given the differences in near- field emissions) are left after this procedure, or if the assumption really is that each site potentially "sees" background air once per week.

Because the weekly minima measured at the reference site will reflect the effective seasonal variation present in the BEACO$_2$N domain, the initial "background subtraction" step and the de-seasonalization are one and the same. To clarify this point, we have revised the text as follows:

"Once the $[CO_2]_{background}$ term is removed, effectively de-seasonalizing the observations, we re-calculate the weekly minima of this new data record and fit the result as a (piecewise, if necessary) linear function of time…"

Furthermore, we have reason to believe that each site samples background air approximately once per week. As described in our response to Referee #1, preliminary analyses revealed that the weekly minima measured at each site roughly track the three-dimensional Pacific boundary "curtain" mentioned in the text. Such comparison with the boundary curtain is of course complicated by the very drift and biases that the subsequent correction procedure aims to remove, so it is not possible to quantify the influence of inter-site gradients independently of systematic uncertainties via this method. We did, however, perform a similar comparison using measurements from the sea level LI-COR LI-820 maintained by the Pacific Marine Environmental Laboratory. As mentioned in the text, this instrument is calibrated against a reference gas prior to every measurement, and so is assumed here to be free of drift and/or bias. Although the LI-820's weekly minima do not agree precisely with the boundary curtain's (residuals ranged from 0–12.7 ppm, with a mean of 1.8 ppm), the deviations from the curtain values were not significantly autocorrelated on timescales greater than one week. From this we conclude that, while the assumption of agreement with a reference site may not be guaranteed for any particular week, the deviations are not in fact persistent across multiple weeks, even for instruments sited well within the roughness sublayer. Because we require at least three months of comparison with the reference for drift correction, the influence of anomalous weeks is minimized, and we are confident that, on these timescales, the network-wide weekly minima agree to within approximately ±2 ppm. This explanation has been added to the text (see below), and the uncertainty associated with this assertion has been factored into a more detailed error propagation (described later in this response).

"BEACO$_2$N's unique location near the Pacific coast results in a relatively consistent wind direction from largely unpolluted over-ocean origins, such that the weekly minima can be assumed to reflect both the seasonal and synoptic variations in network-wide baseline CO$_2$ concentrations while avoiding the influence of shorter term variability in local sources and sinks. This assumption is supported by preliminary analyses comparing observations from a LI-COR LI-820 non-dispersive infrared CO$_2$ gas analyzer with a smoothed, three-dimensional "curtain" of surface CO$_2$ Pacific boundary conditions produced by NOAA's Global Greenhouse Gas Reference Network (Jeong et al., 2013). The LI-COR, positioned at sea level between the EXB and EXE nodes (see Fig. 1), is maintained by NOAA's Pacific Marine Environmental Laboratory and calibrated against compressed gas (400–500 ppm CO$_2$) prior to every hourly measurement and is assumed to have negligible bias. Despite a proximity to local surface-level emissions and complex boundary layer dynamics, the LI-COR's weekly minima are found to generally follow variations in the Pacific curtain, with an average residual of ~2 ppm."

2. Error propagation should be included, propagating errors after pressure correction, temperature correction, water vapour correction, and bias removal (time-varying and constant). In that context it is worth mentioning that the bias error is dominant, not the precision error, when aggregating to yearly signals.

We have added a more detailed accounting of our error propagation in several places throughout the text, including a consideration of short-term drift, as follows:

"…the standard deviation of their differences is tightened from ±1.5 ppm to ±1.4 ppm. This still exceeds the ±1.0 ppm precision one would expect under average conditions given the form of Eq. (1) and (2) and the manufacturer's specifications for the meteorological sensors (see Sect. 3.5), the CarboCap, and the Picarro (Sect. 3.3), suggesting that the combined effect of the lingering temperature and water biases with any unknown factors is ±0.4 ppm."

"Also presented in Fig. 5 is a time series of the running 1 hour means of the differences between the minute-averaged CarboCap and Picarro observations, demonstrating a short-term drift incurred on approximately hourly timescales found to range between 0.01 and 2.9 ppm during any given 6 hour period of the co-location. The upper bound exceeds the ±1 ppm manufacturer–specified 6 hour short-term stability as well as the 1.5 ppm maximum short-term drift observed by Rigby et al. (2008), but in many cases longer averaging times can be used to reduce the influence of short-term drift to well below 1 ppm. Some modeling studies, for example, utilize time steps of 6 hours or more (e.g. Bréon et al., 2015; Wu et al., 2016), and average diurnal cycles can often be assessed across several days. Although some applications require finer temporal resolution, these are typically plume-based analyses that rely on rapidly-varying enhancements above recent background concentrations, essentially eliminating concerns about short-term drift."

"Uncertainties in $\mathrm{U_{temporal}}$ and $\mathrm{U_{atemporal}}$ shown in Table 3 are calculated given ±1.4 ppm random error in the 1 minute averages, ±2.9 ppm short-term drift, and ±2 ppm agreement with the reference site's weekly minima, assumed to add in quadrature. Mapped onto the observations, these uncertainties result in a mean 1 minute error of ±4 ppm. This is the assumed cumulative error used in this study, although longer averaging times could be used to reduce this figure."

Figure 5 and Table 3 have also been updated accordingly.

3. I second the referee #1 comment on the model representation error, which is really crucial as only with a transport model the observations can be quantitatively linked to the fluxes that are of interest.

We agree that the surface-level siting of sensors may enhance their sensitivity to local phenomena and limit their single-handed representativeness of the larger domain. However, the individual sensors' local sensitivity does not necessarily prohibit their collective ability to characterize citywide trends and/or events in the context of atmospheric transport models. Turner et al. (2016), for example, demonstrate the utility of synthetic, BEACO_2N-like observations in constraining mesoscale fluxes, even in the presence of persistent site-to-site biases. To our knowledge, BEACO_2N is the first network with sufficient sensor quantity and density to investigate this issue empirically, a full exploration of which is beyond the scope of this study. To clarify current uncertainties surrounding the representativeness and capabilities of our sensors, we have updated various sections of the text as follows:

"This largely opportunistic siting approach avoids the logistical and financial obstacles associated with tall tower sampling mechanisms, although it does present additional challenges for the quantification of network-wide phenomena in that no low-lying instrument can singlehandedly provide sensitivity to the entire domain. Installing sensors near the surface and/or built environment does ensure heightened sensitivity to individual, ground-level emissions phenomena, but it is currently unknown whether a well-reasoned combination of these locally sensitive signals from a high volume of sensors could nonetheless yield reliable information about the integrated region. A full exploration of this possibility is beyond the scope of this study; the following analyses focus instead on establishing BEACO_2N as a viable platform for investigating such hypotheses."

"Although BEACO_2N demonstrates sensitivity to both highly local fluctuations as well as slowly-varying hemispheric cycles, how best to bootstrap the network's measurements into the analysis of intermediary mesoscale phenomena remains to be determined. Future work will focus on constructing inferred emissions patterns and trends at this scale from the body of observations. In an initial effort in this regard…"

P1 L23: The reference "A.B. 32, 2006" should probably read "Brown et al., 2006"

The Brown et al. publication in the References list refers to California's First Update to the Climate Change Scoping Plan published in 2014, whereas the A.B. 32 citation is intended to refer directly to the California Global Warming Solutions Act passed by the state legislature in 2006 that mandates the creation of such scoping plans. We have revised the format of the latter citation in the References list to clarify this point:

"A.B. 32: California Global Warming Solutions Act, Assemb. Reg. Sess. 2005–2006, (CA 2006)."

P2 L24: Fig. 3 shows gradients in CO2 fluxes, not in CO2

The text has been updated to reflect this correction as follows:

"see Fig. 3 for examples of observed intra-city CO_2 flux gradients"

P3 L10-25: ". . . sensitivity to changes <10 ppm per year are required" this is quite large compared to the 65 ppb/year. On which metric or specification is the focus (mentioned in line 24)?

The focus of the paragraph in question is on the 65 ppb/year metric. The text introducing the <10 ppm/year metric has been revised to clarify as follows:

"The precision requirements at each individual site versus for a network instrument as a whole vary depending on the phenomena of interest. Metropolitan regions produce <10 ppm CO_2 enhancements in the boundary layer (Pacala et al., 2010), requiring sensitivity to changes orders of magnitude smaller for the characterization of citywide integrated inter-annual trends, for example."

P7 L15: The precision of 1-minute averages of the Picarro CRDS systems should be lower than 0.1 ppm, as for a single five second measurement is specified by the manufacturer to be better than 70 ppb (25 ppb for 5 min. averages).

The ±0.1 ppm precision figure for the Picarro CRDS system was obtained via personal communication with the collaborator who maintains said system, and reflects a combination of the fact that this system is based on an older version of the Picarro G2301 instrument (specified by the manufacturer in 2010 to possess 5 second precision better than 150 ppb) and that it contains a myriad of ancillary custom parts that are not necessarily accounted for in the manufacturer's specifications. This being said, the ±0.1 ppm precision does indeed pertain to a 5 second measurement frequency, not a 1 minute average. We have revised the original text to clarify this fact:

"In this case the 1σ precision of the 1 minute averages is ± 1.4 ppm, given by the standard deviation of the differences between the minute-averaged CarboCap and Picarro observations and the Picarro's precision (±0.1 ppm at 5 second measurement frequency)."

P7 L17: Were the different instrument's time response taken into account in the comparison of the CarboCap and the CRDS? As the CarboCap has a diffusion driven exchange of the sample gas, the instrument response function should be quite different from the more or less instantaneous measurement characteristics of the G2301. Taking this into account would potentially improve the precision estimate based on the comparison.

The time series are aligned by eye during analysis to account for any lag between the instruments as well as any differences between their timestamp clocks (which was found to be the dominant factor driving the offset between the two datasets). Manual inspection revealed that the various peaks and other structural features in the two time series were of equal duration, giving no indication of the supposedly slower response time of the CarboCap. This may be partially due to our use of intake/outflow fans, which force air through the sensor faster than it would passively diffuse otherwise. In any case, any small differences in time response that remained after this treatment would be effectively removed by averaging to a one-minute timescale, so this consideration is unlikely to affect our precision estimate.

Fig. 9 should be modified, as it is impossible to discern the different time series. May be a series of time series plots with 3-5 sites per plot, all sharing the same time axis, but with different vertical (CO2) axis.

We agree it is difficult to discern specific trends and features from the original version of Fig. 9, but have found the suggested alternative to be similarly inscrutable. Instead, we have chosen to supplement Fig. 9 with additional subplots depicting progressively shorter time periods at a subset of sites. The goal of this figure is simply to qualitatively represent the variety and sheer volume of atmospheric conditions sampled by the BEACO_2N instrument; more quantitative impressions of the dataset are detailed later in the text. Revisions to the original language and caption reflecting the updated version of Fig. 9 can be found below:

"Figure 9 demonstrates the volume and diversity of urban CO_2 concentrations sampled, exhibiting extreme short-term variability superimposed on a slower, seasonal fluctuation in the minimum values. For clarity, the bottom panels depicting month- and week-long samples of the overall time series show data from six representative sites. Network-wide, daytime (1100–1800 LT) means between 408 and 442 ppm are observed…"

"Figure 9: Time series of drift- and bias-corrected CO_2 dry air mole fractions collected over the course of ~2.5 years at 16 BEACO_2N sites (top), one month at six representative sites (middle), and one week at the same six sites (bottom). The hiatus around Day 600 corresponds to a largescale hardware refurbishment effort that began in mid-2014."

P9 L25-28: the fact that the seasonal cycles agree in summer and not in winter seems mostly related to the choice of July as a reference

Here we intend to refer to the magnitude of the seasonal variation—which is independent of the month chosen as reference—rather than the absolute values of the monthly minima. We have revised the language as follows to clarify this point:

"At many sites, the BEACO_2N minima are seen to exhibit a seasonal variation of a magnitude roughly in keeping with that observed in the curtain, while other sites demonstrate a more exaggerated summer-winter contrast, as might be expected within an urban dome."

Caption figure 11: is the "standard error" the error of the mean, or the standard deviation? This should be made clear.

The caption has been updated to clarify this point:

"Error bars indicate the standard error of the mean (instrument error is negligible at this timescale); thick shaded curves indicate standard deviation."

P10 L9: what was used as lateral boundary condition for the regional WRF-STILT model? This is not specified in the Turner et al. (2016) paper focused on network design.

The lateral boundary conditions for CO_2 were provided by NOAA's three-dimensional Pacific boundary "curtain" (Jeong et al., 2013), while the lateral boundary conditions for meteorology were provided by the North American Regional Reanalysis (Mesinger et al., 2006). We have added text specifying the former, as well as text referring the reader to the supplement of Turner et al. (2016) for further details:

"We then add this local enhancement to a background concentration ($\boldsymbol{y_B}$, from the aforementioned Pacific boundary curtain) to obtain a model estimate of the BEACO_2N observations…"

"…the setup used here follows that of Turner et al. (2016; see their Sect. S1 for details of the WRF setup)."

References:

Jeong, S., Hsu, Y.-K., Andrews, A. E., Bianco, L., Vaca, P., Wilczak, J. M., and Fischer, M. L.: A multitower measurement network estimate of California's methane emissions, J. Geophys. Res. Atmos., 118, 11339–11351, doi:10.1002/jgrd.50854, 2013.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shagran P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. H., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W.: North american regional reanalysis, Bull. Am. Meteorol. Soc., 87, 343–360, doi:10.1175/BAMS-87-3-343, 2006.

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. Discuss., in review, doi:10.5194/acp-2016-355, 2016.