

We thank the referee for their comments and time, which have improved our manuscript in many ways, as detailed below:

This manuscript presents an overview of a city-scale CO₂ monitoring network, based on low-cost instruments. The manuscript is concise and well written and presents an interesting experiment. The instrumentation described appears to be well designed and mostly well tested. My primary concerns center on the setup of the network as a whole, the uncertainty quantification and the interpretation of what is or is not possible with a network of this nature.

1. In an urban network such as this, the details of how the instruments are situated is likely to be critical. However, only a very cursory explanation is given in the text (P5, L28: 2 to 111 m above ground level. . .). I find it a little concerning that the instruments appear to be situated either very close to the ground (2m), or on rooftops. In urban areas, flows will be significantly influenced by near-by obstacles within the “roughness sublayer”. It is typically assumed that this layer extends at least 2 building heights into the atmosphere (Roth et al., 2000). This is important for emissions verification, because: a) in order to calculate fluxes from concentration measurements, we need to be able to accurately simulate flows from source to receptor; b) measurements within the roughness sublayer, or worse, within the urban canopy layer, will be representative of only a very small area around them, rather than the wider region. If the BEACON instruments are all within the roughness sublayer, I suspect that the network will not be able to meet its aim of monitoring changes in city-scale fluxes, because changes will be representative of only very localized areas, and the modelling requirements of simulating complex flows around buildings, etc., will be too demanding (see comment regarding Figure 12). The authors need to provide much more detail on how they plan to deal with these issues, and whether the instruments have already been situated to account for these factors.

We understand the concern regarding the representativeness of sensors located within the roughness sublayer. However, a strong sensitivity to local phenomena does not necessarily preclude a given sensor’s ability to provide information about the larger domain. As this network is the first of its kind with such a high quantity and density of sensors, more investigation is certainly required to definitively describe the relationship between local and citywide sensitivity, a full exploration of which is beyond the scope of this paper. In the meantime, we have added and revised language in several places throughout the paper (see below) to clarify the degree of uncertainty surrounding the potential capabilities and applications of these sensors, and look forward to providing more quantitative evidence in future publications.

“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 BEACO₂N as a viable platform for investigating such hypotheses.”

“Although BEACO₂N 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...”

2. I think that the discussion of uncertainties in the instrumentation could be expanded upon further. In particular, I would like to see further characterization of instrumental drift. More details are given below

We have added a more thorough characterization of the uncertainties associated with our instrumentation throughout the text; see our responses to specific comments below for details.

3. A network such as this is an exciting and important development. However, I think the authors should be a little more self-critical about the potential limitations. In particular, claims such as a 2% potential accuracy on emissions estimates seem overly optimistic to me for reasons given below.

We have revised language throughout the text to clarify the limitations of our original claims; see our response to general comment #1 above and to the specific comment pertaining to the 2% emissions accuracy claim in particular.

P1 L30: I think it’s a bit of an exaggeration to say that national monitoring networks “give no information” on urban emissions. A network of instruments in rural areas can still see integrated signals from nearby cities.

We have revised the original language as follows to provide a more nuanced description of the capabilities of national monitoring networks:

“Traditional strategies for assessing greenhouse gas emissions are limited to a small handful of monitoring instruments scattered sparsely across remote areas, mostly in developed nations (e.g., Worthy et al., 2003; Thompson et al., 2009; Andrews et al., 2014). These stations are capable of measuring regional average and some integrated urban concentrations with extreme accuracy and precision, but are purposefully distanced from and experience reduced sensitivity to urban signals, thus giving little to no spatially resolved information on emissions in the precise areas that the majority of greenhouse gas rules aim to regulate.”

P3 L17: With the assertion that uncertainties scale with \sqrt{N} , the authors are making the assumption that each sensor is an independent estimate of the city-wide concentration. Whilst I agree with the sentiment that an increased number of (well sited and modelled) instruments would lead to a decrease in uncertainty, I suspect that the uncertainty reduction on the urban scale will be nowhere near \sqrt{N} , which must be considered a theoretical limit. In reality, each instrument will see a footprint around it on the order of a few km, superimposed on some signal from the wider region. Even in a somewhat more “box-model”-like limit, the \sqrt{N} argument would assume that all sensors “see” the entire integrated signal of the city, whereas in reality, they would

only see everything upstream. Therefore, at any one time, only some subset of the network will be seeing anything close to the “whole” city.

We agree that a \sqrt{N} improvement in uncertainty is a perhaps unreasonable theoretical limit, and have revised the text as follows:

“If the goal is verification of regional inter-annual emissions targets, we would therefore require N instruments of sufficient individual sensitivity and spatial representativeness such that their combined signals allow us to detect annual changes of ~ 65 ppb year⁻¹ with confidence.”

Section 1.4 and Section 3.4: I’m not sure if the term “bias” is the most appropriate here. It appears from Rigby et al. (2008) and section 3.4 that in addition to potential offsets (which I would class as a bias), these instruments can also drift on a range of timescales. Perhaps “systematic uncertainties” would be more appropriate? Furthermore, I think that the assertion that the instruments can be considered “unbiased” if any systematic uncertainty is smaller than the precision is a little difficult. In practice, any systematic offsets could be much more critical than the repeatability. Even relatively small values could have a major impact on an inversion. I think that the paper would benefit from a more nuanced approach in which the uncertainties are more fully characterized. In particular, I think a discussion of the potential “uncorrected” instrumental drift should be shown (i.e. the authors present a method for correcting \sim weekly drifts. However, to compare the data to a model, one would still need to know what the magnitude of potential sub-weekly drift is likely to be).

We have revised language throughout the original text to characterize temporal drift as a “systematic uncertainty” rather than a “bias” and avoid spurious comparisons between the magnitude of these systematic uncertainties and that of the precision. We have also added more detailed accounting of uncertainties throughout the text, including an exploration of sub-weekly 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 below 1 ppm. Some modelling 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 U_{temporal} and $U_{\text{atemporal}}$ 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.

P5 L14: Following from the discussion above, I think that this comparison would benefit from a plot of the time-varying difference between the two instruments to assess the magnitude of the drift that one would expect in the field.

We have added such a subplot to Fig. 5 in the revised manuscript to aid in the more detailed discussion of instrumental uncertainties described above.

P6 L16: The running costs for a CRDS seem very high here. Furthermore, the sentence sounds like pumps, data loggers, etc are “annual” running costs, which seems erroneous to me.

The characterization of pumps, data loggers, etc. as “annual” costs is indeed erroneous, and the original language has been revised accordingly:

“For comparison, a single commercial cavity ring-down analyzer is priced around \$60,000 USD and the total equipment cost can exceed \$85,000 USD after accounting for pumps, data loggers, etc.”

P7 L4: How has the influence of wind speed and boundary layer height been isolated from other factors? Surely boundary layer height will be strongly correlated with e.g. an emissions diurnal cycle?

We agree that the existing analysis does not allow us to discriminate between wind speed/boundary layer height fluctuations and other diurnal patterns (e.g. in emissions); and have revised the original language as follows:

“The two sensors nonetheless demonstrate remarkable agreement; while typical diurnal CO_2 variations during the same period are on the order of 20–60 ppm, the CarboCaps simultaneously detect CO_2 events as small as 8 ppm, providing preliminary evidence of the suitability of these sensors for high-density urban deployment.”

P7 L31: As I understand it, the correction for weekly drift makes the implicit assumption that all sites see the same minimum CO_2 concentration as the reference sites. Can the authors comment on how robust this assumption is likely to be? I’m particularly concerned that, with a vertical difference of 500masl between the sensors, this procedure could add biases into the network due to persistent vertical gradients within the network in a particular week.

We developed this assumption after preliminary analyses revealed that the weekly minima measured at each site roughly tracked the three-dimensional Pacific boundary “curtain” mentioned in the text. If each site (including the reference) samples background air approximately once a week, the sites’ weekly minima should agree with one another as well. 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 vertical 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, network-wide weekly minima agree to within approximately ± 2 ppm. We have added an explanation of the preceding comments to the text as follows:

“BEACO₂N’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 CO₂ 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 CO₂ gas analyzer with a smoothed, three-dimensional “curtain” of surface CO₂ 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 CO₂) 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.”

P1, first paragraph: The discussion of uncertainty reduction largely focuses on another paper under review (Turner et al., 2016). However, I suspect that the estimate that the “accuracy” of Oakland emissions could be reduced to less than 2% (or even 18%) is wildly optimistic for the following reasons: a) Synthetic data studies of this nature make heavy assumptions of Gaussian PDFs and unbiased statistics; b) systematic model errors are largely ignored. In reality, my suspicion is that models will have a very tough job of accurately simulating flows at these scales. I do not agree with the assertion that “These combined error budgets are typically dominated by transport (model) error, which potentially explains why models based on BEACO₂N-like networks perform comparably to or better than those based on sparser networks of higher quality sensors, for which instrument error may be reduced but accurately representing transport between observation sites is of greater importance.” I suspect that, given the resolution of the flows

involved, it may be even more difficult for a model to accurately simulate concentrations for dense urban monitoring network at present (such difficulties would be impossible to discern in a synthetic data experiment). Furthermore, uncertainties in inversions such as this are likely to be very non-Gaussian, and I suspect that the uncertainty budget is likely to be dominated by systematic factors in both the observations and the model.

The referee raises multiple important points:

Turner et al. (2016) do assume Gaussian PDFs and the error statistics may not in fact be Gaussian. However, the true form of the distribution is typically unknown. Additionally, the assumption of Gaussian prior and likelihood distributions has the benefit of the posterior and prior being conjugate distributions. This means that we have a closed-form expression for the posterior distribution and we can decompose the prior covariance matrix with a Kronecker product (e.g., Yadav & Michalak, 2013; Sect. S2 in Turner et al., 2016). All of this allows us to construct high-dimensional state vectors with fully populated covariance matrices that will give us a better representation of the true statistics than assuming, say, uncorrelated errors. In the absence of knowledge of the true distribution, Gaussian distribution seem to be a fair assumption.

With regards to the effect of systematic errors, these are not ignored, but discussed in Sections 6.1 and S6.2 of Turner et al. (2016).

In terms of the error budget in general and transport error in particular, we agree that the flows will be more difficult to correctly simulate at high resolution in an urban region and will induce a large uncertainty. This term is probably the largest uncertainty, as stated in the text. However, the referee speculates that it would be more difficult to simulate concentrations for a dense network (presumably, compared to a sparse network). This is a point that is discussed in Section 6.3 in Turner et al. (2016):

“In this work we have treated transport error and the number of measurement sites as independent. However, in practice, there would be a relationship between the transport error and measurement network density. This can be understood with a thought experiment using two different observing systems to estimate emissions: a sparse network with a single site and an infinitely dense network (sites at each grid cell in our domain). Estimating emissions with the sparse network would require us to simulate the atmospheric transport with high fidelity if we are to reliably say anything about emissions upwind of our site. This is especially true for point sources. Any errors in the simulated atmospheric transport would adversely impact the estimated emissions, whereas the infinitely dense network could potentially neglect atmospheric transport and use data from only the local grid cell to estimate emissions. This is because the differential signal at each site would be largely governed by the local emissions.”

Figure 11: This figure appears to show model simulations at three sites. However, no details of the model are given in the text. Either the model setup should be explained, or the model runs should be removed from the figure.

Additional details regarding the model setup have been added to the text as follows:

“We simulate hourly CO₂ concentrations ($\hat{\mathbf{y}}$) at each site in the network using the Stochastic Time-Inverted Lagrangian Transport model (STILT; Lin et al., 2003) coupled to the Weather Research and Forecasting model (WRF; Skamarock et al., 2008). The coupled model is known as “WRF-STILT” (Nehrkorn et al. 2010) and the setup used here follows that of Turner et al. (2016; see their Sect. S1 for details of the WRF setup). WRF-STILT advects an ensemble of 500 particles 3 days backwards in time, each with a small random perturbation, from the spatio-temporal locations of the BEACO₂N observations using the meteorological fields from WRF. The trajectories of these 500 particles are then used to construct “footprints” for each observation that represent the sensitivity of the observation to a perturbation in emissions from a given location. The footprints can be represented in matrix form (\mathbf{H}) and multiplied by a set of gridded emissions (\mathbf{x} , from the high-resolution bottom-up CO₂ inventory in Turner et al. 2016) to compute the CO₂ enhancement at each site due to local emissions:

$$\Delta \mathbf{y} = \mathbf{H} \mathbf{x} \quad (5)$$

We then add this local enhancement to a background concentration (\mathbf{y}_B , from the aforementioned Pacific boundary curtain) to obtain a model estimate of the BEACO₂N observations shown as black squares in Fig. 11:

$$\hat{\mathbf{y}} = \Delta \mathbf{y} + \mathbf{y}_B = \mathbf{H} \mathbf{x} + \mathbf{y}_B \quad (6)$$

Figure 12: To me, this figure of a site in a school suggests that representation issues in the current network could be severe. The authors show that concentrations were substantially lower on a day when the school was closed. The magnitude of the signal (~50ppm) shows that this sensor must be completely dominated by the school. Therefore, can we not conclude that the sensor sees little of the wider city, and any long-term changes in concentration at this location will be indicative primarily of change in the school’s emissions? Certainly, separating a city-wide 65ppb decrease from this signal (1000x smaller) would seem highly challenging.

This figure does demonstrate one sensor’s strong sensitivity to its local environment, however, as mentioned earlier in our response, there is not yet any evidence to suggest that a sensor’s local sensitivity is necessarily mutually exclusive with its utility in assessing domain-wide phenomena, although further investigation is clearly needed in this area. We hope that our aforementioned revisions describing the uncertainties surrounding this issue of sensitivity will help to assuage concerns about this figure in particular.

References:

Turner, A. J., Shusterman, A. A., McDonald, B. C., Teige, V., Harley, R. A., and Cohen, R. C.: Network design for quantifying urban CO₂ emissions: Assessing trade-offs between precision and network density, Atmos. Chem. Phys. Discuss., in review, doi:10.5194/acp-2016-355, 2016.

Yadav, V. and Michalak, A. M.: Improving computational efficiency in large linear inverse problems: an example from carbon dioxide flux estimation, Geosci. Model Dev., 6, 583–590, doi:10.5194/gmd-6-583-2013, 2013.

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 BEACO₂N 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 $[CO_2]_{\text{background}}$ term is removed, effectively de-seasonalizing the observations, we recalculate 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).

“BEACO₂N’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 CO₂ 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 CO₂ gas analyzer with a smoothed, three-dimensional “curtain” of surface CO₂ 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 CO₂) 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 U_{temporal} and $U_{\text{atemporal}}$ 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, BEACO₂N-like observations in constraining mesoscale fluxes, even in the presence of persistent site-to-site biases. To our knowledge, BEACO₂N 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 BEACO₂N as a viable platform for investigating such hypotheses.”

“Although BEACO₂N 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 CO₂ 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 CO₂ 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 BEACO₂N 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 CO₂ 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 CO₂ dry air mole fractions collected over the course of ~2.5 years at 16 BEACO₂N 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 BEACO₂N 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 CO₂ 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 (y_B , from the aforementioned Pacific boundary curtain) to obtain a model estimate of the BEACO₂N 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 CO₂ emissions: Assessing trade-offs between precision and network density, *Atmos. Chem. Phys. Discuss.*, in review, doi:10.5194/acp-2016-355, 2016.

The BERkeley Atmospheric CO₂ Observation Network: initial evaluation

Alexis A. Shusterman¹, Virginia Teige¹, Alexander J. Turner², Catherine Newman¹, Jinsol Kim³⁺, Ronald C. Cohen^{1,3}

¹Department of Chemistry, University of California Berkeley, Berkeley, CA 94720, USA

²School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 01451, USA

³Department of Earth and Planetary Science, University of California Berkeley, Berkeley, CA 94720, USA

Correspondence to: Ronald C. Cohen (rccohen@berkeley.edu)

Abstract. With the majority of the world population residing in urban areas, attempts to monitor and mitigate greenhouse gas emissions must necessarily center on cities. However, existing carbon dioxide observation networks are ill-equipped to resolve the specific intra-city emission phenomena targeted by regulation. Here we describe the design and implementation of the BERkeley Atmospheric CO₂ Observation Network (BEACO₂N), a distributed CO₂ monitoring instrument that utilizes low-cost technology to achieve unprecedented spatial density throughout and around the city of Oakland, California. We characterize the network in terms of four performance parameters—cost, reliability, precision, and ~~bias~~systematic uncertainty—and find the BEACO₂N approach to be sufficiently cost-effective and reliable while nonetheless providing high-quality atmospheric observations. First results from the initial installation successfully capture hourly, daily, and seasonal CO₂ signals relevant to urban environments on spatial scales that cannot be accurately represented by atmospheric transport models alone, demonstrating the utility of high-resolution surface networks in urban greenhouse gas monitoring efforts.

1 Introduction

As two-thirds of the human population stand to inhabit cities by 2050 (United Nations, 2014), developing a thorough understanding of urban greenhouse gas emissions is of ever-growing importance. International and local law-making bodies around the world are agreeing to caps on total emissions and enacting multi-faceted regulations aimed at achieving these caps (e.g., A.B. 32, 2006; United Nations, 2015). As of yet there exists no mechanism for judging the efficacy of these individual rules or verifying compliance through direct observations of changes in CO₂ at the scale of cities (Duren et al., 2012).

Traditional strategies for assessing greenhouse gas emissions are limited to a small handful of monitoring instruments scattered sparsely across remote areas, mostly in developed nations (e.g., Worthy et al., 2003; Thompson et al., 2009; Andrews et al., 2014). These stations are capable of measuring regional average and some integrated urban concentrations with extreme accuracy and precision, but are purposefully distanced from and experience reduced sensitivity to urban

signals, thus giving little to no spatially resolved information on emissions in the precise areas that the majority of greenhouse gas rules aim to regulate.

The increasing significance of urban emissions combined with the proliferation of commercial cavity ring-down spectroscopic instrumentation have resulted in a recent trend towards network sensing approaches for constraining greenhouse gas emissions in cities. For example, Ehleringer et al. (2008) maintain a CO₂ monitoring network in the Salt Lake City metropolitan area, the INFLUX network measures CO₂, ¹⁴CO₂, and total column CO₂ across the city of Indianapolis (Turnbull et al., 2015), and NASA’s Megacities Carbon Project has established sensor networks in the pilot cities of Los Angeles (Kort et al., 2013) and Paris (Bréon et al., 2015). These ground-based monitoring efforts are complemented by space-based observations from SCIAMACHY (Burrows et al., 1995), GOSAT (Yokota et al., 2009), and most recently the Orbiting Carbon Observatory-2 (OCO-2) launched in July 2014, which provides total column CO₂ measurements over 1.29 by 2.25 km footprints once every 16 days (Eldering et al., 2012).

Thus far, the urban surface projects have relied on a relatively small number of instruments (between 5 and 15) distributed with sensor-to-sensor distances of 5 to 35 km. Initial efforts suggest this approach may be effective at characterizing average citywide emissions over monthly to annual timescales (McKain et al., 2012), however it has yet to be used to identify and quantify specific emission activities at neighborhood scales. To resolve individual emission sources, much finer spatial resolution is needed. Simple Gaussian dispersion models with total reflection at the surface predict >95% of the one-dimensional footprint of a sensor 10 m above ground level to be within 1.1 km of the sensor under typical conditions (Seinfeld and Pandis, 2006) and prior studies (e.g., Zhu et al., 2006; Beckerman et al., 2007; Choi et al., 2014) have observed e-folding distances of ~100 to 1000 m for urban pollutant plumes mixing into the local background.

Here we propose an alternative approach that strikes a different balance between instrument quality and quantity than in previous CO₂ monitoring efforts. The BErkeley Atmospheric CO₂ Observation Network (BEACO₂N) is a large-scale network instrument that aims to leverage low-cost sensing techniques in order to enable a spatially dense network of CO₂ sensing “nodes” in and around the city of Oakland, California (Fig. 1 and 2). Using commercial CO₂ instrumentation of moderate quality and a suite of low-cost trace gas sensors for additional source attribution specificity, BEACO₂N is able to achieve an unprecedented spatial resolution of approximately 2 km—to our knowledge the only sensor network to date that monitors CO₂ on scale with the heterogeneous patterns of urban sources and sinks (see Fig. 3 for examples of observed intra-city CO₂ flux gradients). We present an initial description and characterization of the instrument, beginning with a description of the nodes, their locations, and the development of various laboratory and in situ calibration techniques. We then evaluate the network in terms ~~for-of~~ four factors—cost, reliability, precision, and ~~bias~~systematic uncertainty; described below—and conclude by demonstrating BEACO₂N’s ability to resolve CO₂ signals of significance to the urban environment.

1.1 Cost

In order to remain cost-competitive with other, less dense networks employing higher-grade instrumentation, a high-density network must utilize sensors with a price 1–2 orders of magnitude lower. However, as sensor price often scales with quality,

low-cost instrumentation may carry associated penalties in other domains, such as diminished precision, persistent bias, or the need for frequent maintenance and/or re-calibration. Thus, we seek to optimize the trade-off between cost and the other considerations.

1.2 Reliability

- 5 Network reliability consists of sensor uptime and continuity of the data stream and is crucial to enabling comparison and averaging across sites as well as improving the statistics of temporal analyses. Poor reliability also has an indirect impact on cost via the resources expended on repeat maintenance visits and/or replacement part purchases.

1.3 Precision

- 10 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 CO₂ 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 integrated over entire metropolitan regions, sensitivity to changes <10 ppm per year are required (Pacala et al., 2010). For example More specifically, according to the First Update to the Climate Change Scoping Plan, the state of California would have to reduce its overall CO₂ emissions by 4.7 million metric tons per year to achieve its goal of
- 15 reaching 1990 emission levels by 2020 (Brown et al., 2014). Assuming a fraction of that total reduction is attributable to the San Francisco Bay Area in proportion to its population (~20% of the California total), this amounts to a change of -2.6×10^6 kg CO₂ day⁻¹ for the San Francisco Bay Area. Given a residence time of air in the region of 1 day, these emissions reductions spread evenly over the 22,681 km² domain and through a 1 km boundary layer would lead to a 65 ppb annual decrease in the daily CO₂ concentrations. If the goal is verification of regional inter-annual emissions targets, wWe would therefore require
- 20 N instruments of sufficient individual sensitivity and spatial representativeness such that their , upon averaging combined their signals together, the resultant \sqrt{N} improvement in the network-wide mean sensitivity allowsallow us to detect annual changes of ~65 ppb year⁻¹ with confidence.

- However, the true strength of the high-density approach lies in the individual sensors' (or sub-group of sensors') sensitivity to intra-city phenomena, which are orders of magnitude larger by virtue of their proximity to sources not yet
- 25 diluted by advection. Larger signal sizes forgive poorer precision in the individual instruments, but demand sufficient temporal resolution to capture these anomalous, often unexpected, events of short duration on top of slowly varying domain-wide fluctuations in the background concentration. Because the BEACO₂N instrument is unique in its sensitivity to these highly local processes, we will focus on this latter specification of the instrument precision in the characterization that follows.

1.4 ~~Bias~~Systematic Uncertainty

~~Bias-Systematic uncertainties~~ can be incurred somewhat abruptly during the initial field installation (~~bias~~) or accrued more gradually ~~as drift away from the laboratory calibration~~ over time (~~drift~~). Systematic ~~bias-uncertainty~~ in the sensor readings is of particular concern in a large-scale network deployment where on-site calibration materials such as reference gases are infeasible and frequent maintenance visits are undesirable. To ensure trustworthy observations, a given network sensing approach must demonstrate some combination of: (a) instrumentation that is ~~sufficiently-reasonably~~ robust against sudden or gradual introduction of ~~bias-in-excess-of-the-instrument-precision~~~~systematic uncertainty~~, (b) a post hoc correction for ~~bias systematic uncertainty~~ in the data record, and/or (c) a procedure for identifying and replacing sensors whose ~~bias-systematic uncertainties~~ cannot be remedied via the prior methods.

10 2 Node Design, Calibration, and Deployment

Each BEACO₂N node contains a non-dispersive infrared Vaisala CarboCap GMP343 sensor for CO₂ as well as SGX Sensortech MiCS-4514 and MiCS-2614 metal oxide-based micro-sensors used to detect CO/NO₂ and O₃, respectively. Following a large-scale node refurbishment and upgrading effort in mid-2014, these core elements are now supplemented with a Sensirion SHT15 and Bosch-Sensortec BMP180 ~~sensor~~ for measuring humidity (SHT15), pressure (BMP180), and temperature (both), a Shinyei PPD42NS nephelometric particulate matter sensor, and a suite of Alphasense B4 electrochemical trace gas sensors for O₃, CO, NO, and NO₂. Discussion of these latter, air quality-related technologies will follow in a forthcoming paper.

All sensors are assembled into compact, weatherproof enclosures as seen in Fig. 4. A Raspberry Pi microprocessor automates data collection via a serial-to-USB converter (for CO₂, every ~2 seconds) and an Arduino Leonardo microcontroller (for everything else, every ~10 seconds), then transmits data to a central server using either: (a) a direct on-site Ethernet connection, (b) a Ubiquiti NanoStation locoM2 Wi-Fi antenna, or (c) an Adafruit FONA miniGSM cellular module. The latter has the unintended consequence of introducing a significant amount of electrical noise into the system. We reduce the impact of this noise by limiting data transmission to two hours per day, on a rotating schedule such that no periods are disproportionately afflicted by elevated noise levels. Battery-powered real time clock modules are also included to ensure timestamp accuracy during planned and unexpected hiatuses in internet connectivity.

Airflow through the node is maintained by two 30 mm fans, one positioned in the “intake” orientation and the other in the “outflow” orientation. An additional, passive air outlet is located adjacent to the AC/DC power supply converter to prevent excessive heating inside the node. Node enclosures measure 90 mm by 160 mm by 360 mm and are made of corrosion-resistant die cast aluminium that minimizes meteorological and magnetic complications. Stainless steel fasteners and a weatherproof seal prevent water intrusion into the enclosure.

Laboratory calibrations are performed on each CarboCap sensor upon receipt of the instrument from the supplier and repeated whenever nodes are retrieved from the field for maintenance, resulting in a re-calibration every 12–18 months.

Reference cylinders of 0 ppm, 1000 ppm, and either 320 ppm or 370 ppm CO_2 ($\pm 1\%$) are used for ~ 10 minute deliveries of each concentration to a chamber containing the sensor, which includes a built-in microprocessor that accepts the results of this multi-point calibration as input and automatically applies the appropriate corrections to the subsequent observations. The CarboCap microprocessor can also be configured to correct for the effects of oxygen, temperature, pressure, and humidity.

- 5 The built-in oxygen compensation is utilized at a constant value of 20.95%, while the latter three compensations are turned off prior to sensor deployment. Instead, a post hoc correction is derived from the ideal gas law and Dalton's law of partial pressures:

$$[\text{CO}_2]_{\text{dry}} = [\text{CO}_2]_{\text{raw}} * \frac{1013.25 \text{ hPa}}{P_{\text{tot}}} * \frac{T}{298.15 \text{ K}} * \frac{1}{1 - (\frac{P_{\text{H}_2\text{O}}}{P_{\text{tot}}})} \quad (1)$$

- Here $[\text{CO}_2]_{\text{dry}}$ is the dry air mole fraction, or the amount of CO_2 that would be measured if the observed air parcel was dried and brought to standard temperature and pressure. $[\text{CO}_2]_{\text{raw}}$, T , P_{tot} , and $P_{\text{H}_2\text{O}}$ are, respectively, the raw CO_2 concentration output by the CarboCap software in ppm, the temperature measured by the internal thermometer of the CarboCap in K, the atmospheric pressure in hPa, and the partial pressure of water in hPa, derived from the dew point temperature (T_{dew} , in $^\circ\text{C}$) using the August-Roche-Magnus approximation of the Clausius-Clapeyron relation as indicated below:

$$P_{\text{H}_2\text{O}} = 6.1094 \text{ hPa} * \exp \left(\frac{17.625 T_{\text{dew}}}{243.04 + T_{\text{dew}}} \right) \quad (2)$$

- 15 For post-2014 observations, we use the pressure and dew point temperature measured inside each node enclosure by the aforementioned BMP180 and SHT15 sensors, respectively. For data collected prior to 2014, Eq. (1) and (2) are calculated from the average sea level pressures (adjusted for altitude) and dew point temperatures measured within ~ 50 km of the BEACO₂N domain by weather stations in the NOAA Integrated Surface Database (<https://www.ncdc.noaa.gov/isd/>).

- Figure 5 compares 1 minute mean CO_2 dry air mole fractions calculated as described above with readings from a custom cavity ring-down reference instrument based on the Picarro G2301 analyzer system co-located with an in-field CarboCap over the course of 2 weeks in January 2016. The ratios between the CarboCap and Picarro observations are then shown in Fig. 6 as a function of temperature, total pressure, and the partial pressure of water. Although most of the impact of these environmental variables is removed by the ideal gas law-based correction in Eq. (1), slight dependencies on each variable remain, likely due to their influence on the vibrational spectra of CO_2 via pressure broadening, etc. Performing similar analyses on observations from in situ co-locations with other reference instruments (see LI-COR LI-820 in Sect. 3.4) reveals the temperature and water dependence to vary in sign and magnitude between individual sensors, while the pressure dependence is found to be quite robust. We therefore apply the following empirical correction to all CO_2 observations with coincident, on-site pressure measurements (i.e. post-2014 data sets):

$$[\text{CO}_2]_{\text{corrected}} = [\text{CO}_2]_{\text{dry}} * (-0.00055 P_{\text{tot}} + 1.5) \quad (3)$$

- 30 The effect of this correction is shown in the histogram of CarboCap–Picarro differences in Fig. 5 (gray bars). The offset between the two instruments is reduced to from -1 ppm to ~ 0 ppm and 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 ~~The~~ lingering temperature and water biases with any unknown factors is ± 0.4 ppm~~are $< 1\%$~~ .

Calibrated nodes are installed on trailers and buildings ~~2– to~~ 111 m above ground level (~~6– to~~ 476 m above sea level), mounted to existing infrastructure or weighted industrial tripods. Rooftop position and intake orientation are chosen to optimize wireless connectivity (if applicable), maximize air exchange with the surrounding area, and minimize sampling of extremely local emission sources (e.g., rooftop ventilation ducts). BEACO₂N nodes are sited on an approximately 2 km square grid across the Oakland metropolitan area (see Fig. 1 and 2 and Table 1), often on top of schools and museums, which possess roughly the desired spatial density and also assist the service of the educational and outreach goals of the project (see <http://beacon.berkeley.edu>). The 2 km spacing is chosen to ensure an approximately 1 km proximity to any significant CO₂ source or sink in the metropolitan area, maximizing coverage without undue overlap between neighboring footprints. Additional sites outside the 2 km grid are also included for sensitivity to potential emission sources of interest, for co-location with useful reference instruments, or as pilots for network expansion.

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 BEACO₂N as a viable platform for investigating such hypotheses.

3 Node Performance

3.1 Cost

The Vaisala CarboCap GMP343 CO₂ sensor in this study is used in its 0 to 1000 ppm measurement range and “diffusion sampling” mode, such that representative air samples passively diffuse into the path of the infrared light beam. With these specifications, each CarboCap costs approximately \$2,800 USD. Although less expensive technologies are available, the CarboCap design has a clear advantage in that the unit contains a digitally ~~-~~controlled Fabry-Perot interferometer to switch on (4.26 μ m) and off (3.9 μ m) of the asymmetric stretching mode of CO₂, generating a baseline intensity measurement for each observation that compensates for variability in the light source.

Additional sensors, ancillary hardware, and labor then bring the total cost per node to ~\$5,500 USD, or \$154,000 USD for the entire 28 node BEACO₂N instrument. For comparison, a single commercial cavity ring-down analyzer is priced

around \$60,000 USD and the annual operating total equipment cost can exceed \$85,000 USD after accounting for pumps, data loggers, etc.

3.2 Reliability

Table 2 gives the percent uptime for nine representative BEACO₂N nodes over the course of 2013, calculated as the fraction of total minutes in the year during which a given node collected valid data. All nine nodes exhibit uptimes in excess of 50% via either hardwired Ethernet connections or Wi-Fi antennas, with six collecting data >80% of the time. Maintenance visits to these sites beginning in mid-2014 revealed little to no incidence of hardware failure. Instead, external issues, such as interruptions in the electricity or Wi-Fi connectivity, are found to be the limiting factors in determining sensor uptime. Transplanting nodes to sites with more dependable electricity supplies and increasing implementation of cellular modules (which are insensitive to interruptions in on-site Wi-Fi networks) continue to enhance network reliability over time. For example, the nine most reliable nodes during the January 2015–April 2016 period all exhibit uptimes >80%, with five collecting data and transmitting it within the next 48 hours ~100% of the time via either Ethernet or cellular data communication.

3.3 Precision

From a qualitative perspective, the Vaisala CarboCap GMP343 demonstrates exceptional sensitivity to CO₂ enhancements on scales typical of an urban environment. Figure 7 compares the 1 minute mean CO₂ dry air mole fractions measured at two nearby in-field BEACO₂N nodes (EXB and EXE in Fig. 1) during 1 week in early October 2015. As these sensors are not precisely co-located (one is stationed approximately 13 m above roadside in downtown San Francisco, while the other sits ~200 m back from the road, near the bay), an exact correlation is not expected. The two sensors nonetheless demonstrate remarkable agreement; while typical diurnal CO₂ variations ~~resulting from boundary layer height fluctuations and wind speed changes~~ during the same period are on the order of 20–60 ppm, the CarboCaps simultaneously detect CO₂ events as small as 8 ppm, providing preliminary evidence of the suitability of these sensors for high-density urban deployment.

More quantitatively, Vaisala advertises the CarboCap as possessing a response time of 75 seconds and a precision of ±3 ppm at 2 second measurement frequency. Here we present our own characterization of the sensors' precision via comparison to: (a) in-laboratory reference gases and (b) a co-located in situ reference instrument.

After exposing an ensemble of CarboCaps to a constant stream of reference gas, we find the 1 minute mean dry air mole fractions to exhibit 1σ precision between ±1.2 and ±2.0 ppm, roughly in keeping with the ±2 ppm precision observed by Rigby et al. (2008). Figure 5 shows the results from our aforementioned co-location with a Picarro G2301 reference instrument, demonstrating near perfect correlation ($R^2 = 0.9999$), slope $\cong 1$, and an offset of approximately 0 ppm after meteorological corrections. In this case the 1σ precision of the 1 minute averages is ±1.4 ppm, given ~~directly~~ by the standard deviation of the differences between the minute-averaged CarboCap and Picarro observations and, as the Picarro's precision (±0.1 ppm at 5 second measurement frequency) ~~contributes negligibly to the noise~~. This presents a slight improvement over

the ± 2.18 ppm in situ precision recorded by van Leeuwen et al. (2010), although still greater variability than would be expected given the manufacturer's 2 second specifications and a 1 minute averaging time ($3 \text{ ppm}/\sqrt{30} = 0.55 \text{ ppm}$). Nonetheless, the agreement between the time series of the Picarro and CarboCap measurements demonstrates this noise level to be effectively negligible on the scale of ambient urban CO_2 fluctuations.

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. Furthermore, we see that averaging the observations from 28 CarboCaps with ± 1.4 ppm precision easily produces the ~ 65 ppb sensitivity required to detect subtler network-wide fluctuations occurring across multiple years (see Sect. 1.3).

3.4 Bias Systematic Uncertainty

Given the limited access to validation and calibration infrastructure, a major concern for a long-term field deployment is **bias systematic uncertainty** resulting from a combination of gradual temporal drift ($UB_{temporal}$, in ppm day^{-1}) and constant **biases or** offsets from the “true” value ($UB_{atemporal}$, in ppm), perhaps incurred abruptly upon installation. The measurement at a given site ($[CO_2]_{node}$, in ppm) is therefore the sum of the real regional and local influences at said site ($[CO_2]_{background}$ and $[CO_2]_{local}$, respectively), as well as these **spurious biases systematic uncertainties**:

$$[CO_2]_{node} = [CO_2]_{background} + [CO_2]_{local} + UB_{atemporal} + (UB_{temporal} \times \text{days}) \quad (4)$$

To derive post hoc corrections for $UB_{atemporal}$ and $UB_{temporal}$ at a given site, we first remove the $[CO_2]_{background}$ signal from the data record by subtracting off the weekly minimum CO_2 concentrations recorded at a reference site within the network domain. BEACO₂N'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.

Once the $[CO_2]_{background}$ term is removed, effectively de-seasonalizing the observations, we re-calculate the weekly minima of this ~~new de-seasonalized~~ data record and fit the result as a (piecewise, if necessary) linear function of time, the slope of which gives the value of $UB_{temporal}$. This linear fit is then itself subtracted from the de-seasonalized data record, yielding a remainder comprised of only the $[CO_2]_{local}$ and $UB_{atemporal}$ terms. While the $[CO_2]_{local}$ component varies rapidly over several orders of magnitude, the contribution of $UB_{atemporal}$ is, by definition, constant in time, so we once again compute the weekly minima of the new data record and define the mean weekly minimum as $UB_{atemporal}$. Having obtained values for $UB_{temporal}$ and $UB_{atemporal}$, we simply subtract these components from the original data record to generate the unbiased observations at each site.

Table 3 gives the results from one iteration of the ~~bias~~-correction procedure outlined above executed using the ELC BEACO₂N node (see Fig. 1) as the reference site needed to calculate $[CO_2]_{background}$. Only sites that enable at least 3 months of comparison to the ELC node are included; multiple values at a single site correspond to the piecewise linear fits employed when $UB_{temporal}$ exhibits discontinuities over the data record. Because we universally define Day 1 to be 1 January 2013 and $UB_{atemporal}$ is strongly influenced by the intercept of the linear fit used to characterize the temporal drift, it should be noted that the magnitude of $UB_{atemporal}$ does not necessarily represent the actual bias present at a node on its deployment date (which may be before or after 1 January 2013), but rather an extrapolation of this initial bias forwards or backwards in time. Uncertainties in $U_{temporal}$ and $U_{atemporal}$ 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.

To evaluate the efficacy of this procedure, we compare the weekly minima of both the raw and ~~bias~~-corrected data records to the weekly minimum CO₂ concentrations measured by ~~the aforementioned~~ LI-COR LI-820 ~~non-dispersive infrared CO₂ gas analyzer positioned at sea level between the EXB and EXE nodes (see Fig. 1). The LI-COR is maintained by NOAA's Pacific Marine Environmental Laboratory and calibrated against compressed gas (400–500 ppm CO₂) prior to every hourly measurement and is assumed to have negligible bias.~~ The results of said comparison are shown in Fig. 8, demonstrating significantly improved agreement (3.7 vs. 9.8 ppm mean residuals) with the LI-COR weekly minima after ~~bias~~-correction. This is likely a conservative estimate of the ~~bias~~-uncertainty reduction achievable with this method, as the ELC node we use to compute our $[CO_2]_{background}$ value is not itself an ~~uncertainty~~-~~bias~~-free reference. Although the raw ELC data record demonstrates the least systematic bias-uncertainty of all the BEACO₂N nodes in an initial comparison with the LI-COR, its observations are nonetheless afflicted by some unknown nonzero drift and/or atemporal ~~offset~~bias. Direct in situ calibration of the reference instrument would allow us to constrain systematic biases-uncertainties even further, ~~possibly to within the precision of the sensors.~~

3.5 Performance of Ancillary Sensor Technology

According to manufacturer documentation, the Sensirion SHT15 provides relative humidity measurements to 0.05% resolution, with an advertised accuracy of $\pm 2.0\%$, a repeatability of $\pm 0.1\%$, an 8 second response time, and a long-term drift of $<0.5\%$ per year. Its temperature measurements are provided to 0.01 °C resolution, with an advertised accuracy of ± 0.3 °C, a precision of ± 0.1 °C, a response time of 5 to 30 seconds, and a long-term drift of <0.04 °C per year. The Bosch-Sensortec BMP180 provides pressure measurements to 0.01 hPa resolution, with an advertised accuracy of ± 0.12 hPa, a precision of ± 0.05 hPa, and a long-term drift of ± 1.0 hPa per year. Its temperature measurements are provided to 0.1 °C resolution, with an advertised accuracy of ± 1.0 °C. An independent verification of these performance specifications is not attempted here. However, the temperature observations from both sensors closely follow the structure of that detected by the internal temperature sensor of the CarboCap, although the CarboCap's readings are consistently slightly elevated, as expected given the heat produced by the instrument itself.

The BMP180 and SHT15 are not intended to reflect local meteorological conditions, but rather to provide a representative picture of conditions inside the node. These internal conditions are integral to various posterior corrections applied to the observations (see Sect. 2).

4 Initial Field Results

The BEACO₂N field campaign is a long-term, ongoing monitoring effort. Here we provide a time series of data collected from 167 BEACO₂N sites between January 2013 and April 2016 (Fig. 9) and some initial descriptive statistics of the bias-corrected dry air CO₂ mole fractions at nine representative sites in 2013 (Table 2).

Figure 9 demonstrates the volume and diversity of extreme short-term variability in urban CO₂ 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, with maximum values between 500 and 820 ppm and minima between 384 and 396 ppm. Standard deviations are seen to range from 9.57 to 22.4 ppm, all of which are lower than the corresponding nighttime (2200–0400 LT) standard deviations due to the reduced convective mixing in the shallow nocturnal boundary layer. Similarly, the majority of nighttime means and maxima exceed the daytime values at the same location, with the exception of three sites: FTK, LAU, and KAI. The dampened or inverted diurnal trends at these sites may be due to unique boundary layer dynamics at those locales or unusually large daytime CO₂ sources nearby. Daytime and nighttime minima do not differ as significantly.

Individual BEACO₂N nodes are observed to capture a number of patterns and cycles typical of ambient CO₂ monitoring. Figure 10 shows the monthly minimum CO₂ concentrations at six select sites in 2013, as the difference from their July value (defined as 0 ppm at each site). A distinct seasonal cycle is observed, with wintertime minima exceeding summertime values by 7 to 24 ppm. For reference, the gray curve presents a similar treatment of the aforementioned Pacific boundary curtain.

~~smoothed, three-dimensional “curtain” of surface CO₂ Pacific boundary conditions produced by NOAA’s Global Greenhouse Gas Reference Network (Jeong et al., 2013). At many sites in the summertime, the BEACO₂N minima are seen to converge 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. the degree of variability within the network increases during the winter months.~~

Figure 11 shows representative diurnal cycles in the bias-corrected CO₂ dry air mole fractions at three different BEACO₂N nodes in September 2013. We observe elevated concentrations at night corresponding to a shallow nocturnal boundary layer, significant enhancements around the morning rush hour when emissions are increasing faster than boundary layer height, and midday minima reflecting mixing into the largest volume of air before the boundary layer collapses again at sunset. However, within this qualitatively well understood average behavior remains a degree of intra-network variability that allows us to discover and probe local scale phenomena of unknown origin. At FTK, for example, concentrations are seen to decrease after an initial rush hour peak (~0800 LT) but remain somewhat elevated until sunset, never achieving the much more pronounced afternoon minimum observed at PAP, 13.5 km away.

Such intra-city heterogeneities are difficult to capture accurately using atmospheric transport models alone. We simulate hourly CO₂ concentrations (\hat{y}) at each site in the network using the Stochastic Time-Inverted Lagrangian Transport model (STILT; Lin et al., 2003) coupled to the Weather Research and Forecasting model (WRF; Skamarock et al., 2008). The coupled model is known as “WRF-STILT” (Nehrkorn et al. 2010) and the setup used here follows that of Turner et al. (2016; see their Sect. S1 for details of the WRF setup). WRF-STILT advects an ensemble of 500 particles 3 days backwards in time, each with a small random perturbation, from the spatio-temporal locations of the BEACO₂N observations using the meteorological fields from WRF. The trajectories of these 500 particles are then used to construct “footprints” for each observation that represent the sensitivity of the observation to a perturbation in emissions from a given location. The footprints can be represented in matrix form (H) and multiplied by a set of gridded emissions (x , from the high-resolution bottom-up CO₂ inventory in Turner et al. 2016) to compute the CO₂ enhancement at each site due to local emissions:

$$\Delta y = Hx \quad (5)$$

We then add this local enhancement to a background concentration (y_B , from the aforementioned Pacific boundary curtain) to obtain a model estimate of the BEACO₂N observations shown as black squares in Fig. 11:

$$\hat{y} = \Delta y + y_B = Hx + y_B \quad (6)$$

~~We calculate the spatial footprint of each site using a Stochastic Time Inverted Lagrangian Transport Model (Lin et al., 2003) coupled to the Weather Research and Forecasting model (Skamarock et al., 2008) (WRF-STILT) in the manner of Nehrkorn et al. (2010) and overlay it on a high-resolution bottom-up emissions inventory (see Fig. 3) as in Turner et al. (2016) to generate the predicted diurnal cycles shown as black squares in Fig. 11. While the model captures midday conditions at FTK and evening levels at PAP, the presence of both over- and under-estimations at other times suggests a need to re-examine the bottom-up emissions inventory as well as the model’s treatment of boundary layer dynamics. BEACO₂N provides the ground truth necessary to identify such deficiencies and potentially improve upon them via inverse~~

modeling, data assimilation, etc.

Comparison of diurnal cycles during noteworthy local scale emission events with averages such as those seen in Fig. 11 gives further insight into the potential application of BEACO₂N observations to CO₂ source attribution. Figure 12 offers one such comparison using hourly averages collected from a BEACO₂N node positioned on top of Oakland High School (OHS in Fig. 1 and Table 1) during a weather-related school closure that occurred on 11 December 2014. Clear reductions in CO₂ concentrations are observed relative to what is typical at this site (and indeed network-wide, although to a lesser extent), as is expected in the absence of emissions related to students' commutes and presence on campus. The sensing technology implemented in the BEACO₂N nodes therefore proves adequate to resolve not only CO₂ patterns typical of an urban environment, but also short-term deviations during anomalous emission events, positioning BEACO₂N as an essential tool for the characterization of current urban conditions as well as the verification of subsequent emissions reductions.

5 Discussion & Conclusion

We have described the design, implementation, and initial observations from a novel high-resolution CO₂ surface monitoring network instrument. We demonstrate that low-cost instrumentation enables an unprecedented level of spatial density, and describe a calibration protocol with post hoc bias corrections that permits the network to operate precisely and reliably enough to characterize variations in ambient concentrations with magnitudes relevant to metropolitan life.

Our preliminary analysis of the first ~3 years of CO₂ observations provides evidence of the expected diurnal and seasonal cycles as well as an encouraging sensitivity to short-term changes in local emission events. Furthermore, we show significant qualitative and quantitative differences among the diurnal cycles at individual nodes on spatial scales that cannot yet be accurately captured by atmospheric transport models, confirming the necessity of a high-density approach when attempting to faithfully represent the variability of a complex urban environment.

Although BEACO₂N 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, Turner et al. (2016) constructed and applied a WRF-STILT inverse model to synthetic observations with density similar to BEACO₂N. For an area source the size of the Oakland metropolitan area, emissions were estimated to within 18% accuracy; for a freeway-sized line source to within 36%; and to within 60% for the sum of six industrial point sources—consistently outperforming a smaller hypothetical network (three sites) with significantly better precision. Using week-long averages, the BEACO₂N-like network was able to further reduce the uncertainty in the integrated urban area source to <2%, a significant improvement over the citywide emissions estimates provided by real and proposed ~10 site sensor networks described by Lauvaux et al. (2016) (25% uncertainty in 5 day averages), Kort et al. (2013) (>10% uncertainty in monthly averages), and Wu et al. (2016) (110% uncertainty in annual monthly averages). These other studies use more conservative estimates of the combined instrument, model, and

representativeness error (≥ 3 ppm, as opposed to ± 1 ppm). These combined error budgets are typically dominated by transport (model) error, which potentially explains why models based on BEACO₂N-like networks perform comparably to or better than those based on sparser networks of higher quality sensors, for which instrument error may be reduced but accurately representing transport between observation sites is of greater importance. Further work is needed to fully assess the efficacy of inverse methods based on the BEACO₂N approach.

In addition, further characterization of the trace gas and particulate matter sensors will allow for more specific source attribution via the exploitation of emissions factors unique to various combustion activities (e.g., Ban-Weiss et al., 2008; Harley et al., 2015), while providing public health-relevant air quality information as well. There is also potential for fine-grained verification of space-based observations or even of personal sensors when their inherent mobility brings them within the geographical area well represented by the fixed BEACO₂N network.

This work constitutes a promising initial infrastructure upon which further advances in high-density atmospheric monitoring can be built, capable of providing research, regulatory, and layperson communities with greenhouse gas and air toxics information on the scale at which emissions and personal exposure actually occur. We are currently planning to expand this validated pilot network into the neighboring locales of San Francisco and Richmond, California, allowing us to characterize other emissions sources, such as oil refining facilities. These efforts will be complemented by modeling studies comparing different sampling resolutions (i.e. 2 km vs. 4 km sensor spacing) and spatial configurations, yielding general network optimization principles that will facilitate future implementations of high-resolution CO₂ monitoring networks in diverse locations.

6 Data Availability

The BEACO₂N data used in this study and all subsequently collected data are available in near real time on the BEACO₂N website: <http://beacon.berkeley.edu>.

Acknowledgements. This work was funded by the National Science Foundation (1035050; 1038191), the National Aeronautics and Aerospace Administration (NAS2-03144), and the Bay Area Air Quality Management District (2013.145). Additional support was provided by a NSF Graduate Research Fellowship to AAS, a Department of Energy Computational Science Graduate Fellowship to AJT, a Kwanjeong Lee Chonghwan Educational Fellowship to JK, and the UC Berkeley Miller Institute to RCC. We acknowledge the use of datasets maintained by NOAA's Integrated Surface Database, Pacific Marine Environmental Laboratory, and Global Greenhouse Gas Reference Network. We thank UC Berkeley Research Computing for access to computation resources, Sebastien C. Biraud for facilitating inter-comparisons with the Picarro G3201, as well as David M. Holstius and Holly L. Maness for their generous contributions to BEACO₂N's code base.

References

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

- Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., Chao, D. Y., Kitzis, D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M., Crotwell, M. J., Fischer, M. L., Parker, M. J., Lee, J. T.,
- 5 Baumann, D. D., Desai, A. R., Stanier, C. O., De Wekker, S. F. J., Wolfe, D. E., Munger, J. W., and Tans, P. P.: CO₂, CO, and CH₄ measurements from tall towers in the NOAA Earth System Research Laboratory’s Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts, *Atmos. Meas. Tech.*, 7, 647–687, doi:10.5194/amt-7-647-2014, 2014.
- Ban-Weiss, G. A., McLaughlin, J. P., Harley, R. A., Lunden, M. M., Kirchstetter, T. W., Kean, A. J., Strawa, A. W.,
- 10 Stevenson, E. D., and Kendall, G. R.: Long-term changes in emissions of nitrogen oxides and particulate matter from on-road gasoline and diesel vehicles, *Atmos. Environ.*, 42, 220–232, doi:10.1016/j.atmosenv.2007.09.049, 2008.
- Beckerman, B., Jerrett, M., Brook, J. R., Verma, D. K., Arain, M. A., and Finkelstein, M. M.: Correlation of nitrogen dioxide with other traffic pollutants near a major expressway, *Atmos. Environ.*, 42, 275–290, doi:10.1016/j.atmosenv.2007.09.042, 2007.
- 15 Burrows, J. P., Hölze, E., Goede, A. P. H., Visser, H., and Fricke, W.: SCHIAMACY–scanning imaging absorption spectrometer for atmospheric chartography, *Acta Astronaut.*, 35, 445–451, doi:10.1016/0094-5765(94)00278-T, 1995.
- Bréon, F. M., Broquet, G., Puygrenier, V., Chevallier, F., Xueref-Rémy, I., Ramonet, M., Dieudonné, E., Lopez, M., Schmidt, M., Perrussel, O., and Ciais, P.: An attempt at estimating Paris area CO₂ emissions from atmospheric concentration measurements, *Atmos. Chem. Phys.*, 15, 1707–1724, doi:10.5194/acp-15-1707-2015, 2015.
- 20 Brown Jr., E. G., Rodriguez, M., Nichols, M. D., and Corey, R. W.: First Update to the Climate Change Scoping Plan: Building on the Framework, California Air Resources Board, Sacramento, CA, 2014.
- ~~[California Global Warming Solutions Act of 2006, A.B. 32, 2006.](#)~~
- Choi, W., Winer, A. M., and Paulson, S. E.: Factors controlling pollutant plume length downwind of major roadways in nocturnal surface inversions, *Atmos. Chem. Phys.*, 14, 6925–6940, doi:10.5194/acp-14-6925-2014, 2014.
- 25 Duren, R. M. and Miller, C. E.: Measuring the carbon emissions of megacities, *Nature Clim. Change*, 2, 560–562, doi:10.1038/nclimate1629, 2012.
- Ehleringer, J., Schauer, A. J., Lai, C.-T., Bowling, D. R., Pataki, D. E., and Stephens, B. B.: Long-term carbon dioxide monitoring in Salt Lake City, AGU Fall Meeting, San Francisco, CA, 15–19 December 2008, B43D-0466, 2008.
- Eldering, A., Solish, B., Kahn, P., Boland, S., Crisp, D., and Gunson, M.: High precision atmospheric CO₂ measurements
- 30 from space: The design and implementation of OCO-2, in: Aerospace Conference, 2012 IEEE, Big Sky, MT, 3–10 March 2012, 1–10, doi:10.1109/AERO.2012.6187176, 2012.
- Harley, R. A., Marr, L. C., Lehner, J. K., and Giddings, S. N.: Changes in motor vehicle emissions on diurnal to decadal time scales and effect on atmospheric composition, *Environ. Sci. Technol.*, 39, 5356–5362, doi:10.1021/es048172+, 2005.

- 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.
- Kort, E., Angevine, W., Miller, C.E., and Duren, R.: Surface observations for monitoring megacity greenhouse gas emissions: minimum requirements for the Los Angeles Megacity: *J. Geophys. Res.–Atmos.*, 118, 1–8, doi:10.1002/jgrd.50135, 2013.
- Lauvaux, T., Miles, N. T., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., Gaudet, B., Gurney, K. R., Huang, J., O’Keefe, D., Song, Y., Karion, A., Tomohiro, O., Patarasuk, R., Razlivanov, I., Sarmiento, D., Shepson, P., Sweeney, C., Turnbull, J., and Wu, K.: High-resolution atmospheric inversion of urban CO₂ emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX), *J. Geophys. Res.*, 121, doi:10.1002/2015JD024473, 2016.
- Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *J. Geophys. Res.*, 108, ACH 2-1–ACH 2-13, doi:10.1029/2002JD003161, 2003.
- McKain, K., Wofsy, S. C., Nehrkorn, T., Eluszkiewicz, J., Ehleringer, J., and Stephens, B. B.: Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region, *P. Natl. Acad. Sci. USA.*, 109, 8423–8428, doi:0.1073/pnas.1116645109, 2012.
- Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., and Freitas, S.: Coupled weather research and forecasting–stochastic time-inverted lagrangian transport (WRF–STILT) model, *Meteorol. Atmos. Phys.*, 107, 51–64, doi:10.1007/s00703-010-0068-x, 2010.
- Pacala, S. W., Breidenich, C., Brewer, P. G., Fung, I., Gunson, M. R., Heddle, G., Law, B., Marland, G., Paustian, K., Prather, M., Randerson, J. T., Tans, P., and Wofsy, S. C.: *Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements*, The National Academies Press, Washington, D. C., 2010.
- Rigby, M., Toumi, R., Fisher, R., Lowry, D., and Nisbet, E. G.: First continuous measurements of CO₂ mixing ratio in central London using a compact diffusion probe, *Atmos. Environ.*, 42, 8943–8953, doi:10.1016/j.atmosenv.2008.06.040, 2008.
- Seinfeld, J. H. and Pandis, S. N.: *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 2nd ed., John Wiley & Sons Inc., Hoboken, NJ, 2006.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: *A Description of the Advanced Research WRF Version 3*, National Center for Atmospheric Research, Boulder, CO, 2008.
- Thompson, R. L., Manning, A. C., Gloor, E., Schultz, U., Seifert, T., Hansel, F., Jordan, A., and Heimann, M.: In-situ measurements of oxygen, carbon monoxide and greenhouse gases from Ochsenkopf tall tower in Germany, *Atmos. Meas. Tech.*, 2, 573–591, doi:10.5194/amt-2-573-2009, 2009.

- 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, I.: Toward quantification and source sector identification of fossil fuel CO₂ emissions from an urban area: Results from the INFLUX experiment, *J. Geophys. Res. Atmos.*, 120, 292–312, doi: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 CO₂ emissions: Assessing trade-offs between precision and network density, *Atmos. Chem. Phys. Discuss.*, in review, doi:10.5194/acp-2016-355, 2016.
- United Nations, Department of Economic and Social Affairs, Population Division: World Urbanization Prospects: The 2014 Revision, Highlights, 2014.
- 10 United Nations Framework Convention on Climate Change Intended Nationally Determined Contributions: http://unfccc.int/focus/indc_portal/items/8766.php/, last access: 4 June 2016.
- van Leeuwen, C. Improved operation of the Vaisala Carbocap GMP343 for the school CO₂-web. Manuscript in preparation, 2010.
- Worthy, D. E. J., Platt, A., Kessler, R., Ernst, M., Braga, R., and Racki, S.: The Canadian atmospheric carbon dioxide measurement program: measurement procedures, data quality and accuracy, in: Report of the 11th WMO/IAEA Meeting of Experts on Carbon Dioxide Concentration and Related Tracer Measurement Techniques, Tokyo, Japan, 25–28 September 2001, 112–128, 2003.
- 15 Wu, L., Broquet, G., Ciais, P., Bellassen, V., Vogel, F., Chevallier, F., Xueref-Remy, I., and Wang, Y.: Atmospheric inversion for cost effective quantification of city CO₂ emissions, *Atmos. Chem. Phys. Discuss.*, 165, ~~30693–30756~~7743–7771, doi:10.5194/acp-~~15-30693-2015~~16-7743-2016, 2016.
- 20 Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., and Maksyutov, S.: Global concentrations of CO₂ and CH₄ retrieved from GOSAT: first preliminary results, *Sola*, 5, 160–163, doi:10.2151/sola/2009-041, 2009.
- Zhu, Y., Kuhn, T., Mayo, P., and Hinds, W. C.: Comparison of daytime and nighttime concentration profiles and size distributions of ultrafine particles near a major highway, *Environ. Sci. Technol.*, 40, 2531–2536, doi:10.1021/es0516514, 25 2006.

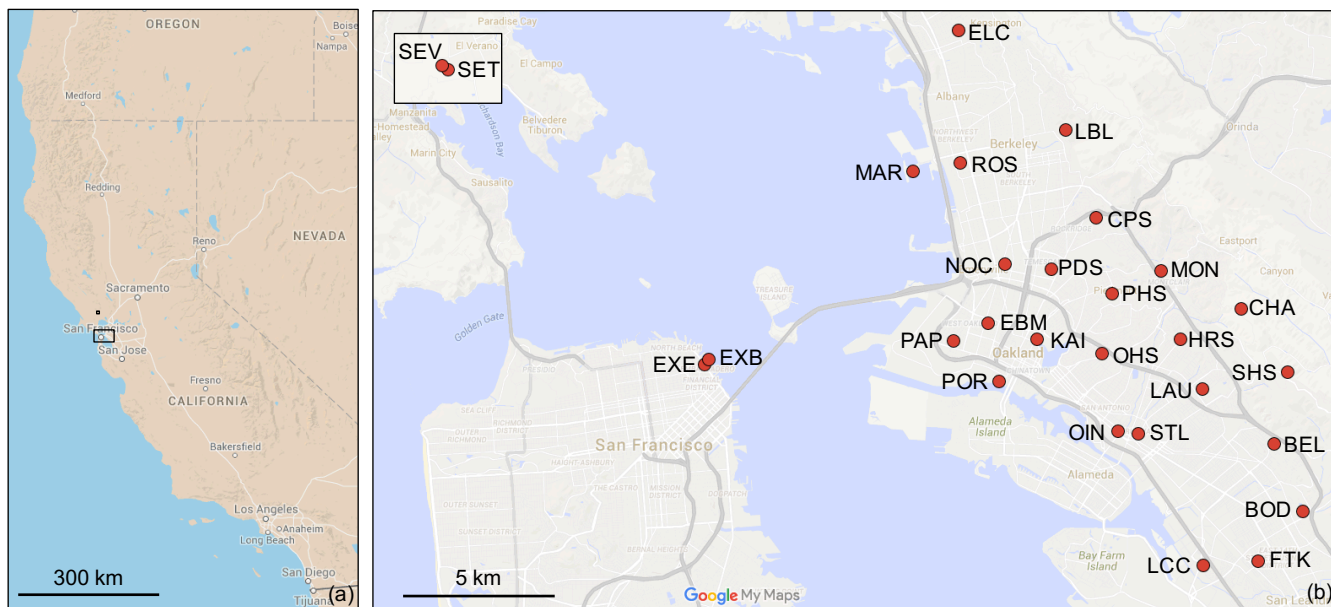


Figure 1: Map of the BEACO₂N domain (a) in the context of the western United States and (b) showing individual node locations. Inset in panel (b) shows the pair of nodes stationed in Sonoma County.

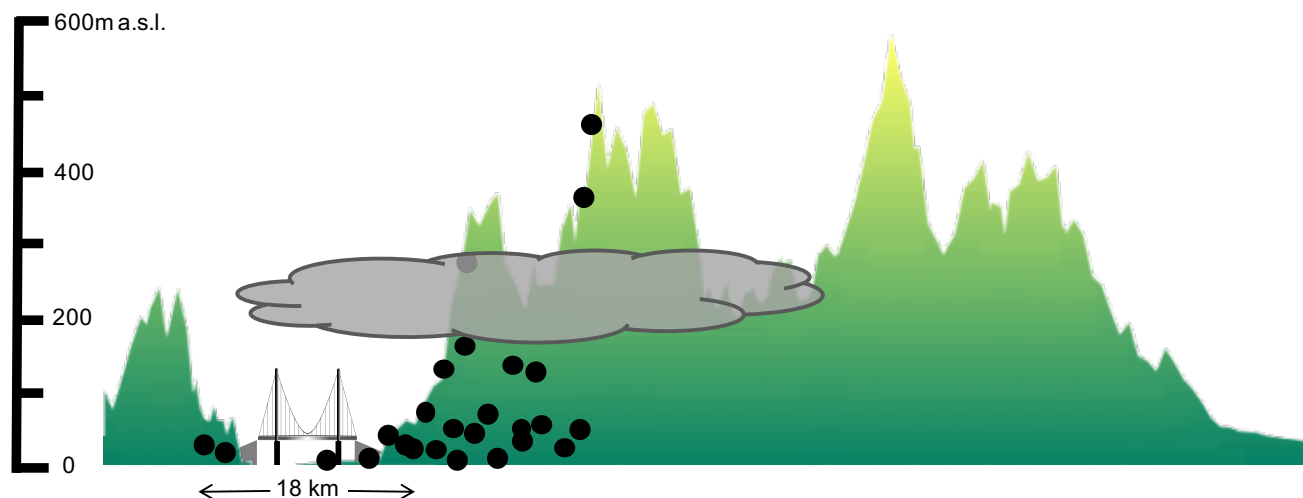


Figure 2: North-facing schematic of Fig. 1 indicating a representative vertical distribution of BEACO₂N node sites (circles) over the topography of Oakland, CA. Cloud marks the altitude and thickness of a typical marine fog layer; bridge delineates the height of the San Francisco-Oakland Bay Bridge. Horizontal placement of nodes has been skewed for visual clarity.

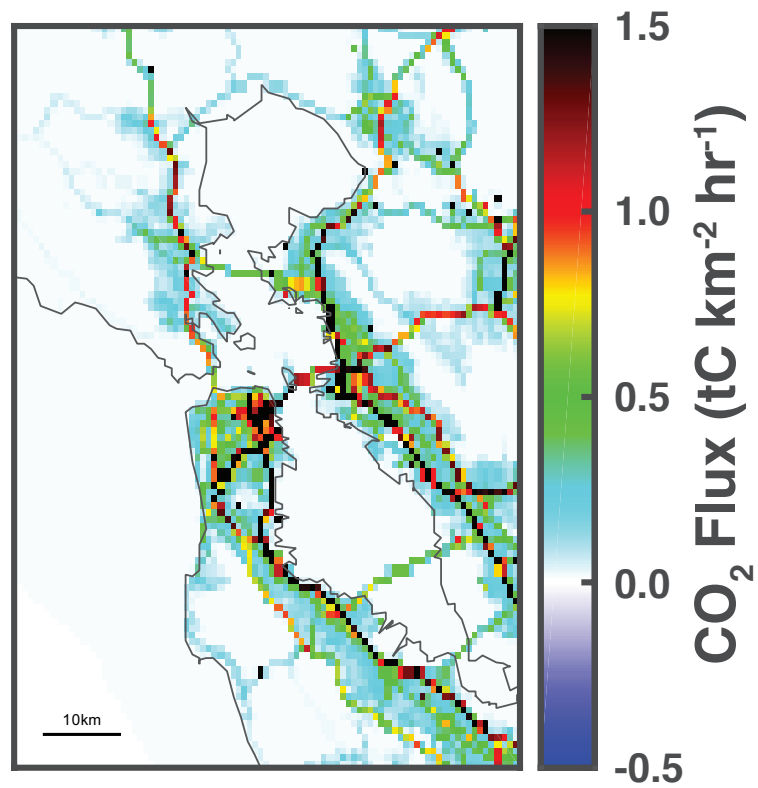


Figure 3: A sample high-resolution bottom-up emissions inventory for the Bay Area adapted from Turner et al. (2016).

5

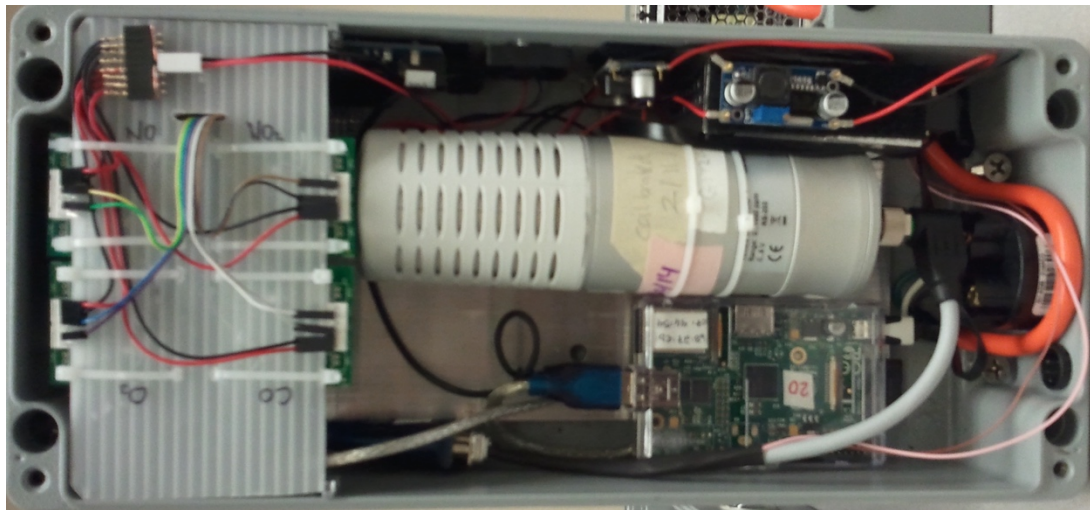


Figure 4: Current BEACO₂N node design.

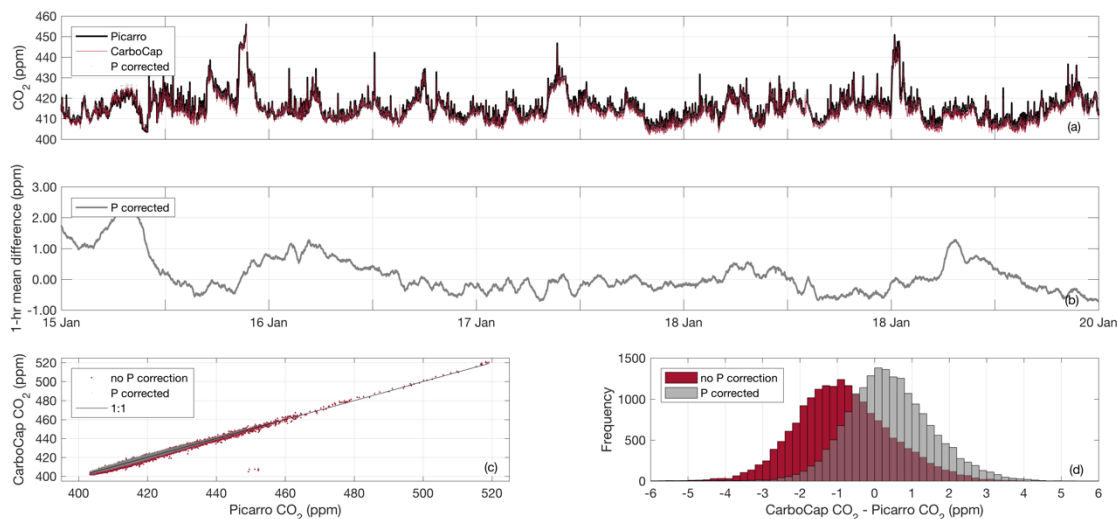


Figure 5: One minute mean results from a 2 week co-location of a Vaisala CarboCap GMP343 and a custom cavity ring-down reference instrument based on the Picarro G2301 system: (a) representative 5 day time series, (b) 1 hr running mean of the differences over the same 5 day period, (c) direct comparison, (d) histogram of the differences. CarboCap observations are dry air mole fractions calculated using Eq. (1) and subsequently pressure corrected with Eq. (3).

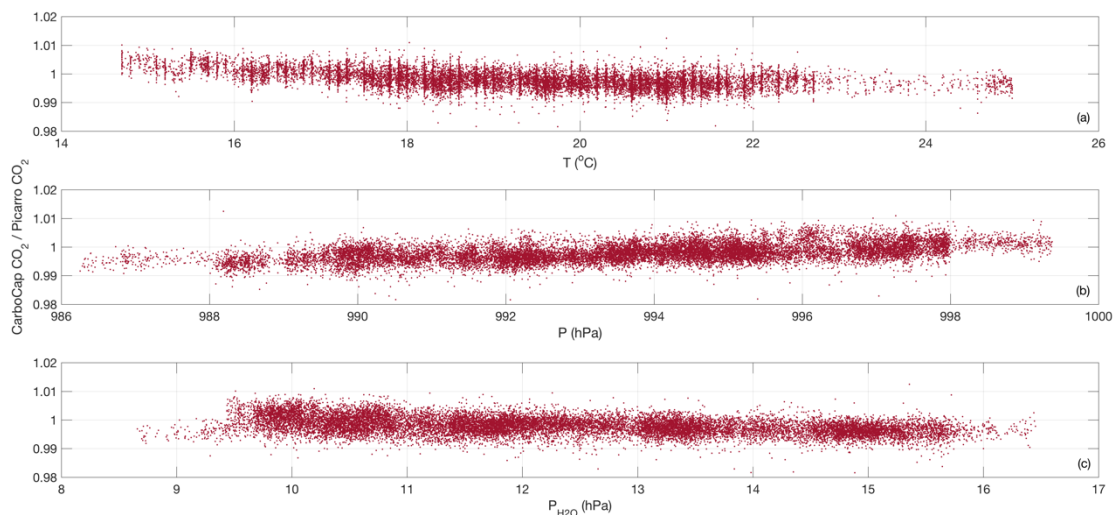


Figure 6: Ratio of 1 minute mean CO_2 dry air mole fractions presented in Fig. 5, shown as a function of temperature (a), pressure (b), and the partial pressure of water (c).

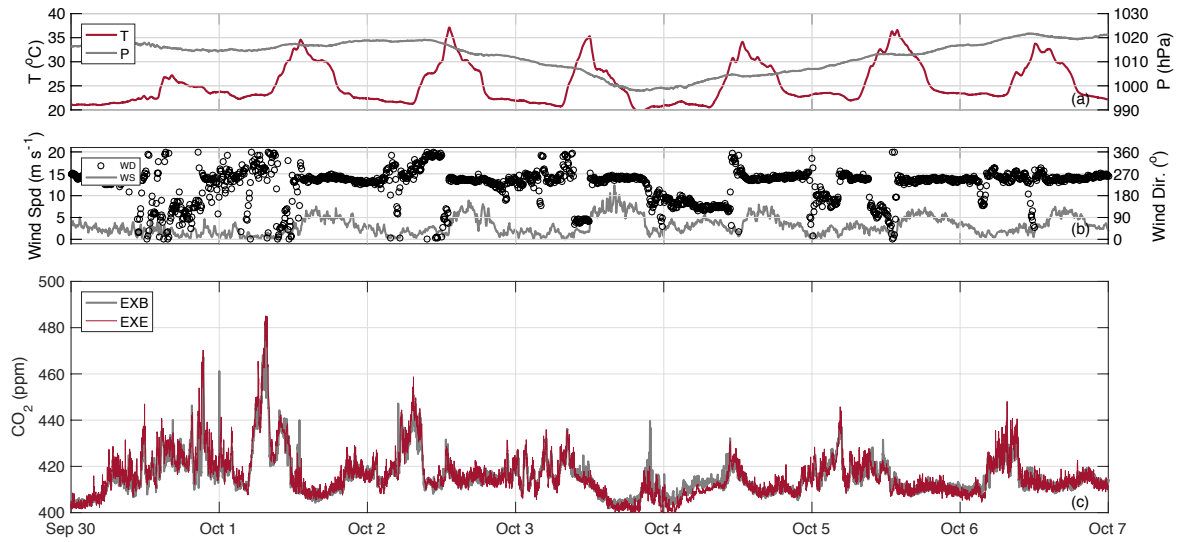


Figure 7: Representative weeklong time series of observations collected at or near two nearby in-field BEACO₂N nodes (EXB and EXE in Fig. 1; ~200m apart) in October 2015: (a) temperature and pressure averaged to 1 minute, (b) wind speed and direction collected once every 6 minutes, (c) drift- and bias-corrected CO₂ dry air mole fractions averaged to 1 minute.

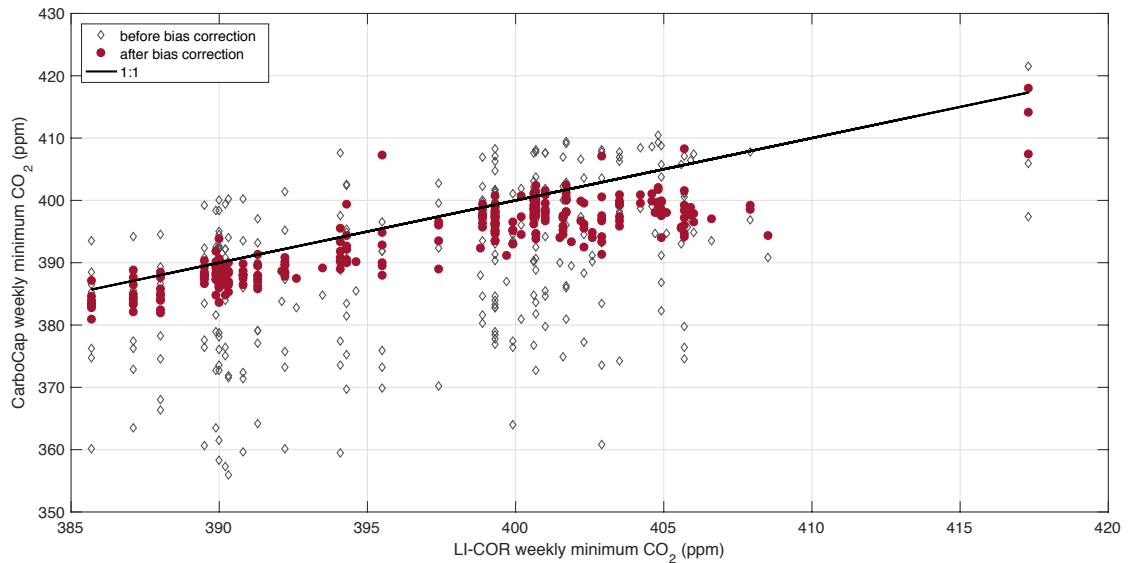


Figure 8: Weekly minimum CO₂ concentrations measured by a LI-COR LI-820 reference instrument compared with weekly minima calculated from the BEACO₂N data record before and after bias-correction for systematic uncertainties.

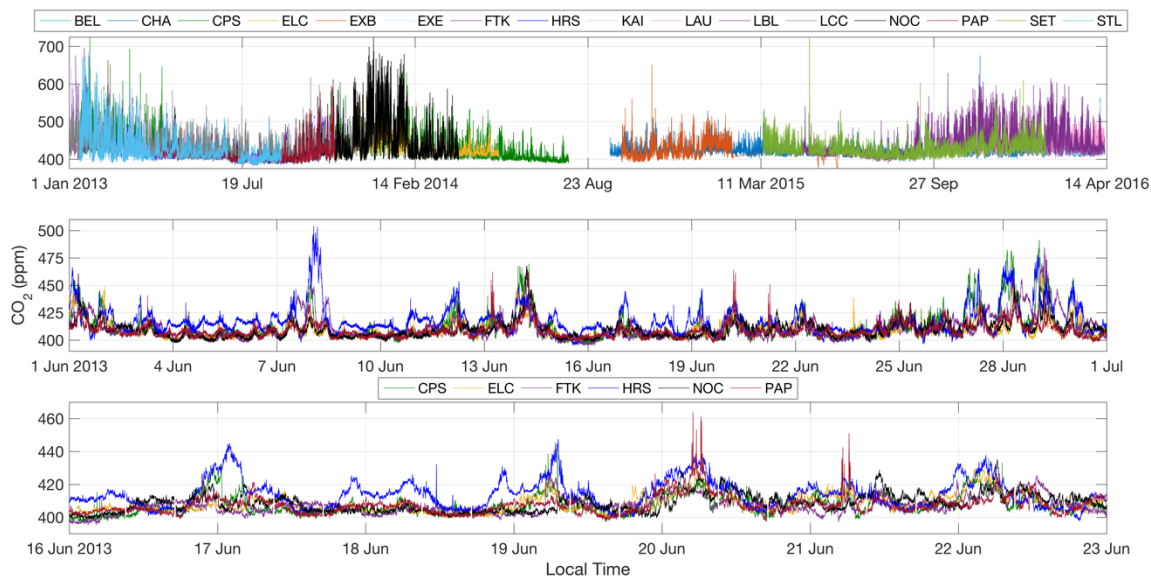


Figure 9: Time series of drift- and bias-corrected CO₂ dry air mole fractions collected over the course of ~2.5 years at 16 BEACO₂N sites (top), one month at six representative sites (middle), and one week at the same six sites (bottom). The hiatus around 23 Aug Day 600 corresponds to a largescale hardware refurbishment effort that began in mid-2014.

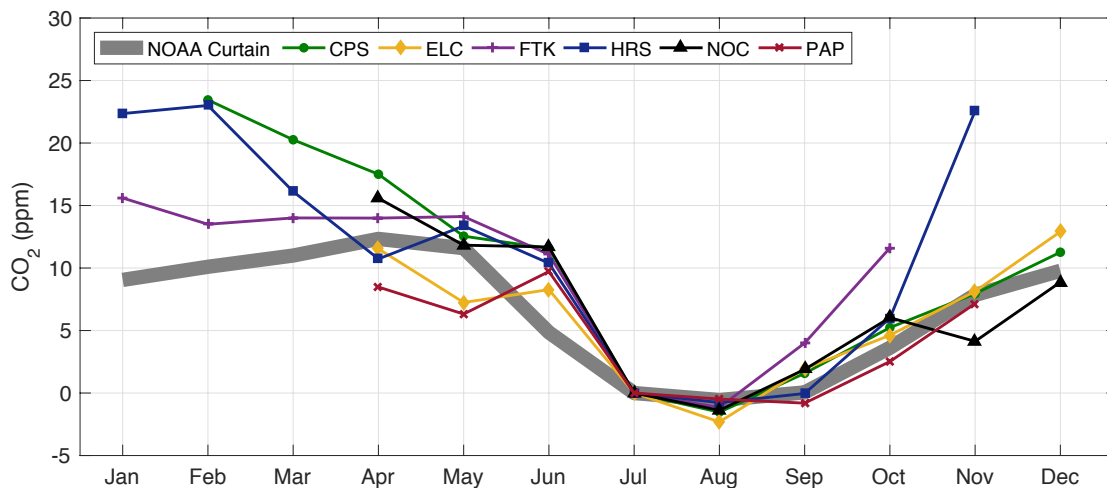


Figure 10: Monthly minimum drift- and bias-corrected CO₂ dry air mole fractions observed during 2013 at the same six representative BEACO₂N sites shown in the bottom panels of Fig. 9, plotted as the enhancement above the July value. Bold gray curve shows a similar treatment of the surface level Pacific Ocean empirical boundary curtain values for 38° N.

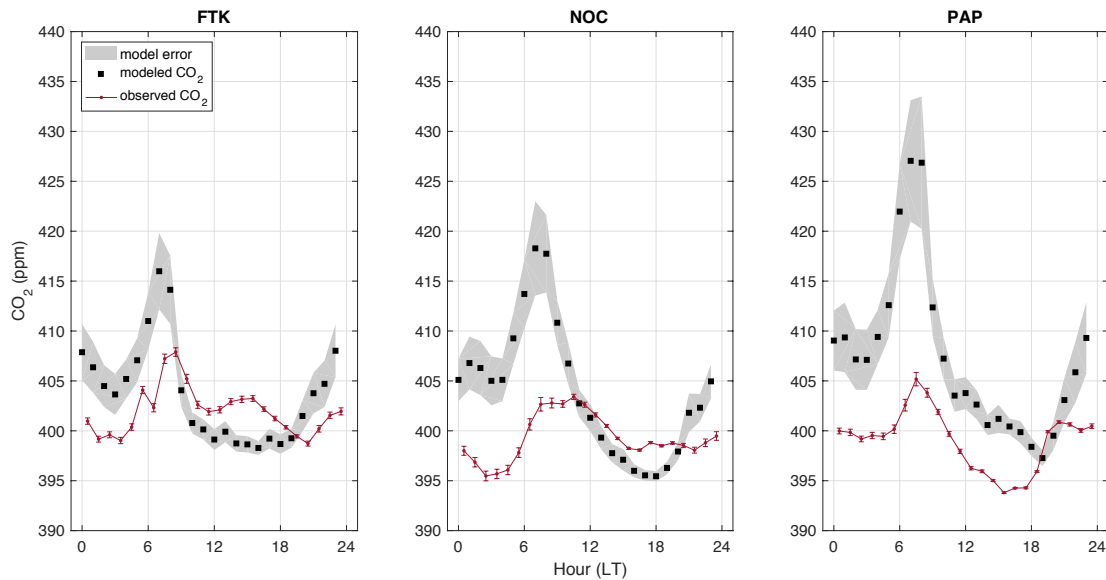


Figure 11: Diurnal variation in **drift- and bias-corrected** CO₂ dry air mole fractions observed and modeled at three representative BEACO₂N sites during September 2013. Error bars **indicate the standard error of the mean (instrument error is negligible at this timescale); and thick shaded curves indicate standard error deviation.**

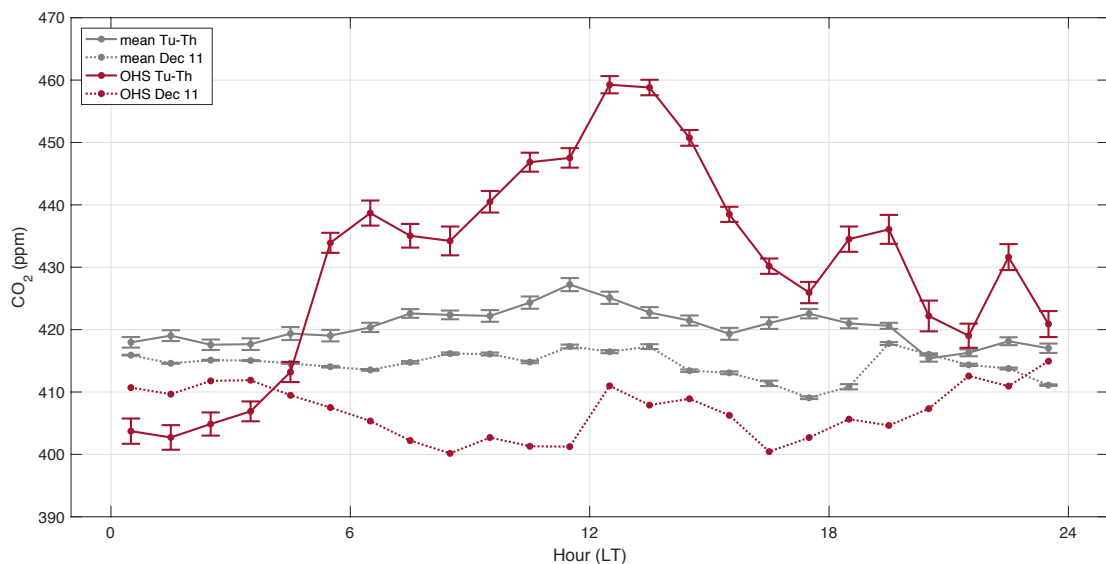


Figure 12. Comparison of diurnal variation in **drift- and bias-corrected** CO₂ dry air mole fractions observed at Oakland High School (OHS in Fig. 1) during a rain-related school closure on 11 December 2014 vs. the mean variation observed on other Tuesdays, Wednesdays, and Thursdays during December 2014 when the school was operating normally. Mean values from five other BEACO₂N sites operational during these time periods are also shown for reference. Error bars indicate standard error **(instrument error is negligible at this timescale).**

CODE	FULL SITE NAME	LAT	LON	ELEV (m asl)	ELEV (m agl)
BEL	Burckhalter Elementary School	37.775	-122.167	97	8
BOD	Bishop O'Dowd High School	37.753	-122.155	82	8
CHA	Chabot Space & Science Center	37.819	-122.181	476	11
CPS	College Preparatory School	37.849	-122.241	102	4
EBM	W. Oakland EBMUD Monitoring Stn.	37.814	-122.282	6	2
ELC	El Cerrito High School	37.907	-122.294	49	13
EXB	Exploratorium (Bay)	37.802	-122.397	13	9
EXE	Exploratorium (Embarcadero)	37.801	-122.399	13	5
FTK	Fred T. Korematsu Discovery Acad.	37.737	-122.173	16	6
HRS	Head Royce School	37.809	-122.204	114	5
ICS	International Community School	37.779	-122.231	19	6
KAI	Kaiser Center	37.809	-122.264	115	111
LAU	Laurel Elementary School	37.792	-122.196	74	6
LBL	Lawrence Berkeley Nat'l Lab, Bldg. 70	37.876	-122.252	246	11
LCC	Lighthouse Community Charter School	37.736	-122.196	9	5
MAR	Berkeley Marina	37.863	-122.314	6	2
MON	Montclair Elementary School	37.830	-122.211	193	4
NOC	N. Oakland Community Charter School	37.833	-122.277	24	6
OHS	Oakland High School	37.805	-122.236	49	7
PAP	PLACE at Prescott Elementary	37.809	-122.298	12	6
PDS	Park Day School	37.832	-122.257	39	7
PHS	Piedmont Middle & High School	37.824	-122.233	86	10
POR	Port of Oakland Headquarters	37.796	-122.279	35	32
ROS	Rosa Parks Elementary School	37.865	-122.295	22	10
SET	Stone Edge Farms (near turbine)	38.289	-122.503	54	2
SEV	Stone Edge Farms (in vineyard)	38.291	-122.506	61	3
SHS	Skyline High School	37.798	-122.161	359	3
STL	St. Elizabeth High School	37.779	-122.222	28	11

Table 1: List of site names, abbreviated codes, geo-coordinates, and elevations.

SITE CODE	UPTIME (%)	MEAN (ppm)	STD (ppm)	MAX (ppm)	MIN (ppm)
CPS	94.6	416	21.6	589	385
		423	24.3	730	384
ELC	90.1	411	18.5	581	387
		415	21.3	567	388
FTK	91.2	415	17.7	609	387
		418	26.4	567	383
HRS	69.1	410	14.7	506	384
		428	18.4	514	398
LAU	91.2	429	22.4	687	392
		421	26.4	603	381
KAI	83.1	442	21.8	820	396
		418	24.7	604	382
NOC	87.3	411	18.4	560	387
		428	50.5	724	384
PAP	55.5	403	9.57	500	387
		411	19.1	548	388
STL	59.1	417	17.5	586	389
		421	36.9	616	383

Table 2: Descriptive statistics for the drift- and bias-corrected CO₂ dry air mole fraction measured at nine representative sites during 2013. Upper row for each site gives the daytime (1100–1800 LT) statistics; lower row gives the nighttime (2200–0400 LT). The ELC node is used as the reference site in Sect. 3.4 and so is presented here without ~~bias~~-correction for systematic uncertainties.

SITE CODE	$\overline{U(\text{CO}_2)}_{\text{temporal-drift}} \text{ day}^{-1} \text{ (ppm day}^{-1}\text{)}$	$\overline{U(\text{CO}_2)}_{\text{temporal-bias}} \text{ (ppm)}$
BEL	0.0327 ± 0.02	-32.7 ± 1
CPS	-0.014 ± 0.002	3.1 ± 0.5
FTK	0.022 ± 0.01	6.1 ± 0.8
HRS	-0.12 ± 0.01	1.2 ± 0.8
LAU	0.10097 ± 0.01	-26.97 ± 0.8
KAI	0.042 ± 0.01 ; -0.0879 ± 0.05	-23 ± 1 ; 65.8 ± 4
NOC	-0.11 ± 0.02 ; 0.030 ± 0.006	22 ± 1 ; -2.7 ± 0.7
PAP	-0.092 ± 0.005	8.7 ± 0.7
STL	-0.034 ± 0.01	9.4 ± 1

Table 3: Results from **drift- and** bias-correction analysis at sites for which at least 3 months of observations are available for comparison with the ELC BEACO₂N node.