*Journal of Geophysical Research: Atmospheres*

Supporting Information for

# Using Short-Term $CO/CO_2$ Ratios to Assess Air mass Differences over the Korean Peninsula during KORUS-AQ

H. S. Halliday[1,2], J. P. DiGangi[1], Y. Choi[1,3], G. S. Diskin[1], S. E. Pusede[4], M. Rana[1,3], J. B. Nowak[1], C. Knote[5], X. Ren[6,7], H. He[6,8], R. R. Dickerson[6,8], Z. Li[6,8,9]

[1]NASA Langley Research Center, Hampton Virginia, USA

[2]Universities Space Research Associate, Columbia Maryland, USA

[3]Science Systems and Applications Inc., Hampton Virginia, USA

[4]Department of Environmental Sciences, University of Virginia, Charlottesville Virginia, USA

[5]Meterological Institute, Ludwig Maximilian University, Munich, Germany

[6]Department of Atmospheric and Oceanic Science, University of Maryland, College Park Maryland, USA

[7]NOAA Air Resources Laboratory, College Park, Maryland, USA

[8]Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland, USA

[9]State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China

## Contents of this file

Figures S1 to S11
Tables S1 to S8

**S1. Vertical Splits on the Free Troposphere Altitude Bin**

The boundary layer height was calculated using the vertical profiles conducted during the DC-8 flights, and linearly interpolated between the individual vertical profiles. Using this method, each observation was tagged as either boundary layer or free troposphere. However, the free tropospheric measurements encompass a large range of vertical altitudes, from ~ 100 meters to 7 km. Figure S1 shows the 60-second correlated $\Delta CO/\Delta CO_2$ slope distributions, in 100 m vertical bins represented by box and whisker plots and separated by boundary layer tag. In the free troposphere at 3 km there is a shift in the distributions. Below 3 km ASL, the correlated slopes in the lower free troposphere have their entire 25th to 50th percentile range above 0% $\Delta CO/\Delta CO_2$; above 3 km ASL most of the vertical bins have their 25th percentile value below zero, indicating that there are significantly more negatively correlated slopes above 3km ASL in the free troposphere. To adequately capture this shift in behavior the free tropospheric bin was split into a low free troposphere (< 3 km ASL) and a high tropospheric bin (≥ 3 km ASL).

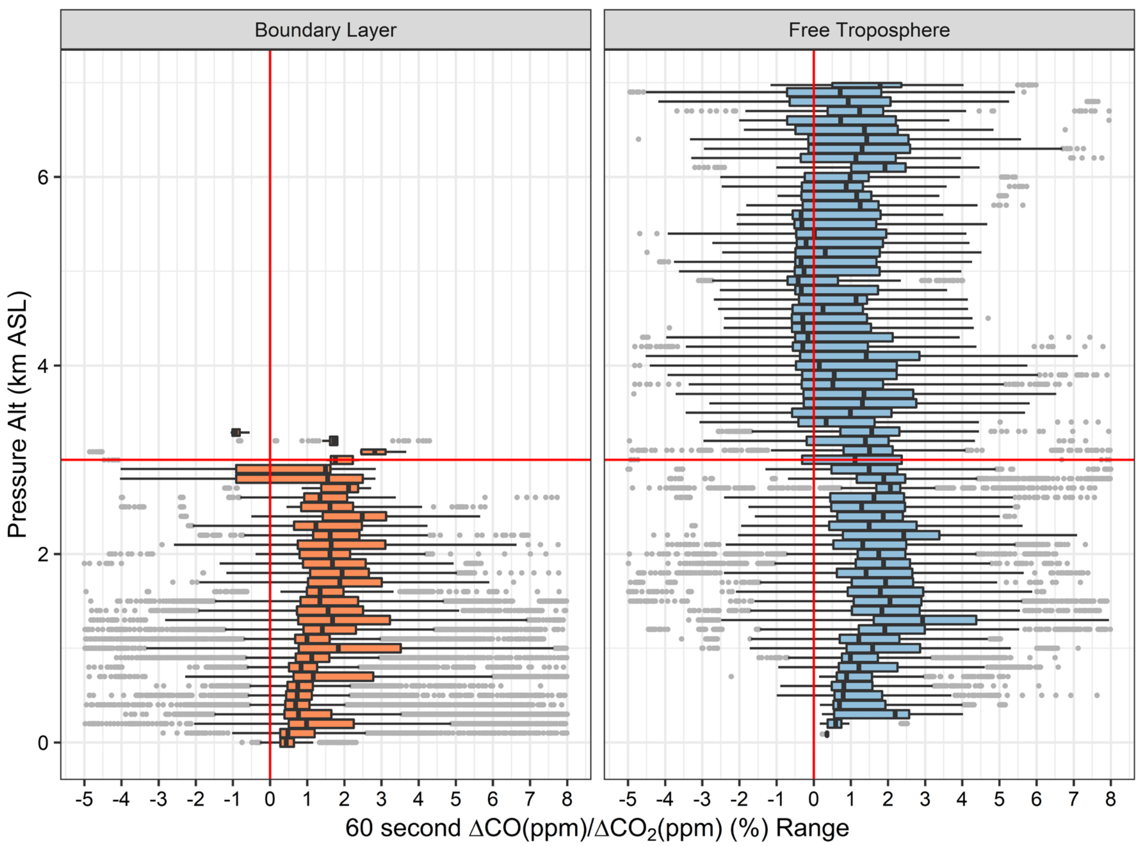


**Figure S1.** 60-second correlated slopes ($R^2 \geq 0.5$) for all KORUS-AQ DC-8 data, organized in 100m altitude bins and represented by box and whisker plots for each vertical bin. Each box and whisker plot shows the 25th, 50th, and 75th percentile; the whiskers terminate at the last observation within 1.5 times the interquartile range, and all other observations are indicated with outlier points. The slope distributions are plotted against the pressure altitude (km ASL), and separated by boundary tag. Two lines have been added, one at $\Delta CO/\Delta CO_2$ = 0, and one at 3 km ASL. In the free tropospheric data there is a shift in the distributions above 3 km ASL, with the 25th percentile of the correlated slopes in most of the high altitude vertical bins falling below 0% $\Delta CO/\Delta CO_2$, and a shift back to slightly higher values above 6 km ASL.

## S2. Method Characterization

### S2.1 Characterization of the Correlated Distributions

Figure S2 shows the explicit comparison between the counts distribution and the probability density distribution by $R^2$ cutoff value for the 60-second rolling window. The relative numbers of observations is summarized in Table S1. Subsetting the calculated slopes by the correlation of determination is necessary, as having a minimum $R^2$ cutoff for inclusion drops periods when both species are steady or otherwise uncorrelated in time during the short time period of the rolling window. This allows the method to appropriately spotlight plumes and other regions of correlated behavior without needing to define and locate plumes in the data, and allows the characterization of nearfield emission signature of $\Delta CO/\Delta CO_2$, rather than the overall regional ratio. This is useful in regions with spatially heterogeneous emissions (e.g., China vs. South Korea).

**Table S1**. The number of correlated slopes calculated for the KORUS-AQ science flights using different $R^2$ cutoff conditions on the rolling correlations calculated with a 30 second window.

| $R^2$ Cutoff Value | # of Observations | % of total Observations |
|---|---|---|
| **0.5** | 192548 | 37.4 |
| **0.6** | 162633 | 31.6 |
| **0.7** | 131212 | 25.5 |
| **0.8** | 95608 | 18.6 |
| **0.9** | 53430 | 10.4 |

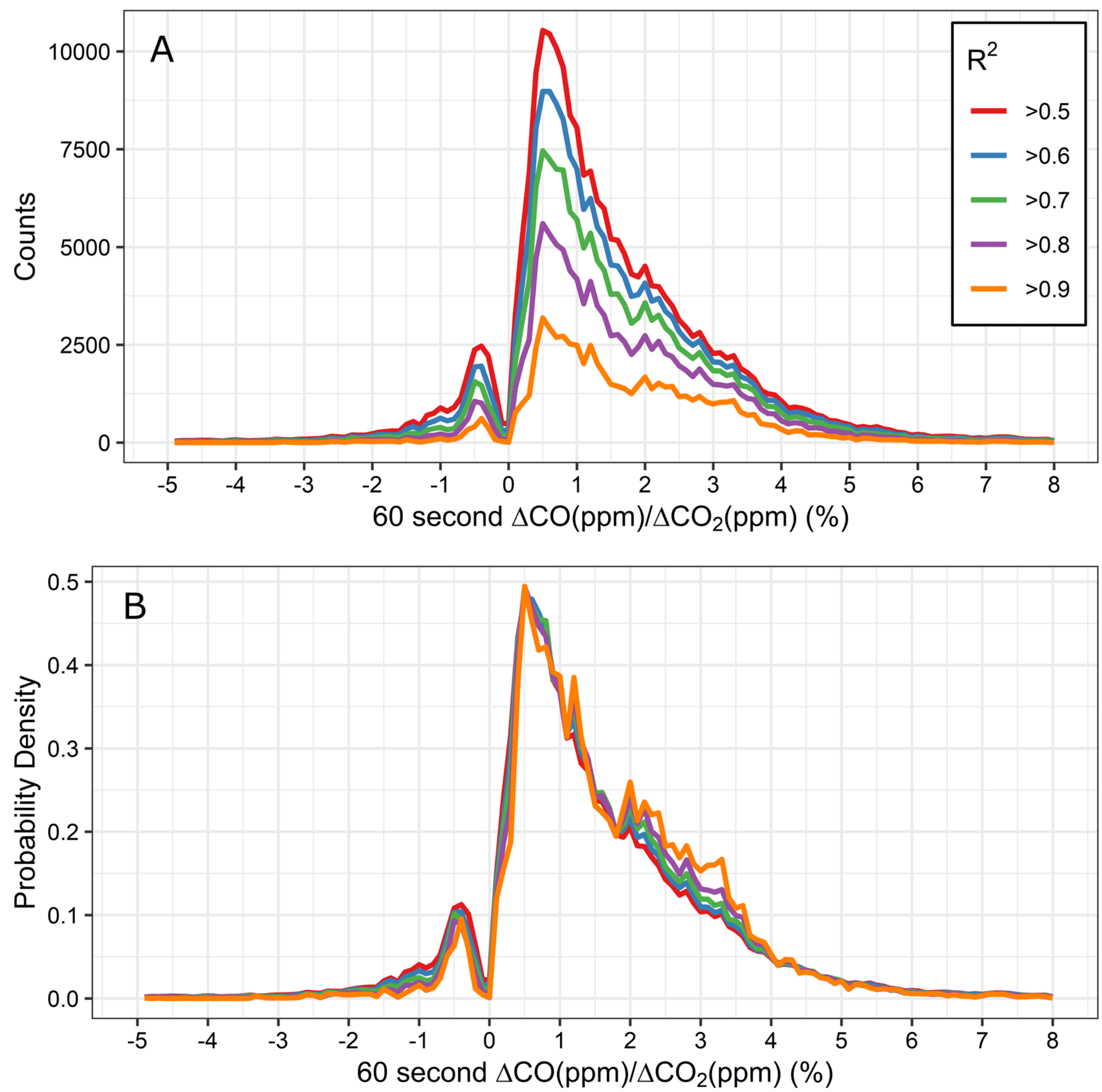


**Figure S2.** (A) Slope distribution for the full dataset for all slopes with a correlation of determination ≥ 0.5, colored by $R^2$ cut off (colors). (B) The normalized distribution of (A), with each distribution normalized to an area of one.

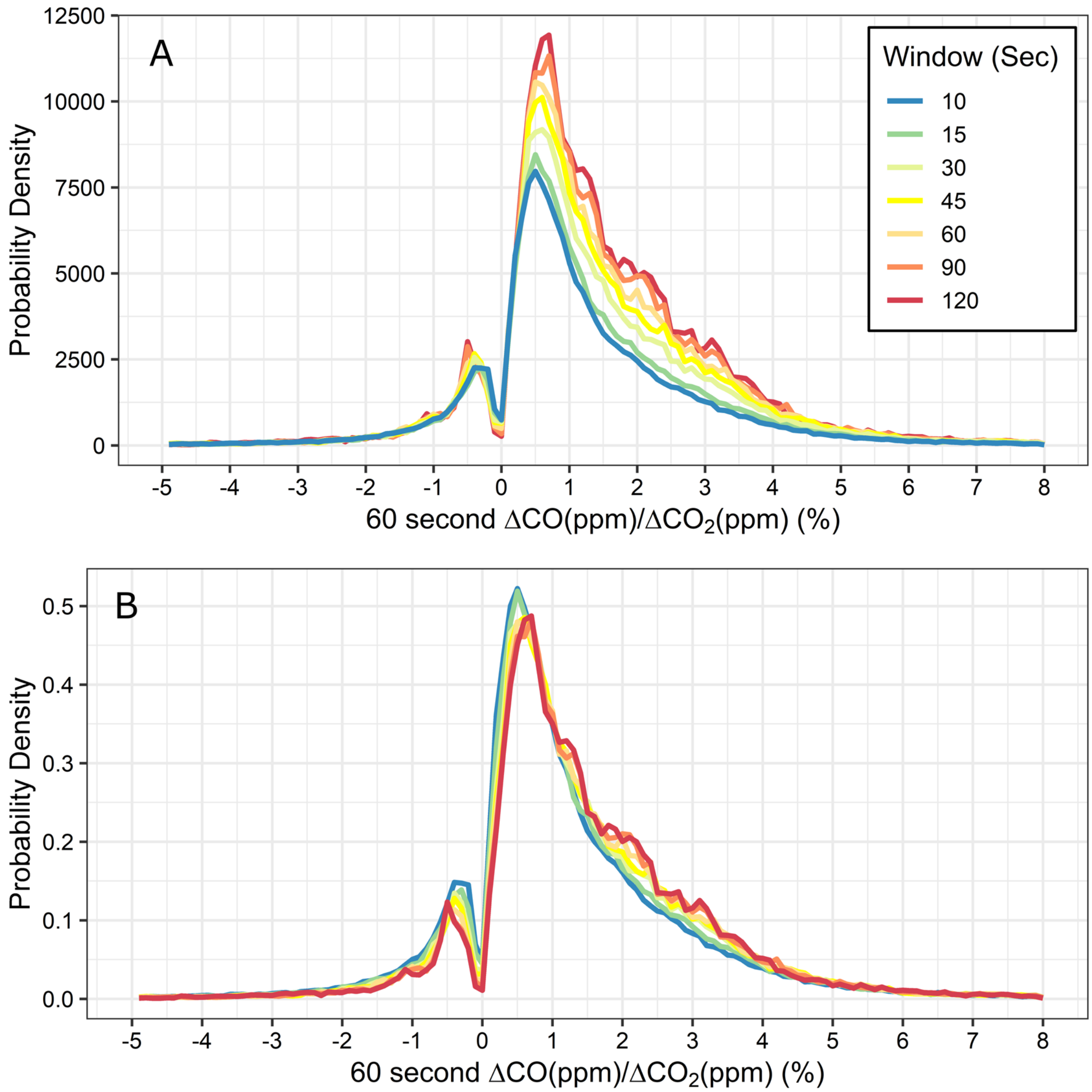


**Figure S3.** Sensitivity of the slope distribution to the window width. Each distribution is the slope distribution for the full 20 KORUS-AQ science flights, with each line showing the distribution for a different rolling window width (in seconds), for all slopes with $R^2 \geq 0.5$. (A) The distributions show as total counts for each bin. (B) The same distributions normalized to a probability density with an area of one.

The rolling window width is an important consideration for interpreting the results. A shorter window will be sensitive short duration events, but miss variations with longer duration in the data and identifies fewer correlated slopes. In contrast, the longer windows will pick out correlations over longer time periods, which will calculate more correlated slopes overall. In cases where there is long range transport or otherwise diffuse plumes the longer windows may pick out features that the shorter windows miss. Figure S3 shows the slope distributions of the full 20 science flights calculated with different rolling window widths, in both counts and normalized probability density. Collapsing the count distributions to a normalized density curve returns similar distributions regardless of window width, and Table S2 summarizes some of the statistics for these different windows widths. However, there are some nuances. First, the locations of the peaks (in both positive and negative slopes) moves away from zero as the window width increases. Second, the longer windows highlight behavior in the higher ratios (> 1% $\Delta CO/\Delta CO_2$) that could indicate transport of more diffuse plumes that are only picked up with the longer windows. Third, the two shortest windows (10 and 15 seconds) are loosely grouped together, yielding both roughly the same number of correlated slopes and loosely the same locations of the peaks. The results in Table S2 are reported to the tenth of a percent as the distributions are binned in 0.1% boxes for. Binning at 0.1% was chosen because it shows the fine scale structure of the distributions without presenting excessive noise.

For the analysis presented in the manuscript we chose a 60-second window. KORUS-AQ sampled over a wide range of scales, and the mid-length 60-second window was used as a compromise that captured most variation well. For future applications of the method, the rolling window width should be tuned appropriately for the dataset and account for both the species in the ratio and the time resolution of the data. For the current work, a 60-second window calculated on a 1 second dataset collected from a fast moving platform worked well for the species used; Alternative scenarios can present cases where a much longer rolling window would be appropriate, such as ground based measurements with continuous monitoring over many weeks or months at one-minute time resolution, or satellite measurements collected over a decadal time scale.

**Table S2.** Summary of the different window widths on the slope distributions.

| Window Width (Seconds) | # of Correlated Slopes ($R^2 \geq 0.5$) | What % of the calculated slopes are correlated | Positive Maximum (%$\Delta CO/\Delta CO_2$) | Negative Maximum (%$\Delta CO/\Delta CO_2$) |
|---|---|---|---|---|
| **10** | 154568 | 32.6 | 0.5 | - 0.4 |
| **15** | 164848 | 33.9 | 0.5 | - 0.3 |
| **30** | 192548 | 37.4 | 0.6 | -0.4 |
| **45** | 210764 | 40.0 | 0.6 | -0.4 |
| **60** | 222444 | 41.8 | 0.5 | -0.4 |
| **90** | 236988 | 44.1 | 0.7 | -0.5 |
| **120** | 246872 | 45.7 | 0.7 | -0.5 |

As the calculation window is rolled over the dataset, the number of points associated with each slope will depend on the coverage of the data within the window. A sufficiently long window will be able to calculate slopes across calibration events or other missing observations in the data, which means that not all slopes will have the expected maximum observations. To determine if there would be an effect on the distributions from including slopes that spanned regions of missing data, the distributions were recalculated with different minimum number of required observations. A hard minimum number of three observations needed for a valid slopes. Figure S4 shows the results of different cut offs for number of points in each slope for the 60-second rolling window, and show clearly that the normalized distributions are not affected by the removal of correlated slopes with fewer than the expected maximum observations. The different cutoff values in Figure S4 were based on what % of the total rolling window width in seconds was the acceptable minimum cutoff (e.g., if we require that half of the rolling window have pairwise observations of CO and $CO_2$ for a valid slope, the minimum number of points would be 15). We used cutoffs at 25%, 50%, 75%, 80%, and 90%, plus the hard minimum of 3 points for a valid slope. As is clear from Figure S4, the method shows negligible sensitivity to the minimum number of observations required. Accordingly, the minimum value was set to 3 in order to include the greatest number of slopes.

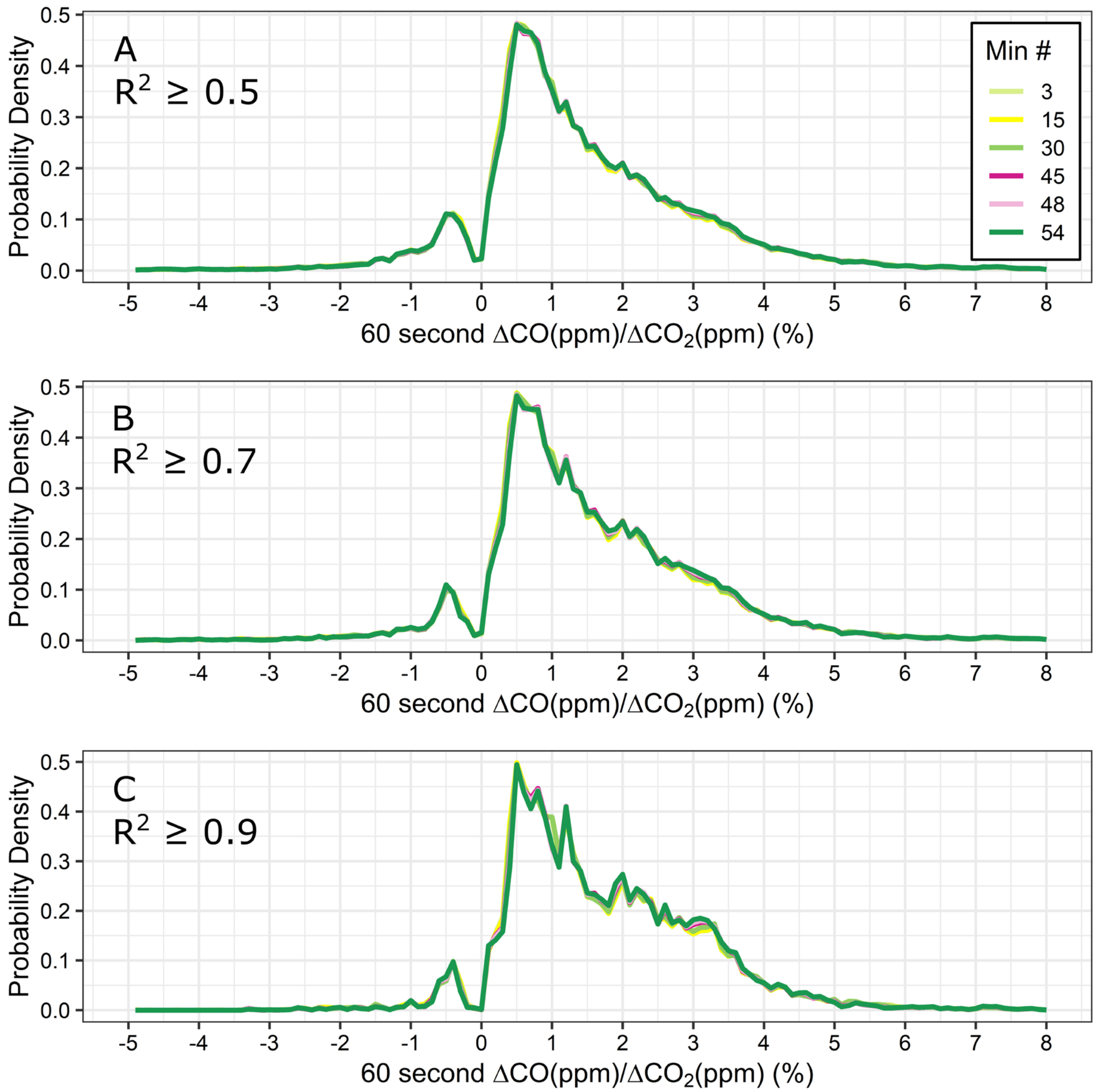


**Figure S4.** Investigating the effects of requiring a minimum number of points in each correlation calculation, using the 30 second rolling window on the full 20 KORUS-AQ science flights. The distributions for the different cutoff values are shown as normalized probability densities, colored by the minimum number of observations in each slope. The three panels show the results for three different $R^2$ cutoff values. (A) $R^2 \geq 0.5$. (B) $R^2 \geq 0.7$. (C) $R^2 \geq 0.9$.

The final check of the data was to assess whether the slopes in the correlated portion of the data are being driven by real variability or by noise. Slopes with Δ values too close to the measurement precision may be unduly influenced by noise or drift in the measurements. The campaign precisions were calculated to be 0.4 ppb for CO and 0.13 ppm for $CO_2$ (1σ, 1 second). We assessed the number of slopes that fail this test at 3, 4, and 5 times the precision (3σ, 4σ and 5σ), with the results shown in Table S3. Very few of the correlated slopes have delta values below our most stringent requirement of 5σ, with a total of 1241 slopes being removed from the distributions (0.56% of the 22444 total correlated slopes).

**Table S3.** The statistics for the calculated standard uncertainties in the correlated short-term slopes. The uncertainties are in units of $ppm^{-1}$.

| | Δ < 3σ | Δ < 4σ | Δ < 5σ |
|---|---|---|---|
| ΔCO | 88 | 147 | 352 |
| $\Delta CO_2$ | 208 | 402 | 985 |
| Both ΔCO & $\Delta CO_2$ | 5 | 16 | 96 |

Based on the testing of the slopes distributions, we apply minimal quality control beyond the correlation cutoff. We used a 60 second rolling window coupled with a $R^2$ cutoff of 0.5, require a hard minimum of 3 observations for a valid slope, and exclude correlated slopes with a ΔCO or $\Delta CO_2$ below 5 times the species precision, which is a conservative threshold for inclusion. This provides a robust, simple calculation of continuous $CO/CO_2$ slopes in a 1 second aircraft dataset.

**S2.3 Understanding the Uncorrelated Behavior at the Zero Slope**

When the slopes are separated by the correlation coefficient ($R^2 \geq 0.5$), the data separate into a correlated distribution, with a positive and negative lobe, and an uncorrelated distribution centered on zero. The bimodal shape of the correlated distribution persists down to a correlation coefficient cutoff of 0.1, shown in Figure S5. This is notable because it indicates that the centering of the uncorrelated data at zero is being driven by the most uncorrelated data ($R^2 < 0.1$, 29.5% of the calculated slopes); additionally, that most uncorrelated artifact is symmetrically oriented to the zero slope.

To ensure that we correctly understand what causes this symmetric uncorrelated behavior at the zero slope, a rolling calculation was performed using synthetic data with no meaningful statistical correlations. The rolling correlation was run with 50,000 observations of synthetic CO (CO_syn) and synthetic $CO_2$ ($CO_2$_syn). The CO_syn was a steady mole fraction of 185 ppb, with added noise from a normally distributed function; the $CO_2$_syn was generated from a sine wave, centered at 400 ppm with an amplitude of 5 ppm (395 – 405 ppm), a frequency of 0.1 (5000 total cycles), with additional noise from a normal distribution. A weighted slope was calculated with a 60 observation window rolled over the synthetic dataset, and the slope distributions were compared for varying levels of noise on CO and $CO_2$.

The results from the tests with the synthetic data are shown in Figure S6. All slopes calculated with this synthetic testing dataset have an $R^2 < 0.5$. The results show that the distributions of the uncorrelated data narrow for increasing noise on the $CO_2$. In contrast, increasing noise levels on the CO with a steady noise level on the $CO_2$ creates broader distributions. The real distribution of the uncorrelated zero slope behavior can be interpreted as a balance between slopes being primarily driven by noise on the CO vs $CO_2$ mole fractions.

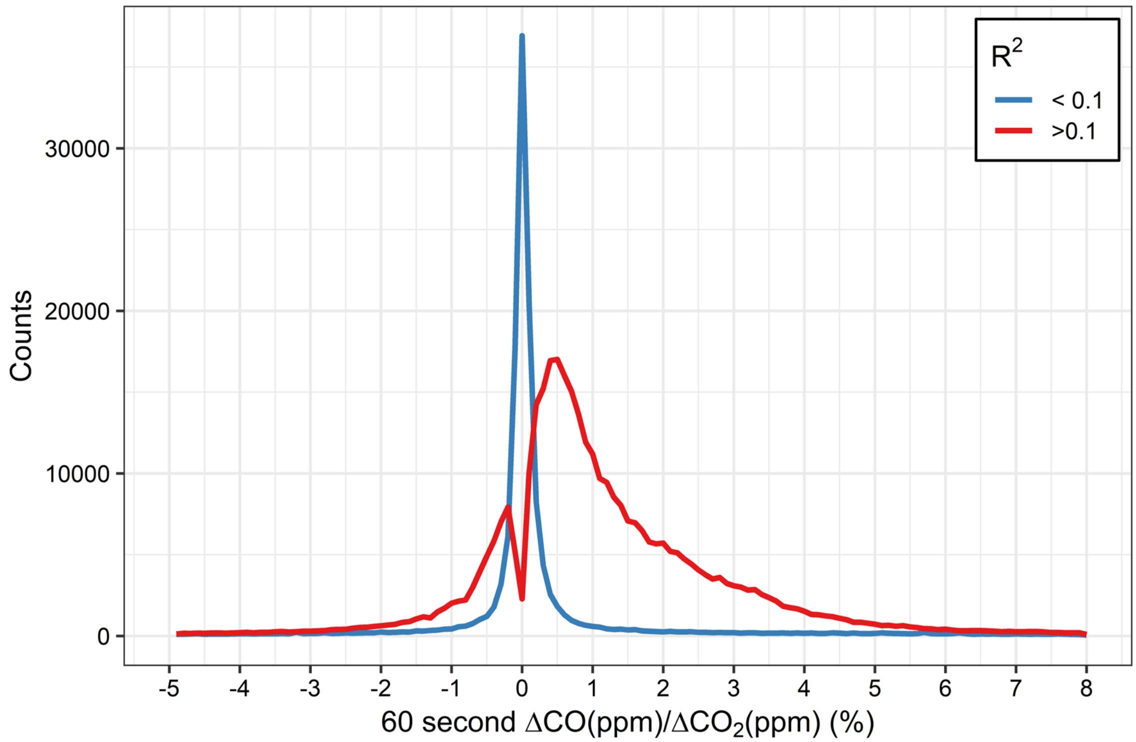


**Figure S5.** Slope distributions for the 20 science flights, showing the distributions split by a correlation coefficient of 0.1. Both lines are shown with total counts, with the data binned to 0.1% $\Delta CO/\Delta CO_2$.

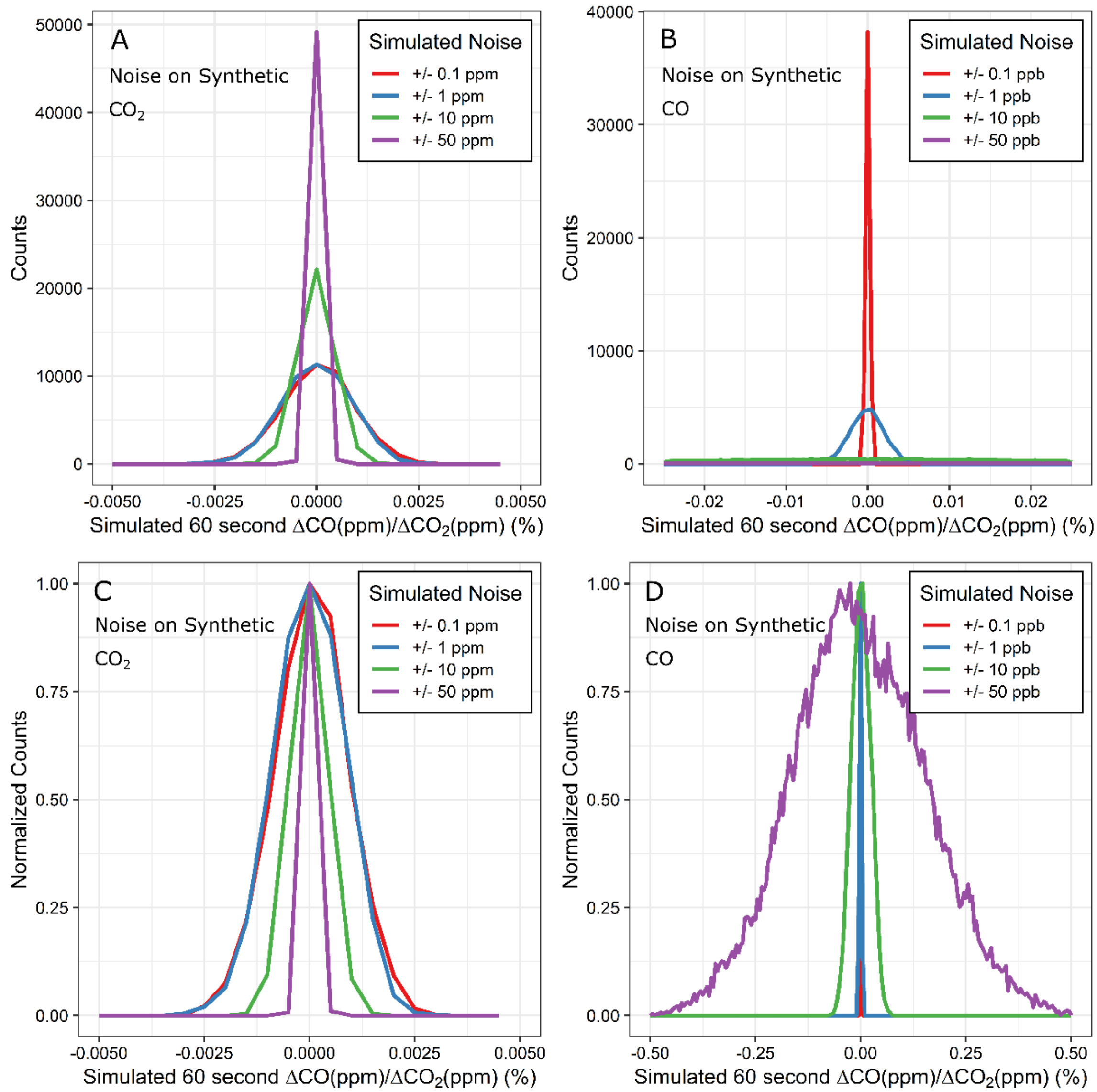


**Figure S6.** Slope distributions created from a 60 observation rolling window calculating slopes from a synthetic testing dataset, designed to have no correlated periods. A. The slope distributions with increasing noise on the simulated $CO_2$ mole fractions, with constant low noise on the simulated CO mole fractions. B. The slope distributions with increasing noise on the simulated CO mole fractions, with constant low noise on the simulated $CO_2$ mole fractions. D. The counts plot from panel A in counts, normalized to a maximum value of 1 to better show the differences in the slopes distributions. C. The counts plot from panel B in counts normalized to a maximum value of 1.

### S2.4 How many Ultra-high Efficiency Slopes are Missed with the Rolling Correlations?

The goal of this method is to find periods of correlated behavior between CO and CO2 while excluding correlations driven by measurement uncertainty. As discussed in the main text, correlated slopes having a delta value less than 5 times the precision of either CO or CO2 were excluded. However, there are ultra-high efficiency combustion processes that produce CO2 with almost no concomitant CO. These events would have an expected $\Delta CO/\Delta CO_2$ value of near zero, and a possibly a poor correlation coefficient due to measurement uncertainties.

In order to assess the impact of the requirement for variability within the 60-second window to exceed five times the measurement precision, data were reprocessed without that requirement, and the results were compared. Over the complete KORUS-AQ dataset, 6811 slopes were removed as a result of this precision criterion, 1.28% of the total number slopes. The fraction removed from each of the analysis sectors was roughly equivalent, with the smallest portion in the Seoul analysis region, and the highest in the Peninsula, possibly driven by the presence the power plants on the northwest coast of South Korea. Table S4 summarizes these findings.

**Table S4.** Assessment of the number of ultra-high efficiency signature slopes, by analysis region. The thresholds for inclusion were set at 5 *species precision, compared to the calculations of $\Delta CO$ and $\Delta CO_2$.

| | All Data | Seoul | Peninsula | West Sea |
|---|---|---|---|---|
| # of slopes | 6811 | 1348 | 3148 | 2315 |
| % of sector slopes | 1.28 | 0.77 | 1.52 | 1.53 |

**S3. Trace Gas Constraints on $\Delta CO/\Delta CO_2$ Slope Distributions**

The use of short-term rolling correlations can be used with any trace gas species measured at a high time resolution, and the use of additional species can be used to gain additional understanding of the atmosphere. While this manuscript focuses on $\Delta CO/\Delta CO_2$ ratios for understanding transport over the Korean Peninsula, additional insight can be gained by including combustion species such as $NO_2$ and $CH_4$.

For example, the use of $CH_4$ and $CO_2$ can be used to differentiate biogenic and anthropogenic behavior. Figure S7 shows a heat map of $\Delta CH_4/\Delta CO_2$ slopes (%, ppm/ppm * 100) plotted against $\Delta CO/\Delta CO_2$ for the full campaign. Air-masses with biogenic influences are expected to have anti-correlated $CO_2$ and $CH_4$ behavior, with plants both producing $CH_4$ and absorbing $CO_2$. Using the assumption that all negative $\Delta CH_4/\Delta CO_2$ slopes are due to a biogenic process, we can check the assumption used in the manuscript that the negative $\Delta CO/\Delta CO_2$ slopes are due to biogenic behavior. The slope distributions shown in Figure S7 are divided by analysis region and boundary layer tag, and have two lines in each subplot – red for the full correlated $\Delta CO/\Delta CO_2$ distributions, and black showing the distributions of the $\Delta CO/\Delta CO_2$ slopes that also have correlated negative $\Delta CH_4/\Delta CO_2$ slopes. If the correlated negative $\Delta CO/\Delta CO_2$ slopes are primarily biogenic, we would expect to see nearly perfect overlap with the negative $\Delta CH_4/\Delta CO_2$ slopes – which is in fact what is seen in the observations.

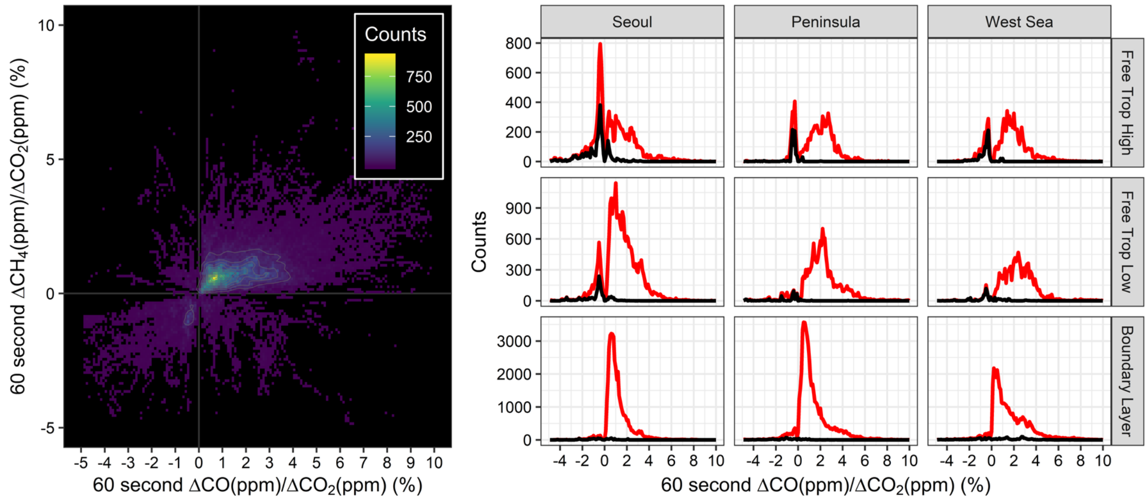


**Figure S7.** Constraining biogenic influences using $\Delta CH_4/\Delta CO_2$ slopes. The heat map shows correlated 60-second $\Delta CH_4/\Delta CO_2$ slopes plotted against correlated $\Delta CO/\Delta CO_2$ slopes, binned at 0.1% ratios, and colored by the number of observations in each bin. The slope distributions show the full correlated $\Delta CO/\Delta CO_2$ slopes over all observations in red, binned by analysis region and boundary layer height. The black curves are the distributions for any correlated $\Delta CO/\Delta CO_2$ slope that is co-located with a correlated negative $\Delta CH_4/\Delta CO_2$ slope, which are associated with biogenic behavior.

Using the negatively-correlated $\Delta CH_4/\Delta CO_2$ slopes as an indicator of biogenic behavior also provides an estimate of how many positive $\Delta CO/\Delta CO_2$ slopes are associated with biogenic processes. If we assume that all correlated $\Delta CO/\Delta CO_2$ slopes that correspond with correlated negative $\Delta CH_4/\Delta CO_2$ slopes are biogenically influenced, this accounts for 5.7% of the total slopes and 41.2 % of the negative $\Delta CO/\Delta CO_2$ slopes. If instead we allow any negative $\Delta CH_4/\Delta CO_2$ slope, well correlated or not, the biogenically attributable component rises to 19.3% of the total slopes and 80% of the negative $\Delta CO/\Delta CO_2$ slopes. Further work is needed to assess the validity of these assumptions.

Another approach to constraining the $\Delta CO/\Delta CO_2$ ratios is to use $NO_2$ mixing ratios. Rather than use a rolling correlation between $NO_2$ and CO or $CO_2$, the 60-second $\Delta NO_2$ can be used as an indicator of emissions variability in the sample. If a correlated slope is accompanied by a $\Delta NO_2$ value $\geq 0.1$ ppb, this can be assumed to be due to a recent, "fresh" emission; otherwise, it is classified as aged.

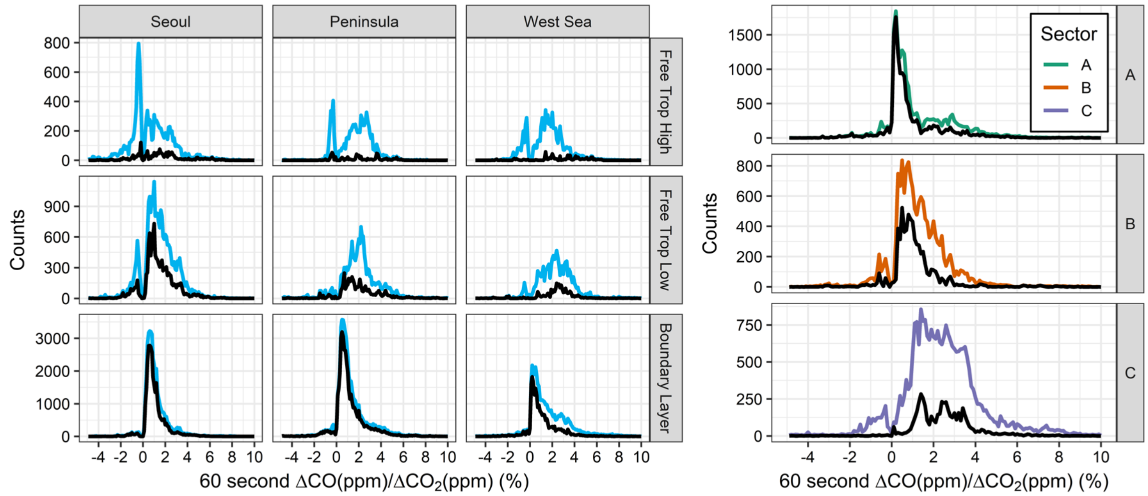


**Figure S8.** Constraining $\Delta CO/\Delta CO_2$ slopes using $NO_2$ variability. The left plot shows the full campaign $\Delta CO/\Delta CO_2$ distributions in blue, divided by analysis region and boundary layer bin. The black curves are the distributions for any correlated $\Delta CO/\Delta CO_2$ slope that is co-located with a 60-second $\Delta NO_2 \geq 0.1$ ppb (recent emissions). The right plot shows the three West Sea analysis sectors, with the total (all altitude) slope distributions shown in color. The black lines show the same distributions filtered for fresh emissions using the same 60-second $\Delta NO_2$ cutoff.

Figure S8 shows the application of this constraint to the correlated $\Delta CO/\Delta CO_2$ distributions. The left side of the plot shows the full correlated $\Delta CO/\Delta CO_2$ distributions in blue, split by analysis region and boundary layer height tag. The distributions of the fresh $\Delta CO/\Delta CO_2$ ratios (60-second $\Delta NO_2 \geq 0.1$ ppb) are shown in black. All three analysis regions show that the slope distributions are primarily driven by recent emissions in the Boundary layer, but that the distributions are aging with increase altitude; this matches the hypothesis that the slopes at higher altitudes are transported air masses rather than emissions originating within the regions.

The right side of Figure S8 shows the same $\Delta NO_2$ constraint applied to the West Sea analysis sectors (all altitudes). The slope distributions in sector A are primarily attributable to fresh emissions, particularly for the lowest ratio (associated with high efficiency combustion). In contrast, sector C is primarily driven by transport from the Chinese mainland, which is validated with both the higher $\Delta CO/\Delta CO_2$ ratios over all altitudes in this area and the low contributions from fresh emissions in this sector.

Further application of these types of constraints is beyond the scope of the current work. However, the use of other trace gas rolling ratios calculated with this technique can be used to gain a better understanding of the origin and characteristics of transported air masses. The constraints shown here are simple applications of additional ratios; more sophisticated techniques such as clustering are a natural next step for bringing additional insights with additional trace gas species.

## S4. KORUS-AQ Regional Analysis Summaries

Figures S9, S10 and S11 present the $\Delta CO/\Delta CO_2$ summaries for the three analysis regions, modeled on Figure 3 which summarized the overall campaign behavior. Table S6 summarizes the fit behavior between $\Delta CO/\Delta CO_2$ for the three analysis regions and for all observations during KORUS-AQ.

**Table S5.** KORUS-AQ analysis region fit information.

| Fit information | All KORUS Data | Seoul | Peninsula | West Sea |
|---|---|---|---|---|
| Total Error Adjusted Slope (% $\Delta CO/\Delta CO_2$) | 1.11 | 1.09 | 0.95 | 1.35 |
| Overall $R^2$ | 0.30 | 0.56 | 0.22 | 0.24 |
| % correlated slopes ($R^2 \geq 0.5$) | 41.7 | 45.5 | 38.1 | 42.1 |
| # correlated slopes ($R^2 \geq 0.5$) | 221203 | 79303 | 78404 | 63496 |

Figure S9 shows the results for the Seoul analysis region. The data collected over Seoul have the highest overall correlation between CO and $CO_2$, with a correlation coefficient above 0.5. Seoul also has the greatest proportion of correlated slopes over the sampling period. Figure S9A shows the full campaign Seoul scatter plot; while there are a handful of outlier observations, the data are reasonably grouped around the fit line compared to the full campaign $CO/CO_2$ plot in Figure 3. Figure S9B shows both the correlated and uncorrelated probability densities. Like the full campaign data in Figure 3, the uncorrelated slopes are centered at 0% $\Delta CO/\Delta CO_2$, while the correlated slopes have both a positive and negative lobe, with the most commonly measured positive slopes around 0.5% $\Delta CO/\Delta CO_2$. Figure S9C shows that the normalized distributions for the different $R^2$ cutoff values have roughly the same behavior for all five correlation cutoffs.

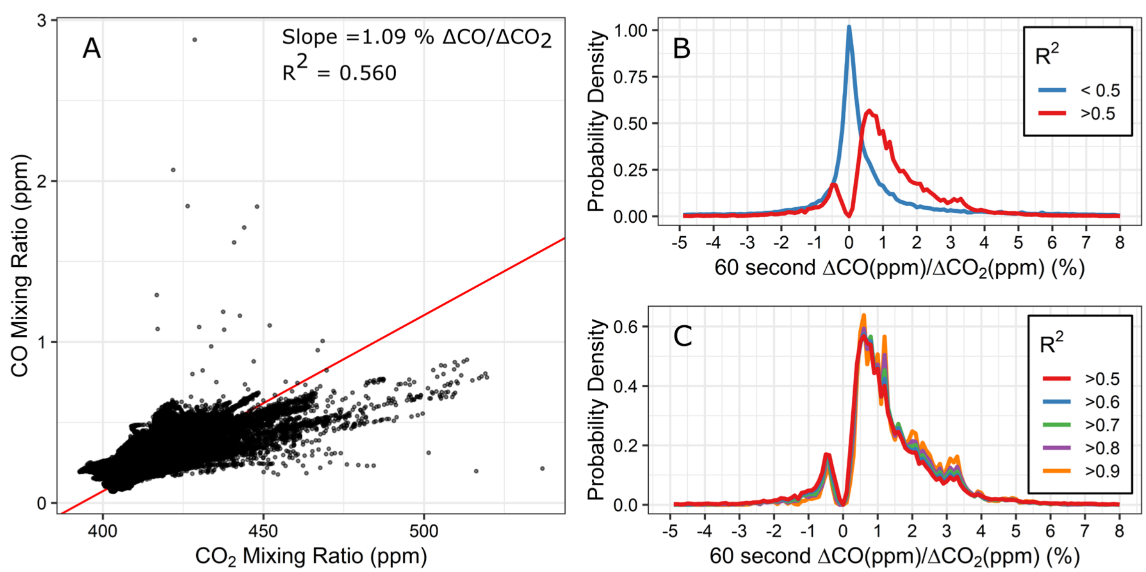


**Figure S9.** Seoul regional analysis characterization plot. A) Full campaign scatter plot between CO and $CO_2$ with the error adjusted bivariate correlation. (B) Slope distributions for the 20 science flights, showing the distributions of the effectively correlated slopes ($R^2 \geq 0.5$) and the effectively uncorrelated slopes ($R^2 < 0.5$). Both lines are normalized to a probability density. (C) The normalized probability distributions of the correlated data, separated by minimum $R^2$ value.

Figure S10 shows the summary of the Peninsula analysis region. This region was not analyzed in detail for the manuscript, as it is the subject of continuing research due to the complicated topography and the variety and mix of sources within this region. The scatter plot in Figure S10 shows the largest spread in the data for the three analysis regions, and also indicates that the overall distribution shape see in the full data scatter plot from Figure 3 is being driven by measurements collected in the Peninsula analysis region. This region has the lowest percentage of correlated slopes, below 40%, and a non-meaningful overall $\Delta CO/\Delta CO_2$ ratio. This is due to the variety of environments that are combined into this analysis region, whereas the Seoul region included the emissions from the main population area, and the West Sea was mostly a receptor location for South Korean inflow or outflow (depending on sector). The Peninsula region covered all of the non-Seoul portions of the South Korean Peninsula and portions of Southern Japan, making it a poor region for the high level analysis that was conducted in this paper.

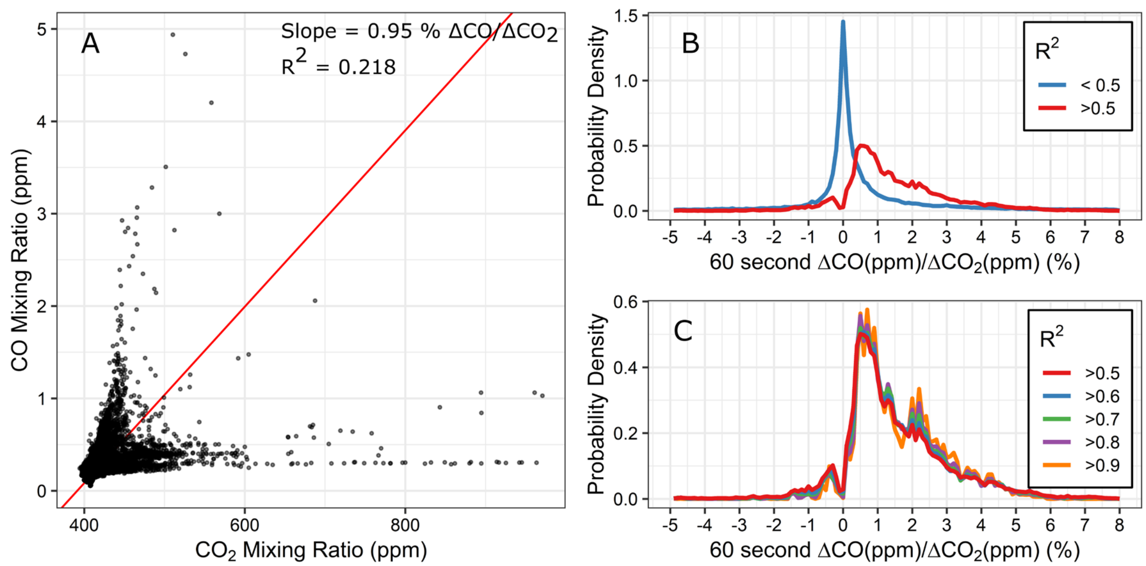


**Figure S10.** Peninsula regional analysis characterization plot. A) The full campaign scatter plot between CO and $CO_2$ with the error adjusted bivariate correlation. (B) The slope distributions for the 20 science flights, showing the distributions of the effectively correlated slopes ($R^2 \geq 0.5$) and the effectively uncorrelated slopes ($R^2 < 0.5$). Both lines are normalized to a probability density. (C) The normalized probability distributions of the correlated data, separated by minimum $R^2$ value.

Figure S11 summarizes the West Sea analysis region. The West Sea was primarily a receptor location for the inflow and outflow from the South Korean peninsula, although measurement of the industrial region on the northwest coast of South Korea were also included. In general, the slope distributions for this region were shifted to higher $\Delta CO/\Delta CO_2$ values compared to the Seoul and Peninsula regions, and this shift can be seen in both Figure S11A, the scatter plot, and Table S5, which shows that the overall $\Delta CO/\Delta CO_2$ ratio for this region was a relatively high 1.35%. Figure S11C shows that while the different $R^2$ cutoff values resulted in similar distribution behavior for all five cutoffs, the more highly-correlated distributions show deviation from the other cutoffs. This is most likely due to the sampling size of the distributions. For regions where there is a large proportion of long range transport, such as over the West Sea where the aircraft sampled Chinese inflow to the peninsula, the plumes measured are expected to be more diffuse, with fewer correlated slopes over the short time windows compared to the longer windows.

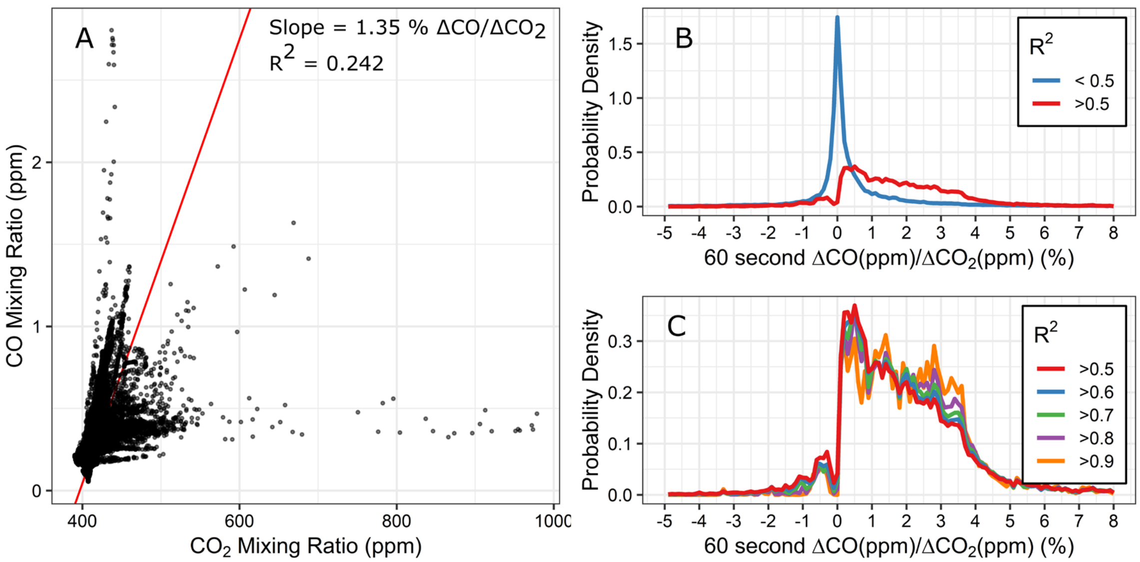


**Figure S11.** West Sea regional analysis characterization plot. A) Full campaign scatter plot between CO and $CO_2$ with the error adjusted bivariate correlation. (B) Slope distributions for the 20 science flights, showing the distributions of the effectively correlated slopes ($R^2 \geq 0.5$) and the effectively uncorrelated slopes ($R^2 < 0.5$). Both lines are normalized to create probability densities. (C) The normalized probability distributions of the correlated data, separated by minimum $R^2$ value.

## S5. Chinese Dataset Summary

The Chinese in situ dataset was comprised of CO and $CO_2$ measurements made over the Chinese mainland during the same time frame as the KORUS-AQ DC-8 campaign. Figure S12 summarizes the $\Delta CO/\Delta CO_2$ characteristics for these measurements. Figure S12A shows the total $CO/CO_2$ scatter plot. The RMA correlation produces an overall slope of 3.33% $CO/CO_2$ and a total correlation coefficient of $R^2 = 0.76$. Figure S12B shows the correlated ($R^2 \geq 0.5$) and uncorrelated ($R^2 < 0.5$) probability density distributions. While the overall correlation returned a total ratio of 3.3% CO/ $CO_2$, the slope distributions show bimodal shape, with peaks at both 1% and approximately 2.5% $\Delta CO/\Delta CO_2$. Figure S12C shows the normalized probability distributions of the correlated data, separated by minimum $R^2$ value. While the KORUS-AQ $\Delta CO/\Delta CO_2$ results indicated that the $R^2$ cut off value did not change the normalized slope distributions significantly, the results from the Chinese data indicate that this observation does not hold for this dataset. This most likely due to the smaller number of data points in the Chinese dataset vs. the KORUS-AQ dataset; The KORUS-AQ dataset includes 20 science flights, which were 6 to 10 total hours per flight over a large spatial and vertical domain, while the Chinese measurements were collected on a smaller aircraft platform with more limited flight capabilities. In the manuscript analysis we use the minimum correlation cutoff value of $R^2 \geq 0.5$ to be consistent with the slope distributions from the KORUS-AQ dataset.

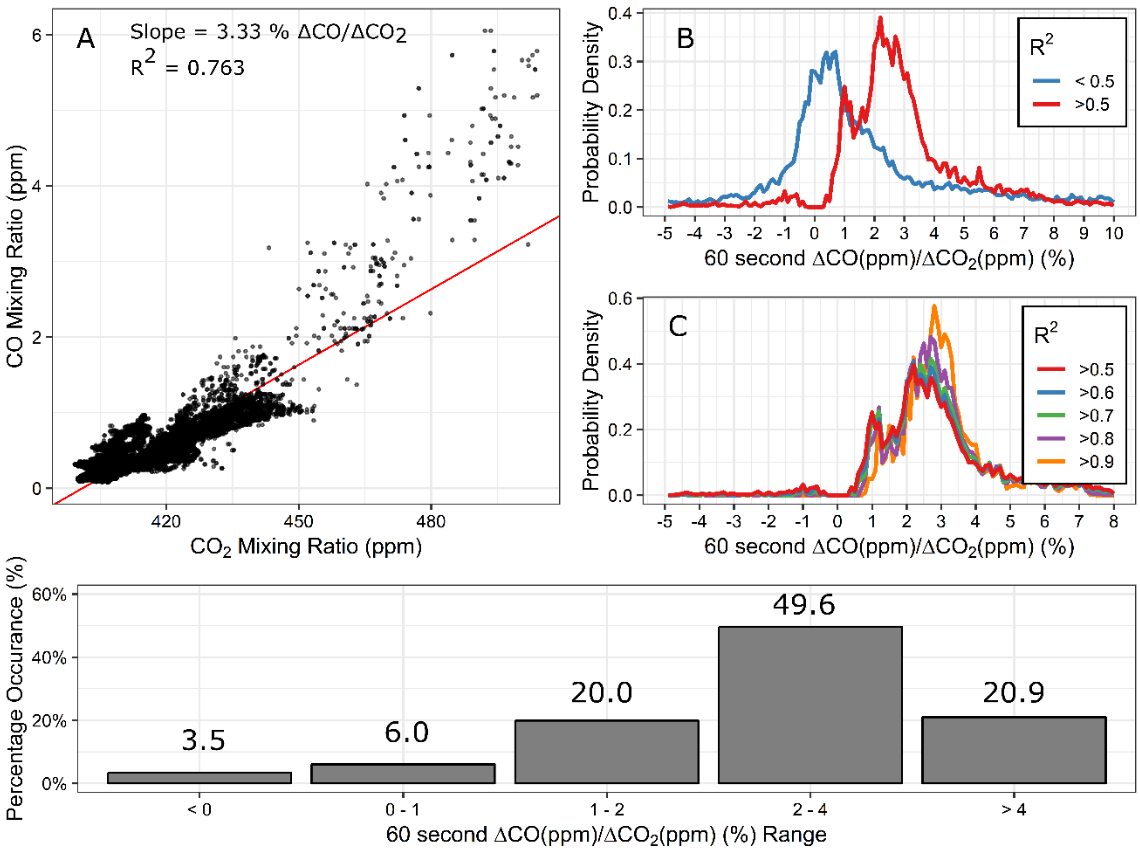


**Figure S12.** Chinese dataset characterization plot. A) The full campaign scatter plot between CO and $CO_2$ with the error adjusted bivariate correlation. (B) Slope distributions for the 20 science flights, showing the distributions of the effectively correlated slopes ($R^2 \geq 0.5$) and the effectively uncorrelated slopes ($R^2 < 0.5$). Both lines are normalized to a probability density. (C) Normalized probability distributions of the correlated data, separated by minimum $R^2$ value. (D) Quantification of the distributions by $\Delta CO/\Delta CO_2$ ratio range, i.e. < 0 % $\Delta CO/\Delta CO_2$, 0 – 1 % $\Delta CO/\Delta CO_2$, etc. Each bar is labeled with the percentage of the distribution that occurs within the listed range.

### S7. Relative Number of Observations between Regions

The slope distributions in the manuscript text are shown as normalized probability distributions, allowing us to capture the distribution behavior differences between regions without the number of samples influencing height of the distributions. Table S6, S7, and S8 show the count statistics for each of the analysis regions. Table S6 breaks down observation statistics for the three primary South Korea analysis regions, Table S7 shows observation statistics for the three West Sea analysis subsectors, and Table S8 shows the observation statistics for the Chinese dataset. The FLEXpart back-trajectories were not calculated for the Chinese measurement dataset.

**Table S6**. Observation Statistics for the three South Korean analysis regions.

| Region | Number of pairwise Observations ($CO + CO_2$) | % of total pairwise observations | Number of Correlated slopes | % of slopes correlated in region | Number of Back-Trajectory Calculations Associated with Region |
|---|---|---|---|---|---|
| | All altitudes | | | | |
| Seoul | 147059 | 32.2 | 79463 | 45.5 | 3098 |
| Peninsula | 179464 | 39.3 | 79052 | 38.4 | 3565 |
| West Sea | 130134 | 28.5 | 63894 | 42.4 | 2568 |
| | High Free Troposphere | | | | |
| Seoul | 43643 | 9.56 | 12326 | 23.6 | 941 |
| Peninsula | 48535 | 10.8 | 8807 | 15.4 | 999 |
| West Sea | 39699 | 8.69 | 8701 | 18.8 | 800 |
| | Low Free Troposphere | | | | |
| Seoul | 42689 | 9.35 | 25610 | 50.2 | 886 |
| Peninsula | 24014 | 5.26 | 14876 | 54.4 | 448 |
| West Sea | 20987 | 4.60 | 12125 | 50.0 | 402 |
| | Boundary Layer | | | | |
| Seoul | 60735 | 13.3 | 41527 | 58.2 | 1270 |
| Peninsula | 105915 | 23.2 | 55369 | 45.6 | 2118 |
| West Sea | 69448 | 15.2 | 43068 | 53.3 | 1366 |

**Table S7**. Observation Statistics for the three West Sea analysis subsections.

| Region | Number of pairwise Observations ($CO + CO_2$) | % of total pairwise observations | Number of Correlated slopes in region | % of slopes correlated in region | Number of Back-Trajectory Calculations Associated with Region |
|---|---|---|---|---|---|
| | All altitudes | | | | |
| Sector A | 41190 | 9.02 | 20662 | 43.7 | 805 |
| Sector B | 29969 | 6.56 | 15894 | 44.7 | 605 |
| Sector C | 58975 | 12.9 | 27338 | 40.0 | 1158 |
| | High Free Troposphere | | | | |
| Sector A | 9524 | 2.09 | 1889 | 17.3 | 189 |
| Sector B | 9563 | 2.09 | 3466 | 30.7 | 199 |
| Sector C | 20612 | 4.51 | 3346 | 13.9 | 412 |
| | Low Free Troposphere | | | | |
| Sector A | 4252 | 0.93 | 1894 | 38.5 | 82 |
| Sector B | 3868 | 0.85 | 2703 | 59.8 | 71 |
| Sector C | 12867 | 2.82 | 7528 | 50.8 | 249 |
| | Boundary Layer | | | | |
| Sector A | 27414 | 6.00 | 16879 | 53.6 | 534 |

| | | | | | |
|---|---|---|---|---|---|
| Sector B | 16538 | 3.62 | 9725 | 49.2 | 335 |
| Sector C | 25496 | 5.58 | 16464 | 55.8 | 497 |

Table S8. Observation Statistics for the Chinese measurement.

| Region | Number of pairwise Observations ($CO + CO_2$) | % of total pairwise observations | Number of Correlated slopes in region | % of slopes correlated in region |
|---|---|---|---|---|
| All Data | 58775 | 100 | 21920 | 36.2 |
| Free Troposphere | 30409 | 51.7 | 9714 | 31.2 |
| Boundary Layer | 27236 | 46.3 | 11884 | 41.0 |

## S8. Emissions Inventory Calculation

In Section 4.6, emissions inventory values for CO and $CO_2$ emissions were converted to mole fractions and the mole fraction ratio of $CO/CO_2$ was calculated for comparison with the short-term $\Delta CO/\Delta CO_2$ ratios from the in-situ observations. The calculation for the conversion from Gg (CO, Tg $CO_2$) to moles is show in eq. S1 (CO) and eq. S2 ($CO_2$). This approach of comparing emissions inventory values to $\Delta CO/\Delta CO_2$ ratios is based on Suntharalingham et al. (2004).

$$\mathrm{CO}\left(\frac{\mathrm{Gg}}{\mathrm{Yr}}\right) * \left(\frac{\mathrm{Tg}}{1000\ \mathrm{Gg}}\right) * \left(\frac{\mathrm{mole}}{28.01\ \mathrm{g}}\right) = \mathrm{CO}\left(\frac{\mathrm{Tmoles}}{\mathrm{Yr}}\right) \qquad \text{(S1)}$$

$$\mathrm{CO_2}\left(\frac{\mathrm{Tg}}{\mathrm{Yr}}\right) * \left(\frac{\mathrm{mole}}{44.01\ \mathrm{g}}\right) = \mathrm{CO_2}\left(\frac{\mathrm{Tmoles}}{\mathrm{Yr}}\right) \qquad \text{(S2)}$$

## S9. Derivation of the Calculation of the Slopes to Facilitate Conceptual Understanding

The short-term ratios between CO and $CO_2$ are calculated with a short rolling window (60-seconds for most of this work), and filtered by correlation coefficient to return a slope distribution that provides some information about the instantaneous ratios between the two species. While we calculate the short-term ratio for all observations in the dataset, the correlation coefficient cutoff ensures that the slopes being analyzed are from areas where the two species have mole fractions that are changing on the same time scale, and with Δ values large enough to produce valid results. For the 60-second window in this analysis, these periods are times when the aircraft samples across an air mass boundary or plumes mixing zones. Times when the concentrations are steady in time for time periods longer than the rolling window are dropped from the analysis, no matter what the overall ratio is for that period.

A note on interpreting the $\Delta CO/\Delta CO_2$ slopes. The $\Delta CO/\Delta CO_2$ ratio can be read as a measure of combustion efficiency, but this interpretation does not fully capture the underlying mathematics in the short-term slope calculations. The slope calculation is an assessment of the change in the mole fractions between two air masses with different characteristics. This means that the measured ratio between CO and $CO_2$ is a measurement of the delta values between air mass 1 and air mass 2; we do not measure the direct emissions, but rather the delta between a plume and the surrounding atmosphere or the difference between two mixing air masses. This means that the ratios we measure and analyze are always some mix of two air masses, either between the background air mass and the plume, or the air masses on either side of a boundary.

This section explicitly lays out this mixing, using two air masses, AM1 and AM2. Each air mass has a district mole fraction of CO and $CO_2$, as stated in Eq S3 (AM1) and S4 (AM2):

$$\mathrm{AM}_1 = \mathrm{A}_{\mathrm{CO}}(CO) + \mathrm{A}_{\mathrm{CO}_2}(CO_2) + (1 - \mathrm{A}_{\mathrm{CO}} - \mathrm{A}_{\mathrm{CO}_2})(Other) \quad \text{(S3)}$$

$$\mathrm{AM}_2 = \mathrm{B}_{\mathrm{CO}}(CO) + \mathrm{B}_{\mathrm{CO}_2}(CO_2) + (1 - \mathrm{B}_{\mathrm{CO}} - \mathrm{B}_{\mathrm{CO}_2})(Other) \quad \text{(S4)}$$

The mixing process can be defined as:

$$Mixing = \alpha \mathrm{AM}_1 + (1 - \alpha)\mathrm{AM}_2 \quad \text{(S5)}$$

Which we can expand with the definitions of the air masses:

$$Mixing = \alpha[\mathrm{A}_{\mathrm{CO}}(CO) + \mathrm{A}_{\mathrm{CO}_2}(CO_2) + (1 - \mathrm{A}_{\mathrm{CO}} - \mathrm{A}_{\mathrm{CO}_2})(Other)] + (1 - \alpha)[\mathrm{B}_{\mathrm{CO}}(CO) + \mathrm{B}_{\mathrm{CO}_2}(CO_2) + \left(1 - \mathrm{B}_{\mathrm{CO}} - \mathrm{B}_{\mathrm{CO}_2}\right)(Other)] \quad \text{(S6)}$$

Rearranging yields:

$$= [\alpha \mathrm{A}_{\mathrm{CO}} + (1 - \alpha)\mathrm{B}_{\mathrm{CO}}]CO + \left[\alpha \mathrm{A}_{\mathrm{CO}_2} + (1 - \alpha)\mathrm{B}_{\mathrm{CO}_2}\right]CO_2 + \left[\alpha\left(1 - \mathrm{A}_{\mathrm{CO}} - \mathrm{A}_{\mathrm{CO}_2}\right) + (1 - \alpha)\left(1 - \mathrm{B}_{\mathrm{CO}} - \mathrm{B}_{\mathrm{CO}_2}\right)\right]Other \quad \text{(S7)}$$

Simplify in terms of α:

$$= [\alpha(\mathrm{A}_{\mathrm{CO}} - \mathrm{B}_{\mathrm{CO}}) + \mathrm{B}_{\mathrm{CO}}]CO + \left[\alpha\left(\mathrm{A}_{\mathrm{CO}_2} - \mathrm{B}_{\mathrm{CO}_2}\right) + \mathrm{B}_{\mathrm{CO}_2}\right]CO_2 + \left[\alpha\left((\mathrm{A}_{\mathrm{co}} - \mathrm{B}_{\mathrm{CO}}) + \left(\mathrm{A}_{\mathrm{CO}_2} - \mathrm{B}_{\mathrm{CO}_2}\right)\right) + \left(1 - \mathrm{B}_{\mathrm{CO}} - \mathrm{B}_{\mathrm{CO}_2}\right)\right] Other \quad \text{(S8)}$$

The concentrations of CO and CO2 for any point in time for this mixing process can be deduced from Eq. S8 to be:

$$\chi_{CO} = \alpha(\mathrm{A}_{\mathrm{CO}} - \mathrm{B}_{\mathrm{CO}}) + \mathrm{B}_{\mathrm{CO}} \quad \text{(S9)}$$

$$\chi_{CO_2} = \alpha\left(\mathrm{A}_{\mathrm{CO}_2} - \mathrm{B}_{\mathrm{CO}_2}\right) + \mathrm{B}_{\mathrm{CO}_2} \quad \text{(S10)}$$

When we measure a slope over a time window, we are measuring the $\Delta CO/\Delta CO_2$ over that time range. So for any two points in time, ti and tj, we can present that mole fraction ratio as:

$$\frac{\Delta CO}{\Delta CO_2} = \frac{\chi_{\mathrm{CO}_\mathrm{j}} - \chi_{CO_\mathrm{i}}}{\chi_{\mathrm{CO}_{2\mathrm{j}}} - \chi_{CO_{2\mathrm{i}}}} \quad \text{(S11)}$$

The characteristic mole fractions of the two mixing air masses don't change, only the extent of the mixing, so Ai and Aj, etc., are equal. With that assumption, we can rewrite Eq. S11 as:

$$\frac{\Delta \mathrm{CO}}{\Delta \mathrm{CO}_2} = \frac{(\alpha_\mathrm{j} - \alpha_i)(\mathrm{A}_{\mathrm{CO}} - \mathrm{B}_{\mathrm{CO}})}{(\alpha_\mathrm{j} - \alpha_i)(\mathrm{A}_{\mathrm{CO}_2} - \mathrm{B}_{\mathrm{CO}_2})} \quad \text{(S12)}$$

Or more simply as:

$$\frac{\Delta \mathrm{CO}}{\Delta \mathrm{CO}_2} = \frac{(\mathrm{A}_{\mathrm{CO}} - \mathrm{B}_{\mathrm{CO}})}{(\mathrm{A}_{\mathrm{CO}_2} - \mathrm{B}_{\mathrm{CO}_2})} \quad \text{(S13)}$$

This derivation of the behavior behind the slope calculation during the mixing of two air masses indicates that as the aircraft samples over a plume within an otherwise chemically homogenous environment, the slope should remain steady during the sampling of that plume. Figure S13 showcases an observation of a plume on May 2, 2016, in the boundary layer over the Seoul analysis region, and shows that the slope does stay at a steady value as the aircraft samples through the CO and $CO_2$ plume, as well as showing how the slope values in the steady state mole fraction region before the plume begins to be sampled are filtered out with the minimum $R^2$ filter. Additionally, for the expression in S13, the calculation of the negative values can be seen to be arising from, as an example, the mixing of an air mass with depleted $CO_2$ with an air mass that has combustion sourced CO and $CO_2$. Other scenarios can also generate negative slopes.

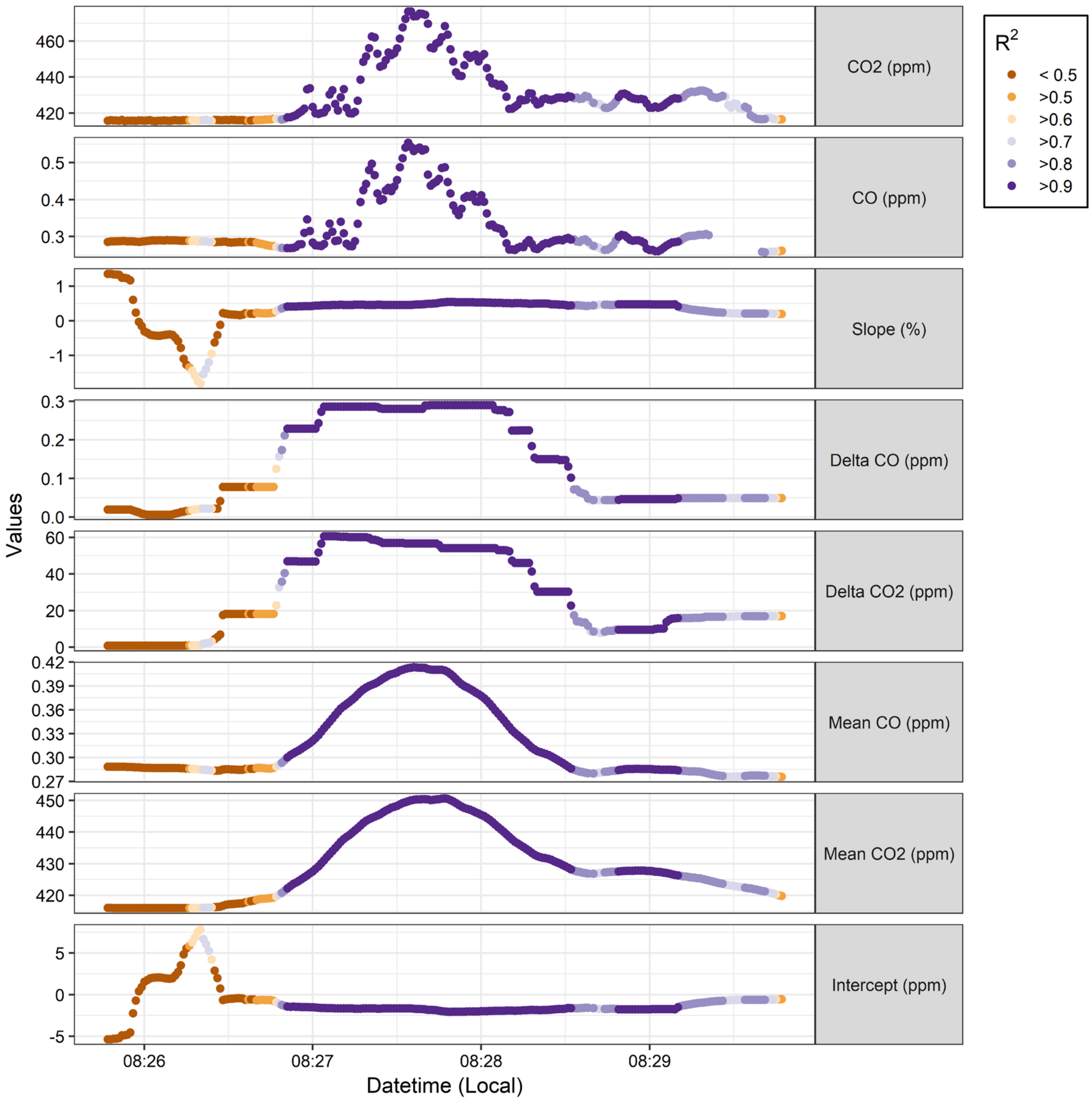


**Figure S13.** A short time series of a CO and $CO_2$ plume measured in the Seoul Analysis region Boundary Layer on 2 May, 2016. The plot shows the CO and $CO_2$ time series, with additional time series showing the calculated slope (60-second rolling window), the 60-second window ΔCO and $\Delta CO_2$ values, and the mean CO and $CO_2$ measured for each step in the rolling window. The time series plots are colored by the $R^2$ value, which has been binned by value. Each time series is noted on the right panel, including units for each individual time series.

Using the definitions of $\chi_{CO}$ and $\chi_{CO_2}$ from equations S9 and S10, we can derive an expression for the intercept that accompanies the slope. Using Eq. S10, we express α in terms of $CO_2$:

$$\alpha = \frac{(\chi_{CO_2} - B_{CO_2})}{(A_{CO_2} - B_{CO_2})} \qquad \text{(S14)}$$

Then use this expression in Eq. S9:

$$\chi_{CO} = \left[\frac{\chi_{CO_2} - B_{CO_2}}{A_{CO_2} - B_{CO_2}}\right](A_{CO} - B_{CO}) + B_{CO} \qquad \text{(S15)}$$

Then rearrange algebraically:

$$\chi_{CO}\left(A_{CO_2} - B_{CO_2}\right) = \left[\chi_{CO_2} - B_{CO_2}\right](A_{CO} - B_{CO}) + B_{CO}(A_{CO_2} - B_{CO_2}) \tag{S16}$$

$$(A_{CO_2} - B_{CO_2})(\chi_{CO} - B_{CO}) = \left[\chi_{CO_2} - B_{CO_2}\right](A_{CO} - B_{CO}) \tag{S17}$$

$$\chi_{CO} = \left[\chi_{CO_2} - B_{CO_2}\right]\frac{(A_{CO}-B_{CO})}{(A_{CO_2}-B_{CO_2})} + B_{CO} \tag{S18}$$

$$\chi_{CO} = \chi_{CO_2}(\frac{(A_{CO}-B_{CO})}{(A_{CO_2}-B_{CO_2})} - B_{CO_2}\frac{(A_{CO}-B_{CO})}{(A_{CO_2}-B_{CO_2})} + B_{CO} \tag{S19}$$

Or more simply,

$$\chi_{CO} = \chi_{CO_2} m - B_{CO_2} m + B_{CO} \tag{S20}$$

Which gives us an expression for the intercept, b (from the linear equation y = mx+b, with $\chi_{CO}$ and $\chi_{CO_2}$ as the x and y variables), and m is the slope ($\Delta CO/\Delta CO_2$):

$$b = B_{CO} - B_{CO_2}\frac{(A_{CO}-B_{CO})}{(A_{CO_2}-B_{CO_2})} \tag{S21}$$

The expressions in Eq. S13 and Eq. S21 give us the slope and intercept in terms of the chemical characteristics of the two air masses. In a situation like the one shown in Figure S11, if we assume the period when the mole fractions are steady is a measurement of the background air mass, we can use those concentrations to give us values for $A_{CO}$ and $A_{CO_2}$; the rolling calculations also give values for the slope and the intercept.