Introduction

This document contains a description of the methodology used to bin HYSPLIT trajectories based on air mass origins in Text S1, with Table S2 and Figure S1 outlining the polygons used to test the paths of trajectories. Additionally, TCEQ sites where VOC data are collected are described in Table S1 and input variables for the Monte Carlo simulation are outlined in Table S3. Figures S2 through S7 support the results in the main text.

Text S1. In order to identify days with mostly southeasterly flow, 48 hour back trajectories were obtained from the HYSPLIT model. The start times for these trajectories were 06:00, 12:00, and 18:00 UTC each calendar day, and 00:00 UTC the following calendar day. The origins of the trajectories were binned by their passage through a series of polygons, as shown in Figure S1. The vertices of the polygons are provided in Table S2.

Back trajectories ending at the San Antonio – Old Highway 90 site were used to identify southeasterly flow when assessing long term alkane trends at San Antonio – Old Highway 90, Clute, and Corpus Christi – Hillcrest. All trajectories that passed through Polygon 1 were assumed to have continental origins and were removed. Polygon 2 was selected to represent the central Texas Coast region, roughly extending from the coastal waters southeast of Corpus Christi to the waters south of Clute. Trajectories that did not pass through this polygon were also removed, leaving generally southeasterly trajectories of maritime origin remaining. Days with 3 out of 4 southeasterly trajectories were used to compare long term alkane trends at these sites.

For the alkane emission calculation using data from Corpus Christi – Oak Park and Floresville, back trajectories ending at Floresville were removed if they passed through Polygon 1. Again, this was to remove air masses which were influenced by continental emissions prior to moving ashore. Trajectories were also removed if they did not pass through Polygon 3, which encompasses Corpus Christi and the surrounding region of the Texas Coast. Again, days with 3 out of 4 trajectories were used to quantify the afternoon alkane enhancement between Corpus Christi – Oak Park and Floresville.

Table S1. Vertices of polygons used to bin HYSPLIT trajectories as described in Text S1.

Polygon	Vertex 1	Vertex 2	Vertex 3	Vertex 4	Vertex 5	Vertex 6
1	29.16° N, 96.12° W	29.16° N, 83.00° W	31.00° N, 83.00° W	31.00° N, 103.00° W	29.50° N, 103.00° W	29.50° N, 96.12° W
2	27.04° N, 96.75° W	27.43° N, 97.07° W	28.67° N, 95.50° W	28.28° N, 95.19° W	-	-
3	27.27° N, 97.48° W	27.81° N, 98.00° W	28.16° N, 97.64° W	27.62° N, 97.12° W	-	-

Table S2. Separate Excel file showing average alkane mixing ratios (ppb) during afternoon hours at Floresville and Corpus Christi – Hillcrest during 68 days with appropriate meteorological conditions. Also shown are input meteorology from the North American Regional Reanalysis (NARR) and production data from the Railroad Commission (RRC) of Texas that were used as input for the mass balance approach and Monte Carlo simulation. Lastly, alkane emission rates and methane emissions relative to production are provided for each day.

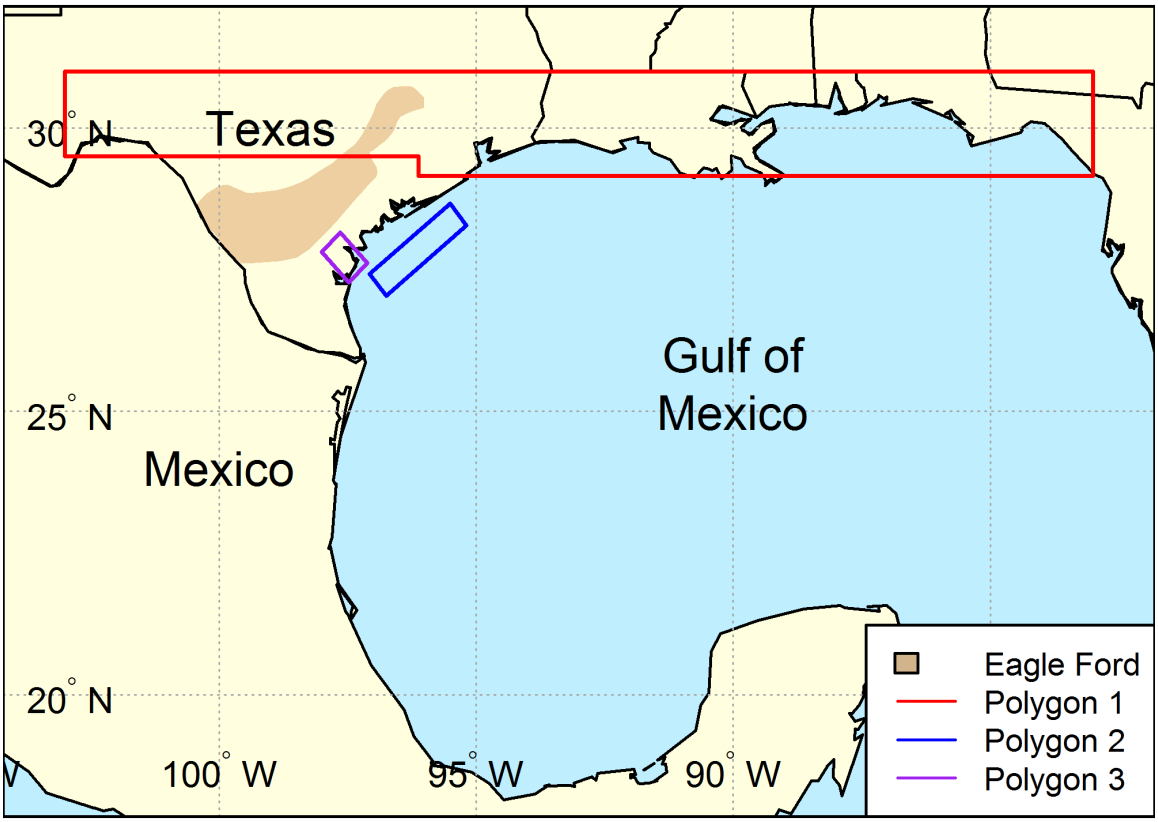

Figure S1. Polygons used to identify trajectories as southeasterly with maritime origins. See Text S1 for a description of the methods and Table S2 for corners of the vertices.

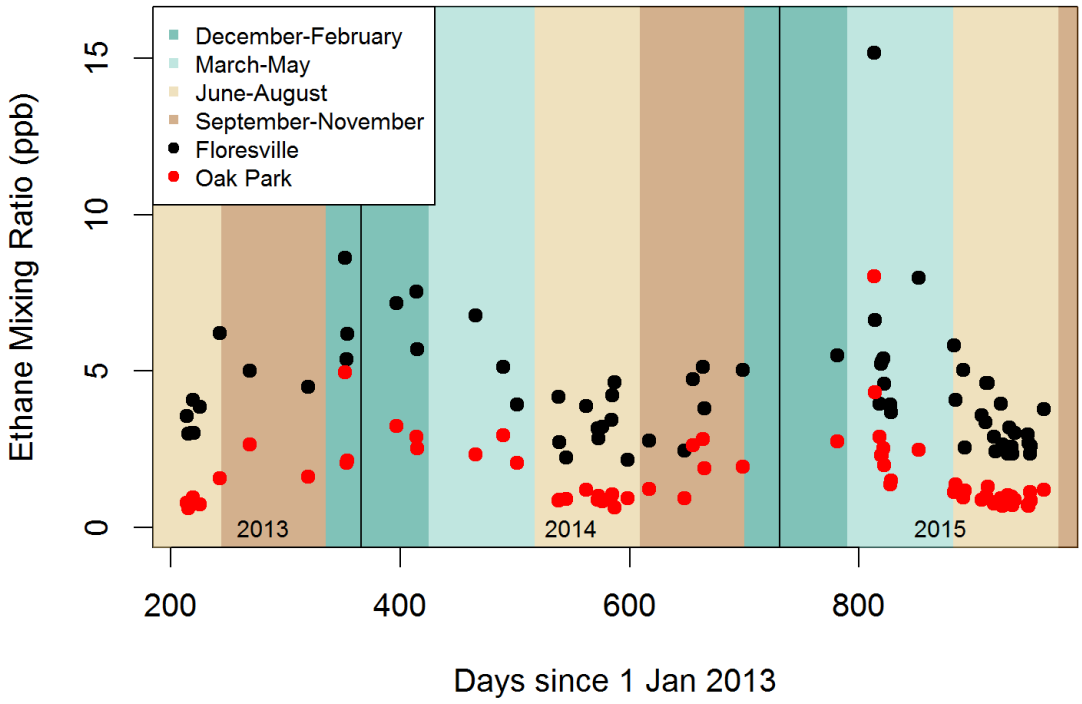

Figure S2. Scatterplot of afternoon ethane mixing ratios at the upwind site (Oak Park), the downwind site (Floresville) over 68 days with southeasterly flow. Note that each mixing ratio has a relative uncertainty of ±5.8% and the uncertainty of the enhancement is equal to the sum of the uncertainties of the upwind and downwind sites. The background colors show the warm and cool seasons. Ethane mixing ratios are generally lower at both the upwind and downwind sites during the summer and fall and higher in the winter and spring.

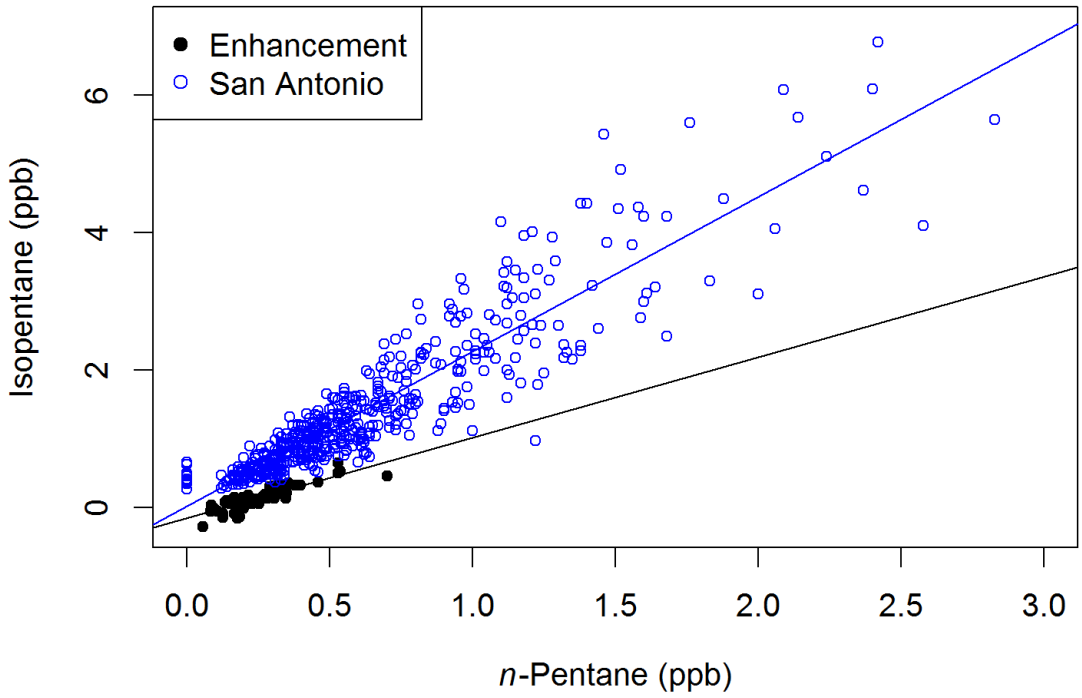

Figure S3. Scatterplot of afternoon isopentane and *n*-pentane enhancements observed between Oak Park and Floresville over 68 days with southeasterly flow. Also shown are isopentane and *n*-pentane mixing ratios in San Antonio for 24 hour canister samples from 2007 to 2015. The slopes of both linear regressions are highly statistically significant ($p < 0.001$). The isopentane-to-n-pentane ratio in San Antonio is 2.25, which is indicative of urban emissions. Meanwhile, the slope of the ratio for the enhancements is 1.17, indicating emissions from petroleum production.

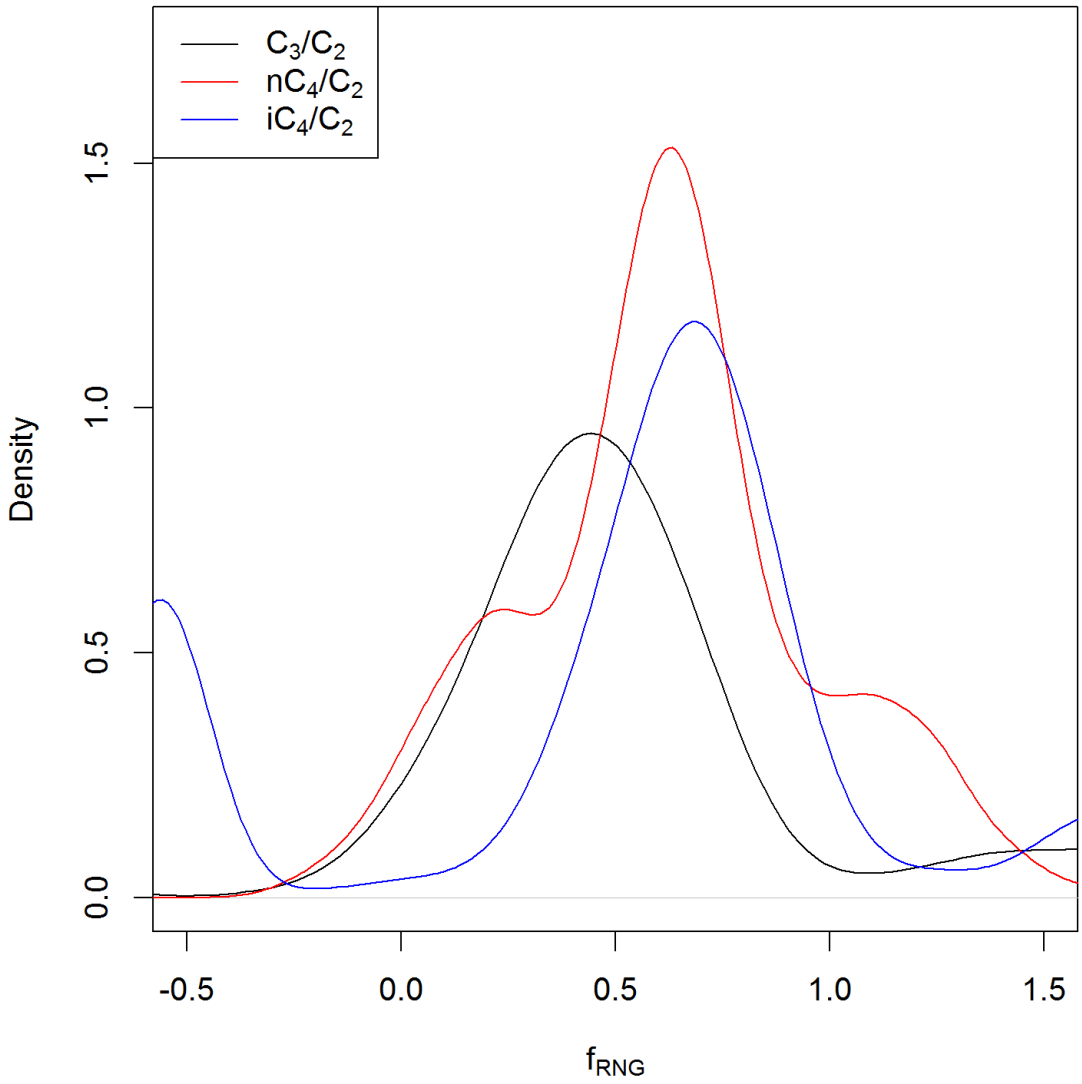

Figure S4. Probability distribution functions for the fraction of observed alkane ratios that can be explained by emissions of raw natural gas (f_{RNG}). Note that the x-axis extends to numbers that are physically unreasonable. This is due to the assumption that alkane ratios can be explained by emissions from raw natural gas and vented tank gas alone (the compositions of which are highly uncertain), and other sources and sinks are neglected. Nonetheless, there is general agreement between the fractions derived from the C_3/C_2 ratio, the nC_4/C_2 ratio, and the iC_4/C_2 ratio, suggesting that the above assumption creates results that are reproducible using these three alkane ratios.

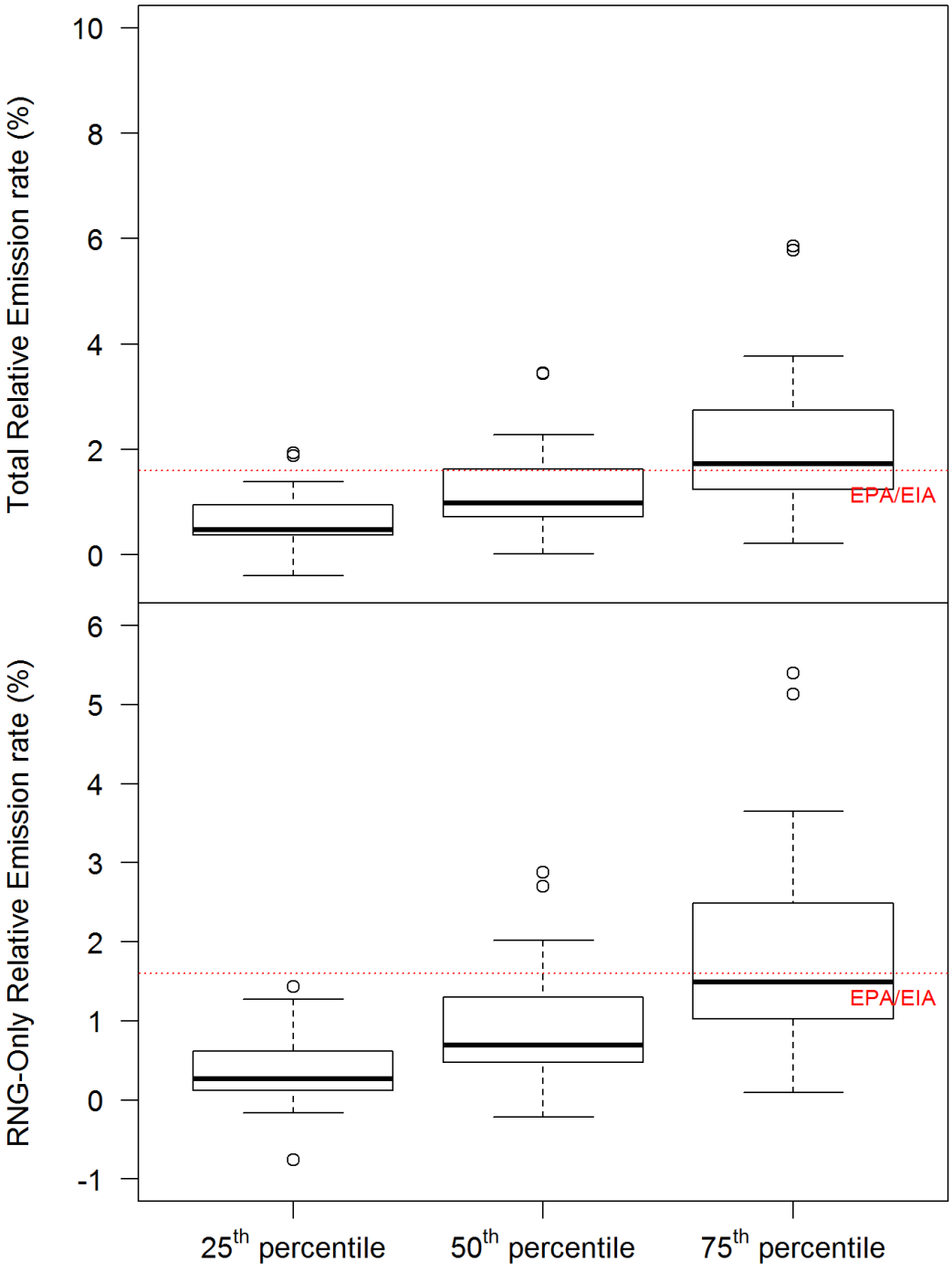

Figure S5. Distributions of the 25th, 50th, and 75th percentile of relative emissions over all 68 days. Methane emissions were converted to a volume of natural gas using methane to natural gas ratios (Table 2 in main text) and compared to natural gas production. The top panel shows total relative emission rate while the bottom panel shows only emissions from raw natural gas (RNG) sources and excludes emissions from liquid storage tanks. Some outliers of the lower

bound were less than zero due to large uncertainties in the methane enhancement which occasionally produced a negative methane flux within the Monte Carlo simulations. The upper bound of the total emission rate is often close to that of the EPA/EIA emission estimate for nationwide natural gas and associated gas production in 2011. The emission rate for raw natural gas emissions alone was generally lower than the EPA/EIA estimate. Note that the emission rate estimate has a slightly skewed distribution, with the upper bound possibly influenced by plumes from large emitters.

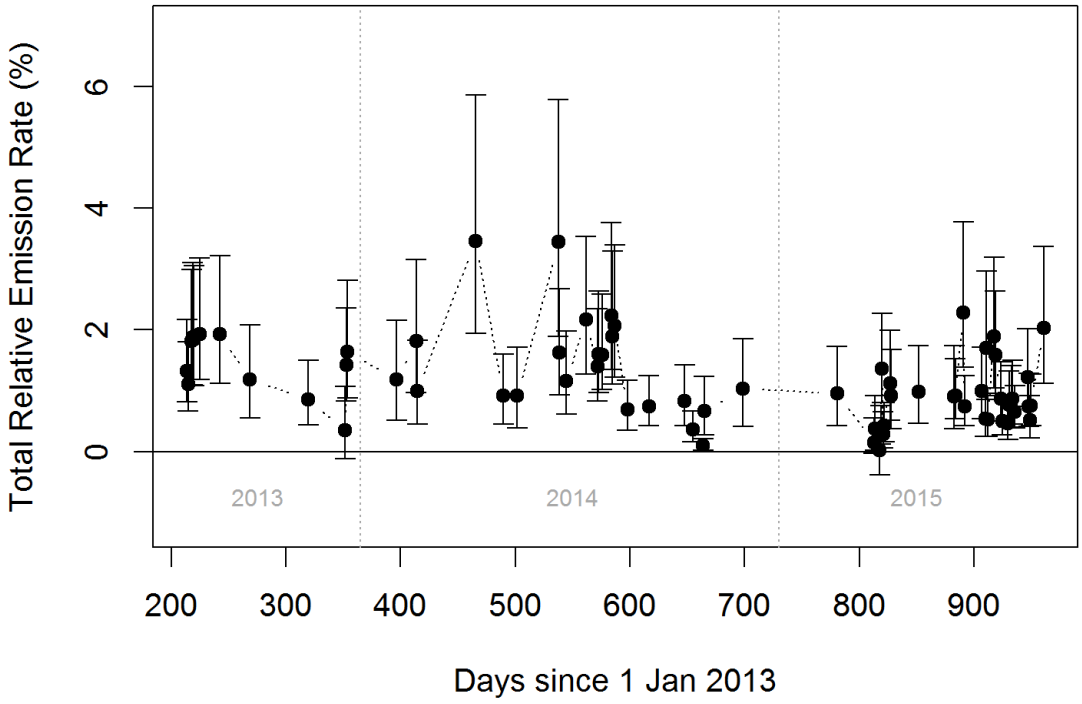

Figure S6. Timeline of the median emission rate for each day with the interquartile range represented by whiskers. The emission rate showed neither apparent seasonality nor trend over time, which is to be expected of a continuous emission source that does not depend on meteorological variables.

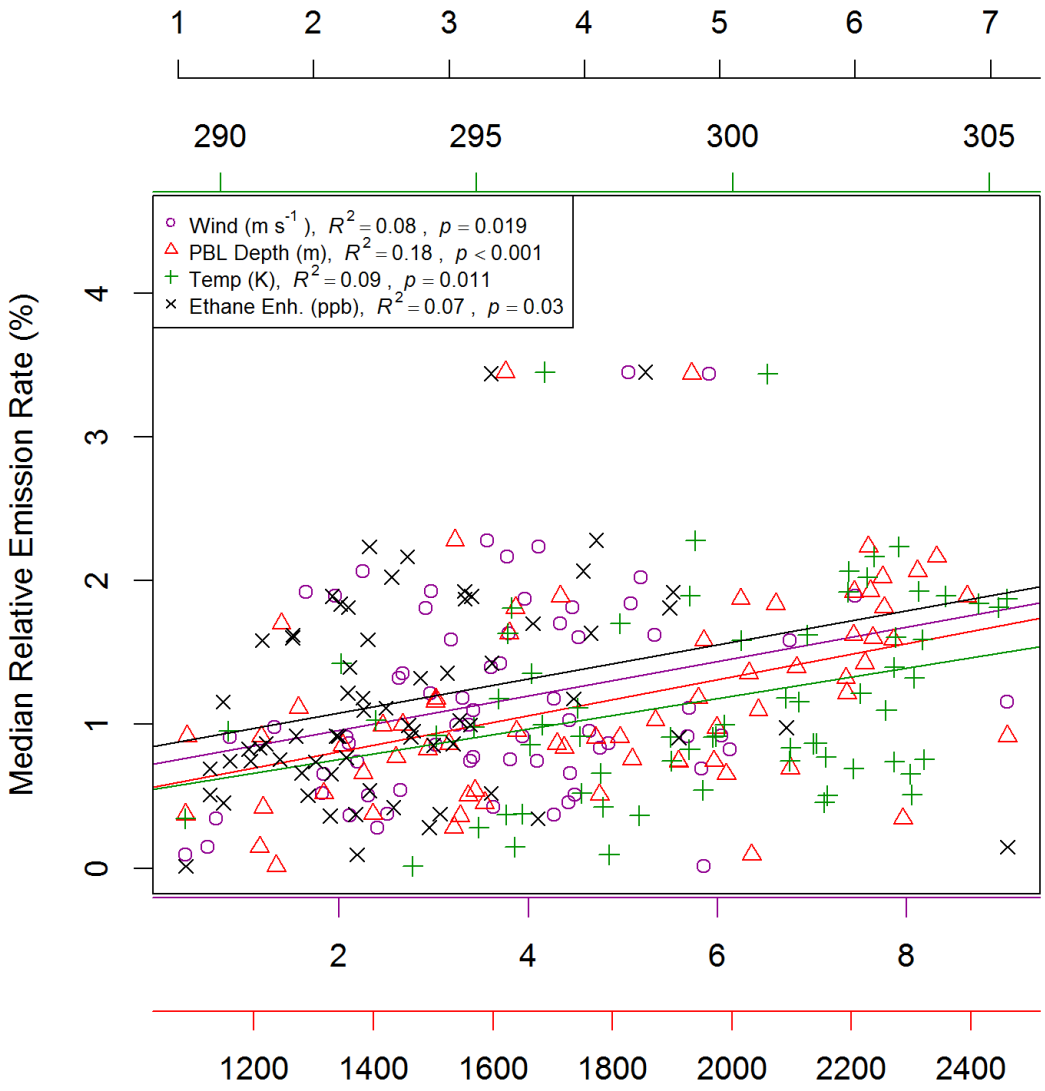

Figure S7. The median emission rate for each of the 68 days plotted against wind speed (component parallel to the transect between Corpus Christi and Floresville), PBL depth, temperature, and ethane enhancement. While there were statistically significant correlations with wind speed, temperature, and ethane enhancement, these correlations were weak. While the PBL depth showed the strongest correlation with the emission rate ($R^2 = 0.18$, $p < 0.001$), the emission rate was not strongly driven by meteorological variability.

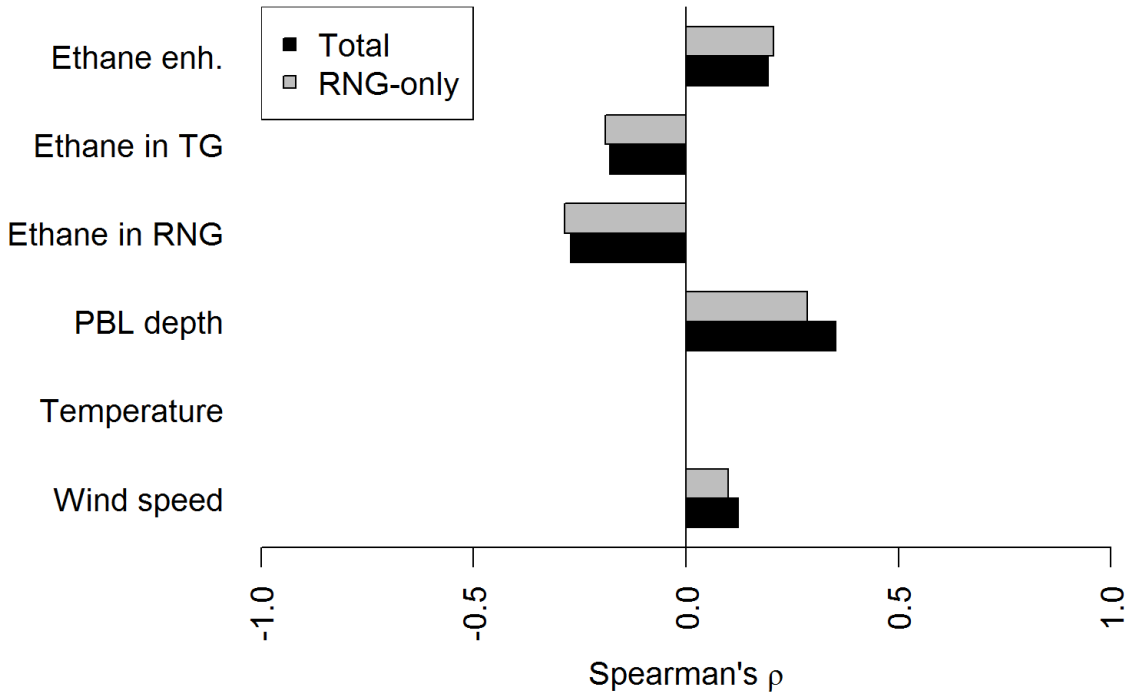

Figure S8. Example of a tornado plot from a Monte Carlo simulation for 2 August 2013. The total emission rate depended largely on the composition of raw natural gas (RNG), with lower ethane content in natural gas (which was used, in part, to estimate methane emissions) resulting in a higher emission rate. The RNG-only emission rate shows a negative correlation with the ethane content in both RNG and vented gas from liquid storage tanks (TG), as these numbers were used to partition emissions between RNG and TG sources. Both the total and RNG-only emission rates showed positive correlations with wind speed and planetary boundary layer (PBL) depth, but not temperature. The rates were also positively correlated with the ethane enhancement.