HEI Research Report 240

Predictive, Source-Oriented Modeling and Measurements to Evaluate Community Exposures to Air Pollutants and Noise from Unconventional Oil and Gas Development

Lea Hildebrandt Ruiz et al.

Additional Materials D: Chapter 6

Correspondence may be addressed to Dr. Lea Hildebrandt Ruiz, The University of Texas at Austin, 200 E. Dean Keeton St., Austin, TX 78712; email: *lhr@che.utexas.edu*.

Although this report was produced with partial funding by the United States Environmental Protection Agency under Contract No. 68HERC19D0010 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and may not reflect the views of the Agency; thus, no official endorsement by the Agency should be inferred. This report also has not been reviewed by private party institutions, including those that support HEI Energy, and may not reflect the views or policies of these parties; thus, no endorsement by them should be inferred.

ACS ES&T Air

Additional Materials D: Chapter 6

pubs.acs.org/estair

Article

Source Attribution of Elevated Ethane Concentrations Detected by Regional Monitors in Oil and Gas Production Regions

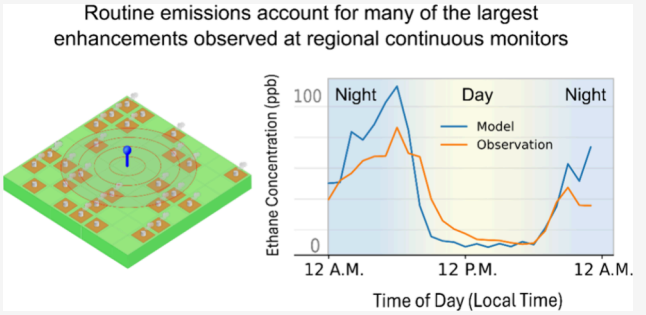
Joel D. Graves, Yosuke Kimura, Mrinali Modi, Shannon Stokes, Morgan Meyer, Lea Hildebrandt Ruiz, and David T. Allen*

Cite This: *ACS EST Air* 2025, 2, 2038−2046

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Measurements of ambient ethane concentrations at a regional air quality monitor in the Eagle Ford oil and gas production region are compared to concentrations predicted using site-level hydrocarbon emission inventories coupled with a Gaussian puff dispersion model (CALPUFF). To account for more than half of mean concentrations due to routine emissions, sites at distances 20−50 km from the receptor site were included in the simulations. Nearly all of the highest observed concentrations were observed at night. For each night in the simulation, the location and magnitude of the maximum predicted concentration and maximum observed concentration were compared, and approximately two-thirds of the highest observed nighttime maximum concentrations were accounted for by routine emissions. In contrast, approximately a third of the highest daytime maxima could be accounted for by routine emissions. Most of the large observed maxima that are attributable to routine emissions are predicted to be caused by sources that were within 10 km of the receptor site, but sources up to 20 or more kilometers from the receptor also contributed to the predicted concentrations. A case study is provided demonstrating the potential of coupling site-level inventories of routine emissions with dispersion modeling for attributing sources of elevated hydrocarbon concentrations.



KEYWORDS: *dispersion modeling, Eagle Ford, emission events, oil and gas emissions, regional monitors*

INTRODUCTION

Ambient concentrations of hydrocarbons in oil and gas production regions show considerable temporal variability.[1,2] Several studies have reported that emission events with large instantaneous emission rates can contribute significantly to total methane emissions from oil and gas sources, leading to complex spatial and temporal variability in ambient hydrocarbon concentrations.[3−5] For example, Cusworth et al.[6] estimated that across multiple oil and gas production regions, sources with instantaneous observed emission rates greater than 10 kg/h accounted for 40% of regional methane emissions. These emission events, which include methane and other hydrocarbons, can be short in duration, often lasting for less than an hour or a day.[7] Additional evidence suggests that the distribution of large emission events is not uniform throughout the day. For example, in an analysis of aircraft-based measurements made in an oil and gas production region in Colorado, Zimmerle et al.[8] concluded that roughly half of large emission events were likely due to maintenance activities. Because operations such as maintenance events occur disproportionately during working hours, there can be diurnal variations in the frequencies of emission events. This complex temporal variability in emissions makes source attribution difficult.

Fixed-point continuously operating monitors which sample ambient concentrations of trace gases with high frequency have become a widely used method for monitoring site-level emissions and performing source attribution in the upstream and midstream segments of the oil and gas sector.[9] These monitors have the potential to provide source attribution information, however, conventional platforms have typically focused on monitoring individual sites using a relatively dense network of sensors (e.g., multiple sensors per site). These configurations have demonstrated promising performance for emissions identification and equipment-level source attribution but scaling these systems to the spatial density necessary to accurately capture the impact of intermittent, large emission events in both time and space would involve large numbers of sensors.[9−11]

Continuously operating monitors with large spatial footprints could greatly reduce the cost of high-frequency regional

Received: June 23, 2025
Revised: July 25, 2025
Accepted: July 28, 2025
Published: August 5, 2025

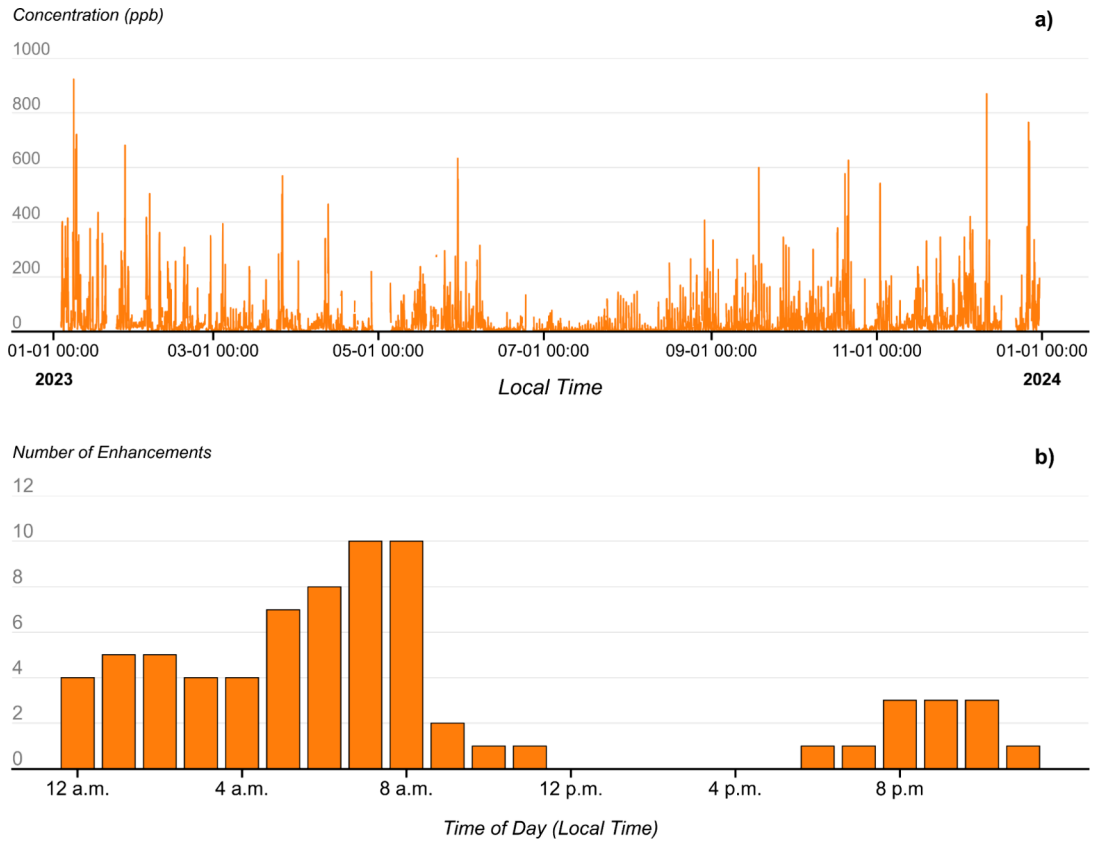


Figure 1. (a) Year-long time series for ethane at a ground site measuring hydrocarbons with hourly time resolution in the Eagle Ford oil and gas production region;[12] (b) Counts of concentrations above the 99th percentile of all hourly averaged concentrations observed at the measurement site in 2023 binned by hour of day.

emissions monitoring by decreasing the number of sensors necessary for comprehensive coverage of oil and gas assets in large production regions. A potential option for improving the spatial coverage of these monitors is to place them further from facility boundaries. Concentration enhancements from emissions at on-site equipment are generally well above background levels when observed by continuously operating monitors deployed along the site boundary (typically ≤ 100m from an individual on-site source). Continuously operating monitors placed further from sites are less influenced by nearby sources allowing them to sample larger effective areas.

To quantitatively characterize emission events detected at continuously operating monitors, the effect of diurnal changes in meteorology on the measurement signals must be separated from the effect of changes in emissions. With on-site sensors, detected emission events are typically more straightforward to identify because the signal is usually greatly enhanced over the background. For regional monitors, this is unlikely to be the case except for very large emission events. This complicates the analysis of measurements from these monitors as observations of large concentration enhancements may also be produced by routine emissions from regional sources during periods with low wind speeds or low mixing heights. For example, Figure 1a shows a typical time series for hourly averaged ethane concentrations recorded by a regional ground monitoring site in the Eagle Ford Shale oil and gas production region in south central Texas.[12] As shown in Figure 1b, peak concentrations, defined as the 99th percentile of observed concentrations, were almost exclusively recorded during early morning hours. The high incidence of elevated ethane concentrations observed in the early morning hours could be attributed to either meteorological effects coupled with routine emissions or oil and gas emission events, but it is unclear, just from the time series, how to accurately characterize the relative importance of routine emissions and large emission events.

This work uses emissions modeling of thousands of oil and gas sources coupled with a Gaussian puff dispersion model (CALPUFF) to evaluate whether the causes of highly elevated concentrations at regional monitors are due to large emission events, routine emissions coupled with wind speeds and atmospheric stability characteristics that lead to elevated concentrations, or a mixture of these causes. The focus of this work will be on source attribution for ethane because it is commonly used as a tracer for oil and gas emissions as it has few significant biogenic sources.[13,14] A variety of species could be considered in these analyses, but results for ethane are used to demonstrate the analysis methods. Ethane observed in the regional measurements is assumed to be produced from either routine or nonroutine emission events from oil and gas sources. This is the first attempt to evaluate signals from regional continuous monitors to distinguish between concentration enhancements produced by routine and nonroutine emission events from oil and gas sources using detailed, site level emission inventories coupled with site level dispersion modeling.

METHODS

Hydrocarbon Concentration Data. Ethane concentrations used in this work were measured by an automated gas chromatograph (auto-GC) operated by the Texas Commission

https://doi.org/10.1021/acsestair.5c00235
ACS EST Air 2025, 2, 2038−2046

on Environmental Quality (TCEQ) in the Eagle Ford Shale oil and gas production region in south central Texas. In 2023, oil production from the Eagle Ford Shale averaged 980,000 barrels per day (7.6% of U.S. production) and gas production averaged 5.5 billion standard cubic feet per day (4.4% of U.S. production).[15−17] The auto-GC is operated at a TCEQ Continuous Ambient Monitoring Station (CAMS) located in Karnes City (CAMS 1070; "measurement site") and is surrounded on all sides by sites associated with oil and gas production (Figure 2). Karnes City is in a county that typically

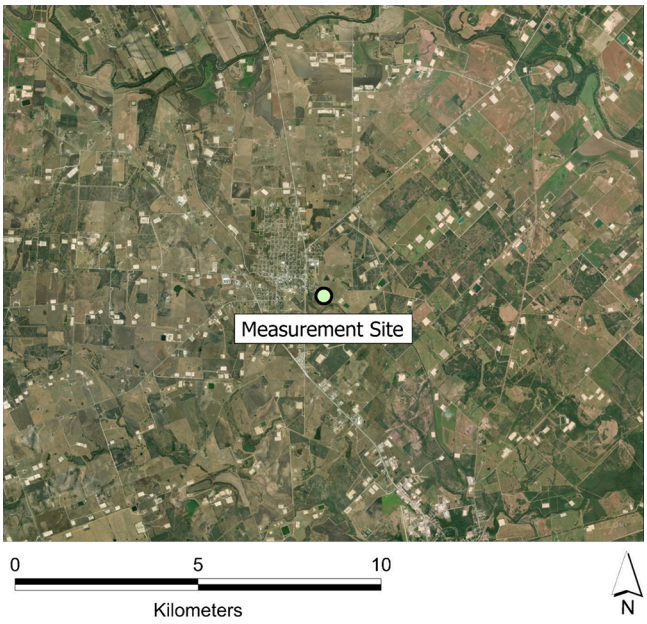


Figure 2. Satellite image of the region within approximately 15 km of the Karnes City auto-GC measurement site[19] Light-colored areas with little vegetation are predominately operations associated with oil and gas. Basemap data provided by Esri.[19]

has some of the highest oil and gas production rates in the Eagle Ford Shale. The auto-GC operates continuously, with a 1 h sampling cycle. Details are available from the TCEQ.[12] Briefly, air collected at approximately 3 m above ground level is continuously drawn over a sorbent, which captures hydrocarbons, over a 40 min period. At the end of the 40 min sample collection period, the sorbent is rapidly heated, releasing the adsorbed hydrocarbons, which are then analyzed in a gas chromatograph. The analysis quantifies the concentrations of approximately 40 different hydrocarbon species. These concentrations are reported on the TCEQ Web site. This work focuses on ethane, which is frequently used as a tracer for emissions from oil and gas operations because it has few biogenic sources.[13,14,18] This simplifies the comparison of the modeling results with observations by focusing on species primarily emitted from oil and gas sources.

Modeling Domain. A 212 km × 208 km domain centered on the auto-GC site, shown in Figure 3, was used for emissions and dispersion modeling. The domain includes three nested regions: an outer modeling domain, divided into 4 km × 4 km grid cells, an intermediate rectangular 48 km × 44 km region divided into 1.333 km grid cells, and an inner modeling domain, spanning 32 km × 32 km, all centered around the auto-GC site (Figure 3).

In the 32 km × 32 km region closest to and centered on the auto-GC site, 3208 individual oil and gas well sites were identified using well location and production data.[20] A majority of the well sites in the region contain very limited equipment, and because most emissions in the production segment are associated with liquids handling equipment, emissions associated with oil, condensate and produced water production from each well were aggregated to 327 tank batteries at centralized liquids handling facilities in the 32 km × 32 km region. These facilities were identified visually with Google Earth imagery. Wells were aggregated to the tank batteries based on the proximity of tanks. All wells that had tanks on the same pad were routed to the tank(s) on site. All tank batteries were assigned an operator based on the operator of the nearest well. Wells without on-site tanks were then mapped to the tank batteries based on the operator and the nearest tank battery. Manual checks were conducted to minimize large discrepancies between the number of tanks and the number of wells mapped to them.

Oil and gas well sites located in the outer two modeling regions were aggregated into point sources representing multiple sites. In the rectangular 48 km × 44 km domain, beyond the inner 32 km × 32 km domain, individual sites within each 1.33 km × 1.33 km grid cell were aggregated and represented as a point source located at the centroid of the well sites within the grid cell. This region included a total of 1782 well site locations, aggregated into centroid locations in 319 grid cells. Well sites located within the outer 212 km × 208 km rectangular domain, but outside of the inner and intermediate domains, were aggregated into point sources representing multiple sites in larger grid cells. All individual sites within each 4 km × 4km grid cell were aggregated and represented as a point source located at the centroid of the sites within the grid cell. This region included a total of 15,520 well site locations, aggregated into centroid locations in 974 grid cells.

The choice of nesting and spatial aggregation was based on a balance between computational intensity and accuracy. Dispersion modeling simulations performed on sites within the region, described in Supporting Information (SI) section S8, predicted that only 38% of average concentrations observed at the auto-GC site, due to emissions from oil and gas sites within 100 km, were accounted for by oil and gas sites within 5 km of the auto-GC. Oil and gas sites within 10, 20, and 50 km of the auto-GC site accounted for 67, 88, and 98% of average ethane concentrations, respectively (SI, Figure S11). This indicated that the large numbers of sites at significant distances from the auto-GC site could have a measurable impact on average concentrations. In contrast, as the distances between sources and the auto-GC site increased, the exact location of the individual sources became less important in predicting concentrations, so emissions from multiple nearby sites were aggregated into single point sources.

Emission Estimation. Routine emissions (i.e., emissions not associated with process disruptions or equipment malfunctions) were modeled for upstream and major midstream sources in the oil and gas sector. Upstream (well site) emissions were estimated using the Methane Emission Estimation Tool (MEET)[3] for all individual wells. Midstream emissions were estimated at the facility level using throughput-scaled methane emission factors[21,22] and facility-specific emission composition estimates based on throughput compositions.

https://doi.org/10.1021/acsestair.5c00235
ACS EST Air 2025, 2, 2038−2046

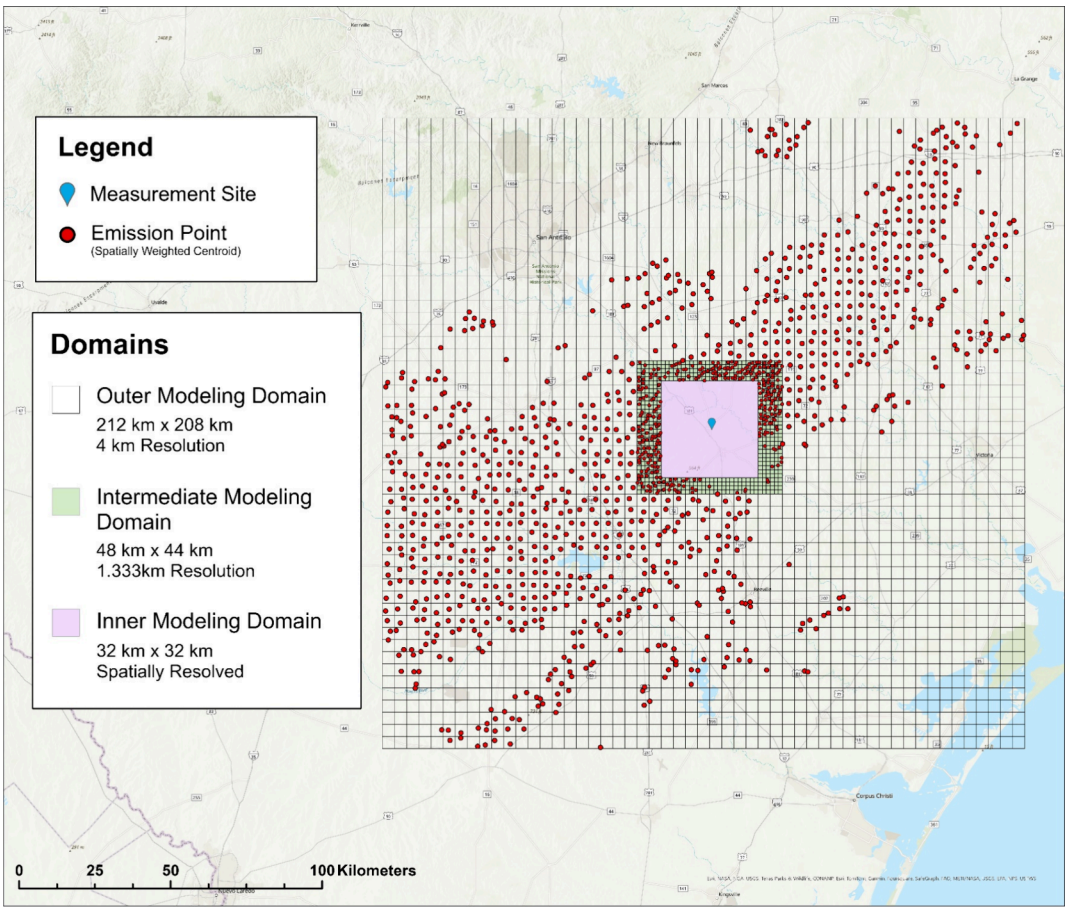


Figure 3. Nested modeling domain. Basemap data provided by Esri.[19]

Upstream Emissions Estimates. MEET simulates methane and other hydrocarbon emissions from various source categories using activity and emission factors that can be either user-defined, calculated by MEET, or selected from MEET's default modeling settings. Using these activity and emission factors, MEET then calculates an emissions time series for each equipment category at a user-supplied temporal resolution.[3] For this work, ethane emissions from seven source categories representative of upstream oil and gas facilities were modeled. These categories are

1. Pneumatic controllers (PC)
2. Chemical injection pumps (CIP)
3. Equipment leaks
4. Condensate tank flashing
5. Liquids unloading
6. Water tank flashing
7. Well Completion

Information about the inputs and modeling configuration used in this work for each of these seven source categories is provided in Supporting Information (SI).

Gathering Compressor Stations. Emissions from gathering compressor stations were estimated based on whole gas throughput rates and compositions at each facility. In this approach, a throughput-scaled emission factor (kg CH_4 emitted per kg CH_4 in throughput) is estimated at each facility using a linear fit of methane emissions as a function of whole gas throughput for gathering compressor facilities reported by Zimmerle et al.[22] The throughput-scaled emission factor (mass fraction emitted) is multiplied by the ethane throughput to estimate ethane emissions. Whole gas throughput rates and compositions were attributed to each facility as the sum of the produced gas from the nearest wells. Additional information on gathering compressor emissions estimates is provided in the SI.

Natural Gas Processing Plants. Emissions from natural gas processing plants were estimated on a facility-level basis using whole gas throughput rates and a throughput normalized emission factor in an approach similar to the approach used for estimating gathering compressor station emissions. An average throughput normalized facility-level emission rate was used to estimate ethane emissions from natural gas processing plants based on ethane throughput at each facility.[21] Whole gas throughput estimates for individual facilities were calculated as the sum of the produced gas streams from the nearest wells to that facility. Additional information on natural gas processing plant emissions estimates is provided in the SI.

Emissions Speciation. MEET predicts speciated emissions for each equipment category at upstream facilities based on the composition of one of four process streams associated with the production separator: the wellstream, the separator overhead (produced gas), the separator exit to condensate tanks, and the separator exit to water tanks. In this work, speciated emissions from midstream facilities were estimated as a composite of the produced gas streams from each of the multiple wells associated with the facility. A summary of process streams and associated equipment is provided in Table 1.

https://doi.org/10.1021/acsestair.5c00235
ACS EST Air 2025, 2, 2038−2046

Table 1. Production Separator Process Streams and Associated Production Site Equipment[a]

Process Stream	Associated Equipment
Wellstream	Liquids Unloading
	Well Completion
Separator Overhead (Produced Gas)	Pneumatic Controllers
	Leaks
	Chemical Injection Pumps
	Gathering Compressor Stations
	Natural Gas Processing Plants
	Flares
Separator Exit to Condensate Tank	Condensate Tank Flash
Separator Exit to Water Tank	Water Tank Flash

[a]MEET estimates equipment-specific emissions compositions based on these associations.

MEET can estimate process stream compositions from operational and production data using a thermodynamic model of the production separator provided in the accompanying Emissions Composition Tool (ECT).[23] To estimate the operational parameters at each facility, emission sources were categorized by their location within characteristic production regions in the Eagle Ford Shale.[24] For each region, representative operating conditions and production compositions were tabulated based on composition reports available for that region.[25,26] These representative parameters are then applied to all sources within the sub-basin. Additional information on operational and production data inputs for the emission speciation can be found in the SI.

Dispersion Modeling. Dispersion modeling was performed using CALPUFF v7.2. Meteorological modeling inputs with 4 km horizontal spatial resolution and 10 vertical layers were generated at a1-h temporal resolution using CALMET v6.5. The primary meteorological inputs to CALMET were regional surface observations of wind and temperature (5 min temporal resolution) from the Texas Commission on Environmental Quality (TCEQ) CAMS 1070 site in Karnes City. Surface-based meteorological observations from 21 National Weather Service stations (6 ASOS 1 min stations, 15 ISD hourly stations) were also incorporated (SI Table S22). Hourly vertical profiles of meteorological variables (e.g., wind speed, temperature) from the high-resolution rapid refresh (HRRR) atmospheric model were extracted for eight horizontal grid cells and used as upper-air data inputs. The modeling period selected was March 5, to May 24, 2023 (local time) to coincide with a three-month measurement campaign being conducted in the Eagle Ford Shale production region at this time. Dispersion of intermittent and continuous ethane emission sources was modeled. Concentration enhancements at the receptor location were simulated with 1 h resolution.

Monte Carlo Analysis. MEET provides representative estimates of the behavior of oil and gas emissions for specific equipment types through process modeling, but the start times and durations of intermittent emission events are uncertain. To account for these uncertainties, multiple simulations with varying assumptions about the beginning and end of intermittent emission events (e.g., the start time and durations for liquid unloadings) and the location and timing of leak events were prepared as 10 Monte Carlo emission time series. These emissions time series were coupled with individual CALPUFF runs and the results were used to generate distributions of predicted ethane concentrations at the Karnes City measurement site associated with each source site at each hour. Individual CALPUFF runs were necessary as the location and timing of emission events changed between Monte Carlo simulations. After all simulations were run, the median concentration and range of concentrations at each point in time from the 10 resulting concentration time series were used for analysis, unless otherwise noted. Ten Monte Carlo simulations were used because of the significant computational demands required for each set of simulations.

RESULTS AND DISCUSSION

The analyses in this work compare predicted and observed concentrations of hydrocarbons emitted by oil and gas operations. A variety of species could be considered, but results for ethane are used to demonstrate the analysis methods. Ethane in this region is emitted primarily from oil and gas sources with limited interference from other anthropogenic and biogenic sources. Ethane is also observed at relatively high concentrations compared to other hydrocarbons in the region.[27] Sources from up to 100 km from the

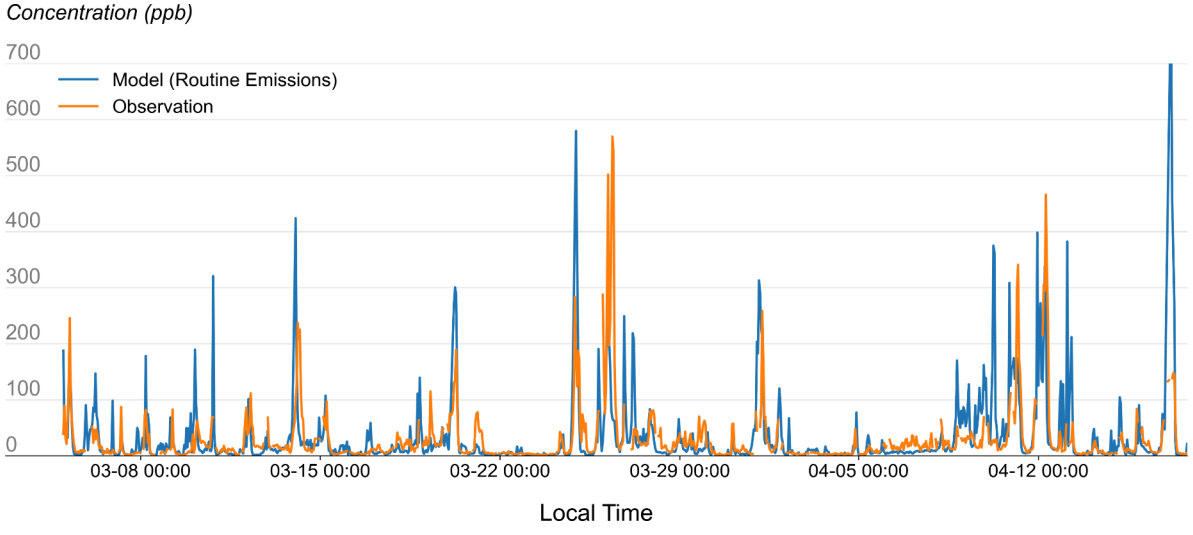


Figure 4. Predicted and observed ethane time series during one month of the modeling period.

https://doi.org/10.1021/acsestair.5c00235
ACS EST Air 2025, 2, 2038−2046

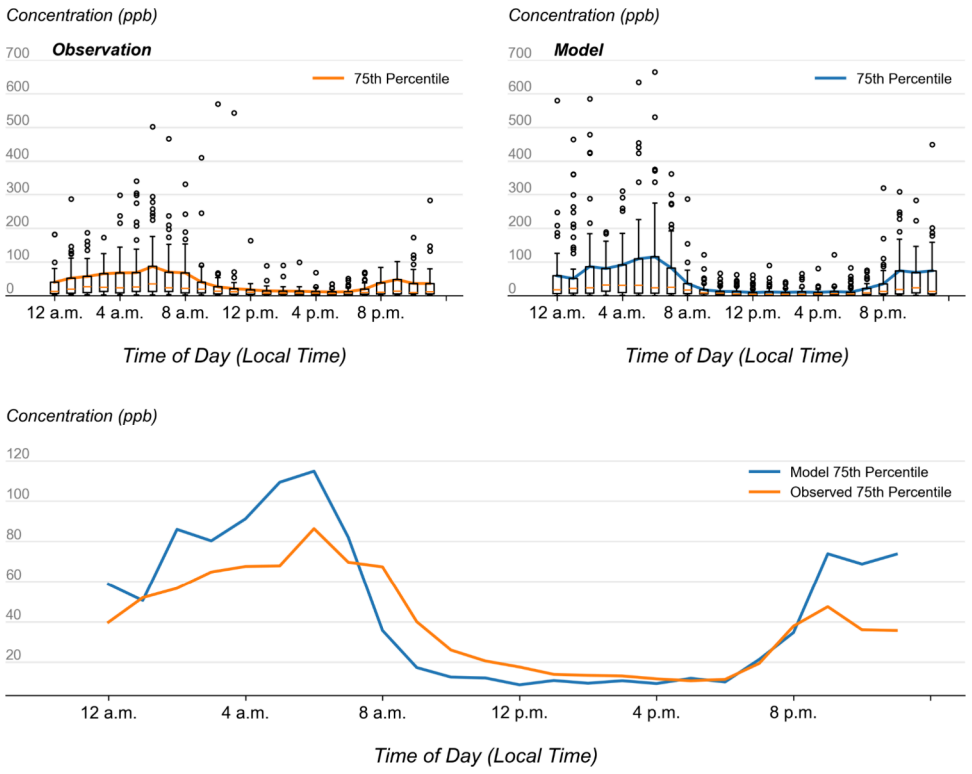


Figure 5. : Time of day distributions for predicted and observed data sets.

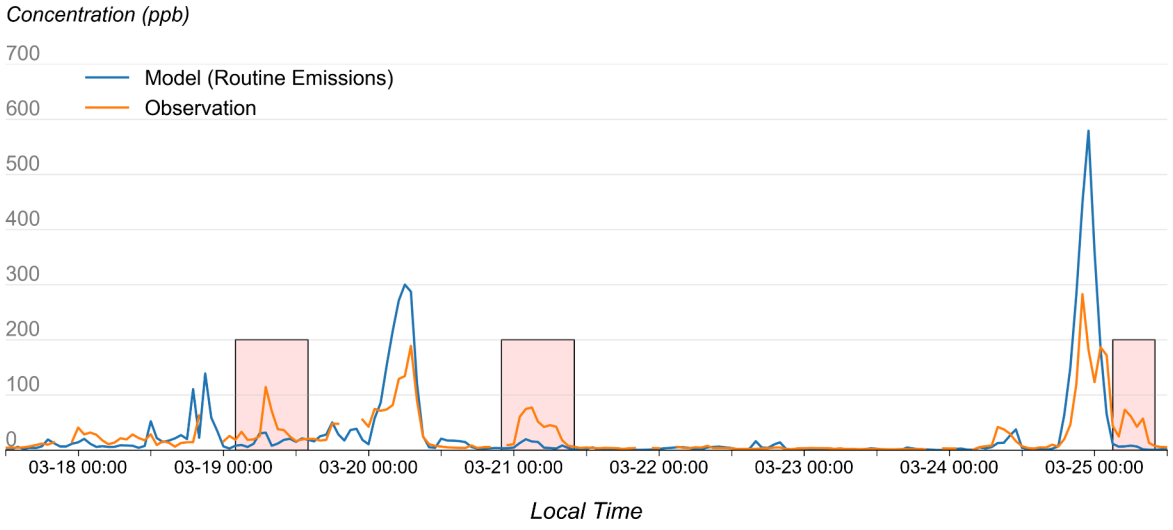


Figure 6. One week modeling period with multiple unmatched observations. The orange line shows the observed data from the Auto-GC, the blue line shows the model predictions using routine emissions, and the pink shaded boxes highlight "unmatched" concentration enhancements (i.e., prominent peaks in the observed data set that the model does not predict).

receptor were included to capture a majority of mean ethane emissions at each hour of the modeling period. Analyses of source contributions from different distances are provided in SI Section S8.

Figure 4 shows one month of ethane concentrations predicted at the Karnes City measurement location by the coupled routine emissions and dispersion modeling. Observed ethane concentrations for the same time period are overlaid. Qualitative agreement between the timing and magnitude of many of the enhancement events is observed. Figure 5 shows the distributions of predicted and observed ethane concentrations for each hour of the day over the course of the three-month modeling period. The reproduction of the diurnal trend suggests that a majority of the largest ethane concentration enhancement episodes observed and predicted during this time period are the result of night-time meteorological conditions coupled with routine emissions.

Several additional analyses of model performance using statistical metrics (root-mean-square, fractional bias, and correlation, among others) and data exploration techniques (quantile-quantile plots, parity plots) are provided in SI Section S9. Notably, a linear covariance is observed between the observed and predicted enhancements when plotted as log–log parity plots (SI Figures S14 and S17). The ranked and

https://doi.org/10.1021/acsestair.5c00235
ACS EST Air 2025, 2, 2038−2046

Table 2. Summary of Episode Screening Analysis

Day			Night		
Total Episodes Identified	Top episodes used for analysis	Maximum enhancement range (min-max; ppbV)	Total Episodes Identified	Top episodes used for analysis	Maximum enhancement range (min-max; ppbV)
6	6	17.8−569.3	63	10	209−466

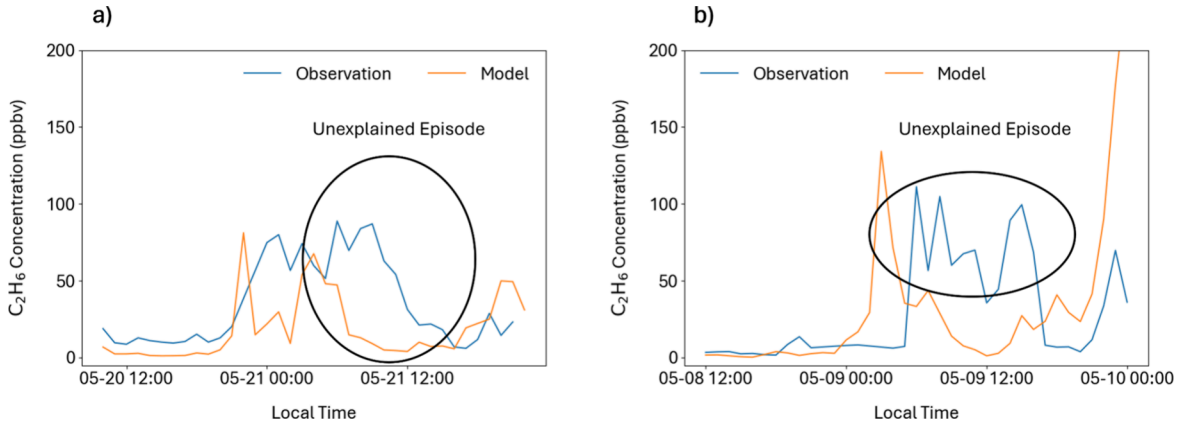


Figure 7. Manually identified unexplained enhancements.

unranked correlations (Spearman's ρ and Pearson's r respectively) between these variables (SI Table S25) suggest a monotonic, but not necessarily linear, relationship exists (i.e., the ranked correlation is much stronger than the unranked correlation). This indicates that the timing of the observed enhancement episodes and their relative rankings are reasonably well predicted by the model, even if the magnitude of the enhancements may be over or underestimated. This behavior may have several explanations including, among others, uncertainty in the emissions and dispersion modeling. To quantify the influence of these uncertainties on the modeling predictions, uncertainty and sensitivity analyses were conducted using several different meteorological and emissions scenarios. The details of these analyses are provided in SI Sections S9 and S10.

While many periods of elevated ethane concentrations are accounted for by routine emissions, several periods are not. Figure 6 shows a one-week period with observed concentration enhancements that are unmatched in the predicted data. These "unmatched" enhancement episodes are presumed to be driven by nonroutine emission events. To systematically identify observations of elevated concentrations that were not predicted by the simulations, a gradient-based method for identifying enhancement episodes in the observed and predicted enhancement timeseries, based on methods developed by Daniels et al., was employed.[10] Briefly, the observed and predicted concentration time series are individually marched in time until encountering a concentration gradient between successive hours that is above a specified threshold; at this point, an episode begins. The time series continues to march in time, and the maximum enhancement encountered in the episode is continuously updated until encountering an absolute concentration (ppb) that is some fraction of the maximum enhancement in the episode, at which point the episode ends. It is common for the auto-GC to have missing data for a few hours each night when the instrument undergoes routine calibration. In the instance that the episode identification algorithm is in an in-episode state when encountering these hours, it will skip over the missing hours and continue marching in time to the nearest hour with observational data. These hours are included in the episode. In this analysis, the threshold for beginning an episode was set at a gradient of 10 ppb per hour (ethane) and the exit threshold was set at 30% of the maximum enhancement in the episodes. Using this analysis, 69 episodes were identified in the 81-day analysis period. After episodes are identified, the overall modeling period is split into individual daytime (hourly sampling beginning at 9 A.M. − hourly sampling beginning at 5 P.M., 9 h) and nighttime (6 P.M. − 8 A.M., 15 h) intervals and episodes are categorized as daytime or nighttime based on the location of their maximum concentration. For example, an episode occurring between the hours of 3 A.M. and 11 A.M. with a maximum concentration observed at 5 A.M. would be considered a nighttime enhancement. Using the peak identification thresholds, 6 daytime episodes and 63 nighttime episodes were identified. This is representative of a broader trend in both the observed and predicted data sets in which nearly every night experiences a relatively significant enhancement episode while daytime episodes are less common.

The focus of the analysis is on large observed concentrations, so the top episodes from each time-of-day category (day or night) were evaluated to determine how frequently these episodes could be explained by the model predictions. Because only 6 daytime episodes were identified, each of these episodes was considered in the matching analysis. For the nighttime episodes, the maximum concentration from each observed episode was ranked compared to all the other nighttime episode maxima and top 10 episodes were selected for further analysis (SI Figures S29 and S30). Table 2 summarizes the results of the daytime and nighttime episode screening.

To evaluate potential sources of unmatched enhancements, a case study was developed to demonstrate an analysis framework for constraining the location and processes associated with the observation of an unmatched enhancement episode. The case study uses the episode shown in Figure 7(a) as an example. In this approach, back trajectory analyses were conducted using the NOAA HYSPLIT trajectory model to

https://doi.org/10.1021/acsestair.5c00235
ACS EST Air 2025, 2, 2038−2046

determine a subset of potential source locations. Parcel trajectories were initialized at the time of the maximum enhancement in the observed episode and were evaluated out to 50 km from the receptor point. If a parcel trajectory overlapped with a potential source, that source was considered a candidate for analysis. Candidate sources included upstream, midstream, and flaring operations, and were further constrained in each episode by evaluating the chemical fingerprint of the observed enhancements for characteristic ratios of combustion and noncombustion tracer species. Uncertainty was considered in the parcel trajectories using ensemble trajectory approaches. Additional details are provided in section S11 of the Supporting Information.

To estimate the magnitude of the emission rates necessary to produce the observed, unmatched enhancements, each candidate source is simulated with a constant emission rate (typically using a unit rate; here 1 kg/h) using the same model configuration to produce a quantitative relationship between emission rates at potential sources and observed concentration enhancements at the receptors. All sources are modeled with the stack parameters as used in the initial simulations. Candidate sources can be further constrained based on assumptions about the maximum magnitude of expected emission rates. For example, if a threshold of 1 ton/h is set as the maximum expected emission rate for a region or equipment group, sources that have an estimated emission rate needed to generate the observed enhancement greater than 1 ton/h can be removed from consideration. In this case study, the chemical fingerprinting analysis indicated the source of the enhancement was relatively high in heavier hydrocarbons (propane) suggesting an upstream source (e.g., tanks, unloadings). Based on the back trajectories, several sources can be identified as potential candidates for reconciliation (SI Figure S28(a)). The predicted emission rates at each of these candidates necessary to reconcile the observed and predicted episodes range from approximately 25 kg/h − 1 ton/h (Table S26). Additional details are provided in SI Section S11.

IMPLICATIONS

Routine emissions from oil and gas operations coupled with diurnal meteorological patterns can produce some of the largest enhancements in ambient hydrocarbon concentrations observed at regional ground monitors in oil and gas production regions. To use these monitors for large emission event detection, it will be necessary to develop methods for decoupling the influence of routine and nonroutine emission sources on ambient measurements. Detailed emissions and dispersion modeling provide a mechanism to make these distinctions. For observed enhancements that cannot be explained by meteorology and routine emissions alone, coupled emissions and dispersion modeling frameworks can provide a base-case estimate for source-attribution analyses. These analyses can be facilitated using back trajectory analyses and (when available) additional measurements of ambient chemical species.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsestair.5c00235.

Details of emission inventory development, dispersion modeling, contributions of sources to predicted concentrations (PDF)

AUTHOR INFORMATION

Corresponding Author

David T. Allen − *Center for Energy and Environmental Systems Analyses, University of Texas at Austin, Austin, Texas 78759, United States;* orcid.org/0000-0001-6646-8755; Phone: 512-475-7842; Email: allen@che.utexas.edu

Authors

Joel D. Graves − *Center for Energy and Environmental Systems Analyses, University of Texas at Austin, Austin, Texas 78759, United States;* orcid.org/0000-0002-1595-4996

Yosuke Kimura − *Center for Energy and Environmental Systems Analyses, University of Texas at Austin, Austin, Texas 78759, United States*

Mrinali Modi − *Center for Energy and Environmental Systems Analyses, University of Texas at Austin, Austin, Texas 78759, United States;* orcid.org/0000-0002-9501-0513

Shannon Stokes − *Center for Energy and Environmental Systems Analyses, University of Texas at Austin, Austin, Texas 78759, United States;* orcid.org/0000-0003-2546-9617

Morgan Meyer − *Center for Energy and Environmental Systems Analyses, University of Texas at Austin, Austin, Texas 78759, United States*

Lea Hildebrandt Ruiz − *Center for Energy and Environmental Resources, University of Texas at Austin, Austin, Texas 78758, United States;* orcid.org/0000-0001-8378-1882

Complete contact information is available at: https://pubs.acs.org/10.1021/acsestair.5c00235

Notes

The authors declare the following competing financial interest(s): D.T.A. and L.H.R have served on the Environmental Protection Agency's Science Advisory Board; in this role, they were Special Governmental Employees. L.H.R's research is currently supported by the Texas Commission on Environmental Quality, the Health Effects Institute, and ExxonMobil. D.T.A.s research is currently supported by the National Science Foundation, the Department of Energy, the Texas Commission on Environmental Quality, Chevron, ExxonMobil, Pioneer Natural Resources, and the Environmental Defense Fund. He has also worked on air quality projects that have been supported by oil and gas producers and the Environmental Defense Fund. D.T.A. has worked as a consultant for multiple companies, including British Petroleum, Cheniere, Eastern Research Group, KeyLogic, and SLR International. M.Meyer has worked as a consultant for multiple companies under Trinity Consultants. J. Graves has worked as a consultant for the United Nations International Methane Emissions Observatory (U.N. IMEO).

ACKNOWLEDGMENTS

Research described in this work was conducted under contract to the Health Effects Institute (HEI), an organization jointly funded by the United States Environmental Protection Agency (EPA) (Contract No. 68HERC19D0010) and certain oil and natural gas companies. Although the research was produced

with partial funding by EPA and industry, they have not been subject to their review, and therefore, the research does not necessarily reflect the views of the agency or the oil and natural gas industry, and no official endorsement by the agency or the industry should be inferred.

REFERENCES

(1) Allen, D. T.; Cardoso-Saldaña, F. J.; Kimura, Y. Variability in Spatially and Temporally Resolved Emissions and Hydrocarbon Source Fingerprints for Oil and Gas Sources in Shale Gas Production Regions. *Environ. Sci. Technol.* **2017**, *51* (20), 12016−12026.

(2) Cardoso-Saldaña, F. J.; Allen, D. T. Projecting the Temporal Evolution of Methane Emissions from Oil and Gas Production Basins. *Environ. Sci. Technol.* **2021**, *55* (5), 2811−2819.

(3) Allen, D. T.; Cardoso-Saldaña, F. J.; Kimura, Y.; Chen, Q.; Xiang, Z.; Zimmerle, D.; Bell, C.; Lute, C.; Duggan, J.; Harrison, M. A Methane Emission Estimation Tool (MEET) for Predictions of Emissions from Upstream Oil and Gas Well Sites with Fine Scale Temporal and Spatial Resolution: Model Structure and Applications. *Sci. Total Environ.* **2022**, *829*, No. 154277.

(4) Brandt, A. R.; Heath, G. A.; Kort, E. A.; O'Sullivan, F.; Pétron, G.; Jordaan, S. M.; Tans, P.; Wilcox, J.; Gopstein, A. M.; Arent, D.; Wofsy, S.; Brown, N. J.; Bradley, R.; Stucky, G. D.; Eardley, D.; Harriss, R. Methane Leaks from North American Natural Gas Systems. *Science* **2014**, *343* (6172), 733−735.

(5) Allen, D. T.; Pacsi, A. P.; Sullivan, D. W.; Zavala-Araiza, D.; Harrison, M.; Keen, K.; Fraser, M. P.; Daniel Hill, A.; Sawyer, R. F.; Seinfeld, J. H. Methane Emissions from Process Equipment at Natural Gas Production Sites in the United States: Pneumatic Controllers. *Environ. Sci. Technol.* **2015**, *49* (1), 633−640.

(6) Cusworth, D. H.; Thorpe, A. K.; Ayasse, A. K.; Stepp, D.; Heckler, J.; Asner, G. P.; Miller, C. E.; Yadav, V.; Chapman, J. W.; Eastwood, M. L.; Green, R. O.; Hmiel, B.; Lyon, D. R.; Duren, R. M. Strong Methane Point Sources Contribute a Disproportionate Fraction of Total Emissions across Multiple Basins in the United States. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (38), No. e2202338119.

(7) Stokes, S.; Tullos, E.; Morris, L.; Cardoso-Saldaña, F. J.; Smith, M.; Conley, S.; Smith, B.; Allen, D. T. Reconciling Multiple Methane Detection and Quantification Systems at Oil and Gas Tank Battery Sites. *Environ. Sci. Technol.* **2022**, *56* (22), 16055−16061.

(8) Zimmerle, D.; Dileep, S.; Quinn, C. Unaddressed Uncertainties When Scaling Regional Aircraft Emission Surveys to Basin Emission Estimates. *Environ. Sci. Technol.* **2024**, *58* (15), 6575−6585.

(9) Chen, Q.; Schissel, C.; Kimura, Y.; McGaughey, G.; McDonald-Buller, E.; Allen, D. T. Assessing Detection Efficiencies for Continuous Methane Emission Monitoring Systems at Oil and Gas Production Sites. *Environ. Sci. Technol.* **2023**, *57* (4), 1788−1796.

(10) Daniels, W. S.; Jia, M.; Hammerling, D. M. Detection, Localization, and Quantification of Single-Source Methane Emissions on Oil and Gas Production Sites Using Point-in-Space Continuous Monitoring Systems. *Elem Sci. Anth* **2024**, *12* (1), 00110.

(11) Cheptonui, F.; Emerson, E.; Ilonze, C.; Day, R.; Levin, E.; Fleischmann, D.; Brouwer, R.; Zimmerle, D.Assessing the Performance of Emerging and Existing Continuous Monitoring Solutions under a Single-Blind Controlled Testing Protocol*ChemRxiv*, **2025**.

(12) Texas Commission on Environmental Quality*Texas Air Monitoring Information System (TAMIS) [Dataset]*. https://www17.tceq.texas.gov/tamis/index.cfm?fuseaction=home.welcome

(13) Allen, D. Attributing Atmospheric Methane to Anthropogenic Emission Sources. *Acc. Chem. Res.* **2016**, *49* (7), 1344−1350.

(14) Simpson, I. J.; Sulbaek Andersen, M. P.; Meinardi, S.; Bruhwiler, L.; Blake, N. J.; Helmig, D.; Rowland, F. S.; Blake, D. R. Long-Term Decline of Global Atmospheric Ethane Concentrations and Implications for Methane. *Nature* **2012**, *488* (7412), 490−494.

(15) Texas Railroad Commission (RRC)*Eagle Ford Shale 2008 to Current [Dataset]*. https://www.rrc.texas.gov/media/0eulto5v/eagle-ford-shale-2008-to-current.xlsx

(16) U.S. Energy Information Administration. U.S*Monthly Crude Oil and Natural Gas Production*. https://www.eia.gov/petroleum/production/#ng-tab

(17) U.S. Energy Information Administration. Monthly U.S*Field Production of Crude Oil (Thousand Barrels per Day) [Dataset]*. https://www.eia.gov/dnav/pet/hist_xls/MCRFPUS2m.xls

(18) Helmig, D.; Rossabi, S.; Hueber, J.; Tans, P.; Montzka, S. A.; Masarie, K.; Thoning, K.; Plass-Duelmer, C.; Claude, A.; Carpenter, L. J.; Lewis, A. C.; Punjabi, S.; Reimann, S.; Vollmer, M. K.; Steinbrecher, R.; Hannigan, J. W.; Emmons, L. K.; Mahieu, E.; Franco, B.; Smale, D.; Pozzer, A. Reversal of Global Atmospheric Ethane and Propane Trends Largely Due to US Oil and Natural Gas Production. *Nat. Geosci.* **2016**, *9* (7), 490−495.

(19) Esri*"World Imagery Hybrid" [Basemap]. Scale Not Given. "Imagery Hybrid."* https://www.arcgis.com/home/item.html?id=86265e5a4bbb4187a59719cf134e0018

(20) IHS Markit*U.S. Well and Production Data [Dataset]*. https://ihsmarkit.com/index.html

(21) Mitchell, A. L.; Tkacik, D. S.; Roscioli, J. R.; Herndon, S. C.; Yacovitch, T. I.; Martinez, D. M.; Vaughn, T. L.; Williams, L. L.; Sullivan, M. R.; Floerchinger, C.; Omara, M.; Subramanian, R.; Zimmerle, D.; Marchese, A. J.; Robinson, A. L. Measurements of Methane Emissions from Natural Gas Gathering Facilities and Processing Plants: Measurement Results. *Environ. Sci. Technol.* **2015**, *49* (5), 3219−3227.

(22) Zimmerle, D.; Vaughn, T.; Luck, B.; Lauderdale, T.; Keen, K.; Harrison, M.; Marchese, A.; Williams, L.; Allen, D. Methane Emissions from Gathering Compressor Stations in the U.S. *Environ. Sci. Technol.* **2020**, *54* (12), 7552−7561.

(23) Cardoso-Saldaña, F. J.; Pierce, K.; Chen, Q.; Kimura, Y.; Allen, D. T. A Searchable Database for Prediction of Emission Compositions from Upstream Oil and Gas Sources. *Environ. Sci. Technol.* **2021**, *55* (5), 3210−3218.

(24) Gherabati, S. A.; Browning, J.; Male, F.; Ikonnikova, S. A.; McDaid, G. The Impact of Pressure and Fluid Property Variation on Well Performance of Liquid-Rich Eagle Ford Shale. *J. Nat. Gas Sci. Eng.* **2016**, *33*, 1056−1068.

(25) Texas Railroad Comission*Oil and Gas Well Records*. https://rrcsearch3.neubus.com/esd3-rrc/index.php?_module_=esd&_action_=keysearch&profile=17

(26) Texas Railroad Comission*Oil and Gas Completions*. https://webapps.rrc.state.tx.us/CMPL/publicSearchAction.do

(27) Cardoso-Saldaña, F. J.; Kimura, Y.; Stanley, P.; McGaughey, G.; Herndon, S. C.; Roscioli, J. R.; Yacovitch, T. I.; Allen, D. T. Use of Light Alkane Fingerprints in Attributing Emissions from Oil and Gas Production. *Environ. Sci. Technol.* **2019**, *53* (9), 5483−5492.

https://doi.org/10.1021/acsestair.5c00235
ACS EST Air 2025, 2, 2038−2046

Supporting Information

Source attribution of elevated ethane concentrations detected by regional monitors in oil and gas production regions

Joel Graves [a], Yosuke Kimura [a], Mrinali Modi [a], Shannon Stokes[a], Morgan Meyer[a], Lea Hildebrandt Ruiz[b], David T. Allen [a, *]

[a] Center for Energy and Environmental Systems Analyses, University of Texas at Austin, 3925 West Braker Lane, Austin, TX 78759, United States

[b]Center for Energy and Environmental Research, University of Texas at Austin, 10500 Exploration Way, Austin, TX 78758, United States

*Corresponding author: email: allen@che.utexas.edu; tel.: 512-475-7842

Table of contents

Total pages: 66

Number of Figures: 30

Number of Tables: 28

S1: Production Information and Well Categorization

Monthly production rates of gas, condensate, and water between March and May 2023 were obtained for 20,510 wells in the Eagle Ford Shale production region (EFS).[1] An average daily production rate (i.e. Mcf/day gas; bbl/day oil; bbl/day water) was calculated for each site as the three-month average (92 days) of the well stream gas, condensate, and water production rates for that site. Any well that had a non-zero gas production stream was labeled a "gas" well and all others were labeled "oil". The characterization of each well as gas or oil producing is used to assign the number of pneumatic controllers and chemical injection pumps at individual sites and to decide whether to consider liquids unloading events for specific wells (Sections S2.1 and S2.3).[2] Well counts by category for each of the nested modeling domains are summarized in Table S1.

Well Type / Modeling region (nest)	Extent	Gas wells	Oil wells	Total
Inner Modeling Domain	32 km x 32 km	3041	147	3208
Intermediate Modeling Domain	48 km x 44 km	1722	60	1782
Outer Modeling Domain	212 km x 208 km	11760	3760	15520
Total	-	16523	3987	20510

Table S1*:* Well count by category and modeling region (nest)

S2: Well Completion Data

Well completion dates[1] were used to constrain the location and duration of emissions when modeling well completions (Section S3.6). The day after the most recent completion date was assumed to be the first date of production. After a completion, constant production rates for each fluid were assumed for the remaining modeling period.

S3: Activity Data and Emission Factors

Emission rates were estimated for each source by multiplying activity data by emission factors. Activity data describes the number of sources or events within some spatial extent or equipment category, and emission factors describe the corresponding emission rates associated with each of these sources or activities. Depending on the equipment category, the Methane Emission Estimation Tool (MEET) used in this work will either calculate activity and emission factors based on production data and process modeling (e.g., tank flashing) or use pre-defined values selected during model set-up (e.g., pneumatic controller emissions). For the latter case, an archive of activity and emission factors reported in national and regional measurement surveys is included with MEET. Optionally, user-defined activity and emission factors can also be used. Activity and emission factors for this study are summarized for each source category below.

S3.1: Pneumatic Controllers

The activity factor for pneumatic controllers (PCs) is the number of controllers per well site. MEET defaults the number of controller counts to an average estimate of pneumatic controllers per gas well (2.02 controllers/gas well) and per oil well (1.11 controllers/oil well) derived from national estimates of pneumatic controllers and active wells reported in Annex 3.6 and 3.5 of 2017 U.S. Inventory of Greenhouse Gas Emission and Sinks (GHGI).[3,4] In this work, controllers were randomly assigned to wells such that the average controller count per well in the modeling region matched the national averages from the 2017 GHGI. For example, gas producing wells were randomly assigned 2 or 3 controllers at a frequency of 98% and 2% respectively to arrive at an average value of 2.02 controllers across all gas wells in the model. Oil wells were treated similarly.

MEET further categorizes pneumatic controllers based on their operating mode and expected emission rate. Three categories of pneumatic controllers are specified: continuous low-bleed, continuous high-bleed, and intermittent. Low- and high-bleed controllers are differentiated by a flow rate threshold of 6 scfh (i.e., low-bleed controllers are expected to have whole gas emission as rates < 6 scfh during normal operation).[5] In addition to an overall national pneumatic controller count estimate, GHGI reports estimated counts of high-bleed, low-bleed, and intermittent controllers.[3,4] These values were used to derive estimated distributions of controller types at oil and gas wells (Table S2).

Oil			Gas		
Intermittent	Low Bleed Continuous	High Bleed Continuous	Intermittent	Low Bleed Continuous	High Bleed Continuous
71.7%	25.3%	3.0%	79.2%	18.9%	1.9%

Table S2*:* National distributions of pneumatic controller types reported by GHGI (2017) for oil and gas wells[3,4]

Each pneumatic controller in the modeling domain was categorized as either intermittent, low bleed continuous, or high bleed continuous based on the data in Table S2. For example, every pneumatic controller assigned to a gas well was individually classified as an intermittent, low bleed continuous, or high bleed continuous controller with 79.2%, 18.9% and 1.9% probability of classification respectively. Pneumatic controllers at oil wells were treated similarly. Because each controller is individually classified by type, the overall distribution at individual sites varied between simulations.

Allen et al. observed the distribution of emissions from pneumatic controllers was highly skewed with a large fraction of overall emissions coming from a few high emitting controllers.[6] Many of these high emitting controllers were observed to be low-bleed or intermittent devices operating in an abnormal state where emission rates were much greater than expected. MEET accounts for these abnormal operating modes in intermittent and low-bleed controllers by using a mode switching model. Briefly, this model separates the low-bleed and intermittent pneumatic controller emission rates reported by Allen et al.[6] into “normal” and “abnormal” records. Abnormally operating devices are defined as those where whole gas emission rates were

observed to be greater than 2 scfh for low-bleed controllers and greater than 15 scfh for intermittent controllers.[6] MEET then estimates whether each low-bleed or intermittent controller is in a normal or abnormal state using a discrete event simulator and a specified mean time to repair (MTTR) which sets the amount of time a controller spends in the abnormal emission mode. The MTTR can be user-defined or MEET will use a default duration that ensures the average fraction of abnormally operating devices for intermittent and low bleed controllers at any time are 19% and 7.1% respectively based on observations from Allen et al.[6]

At the beginning of the simulation, each intermittent and low bleed pneumatic controller is independently assigned a pair of emission factors corresponding to a normal and abnormal operating mode. These emission factors are randomly selected from the distribution of measurements in the disaggregated low bleed and intermittent controller inventories that correspond to the controller type and operating mode (i.e., an emission factor for the abnormal mode of an intermittent controller is sampled from the abnormal records in the intermittent controller inventory from Allen, et al.[6]). As the simulation progresses, controllers switch between these emission factors based on the state of operation as determined by a discrete event simulator. State-specific (normal and abnormal) emission rates are fixed for individual controllers throughout the simulation. MEET does not differentiate between normal or abnormal operation for high-bleed controllers. Instead, site specific high-bleed pneumatic controller emission factors are randomly sampled from the overall high-bleed emissions inventory reported by Allen et al..[6] Table S3 summarizes the pneumatic controller observations from Allen et al..[6]

	Number of observations	Emission rate (whole gas emission rate in scf/hr)			Abnormally operating devices		Transition of operating states	
		Mean	median	min – max	Threshold (scf/hr)	Count/percentage of abnormal	MTTR (days)	MTBF (days)
Intermittent	126	3.9	0	0 – 111	15	9 (7.1%)	30	390
Continuous Low Bleed	32	10.8	.0015	0 – 149	2	6 (19%)	30	130
Continuous High Bleed	7	22.8	16.5	14.5 – 37.4	+Inf	0 (0%)	n.a.	n.a.

Table S3: Summary of pneumatic controller observations from Allen et al.[6]

S3.2 Chemical Injection Pumps

The activity factor for pneumatic chemical injection pumps is the pump count per well multiplied by the number of wells. The 2017 GHGI estimates 0.178 pneumatic chemical injection pumps per gas well and 0.095 pneumatic chemical injection pumps per oil well.[3,4] Individual pneumatic chemical injection pumps were randomly assigned to wells such that the average pump count per well in the modeling region matched the GHGRP averages. For example, gas producing wells were randomly assigned 0 or 1 pneumatic chemical injection pumps at a frequency of 82.2% and 17.8% respectively to arrive at an average value of 0.178 chemical injection pumps per gas well across the entire modeling domain. Oil wells were treated similarly. Emission factors for each pneumatic chemical injection pump were assigned randomly from observations.[7] MEET treats pneumatic chemical injection pumps as emitting continuously.

Number of observations	Emission rate (kg CH_4 / hr)		
	Mean	Median	Min – Max
63	0.222	0.0616	0.00238 – 2.31

Table S4: Summary of pneumatic chemical injection pumps[7]

S3.3 Leaks

The activity factor for leaks is the number of leaking components per wellsite. Typical leak sources include valves and flanges on process equipment such as separators and compressors. MEET treats six specific equipment categories at each site as having the potential for leaks. These categories are:

1. Wellheads
2. Separators
3. Meters/piping
4. Compressors
5. In-line heaters
6. Dehydrators

This work only includes leaks from wellheads, separators, and meters/piping because compressors and dehydrators are assumed to be located on gathering and boosting sites and heaters are not expected to be operational during the simulation period. In MEET, component counts can be user defined. Alternatively, MEET will default to multiplying qualifying equipment counts by equipment-specific average component (valves, flanges, etc.) counts reported in EPA 40 CFR 98 Table W1-B (Table S4) to arrive at a total count of potential leak sources.[8] Emission factors and the probability of occurrence for each potential leak are component-specific and are derived from the 2018 GHGRP[9] and EPA 40 CFR 98 Table W1-E respectively (Table S5).[10]

MEET treats leaks using a mode-switching model similar to the model used for pneumatic controllers; components are either in a state of leaking or normal operation with some prescribed time-in-state controlling the relative frequency of events. This behavior is managed by a discrete event simulator; a description of the method is as follows:

At the beginning of each simulation, for each site, the probability of leak occurrence for each component (i):

$$p_i = \frac{n_{obs,i}}{N_{obs,i}}$$

derived from the 2018 GHGRP[2,9] (Table S6) is multiplied by the site-specific count of component (i) to arrive at a site-specific number of initial leaks of component (i):

$$\chi_i = p_i \cdot N_{obs,i}$$

Where $n_{obs,i}$ is the number of leaks observed for component type (i) and $N_{obs,i}$ is the total number of components of that type on the site.

MEET then randomly assigns χ_i components at the site to be leaking at the beginning of the simulation. The duration of each leak event is drawn from an exponential distribution with a mean value of 1/MTTR, where MTTR is a user-specified mean time to repair. Once the leak duration has been reached, leaking components switch to a non-leaking mode and emissions are halted until the component switches back into a leaking mode. The time each component spends in the non-leaking mode is selected from an exponential distribution with a mean of the Mean Time Before Failure (MTBF) which is defined relative to MTTR as:

$$MTTR = \frac{p}{1-p} * MTBF$$

An MTTR of 182.5 days (half year) was assigned to all leaks. The discrete event simulator handles mode switching, and each component continues to switch between a leaking and non-leaking operating mode, sampling new time-in-mode durations for each interval, for the duration of the simulation. MEET assigns component-specific emission factors to leaks from the average emission rates reported in EPA 40 CFR 98 Table W1-E (Table S5).[10] Each component is assigned a single emission factor for its leaking state which does not change throughout the course of the simulation. While MEET differentiates between components when assigning emission rates and leak probabilities, it reports aggregated site-specific leak emission rates. Thus, at any given time the facility-level leak emission rate is the number of a given component in the leaking mode multiplied by the emission factor for that component, summed over all component types. Equipment counts and leaks predicted by MEET for this work are summarized in Tables S7 and S8.

Major Equipment	Valves	Connectors	Open-ended lines	Pressure Relief Valves
Wellheads	11	36	1	0
Separators	34	106	6	2
Meters/Piping	14	51	1	1
Compressors	73	179	3	4
In-line heaters	14	65	2	1
Dehydrators	24	90	2	2

Table S5: Average component counts for major equipment categories at natural gas well sites in the Western U.S.[8]

Component	Fraction of Leaking (p)[a]	Average Emission Rate[b]
Valve	0.00191	4.9 scf whole gas/hour/component
Connector	0.000665	1.3 scf whole gas/hour/component
Open-Ended Line	0.00646	2.8 scf whole gas/hour/component
Pressure Relief Valve	0.0272	4.5 scf whole gas/hour/component

Table S6: Fraction of leaking components at any time and average emission rates – averages from (a) national survey (GHGRP) and (b) EPA 40 CFR 98 Table W1-E respectively[9,10]

Equipment	Count of equipment per well	Count of device per equipment			
		Valves	Connectors	Open-ended lines	Pressure relief valves
Well head	1	11	36	1	0
Separator	1; two or more fluids produced	34	106	6	2
	0; one or less fluids produced				
Meter/piping	1 per well	14	51	1	1

Table S7: Equipment Counts per well in EFS simulation

Modeling region	Extent	Count of equipment that have one or more leaks during the 3month modeling period			Whole-gas hourly emission rate summed across region (scf / hr)		
		wellhead	separator	Meters/piping	Wellhead	separator	Meters/piping
Inner Modeling Domain	32 km x 32 km	243.8 (200 – 284)	591.5 (559 – 616)	419.9 (397 – 444)	497 (362 – 599)	1544 (1366 – 1761)	990 (890 – 1108)
Intermediate Modeling Domain	48 km x 44 km	133.3 (124 – 145)	407.2 (386 – 430)	240.8 (230 – 253)	259 (212 – 309)	1055 (906 – 1178)	568 (466 – 652)
Outer Modeling Domain	212 km x 208 km	1162.1 (1096 – 1230)	4550.6 (4384 – 4685)	2049.0 (1980 – 2110)	2368 (2121 – 2594)	11865 (11293 – 12460)	4781 (4500 – 5027)
All Regions	-	1539.2 (1488 – 1630)	5549.3 (5389 – 5647)	2709.7 (2654 – 2760)	3124 (2894 – 3376)	14463 (13884 – 15032)	6339 (6013 – 6682)

Table S8: Leak statistics by source group and nest across 10 Monte Carlo simulations. Table shows equipment counts with one or more leaks over the entire duration of the simulation averaged across the 10 Monte Carlo simulations (minimum and maximum individual Monte Carlo averages shown in parenthesis). Also shown is the equipment-level hourly whole-gas emission rates averaged over the 10 Monte Carlo simulations (minimum and maximum individual Monte Carlo averages shown in parenthesis).

S3.4 Condensate Tank Flash

The activity factor for condensate tank flashing is the volume of liquid production at the wellsite. Condensate tank flashing occurs when hydrocarbons travel from the pressurized separator (~100 – 1500 psi) to atmospheric holding tanks and depressurization of the hydrocarbon stream causes high vapor pressure components to flash out of the liquid phase. Separator liquids are usually sent to storage tanks in batches called dumps. Because the tank emissions are a function of the dump volume, frequency, and composition, the tank emission rates are controlled by the overall liquid production rate and separator conditions and the periodicity of the emissions is controlled by the volume of the individual dumps.

MEET estimates the volume and duration of each dump by sampling the values from a truncated normal distribution (Figure S1). Parameters for these distributions are summarized in Table S9. Once a dump volume is assigned to a separator, it remains constant throughout the duration of the simulation. The frequency of dumps is calculated by dividing the liquid production rate of well by the dump volume in the separator associated with the well.

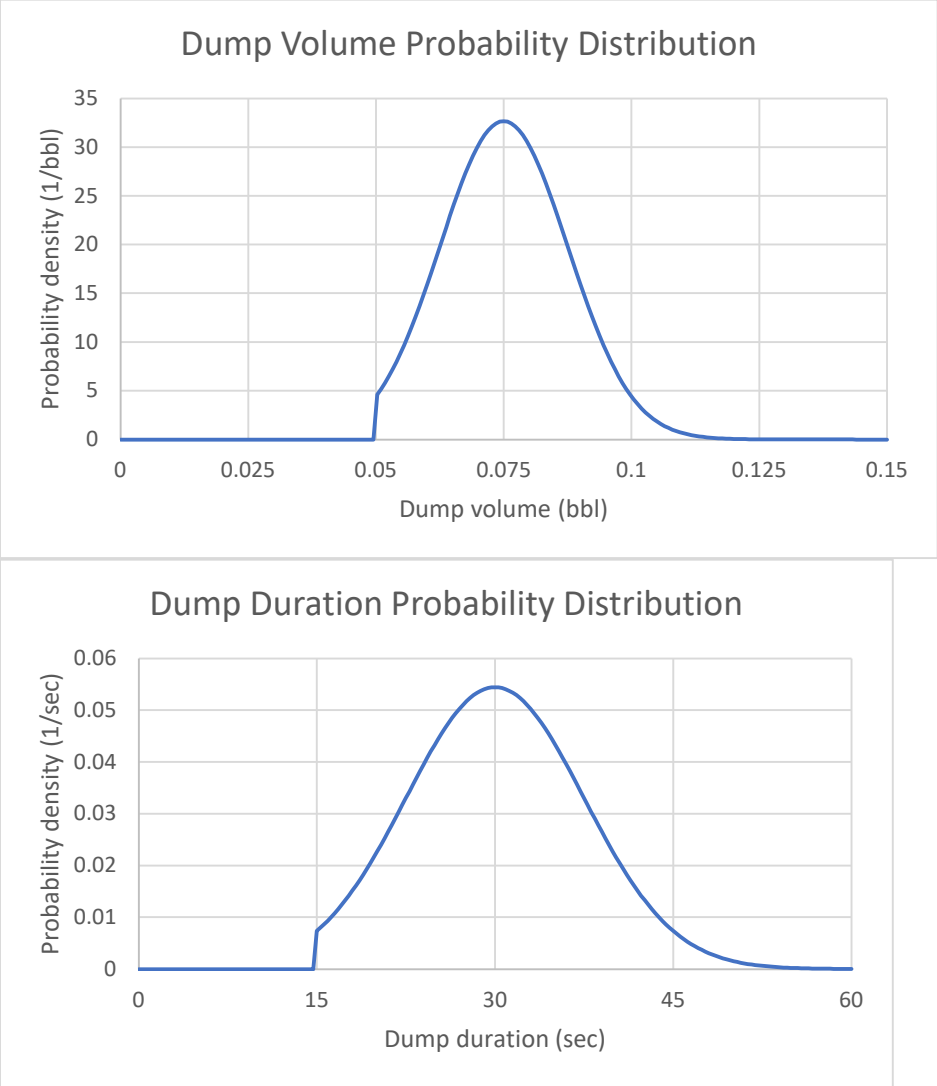


Figure S1: Dump volume and duration probability distributions

Parameter	Mean	Standard deviation	Lower Bound
Dump Volume (bbl)	0.075	0.0125	0.05
Dump Duration (sec)	30	7.5	15

Table S9: Tank flash dump volume and duration probability distribution parameters[2]

With each tank flash, the thermodynamic model embedded in MEET calculates an associated mass of flashed gas based on the separator compositions, operating conditions, and dump volume. For this reason, MEET requires production and operational information about the well to estimate tank flash emissions. Production data[1] and average daily production rates were calculated for each site as the three-month (92 day) average of the well stream gas, condensate, and water production rates. During each dump, emissions are assumed to follow a triangular wave form. However, for this work, because emissions are aggregated at an hourly level, tank flash emissions are effectively constant since most wells have one or more separator dumps per hour.

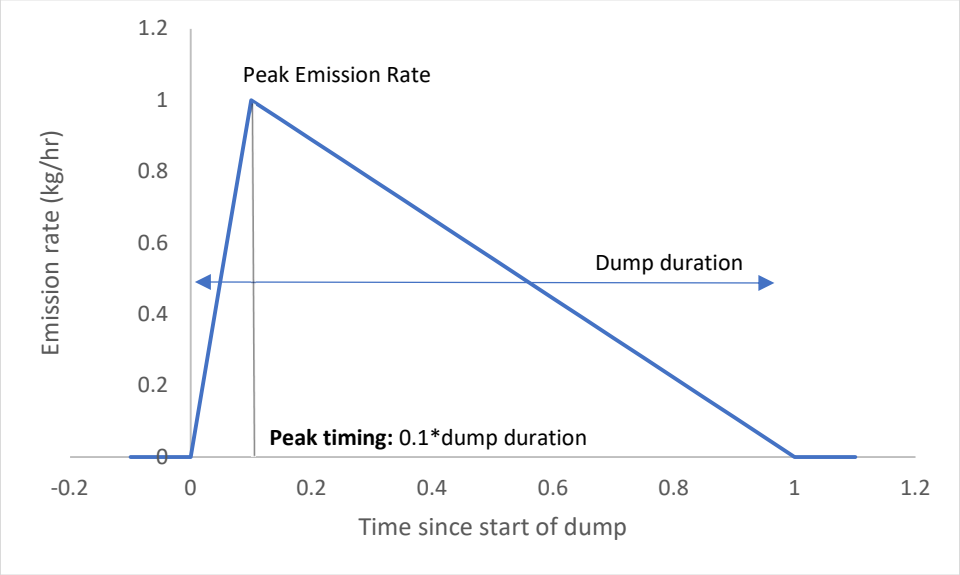


Figure S2: Triangular wave form of typical tank emissions rate. Hourly averaging of emissions distributes the total emissions across the hour leading to an apparent continuous emission rate. In the instance that a dump event starts and finishes in two different hours, the emissions are distributed across the hours based on the fraction of the dump occurring in each hour. A majority of tanks in the simulation had dump frequencies high enough that they displayed emissions at every hour of the simulation, effectively becoming continuous emission sources. A small fraction of tanks had dump frequencies on the order of hours leading to an intermittent emissions profile.

MEET provides the option to model condensate tanks using an "all-flash" method to account for emissions that may occur beyond the period of the initial flash. This method assumes all light alkanes (C_1-C_4) that reach the condensate tank are flashed. This is a reasonable assumption for species which have relatively high vapor pressures (e.g., light alkanes) at tank conditions. The all-flash assumption was used for all condensate tanks. Emissions control systems were assumed

to be installed on all tanks and 95% control efficiency (i.e., 5% of whole gas flashed is emitted) was assumed. All reported tank emissions are post-control. Table S10 shows counts of condensate tanks grouped by intermittency for the modeling time period. Tank flash emissions are estimated on a per-barrel of liquid hydrocarbon produced or per-barrel of water produced basis

Modeling region	Extent	Count			Average fraction of all condensate tanks that flash continuously
		Average number of continuously emitting condensate tanks	Average number of intermittently emitting condensate tanks	All condensate tanks	
Inner Modeling Domain	32 km x 32 km	3009.7 (3005 – 3020)	162.3 (152 – 167)	3172	94.9%
Intermediate Modeling Domain	48 km x 44 km	1617.6 (1612 – 1621)	124.4 (121 – 130)	1742	92.9%
Outer Modeling Domain	212 km x 208 km	10463.8 (10445 – 10480)	4129.2 (4113 – 4148)	14593	71.7%
All Region	-	15091.1 (15070 – 15109)	4415.9 (4398 – 4437)	19507	77.4%

Table S10: Condensate tank statistics by nest across 10 Monte Carlo simulations (minimum and maximum individual Monte Carlo averages shown in parenthesis).

S3.5 Water tank flash

Water tank flashes are treated similarly to condensate tank flashes except the liquid phase is water. Dump volume and duration are assigned using the approach outlined for condensate tanks. Table S11 summarizes counts of water tanks. Every well that produces water is assumed to have a water tank. 95% emissions control was assumed.

Modeling region	Count			Average fraction of all water tanks that flash continuously
	Average number of continuously emitting water tanks	Average number of intermittently emitting water tanks	All water tanks	
"<15km"	1434.0 (1428 – 1442)	663.0 (655 – 669)	2097	68.4%
"15 to 20km"	964.4 (957 – 972)	484.6 (477 – 492)	1449	66.6%
">20km"	6858.0 (6843 – 6881)	4976.0 (4953 – 4991)	11834	58.0%
All Region	9256.4 (9239 – 9286)	6123.6 (6094 – 6141)	15380	60.2%

Table S11: Water tank counts and emissions summary for modeling period

S3.6 Liquids Unloading

Production in gas wells that also produce oil and water can be disrupted by the accumulation of liquids in the wellbore. This accumulation is often the result of changes in well properties (e.g., gas to oil ratio) or changes in gas velocity in the wellbore. Unloading these liquids from the wellbore can be done using a variety of techniques, both with and without associated emissions. MEET estimates emissions from unloadings that result in emissions. Three techniques are primarily used for performing unloadings that result in emissions: manual unloadings without plunger lift, manual unloadings with plunger lift, and automated unloadings with plunger lift.

MEET first selects specific wells to be modeled with unloadings and then characterizes these unloadings as using one of the three techniques above. In this study, the number of wells with unloadings was based on reported unloadings statistics for unconventional wells in Karnes County in 2018 (GHGRP, Table S12).[2,9] The total number of wells selected by MEET to model with unloadings follows the fraction of reported wells with unloadings (5.83%) from these reports. Wells for modeling unloadings are selected randomly and independently from the overall population of wells provided to MEET. Unloadings techniques are then randomly assigned to the selected wells based on the distribution of technologies from the report. In this study, all unloadings were simulated as manual unloadings. Independent samples are selected at the beginning of each Monte Carlo instance, so the location and type of each unloading varies between instances.

After a well is selected for simulated unloadings and an unloading type is assigned, MEET requires six unloading properties to estimate associated unloadings emissions: an unloading frequency, a vent duration, a shut-in duration, an unloading onset age, a time of day for the beginning of venting, and an emission factor.[2] All unloadings in this study were simulated as manual unloadings, consistent with the reported unloadings statistics for unconventional wells in Karnes County (Table S12). For wells with simulated unloadings, unloading frequencies, venting durations, and emission factors were individually sampled from a measurement survey conducted by Allen et al. (Table S13)[11] Unloading start times were constrained using data from a distribution of daytime measurements reported by Vaughn, et al.[12,13] and Allen et al.[11] Shut in durations and unloading onset ages were calculated by MEET. A description of the unloading process, and how each of these parameters contributes to the emissions profile is as follows: prior to each unloading event, the well is temporarily sealed (shut-in) to allow pressure to build in the wellhead. During this time production is suspended. After the pressure has risen enough to flush the accumulated liquids out of wellbore, the well is opened to atmosphere and allowed to vent. Once the accumulated liquids have been sufficiently purged, production is redirected to the process stream. MEET simulates the shut-in process by halting all on-site emissions (since production is paused) for a period of time before allowing the unloading to begin. The shut-in duration for all wells was assumed to be 24 hours or 1/4 of the interval between the end of an event and the beginning of the next event, whichever is the shortest. Event durations are likewise constrained to not exceed 1/4 of the interval between events.

All unloadings in this work were simulated as manual unloadings which require operator supervision and, as a result, generally occur during the work day. MEET can consider this temporal bias when modeling manual unloadings by sampling unloading start times only from daytime distributions. In this work, manual unloading start times were estimated using data from

daytime unloadings measurements reported by Vaughn, et al.[12,13] and Allen, et al.[11] In this approach, a measured event with a reported start and end time is randomly selected from Vaughn, et al.[12,13] and the middle of the time period is selected as the midpoint for the simulated unloading event (e.g. an event lasting between 11 A.M. and 2 P.M. has a midpoint at 12:30 P.M). Next, a measured event from Allen, et al.[11] with a reported vent duration, emission factor, and annual unloading frequency is randomly selected and centered in time on the midpoint determined by sampling from the Vaughn, et al.[12,13] dataset. The resulting simulated unloadings at the well have durations, emission rates, and a frequency defined by data from the Allen, et al.[11] dataset but are constrained to a specific time-of-day by the Vaughn, et al.[12,13] data.

Unloadings are usually unnecessary early in the production period of gas wells because reservoir conditions (e.g. pressure, GOR) are such that production flows are self-sustaining. Over time, these properties gradually change until eventually, liquid accumulation in the production stream can begin to impede gas flow up the borehole. This is generally when unloadings begin. MEET uses the unloading onset age to determine the time after completion to begin simulating unloadings at a well.[14] MEET samples unloading onset ages from a triangular distribution with a mode at 31.3 months, a minimum at 21.1 months, and a maximum of 41.5 months. MEET then calculates the day of liquids unloading onset as the date of last completion plus the unloadings onset age. The day of last completion for each well was sourced from IHS Markit (See Section S.2).[1]

Table S14 summarizes domain-wide statistics for simulated unloadings events across the 10 Monte Carlo instances. Figures S3-S5 summarize the time-of-day distribution of unloading events and associated ethane and methane emissions across all 10 Monte Carlo simulations.

County	Basin	Formation type	Total Wells	Wells with venting			
				Total	Manual Without Plunger	Manual With Plunger	Automated Plunger
Karnes, TX	Gulf Coast Basin (LA, TX)	Unconventional	1115	65 (5.83%)	65 (5.83%)	0 (0%)	0 (0%)

Table S12: Summary of wells with unloadings in the Eagle Ford Shale in 2018[2,9]

Unloading method	Count of wells observed	Frequency (events/year)	Venting duration (min)	Emission factor (whole gas scf / event)
Manual without plunger	32	33.4 (1 – 185)	82.9 (10.5 – 269)	26557 (599 – 137867)

Table S13: Unloading properties sampled from Allen, et al., (2015b).[11] Values are: mean (minimum-maximum).

	Per Event				Per Site**		Domain wide
	Unloading Event Duration (hr)	Count of Event in 3 Month modeling period	Ethane Emission per Event (kg/event)	Methane Emission per Event (kg/event)	Hourly Ethane Emission (kg/hr)	Hourly Methane Emission (kg/hr)	Count of wells unloading *
Mean	1.89	9.9	35.79	96.66	19.61	52.89	692.9
Minimum	1.00	1	0.002	0.005	0.001	0.00	648
Median	1.89	3	16.50	45.37	9.70	26.48	696
Maximum	4.49	46	1357.75	1632.55	549.42	1357.75	717

Table S14: Unloading statistics across 10 Monte Carlo instances in the emissions modeling

* On average, 945 wells of the 16436 gas wells in the region were assigned to have unloadings. However, approximately 100 wells with assigned unloadings were younger than the assigned unloading onset age and did not simulate any unloadings. Approximately 220 additional wells have infrequent unloadings (< 3 unloadings per year), potentially leading to a situation where no unloading events were simulated for these wells during the simulation period (3 months). These wells have been omitted from the counts reported in this table.

** Hourly statistics do not include periods of zero emissions.

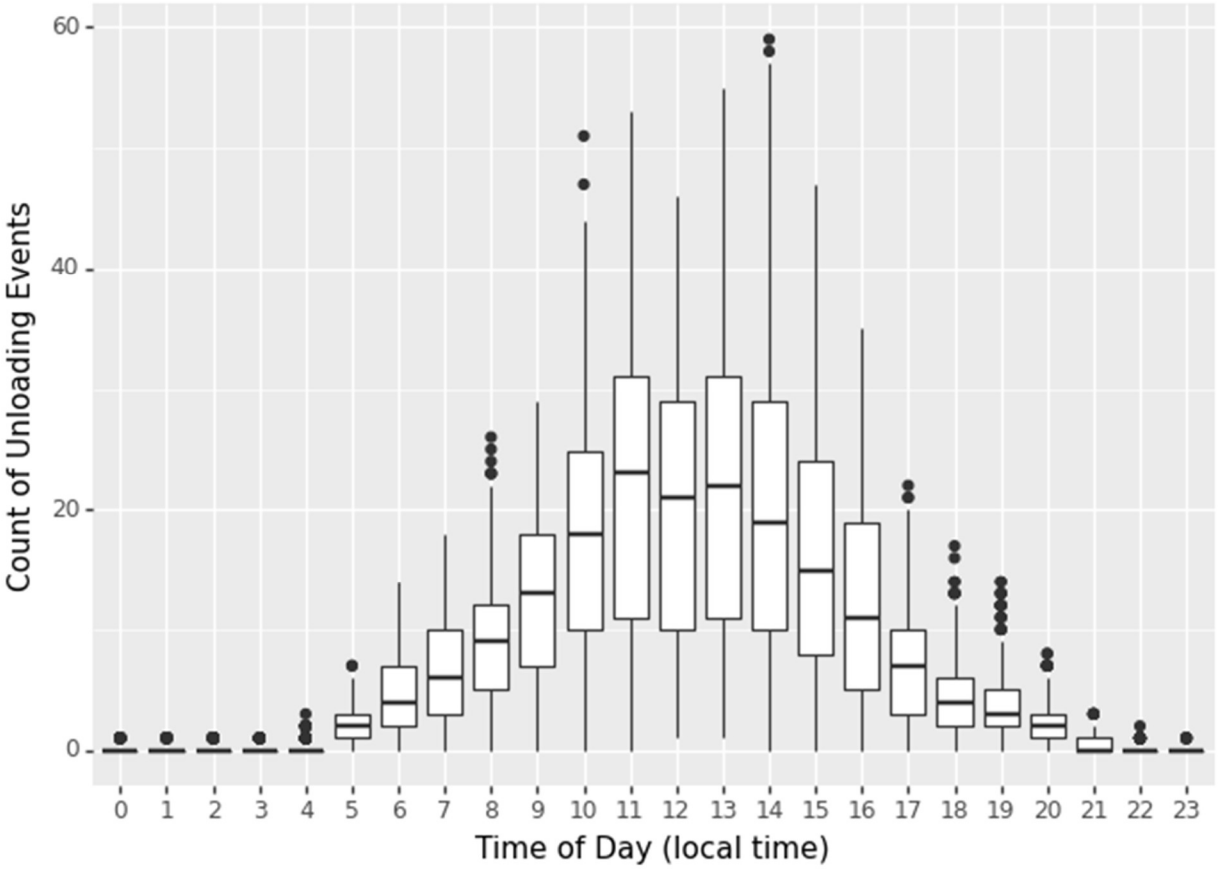


Figure S3: Distributions of unloading event counts for facilities in the entire modeling domain as a function of the time of day across 10 Monte Carlo instances of 81 modeling days. All unloadings were modeled as manual unloadings which require operator supervision and occur typically during working hours.

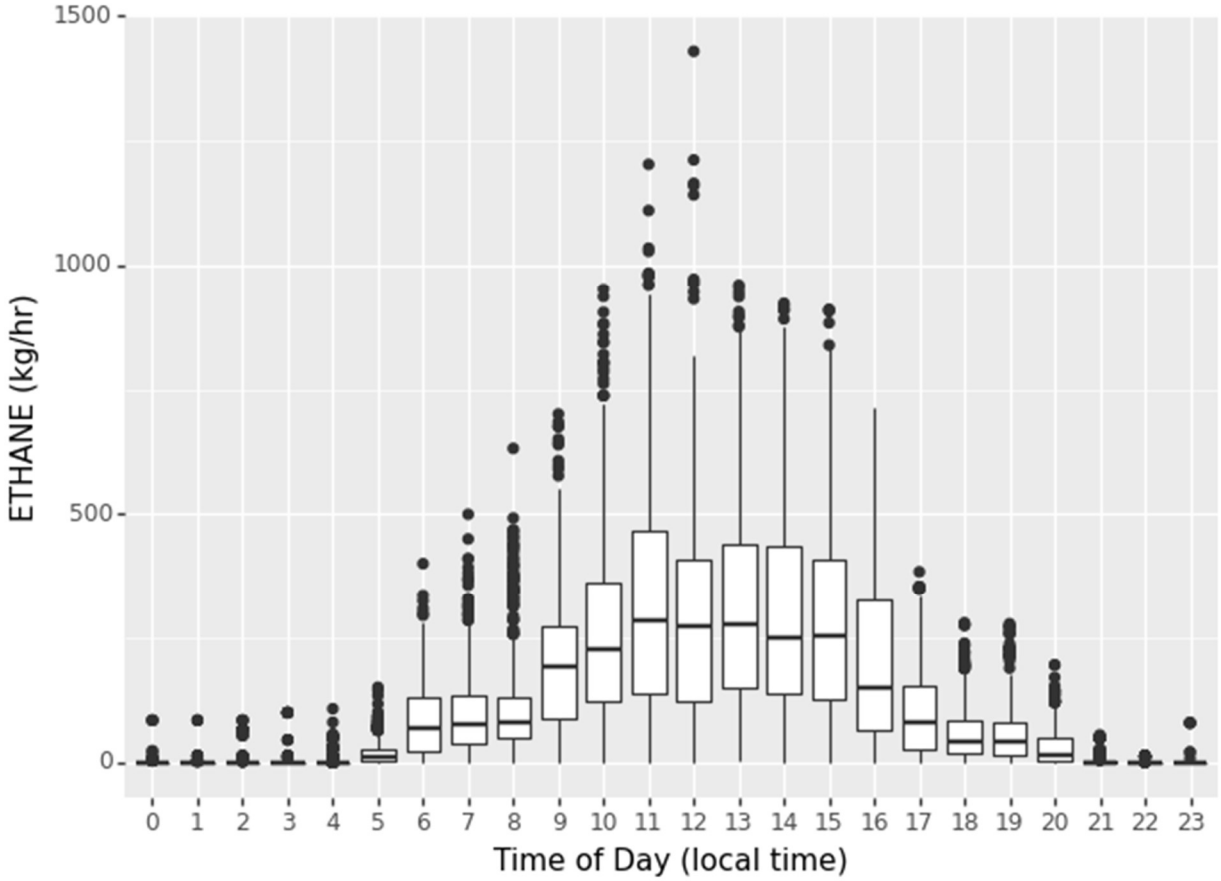


Figure S4: Distributions of ethane emissions from manual unloadings for the entire modeling domain as a function of the time of day across 10 Monte Carlo instances of 81 modeling days. All unloadings were modeled as manual unloadings which require operator supervision and occur typically during working hours.

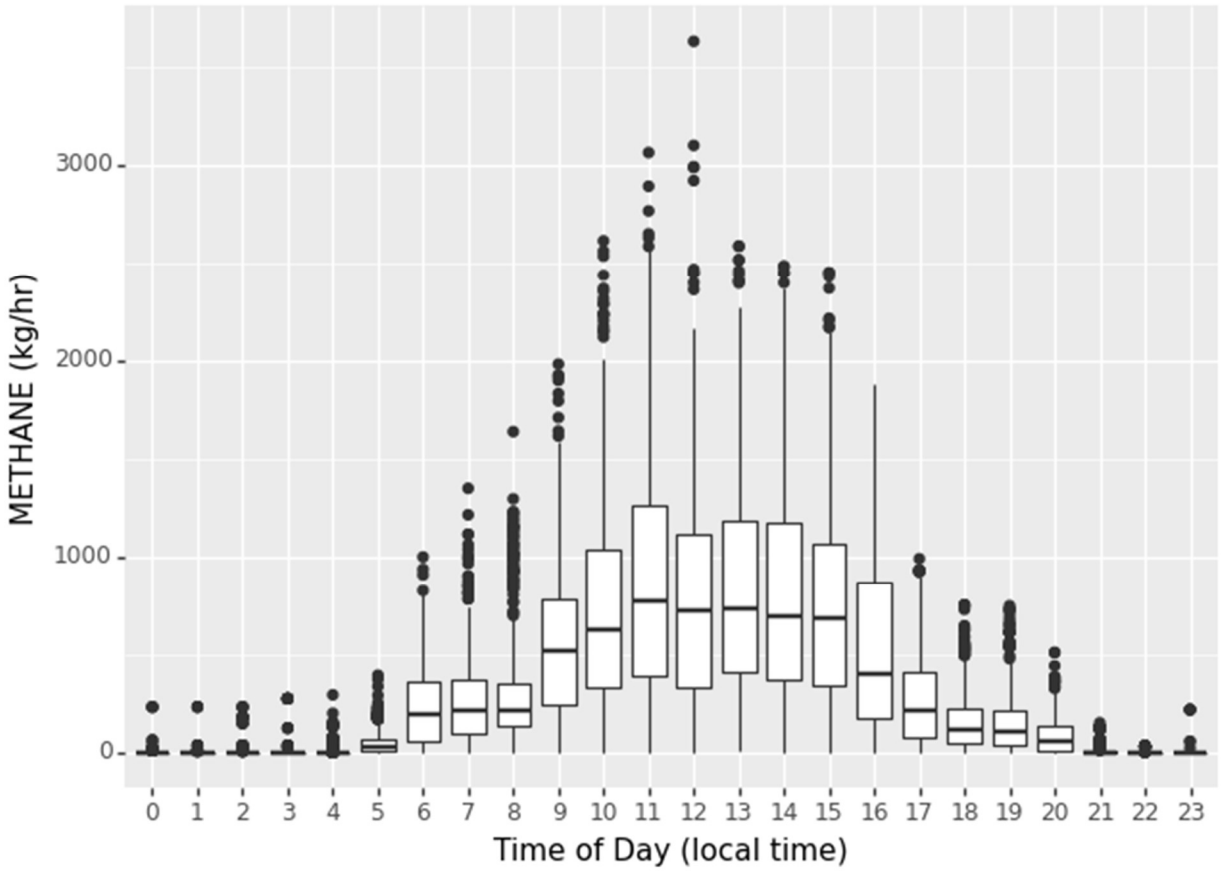


Figure S5: Distributions of methane emissions from manual unloadings for the entire modeling domain as a function of the time of day across 10 Monte Carlo instances of 81 modeling days. All unloadings were modeled as manual unloadings which require operator supervision and occur typically during working hours.

S3.7 Well Completions

MEET models completion activities for wells with user-supplied completion dates that fall within the simulation period. During a completion, MEET ignores all on-site equipment and attributes a constant emission rate to the site for the duration of the completion. Completion conditions including duration and emission rate vary considerably between cases depending on well design and operator-specific operating procedures. Data used in this work were sampled from measurements of well completions from multiple U.S. production regions by Allen, et al.[7] MEET allows the user to select either one of (1) an EPA GHGRP emission factor, (2) measured values from Allen et al.,[7] (3) potential values from Allen et al.,[7] or (4) a user specified emission rate. In this work, it was assumed that all completions would have emission controls. A total of 135 completions were modeled, all with controls. Emission rates and event durations were sampled from the measurements of controlled completions reported by Allen, et al.[7] Completion statistics for the modeling period are summarized in Table S15.

	Per Event Statistics			Domain Total Statistics		
	Event Duration (hr)	Ethane emission rate (kg/hr)	Methane emission rate (kg/hr)	Simultaneous Event counts	Ethane emission rate (kg/hr)	Methane emission rate (kg/hr)
Mean	80.9	11.61	30.95	5.5	48.38	128.86
Minimum	5	0.11	0.29	1	0.11	0.29
Median	45	6.01	16.02	5	36.52	98.36
Maximum	340	72.22	178.48	21	341.34	935.04

Table S15: Summary of completion events and emissions for the modeling period. 135 wells were simulated with completion events during the 81-day modeling period. The per-event statistic summarizes hourly emissions for individual wells aggregated across time and Monte Carlo instances (e.g., every simulated completion across the 10 Monte Carlo instances is considered in a single set and the mean ethane emission rate per event is the average ethane emission rate of all the emission events in this overall set). Domain total statistics summarize the relative occurrence and impact of completions within the domain during every hour of the simulation across the 10 Monte Carlo instances (e.g., at each hour, every simulated completion across the 10 Monte Carlo instances occurring during that hour is aggregated into a single set with the statistics describing the aggregated emissions so the mean ethane emission rate is the average of the overall ethane emission rate from unloadings occurring during each hour of the simulation).

The number of well completions and last completion date for each well during the modeling period were determined using reported completion dates.[1] The day after the last completion date was assumed to be the first date of production. Constant production rates for each fluid were assumed for the remaining modeling period.

S4. Mid-stream Site Emission Modeling

S4.1: Natural Gas Processing Plants

A base case of 60 natural gas processing plants (NGPPs) in the Eagle Ford Shale (EFS) were identified using data from the GHGRP and EIA.[15,16]. A list of facilities from each data source were obtained and merged into a single aggregated inventory and duplicate records, arising from overlap between the individual inventories, were removed. NGPPs within and up to 50 km from the EFS were considered. The location of each NGPP was cross-referenced using ArcGIS and Google Maps to confirm facilities were appropriately identified. Coordinates reported for 12 NGPPs did not correspond with an oil and gas facility during manual cross-referencing. For these 12 entries, the facility location was further cross-referenced using an address search (if available) and updated. In the instance that an address search did not return a match, a manual search within 10 km of the reported location was used to identify an NGPP or other large gathering facility. If one could be found, the inventory entry was updated to this location. Three

reported facilities could not be identified using any of these methods and were removed from the dataset. An additional 4 NGPPs were manually identified in the innermost modeling domain using Google Maps. These facilities may have been unlisted in the inventories if they were constructed after the EIA dataset was released or their estimated emissions were below the threshold (< 25,000 kg CO_2e/year) required for GHGRP reporting.

Methane emissions from each NGPP were estimated using throughput normalized weighted average facility-level emission rates from Mitchell, et al.[17,18] Ethane emissions were then calculated assuming facility throughput and emissions molar compositions were identical (i.e. methane to ethane mole fractions in the throughput and emission streams are identical; Equation S1). Throughput at each NGPP was calculated as the sum of all produced gas from the nearest wells to a given facility. Individual wells within the nested modeling domain were assigned to the nearest NGPP using a k-nearest neighbor algorithm. 24 NGPPs were not returned as a nearest neighbor to any wells and were removed from further analysis. Total whole-gas throughput (Mcf/day) for each natural gas processing plant was then calculated as the sum of the produced gas flow rates from each of the individual wells assigned to the plant. Whole-gas throughput compositions at each NGPP were calculated as the flow-weighted average of the produced gas compositions from each well assigned to the plant. Wellsite produced gas compositions and flow rates were calculated from MEET. Mitchell, et al.[17,18] report throughput normalized weighted average methane emission rates for 132 midstream gathering and processing facilities obtained from tracer flux measurements in a national survey of midstream facilities. The average emission factor for natural gas processing plants was 0.00136 kg CH_4 emitted per kg CH_4 throughput. Ethane emissions were then calculated following equation S1.

$$E_{C_2H_4} = \frac{F_{C_2H_4}}{F_{CH_4}} \cdot E_{CH_4} = \frac{n_{C_2H_6} * mw_{C_2H_6}}{n_{CH_4} * mw_{CH_4}} \cdot E_{CH_4} = \frac{\frac{Pv_{C_2H_6}}{RT} * mw_{C_2H_6}}{\frac{Pv_{CH_4}}{RT} * mw_{CH_4}} \cdot E_{CH_4} = \frac{v_{C_2H_6}}{v_{CH_4}} * \frac{mw_{C_2H_6}}{mw_{CH_4}} \cdot E_{CH_4}$$

where:

$E_i = Emission\ rate\ of\ species\ i\ \left(\frac{kg}{hr}\right)$

$F_i = Throughput\ flow\ rate\ of\ species\ i\ \left(\frac{kg}{hr}\right)$

Equation S1: Estimating C_2H_6 emissions from CH_4 emissions at midstream facilities assuming the composition of emissions and whole gas throughput streams are equivalent on a molar basis

Statistics for emissions from natural gas processing plants are summarized in Table S16.

	Methane emissions (kg/hr)	Ethane emissions (kg/hr)
Mean	64.3	18.8
Min	0.05	0.01
Max	310.6	87.6

Table S16: Natural gas processing plant emissions statistics for the modeling period. Natural gas processing plants are modeled as continuously emitting with a constant emission rate, so emissions are unchanged in time and between Monte Carlo simulations.

S4.2 Gathering Compressor Stations

Emissions from gathering compressor stations were estimated using a throughput normalized approach similar to emissions estimates for natural gas processing plants. Locations of gathering compressor stations were manually identified within the inner modeling domain using Google Earth and ArcGIS and wells within the inner modeling domain were assigned to each facility using a k-nearest neighbor approach. Only wells within 5 km of each facility were considered for assignment. Gathering compressor stations were only considered for the innermost modeling domain because a data repository does not exist for these facilities and manual identification of facilities across the entire domain is prohibitive. 38 compressor stations were identified. Whole-gas throughput rates and compositions for individual facilities were calculated as the sum and flow weighted average composition of produced gas streams from wells assigned to that facility respectively. Facility-level throughput normalized emission factors kg (CH_4 emitted per kg CH_4 in throughput) were estimated for each gathering compressor station using a linear fit of emissions as a function of whole gas throughput reported Figure 3 of Zimmerle, et al.[17] Methane emissions at each facility were then estimated by multiplying the throughput normalized emission factor by the mass of methane in the throughput. Ethane emissions were then estimated as the mass ratio of ethane to methane in the throughput stream multiplied by the methane emission rate (Equation S1). Gathering compressor station emissions in the modeling period are summarized in Table S17. Emissions from gathering compressor stations were modeled as continuous.

	Methane emission rate (kg/hr)	Ethane Emission rate (kg/hr)
Mean	15.4	4.6
Min	3.6	1.04
Max	33.04	9.2

Table S17: Gathering compressor station summary statistics for the modeling period. Gathering compressor stations are modeled as continuously emitting with a constant emission rate, so emissions are unchanged in time and between Monte Carlo simulations.

S5: Production Site Emission Modeling

Emission and activity factors for each of emission source in the model are summarized in table S18.

Emission sources	Activity Factor	Emission Factor
Pneumatic controllers	GHGI 2017[3,4]	Allen , et al., (2015a)[6]
Chemical Injection pumps	GHGI 2017[3,4]	Allen, et al. (2013)[7]
Condensate tank flash	Condensate production rate[1]; One separator for each oil/condensate producing well is assumed	Calculated by MEET
Water tank flash	Water production rate[1]; One separator for each water producing well is assumed	Calculated by MEET
Equipment leaks	GHGRP RY2018 for leak frequency[2,9]	EPA (2024)[9,10]
Liquid unloading	Number of wells with unloadings: GHGRP RY2018,[2,9] Frequency: Allen, et al., (2022)[2] Timing: Allen, et al, (2015b)[11], and Vaughn, et al., (2018)[12,13]	Allen, et al., (2015b)[11]
Well completion	Completion date[1] Duration: Allen, et al., (2013)[7]	Allen, et al., (2013)[7]
Gathering Compressor Stations	Manual Identification	Mitchell, et al. (2015)[18,19]
Natural Gas Processing Plants	GHGRP (2022)[15], EIA (2017)[16], Manual Identification	Zimmerle, et al. (2020) [17]

Table S18: Emissions estimation criteria for each source type

Because the specific configuration of each site is unknown and emission sources can be episodic, emissions from oil and gas operations have a stochastic component. To capture this, 10 Monte Carlo instances of the emission inventory were created for production site emissions. Figure S6 shows the distribution of the aggregated emission rate for all sources within the modeling domain for each hour of the simulation across all 10 Monte Carlo instances. Figure S7 shows a similar aggregated emission distribution for tank batteries within 5 km of the receptor (38 tank batteries with production assigned from 251 wells). Comparing the two figures reveals a stronger degree of skewness when considering only the nearby sources. For example, the ratio of the 99th percentile to the median in the Figure S6 is 1.21 (i.e., the 99th percentile is only 20% larger than the median) while the same ratio in Figure S7 is 2.37 (i.e., the 99th percentile is 137% greater than the median). This indicates that the total emissions in the overall modeling domain are less variable because the effect of considering numerous sources as an ensemble in the outer domain averages out any significant outliers in individual emission rates. However, this means nearby sources may have a more variable impact on local concentration enhancements as individual sources can have a stronger influence, despite similar underlying emissions behavior.

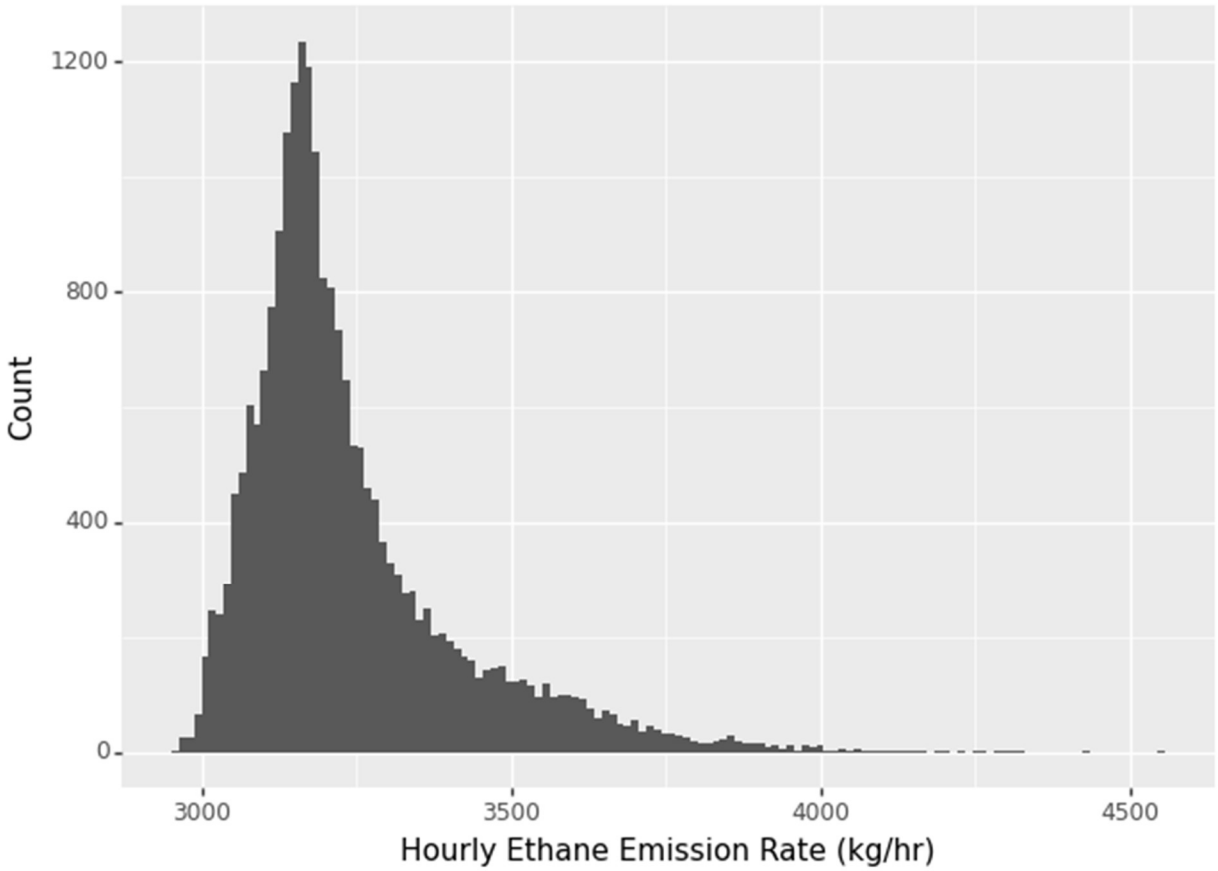


Figure S6: Distribution of total hourly emissions (kg/hr) for all sources in the modeling domain across 10 Monte Carlo instances.

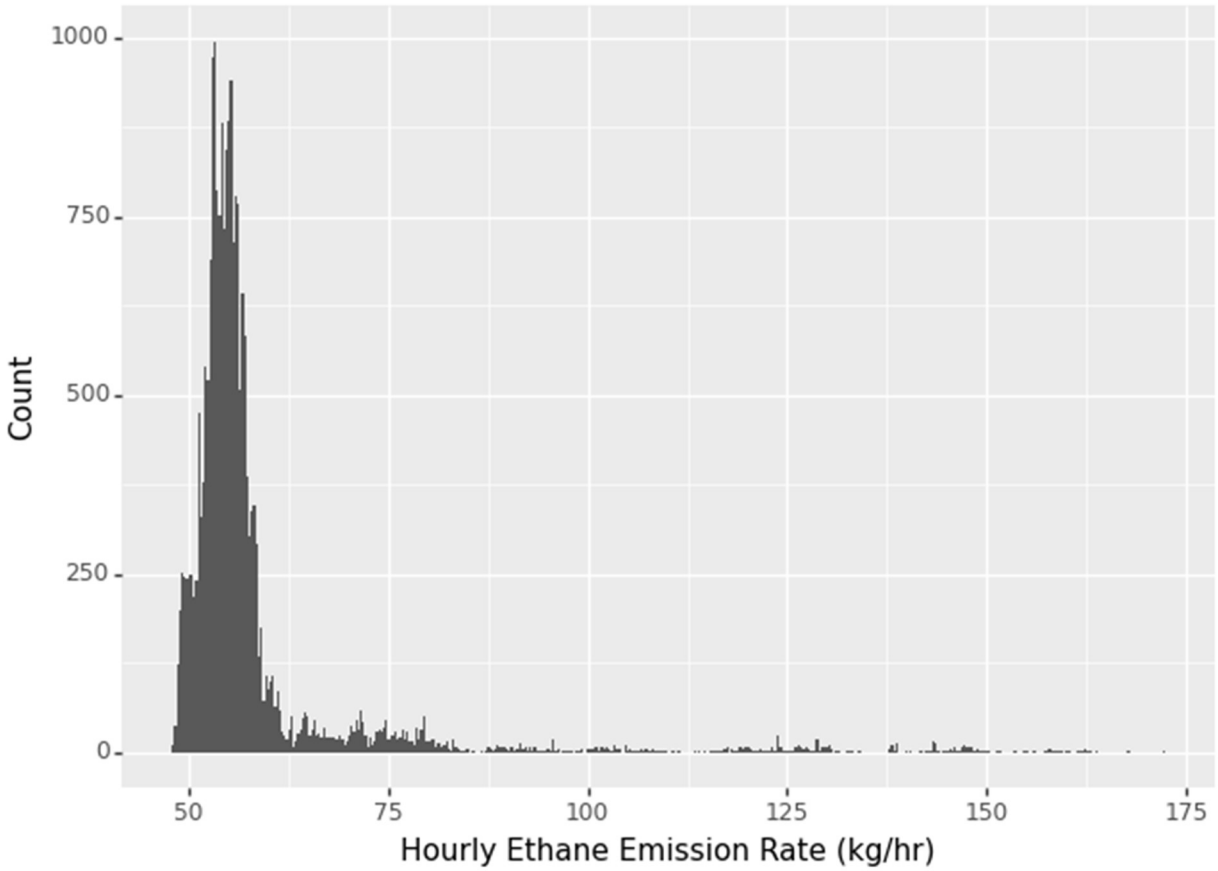


Figure S7: Distribution of total hourly emissions from tank batteries within 5km from the receptor for each hour of the modeling period, for each of 10 MC instances.

Figure S8 and S9 summarize the time-of-day distribution of predicted emission rates from sources within the entire modeling domain and sources near the measurement site respectively. Median aggregated hourly emissions for sources within the entire domain are consistent throughout the night (hours 21-23 and 0-5) but vary slightly (<5%) during the daytime. This variation is driven primarily by unloadings events (Figure S4). This indicates that the large nighttime concentration enhancements observed at the regional continuous monitor, which are also predicted by the model, are not explained by diurnal variations in emission rates from routine sources. Instead, these enhancements are most likely explained by a separate phenomenon (variable mixing heights, variable rates of dispersion, etc.). Median aggregated hourly emissions profiles for the nearby sources show similar behavior, with daytime variations largely stemming from unloadings and completion emissions.

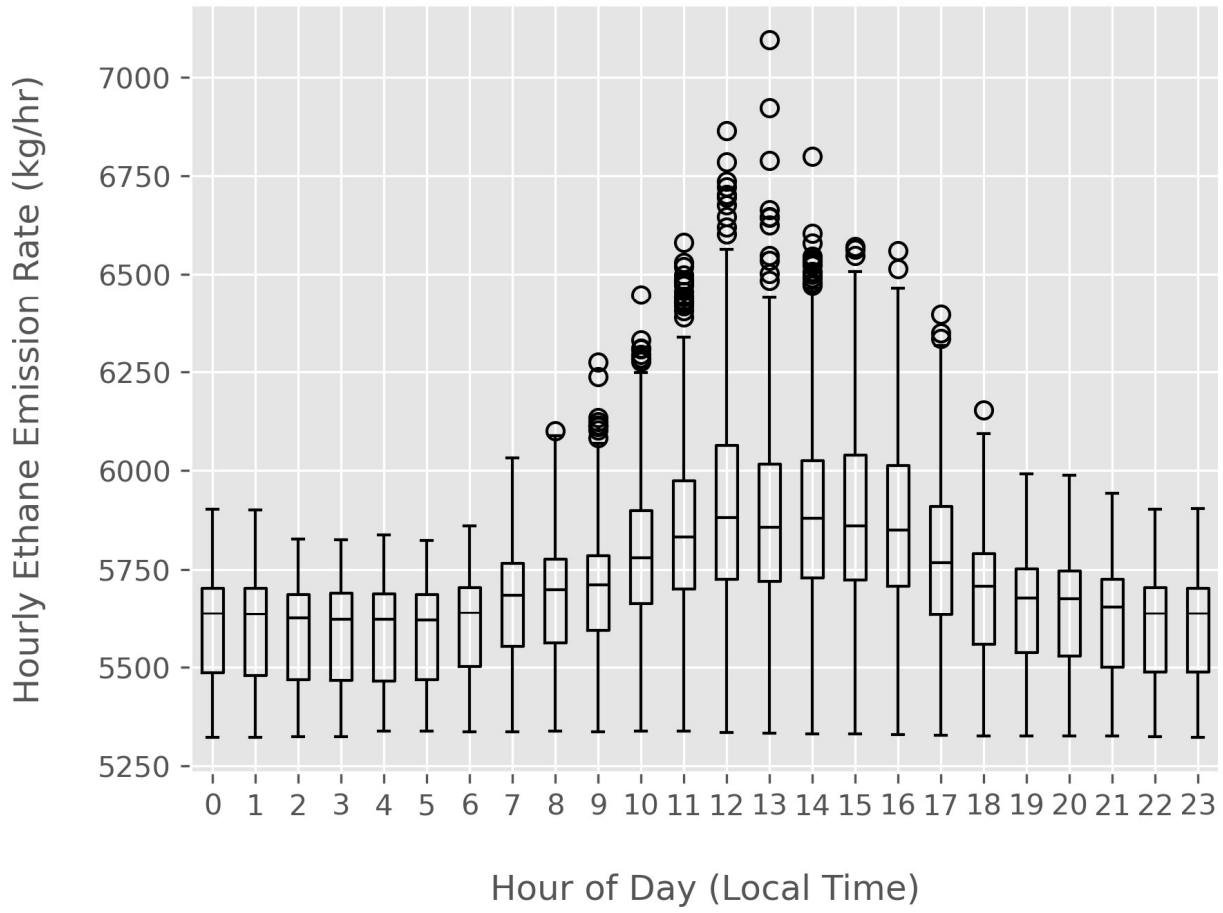


Figure S8: Distributions of total hourly emissions from all sources within 100km of the measurement site.

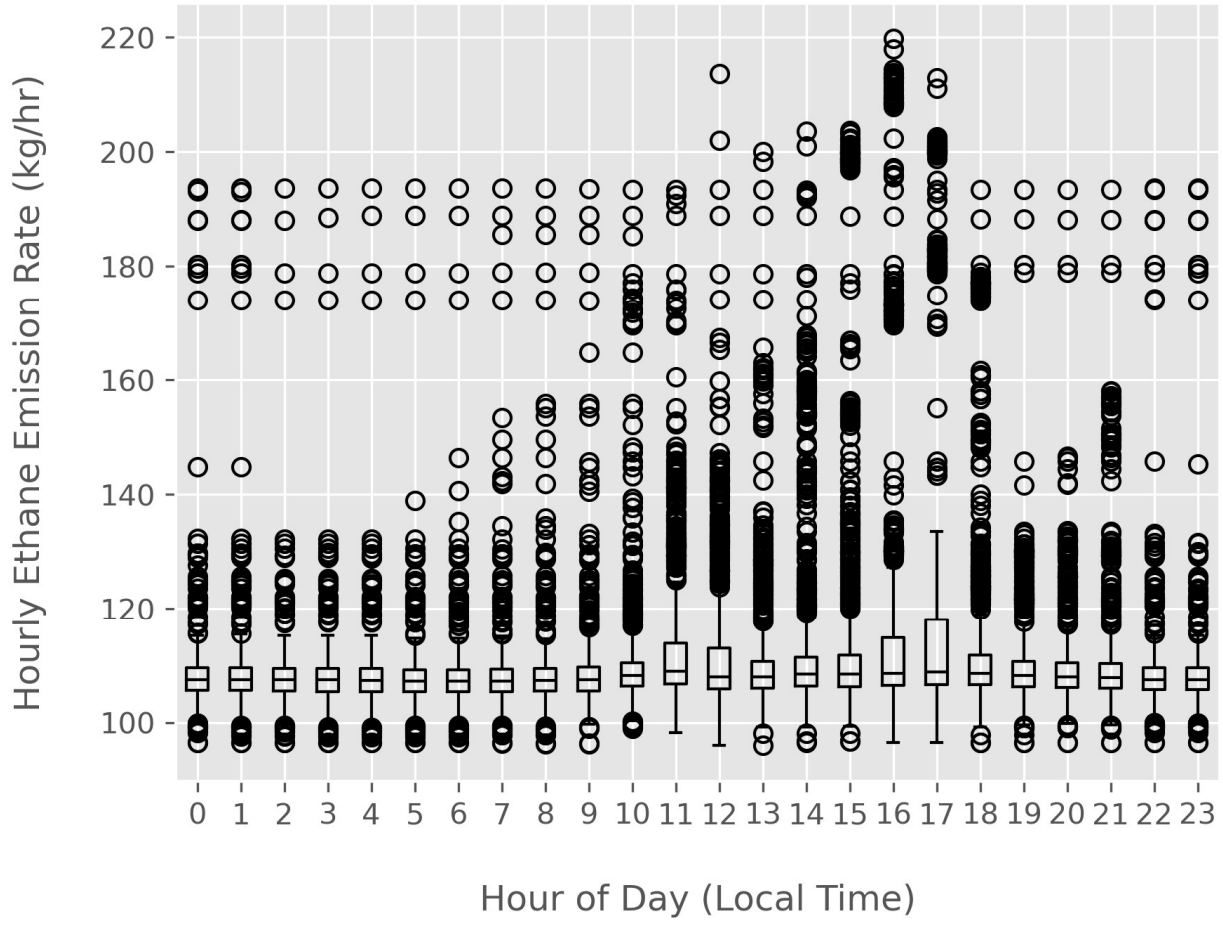


Figure S9: Distributions of total hourly emissions from all sources within 5 km of the auto-GC site. The presence of consistent outliers around 180-200 kg/hr are the result of multi-day completion events, modeled with continuous emission rates.

S6: Emissions Speciation

Twelve regions exhibiting distinct stratification in process conditions and production composition were identified by Gherabati, et al.[20] in the Eagle Ford Shale based on production gas-to-oil ratios (GORs). These regions were then aggregated by fluid type into three composite regions: oil, condensate, and dry gas (Figure S10). For each of the three regions, median values of GOR and separator pressure from were calculated from emission composition reports available through the Texas Railroad Commission.[20,21]. The median GOR and median separator pressure were then paired and compared to the GOR and separator pressure from reports in the respective region. The report with the shortest Euclidian distance to the median GOR and separator pressure value pairing in each region were then selected as being representative of the entire region. A fourth category, "other", was created for wells outside of the twelve original regions. A representative report for this region was selected using the same methods outlined above. Separator compositions for all wells in each region were then queried from the Emissions

Composition Tool (ECT) using the operational and production parameters associated with the region-specific representative well.[22] Representative compositions are summarized in Table S19.

Region	Separator Pressure (psia)	API Gravity	Gas to Oil Ratio (scf/bbl)	Separator Temperature (F)	Wellstream Methane (molar fraction)	Wellstream Ethane (molar fraction)	Wellstream Propane (molar fraction)
Dry Gas	109	53.1	14266	91	0.717	0.119	0.065
Wet Gas	933	52.0	4130	173	0.798	0.114	0.037
Oil	148	47.3	4203	100	0.775	0.127	0.054
Other	185	44.6	6381	133	0.790	0.120	0.047

Table S19: Representative compositions and operational parameters for each of four production regions identified in the Eagle Ford Shale.

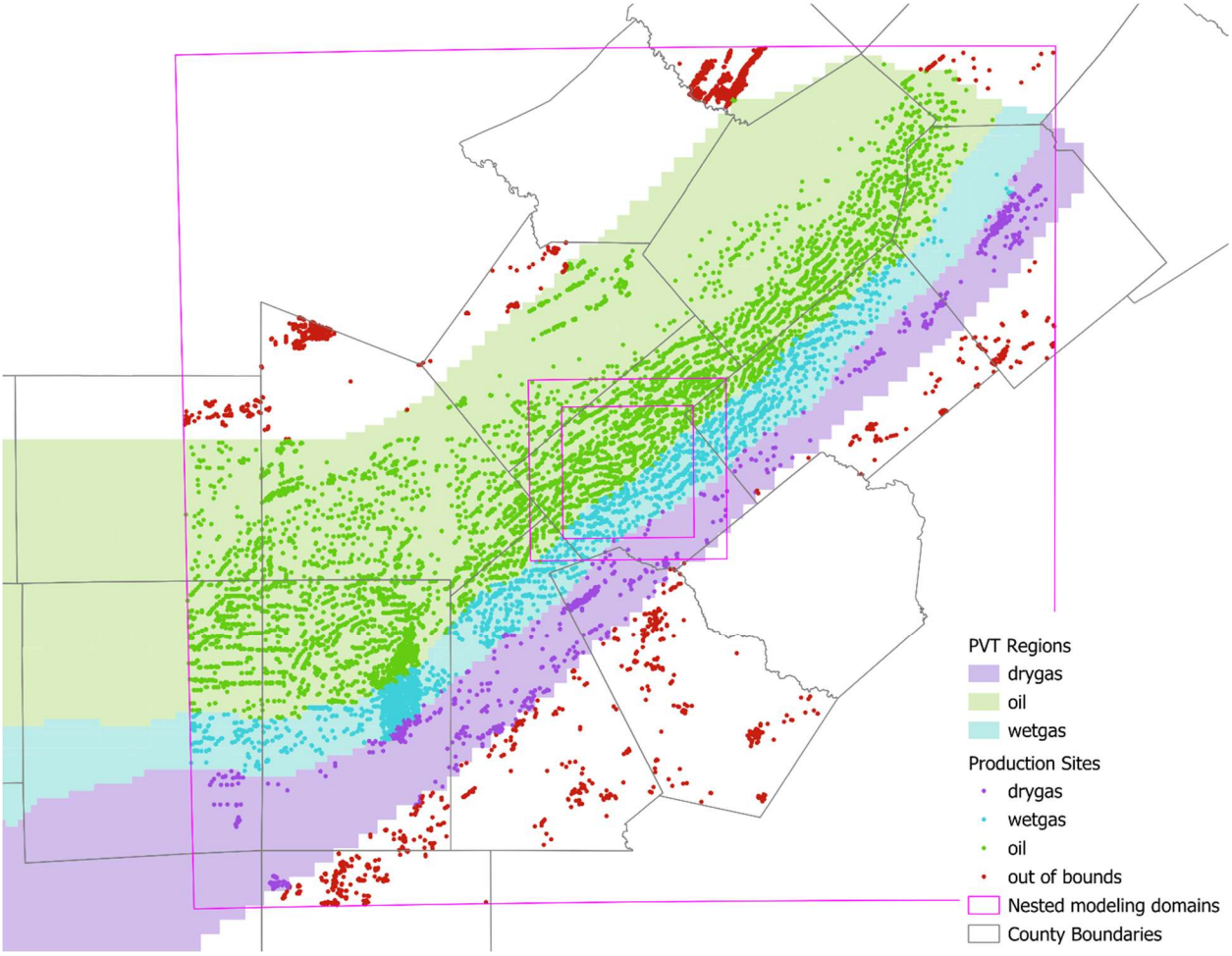


Figure S10: PVT region boundaries and well assignments

A majority of the equipment types have simulated emissions compositions equivalent to the separator overhead (produced gas) or wellstream compositions. However, emissions simulated from tank flashes are derived from the separator liquid and water fractions. The volume of the flash gas in each separator dump to the atmospheric tanks is calculated using production and operational data and the thermodynamic models embedded in MEET. First, stability testing determines whether one or two hydrocarbon phases are present in separator. In this study, two phases were predicted in every case. The Peng-Robinson equation of state (PR-EOS) is then used to determine the composition of the produced gas and liquid hydrocarbon streams. The composition in the gas and condensate phases in the atmospheric condensate storage tank can then be calculated using similar stability and EOS approaches. Tanks are typically in an open system configuration where headspace gas can be exchanged with the atmosphere through various processes including handling and breathing losses in addition to the flashing of high vapor pressure components as separator dumps undergo reductions in pressure. Since ethane is highly volatile, it was assumed that dissolved ethane in the separator liquid will be released completely to the atmosphere. Ethane emission (kg/day) rates can thus be estimated using the condensate production rate (bbl/day) and the ethane fraction (kg ethane / bbl condensate) in the separator liquid. For water tank flashing, the fraction of dissolved hydrocarbons in the water within the separator is estimated with Henry's law. The dissolved hydrocarbons are assumed to be released to atmosphere because of the low solubility of hydrocarbons in water.

Table S20 summarizes relevant compositions and emission factors for separator and wellstream compositions for each PVT region.

	Species	Dry Gas Region	Wet Gas Region	Oil Region	Other
Wellstream Composition (Molar Fraction)	Methane	0.674	0.645	0.669	0.659
	Ethane	0.128	0.116	0.114	0.100
	Propane	0.060	0.051	0.050	0.048
Produced Gas Composition (Molar Fraction)	Methane	0.716	0.794	0.761	0.763
	Ethane	0.135	0.118	0.126	0.111
	Propane	0.063	0.039	0.050	0.048
Separator Liquid Composition (Molar Fraction)	Methane	0.010	0.222	0.047	0.054
	Ethane	0.010	0.109	0.038	0.035
	Propane	0.015	0.085	0.047	0.044
Separator Water Composition (Molar Fraction)	Methane	3.99E-05	8.71E-04	1.91E-04	1.94E-04
	Ethane	9.67E-06	1.54E-04	4.05E-05	3.35E-05
	Propane	3.43E-06	3.80E-05	1.23E-05	1.07E-05
Condensate Tank Flash Emission Factor (kg/bbl condensate)	Methane	0.149	8.043	0.780	0.904
	Ethane	0.278	7.411	1.175	1.079
	Propane	0.630	8.469	2.171	2.016
Water Tank Flash Emission Factor (kg/bbl water)	Methane	0.0056	0.1233	0.0270	0.0274
	Ethane	0.0026	0.0410	0.0107	0.0089
	Propane	0.0013	0.0148	0.0048	0.0042

Table S20: $C_{1\text{-}3}$ Emission compositions for each separator stream in each region

Condensate and water tank flash emission rates are estimated based on the frequency and volume of separator dumps which is controlled by wellstream production rates, separator size and operating conditions, and the process stream compositions. Emissions from tanks are typically controlled. To simulate emissions controls, tank emission rates are multiplied by a constant recovery efficiency factor (95% in this this study) and the difference is assumed to be released into the atmosphere (i.e., 5% of whole gas emissions are released). Emission rates from other categories are estimated as either whole gas or methane emissions in unit time by the discrete event simulator in MEET and the wellstream or produced gas composition is applied to estimate emission rates of ethane. Table S21 summarizes emission source categories and their predicted gas phase compositions.

Process Stream	Associated Equipment
Wellstream	Liquids Unloading
	Well Completion
Separator Overhead (Produced Gas)	Pneumatic Controllers
	Leaks
	Chemical Injection Pumps
	Gathering Compressor Stations
	Natural Gas Processing Plants
	Flares
Separator Exit to Condensate Tank	Condensate Tank Flash
Separator Exit to Water Tank	Water Tank Flash

Table S21: Separator process stream and associated equipment

S7: Dispersion Modeling

Terrain Data:

Terrain data was sourced from the United States Geological Survey's National Elevation Database (NED).[23] Twelve, 1 arc-second (30x30m) Digital Elevation Model (DEM) files (US Geological Survey, 2019) covering individual subregions of the 200 × 200km region were decompressed before input into the terrain pre-processor. Data from the year 2020 were used for all files.

Land Cover Data:

Land cover data were sourced from the Multi-Resolution Land Characteristics (MRLC) Consortium's National Landcover Database (NLCD).[24] Land cover data from the NLCD CONUS land cover product was ingested in the ctgproc preprocessor. All data were from the year 2021 with a resolution of 30 m × 30 m.

Meteorological Inputs

CALMET, the meteorological pre-processor for CALPUFF, requires a combination of surface and upper air data inputs to prepare a gridded 3-dimensional meteorological field for input into CALPUFF. On-site surface observations of wind speed (5-minute averaged wind; 10 m AGL), wind direction, temperature (2m AGL), and relative humidity were sourced from the continuous ambient monitoring station (CAMS 1070) operated by the TCEQ in Karnes City.[25] Additional

surface meteorological data from across the modeling domain were obtained from 15 surface observation stations (1 hour resolution) and 6 Automated Surface Observation System (ASOS1; 1 minute resolution) stations reporting to the Global Hourly Integrated Surface Database (ISD).[26–28] Typical CALMET applications utilize twice daily radiosonde observations for characterizing upper air meteorological fields. However, the nearest upper air station in the region is located in Corpus Christi, TX, approximately 150 miles away from the receptor point in the center of the domain and much closer to the coastline. Instead, upper air data variables were extracted from the High-Resolution Rapid Refresh (HRRR) archive at 8 locations across the domain.[29] HRRR is a prognostic weather forecasting model with data assimilation capabilities that models atmospheric dynamics hourly at a gridded resolution of 3 km^2. For each of the eight selected grid cells, temperature, pressure, elevation, wind speed, and wind direction were extracted for 25 sounding levels between the surface (approximately 1000 mb) and 500 mb pressure levels. Meteorological data sources are summarized in Table S22.

Level	Station ID	Name	Source
Surface	722416	New Braunfels Municipal Airport	ASOS1
Surface	722533	Hondo Municipal Airport	ASOS1
Surface	722524	Aransas County Airport	ASOS1
Surface	722530	San Antonio International Airport	ASOS1
Surface	722523	Stinson Municipal Airport	ASOS1
Surface	722550	Victoria Regional Airport	ASOS1
Surface	720391	Beeville Municipal Airport	ISD
Surface	720395	Pleasanton Municipal Airport	ISD
Surface	722209	Calhoun County Airport	ISD
Surface	722369	T P MC Campbell Airport	ISD
Surface	722518	Naval Auxiliary Landing Field	ISD
Surface	722535	Lackland Air Force Base (Kelly Field Annex)	ISD
Surface	722536	Randolph AFB Airport	ISD
Surface	722537	Kerrville Muni/Louis Schreiner Field Airport	ISD
Surface	722539	San Marcos Municipal Airport	ISD
Surface	998165	Mission Aransas Reserve	ISD
Surface	998185	Seadrift	ISD
Surface	998226	Rockport	ISD
Surface	999999	Roger M Dreyer Memorial Airport	ISD
Surface	989999	Port Aransas 32 NNE	ISD
Surface	900008	Fayette Regional Air Central Airport	ISD
Surface	482251070	Karnes County	TCEQ
Upper	NA	HRRR extract	HRRR
Upper	NA	HRRR extract	HRRR
Upper	NA	HRRR extract	HRRR
Upper	NA	HRRR extract	HRRR
Upper	NA	HRRR extract	HRRR
Upper	NA	HRRR extract	HRRR
Upper	NA	HRRR extract	HRRR
Upper	NA	HRRR extract	HRRR

Table S22: Description of meteorological data sources for inputs CALMET. Station ID refers to United States Air Force USAF station catalogue ID (USAF ID) for ASOS1 and ISD sites, and the US Environmental Protection Agency Air Quality System ID (AQS ID) for TCEQ monitors.

CALPUFF

The default settings for CALPUFF v7.2.1 were used with the exception of the items listed in Table S23

	Used in this work	Default	Note
Chemical Transformation	Not modeled (MCHEM=0)	MESOPUFF II scheme (MCHEM=1)	A sensitivity analysis was performed assuming an OH concentration of $1.0*10^7$ molecules/cm^3, (daytime urban environment) and the impact was negligible.
Dry and Wet Deposition	Not modeled (MDRY=0, MWET=0)	Modeled (MDRY=1, MWET=1)	Similar to the reactivity, interactions of light alkanes with the surface and air are expected to be minimal. Dry and wet deposition are expected to be negligible as a result.

Table S23: Non-default CALPUFF settings selected for modeling.

Source Stack Parameters
Stack parameters for each source category are summarized in Table S24

	Physical Height (m)	Plume Rise * (m)	Effective Height (m)	Diameter (m)	Temperature (K)	Exit Velocity (m/sec)	Initial σ_y (m)	Initial σ_z (m)
Wellhead Leak	0.2	0	0.2	0.1	298.15	0	0	0
Condensate/Water Tank Flash	5.5	0	5.5	0.1	298.15	0	0	0
Other Sources at Production Sites	2.0	0	0.2	0.1	298.15	0	0	0
Any sources at Midstream sites **	6.1	5.0	11.1	0.1	298.15	0	0	0

Table S24: Source category stack parameters

* A fixed height of 5 meters was assumed for plume rise in buoyant/jet sources instead of using the plume rise algorithm implemented in CALPUFF
** Compressors are assumed to be the dominate source of emissions at midstream facilities, so parameters were selected that best represented these sources.

S8. Contribution of Sources by Distance from the Receptor

Mean ethane concentrations predicted by the modeling of routine emissions are influenced by both near and distant sources. Figure S11 shows the fraction of mean concentrations over all days of the simulation period, due to sources within 100 km of the receptor site, that are accounted for by sources within 5 km, 10 km, 20 km and 50 km. The values reported are the ratios, for each hour, of the total concentration predicted at the auto-GC site from sources within a specified distance, to the total concentration predicted from sources in the modeling domain extending to 100 km from the auto-GC site (Equation 2). Sources within 5 km contributed 38%, on average to the mean concentrations predicted when all sources within 100 km from the auto-GC site were accounted for. There was a high degree of variability in this ratio from hour to hour, as shown in Figure 4, depending on whether the nearest sources were upwind of the receptor site. Sources within 10, 20, and 50 km contributed 67, 88, and 98% respectively, on average, of mean concentrations predicted when all sources within 100 km of the auto-GC site were accounted for. As the distance from the auto-GC site increased the variability in these ratios decreased.

$$\text{Concentration Ratio } \left(\frac{x}{100}\right) = \frac{\sum \text{Concentration contribution from sources} \leq \text{x km from measurement site}}{\sum \text{Concentration contribution from sources} \leq \text{100 km from measurement site}}$$

Equation S2: Concentration ratio calculation for individual hours

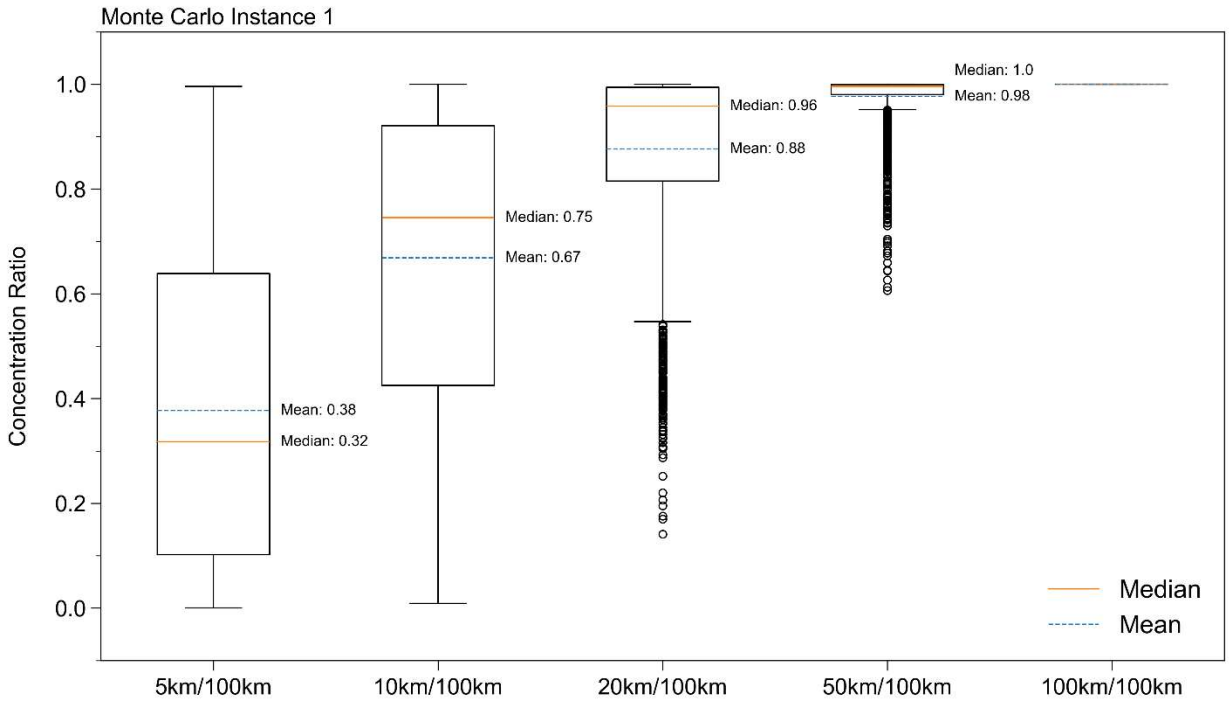


Figure S11: Hourly concentrations by contributions from different source distances. Each record in each of the boxplots represents the concentration contribution from sources within x km from the receptor point divided by the concentration contribution from sources within 100 km from the receptor point for each hour of the entire modeling period.

To evaluate the influence of distant sources on large enhancement events, fractional ethane concentration contributions from sources out to 100 km from the receptor point were evaluated for the top 20 predicted enhancement events for both the nighttime and daytime intervals of the overall modeling period (Figure S12). The largest enhancement events during the day are more strongly influenced by more distant sources than the largest enhancement events at night with many of the largest nighttime sources having minimal contributions from sources beyond 20km. This indicates large enhancement episodes during the daytime and nighttime can be driven by different types of sources and underscores the need to simulate emissions from 50+ km away.

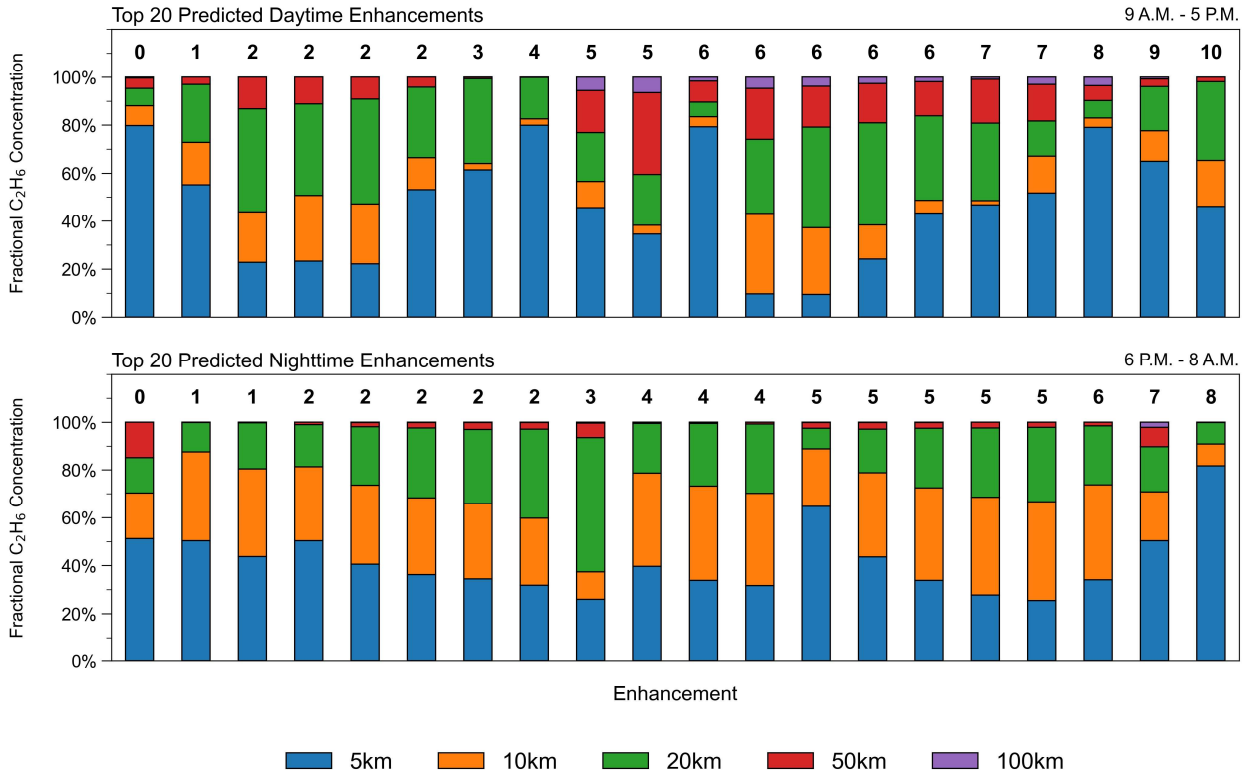


Figure S12: Fractional ethane concentrations for sources out to 100km for the top 20 predicted daytime and nighttime enhancement episodes. Numbers 0-10 and 0-8 in the daytime and nighttime figures indicate clusters of enhancements that occur during adjacent hours (e.g., the top 20 predicted daytime enhancements show 11 distinct episodes, with a few of the top enhancements occurring within the same episode).

S9. Statistical Performance Analyses of Dispersion Modeling

Qualitative and quantitative analyses were conducted to evaluate model performance. Quantitative performance metrics included bivariate correlation coefficients (Pearson's r and Spearman's ρ), error analyses (normalized mean squared error and root-mean squared error), and fractional and geometric mean bias (Equations S3 – S7). Pearson (Equation S3) and Spearman's rank (Equation S4) correlation coefficients (PCC and Spearman's ρ herein), were derived using observed and predicted concentrations as the analysis variables.

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 (y_i - \overline{y})^2}}$$

Equation S3: Pearson Correlation Coefficient

r = correlation coefficient

x_i = values of x-variable in the sample

$\overline{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in the sample

$\overline{y}$ = mean of the values of the y-variable

$$\rho = \frac{\sum(x_{i,r} - \overline{x_r})(y_{i,r} - \overline{y_r})}{\sqrt{\sum(x_{i,r} - \overline{x_r})^2 (y_{i_r} - \overline{y}_r)^2}}$$

Equation S4: Spearman's Rank Correlation Coefficient

ρ = Spearman's rank correlation coefficient

$x_{i,r}$ = ranked values of the x-variable in the sample

$\overline{x}_r$ = mean of the ranked values of the x-variable

$y_{i,r}$ = ranked values of the y-variable in the sample

$\overline{y_r}$ = mean of the ranked values of the y-variable

In each case, records from the observed and predicted time series are paired in time and the strength of their covariance is evaluated. With the Pearson Correlation Coefficient (PCC), the observed and predicted concentration values are compared directly. With Spearman's ρ, each reported concentration in the observed and predicted timeseries are first ranked relative to the other records in that timeseries (e.g., observed records are ranked compared to other observed records) and, while maintaining the original time-pairing, the covariance of these ranks is compared. Spearman's ρ allows for monotonic, but not necessarily linear relationships between

two variables to be evaluated. It is also less sensitive to the influence of a small number of extreme values. Here, the PCC is used as a measure of how well each model reproduces the absolute concentrations measured at the ground-truth receptor at each point in time, and to evaluate conditions where the model may be more or less representative of the observations (e.g., day or night). Spearman's ρ is used here as a measure of how well the models predict the ranking of the enhancements (e.g., are the highest predicted concentrations associated with the highest observed concentrations), regardless of the magnitude of the enhancement. This gives an indication of whether the model is reproducing more general trends in the data, regardless of its tendency to over or underestimate these enhancements. Bivariate analysis was conducted for the entire timeseries, as well as daytime and nighttime periods individually, to evaluate whether diurnal effects may strongly influence the association between model predictions and observations.

Fractional bias (FB; Equation S5) is computed for the average of the entire observed and predicted datasets. This provides a more comprehensive estimate of model bias over the range of estimated values when compared to the screening analysis. A geometric mean bias (MG; Equation 4S6is also calculated using the entire dataset. The geometric mean is like the arithmetic mean, but it uses logarithmic versions of the differences between observed and predicted concentrations and is useful for datasets that span several orders of magnitude, especially when ratios of predicted and observed concentrations display a large range of values.

$$FB = 2 \cdot \frac{(\overline{C_o} - \overline{C_p})}{(\overline{C_o} - \overline{C_p})}$$

C_p = model predictions
C_o = observations
$\overline{x}$ (overbar on variable) = average of variable

Equation S5: Fractional Bias

$$MG = exp\left(\overline{ln(C_o)} - \overline{ln(C_p)}\right)$$

C_p = model predictions
C_o = observations
$\overline{x}$ (overbar on variable) = average of variable

Equation S6: Geometric Mean Bias

Normalized mean square error (NMSE; Equation S7) and root mean square error (RMSE; Equation S8) are calculated as error metrics. NMSE and provide information about the spread of the residuals between predicted and observed concentrations and they are commonly used metrics for evaluating model performance. The squared error gives an indication of how closely a model predicts an individual observation. The mean of the squared error for every pair of

predictions and observations (NMSE) in a test dataset indicates how well, on average, the model predicts the range of test data. The absolute error is a similar metric but is less sensitive to outliers. Both metrics are normalized by representations of the means of the observed and predicted values to facilitate more straightforward cross-model comparisons.

$$NMSE = \frac{\overline{(C_o - C_p)^2}}{\overline{C_o} \cdot \overline{C_p}}$$

C_p = model predictions
C_o = observations
$\overline{x}$ (overbar on variable) = average of variable

Equation S7: Normalized Mean Squared Error

$$RMSE = \sqrt{\overline{(C_o - C_p)^2}}$$

C_p = model predictions
C_o = observations
$\overline{x}$ (overbar on variable) = average of variable

Equation S8: Root Mean Square Error

The fraction of data within a factor of two of observations (FAC2) is also evaluated.

$$\text{FAC2} = \text{fraction of data that satisfy } 0.5 \leq \frac{C_p}{C_o} \leq 2$$

C_p = model predictions
C_o = observations
$\overline{x}$ (overbar on variable) = average of variable

A summary of these statistics for nighttime, daytime, and all enhancements is provided in Table S25.

Time of Day	FRAC2	NMSE	NAE	RMSE	MG	FB	PCC (r)	Spearman's ρ
All	0.4	3.87	0.86	61.93	0.84	-0.24	0.54	0.67
Night	0.38	2.81	0.85	74.73	0.7	-0.36	0.57	0.69
Day	0.42	9.28	0.93	36.5	1.09	0.39	0.43	0.57

Table S25: Statistics comparing model predictions and observed enhancements by time-of-day

More qualitative data-exploration analyses can supplement statistical model performance metrics and provide additional information about where model performance issues may arise. Quantile-quantile and parity plots were prepared to evaluate the agreement between the observed and predicted datasets. Figure S13 shows the observed concentrations paired in time with the corresponding model prediction and plotted as a log-log parity plot. Concentrations are relatively well represented (within a factor of 2x) across the range of observed concentrations. The linearity observed in the log-log parity plot suggests a non-linear association exists between the observed and predicted enhancements, but that the lowest and highest observed enhancements are typically predicted as some of the lowest and highest predicted enhancements. This is supported by the bivariate correlation coefficients (Pearson correlation coefficient and Spearman's ρ). The model has a weaker linear association (Pearson's r) than monotonic association (Spearman's ρ) suggesting that the model is performing reasonably well at identifying the timing and ranking of peaks, but the degree of over and under prediction of the magnitudes of each peak is non-linearly associated with the observed concentrations.

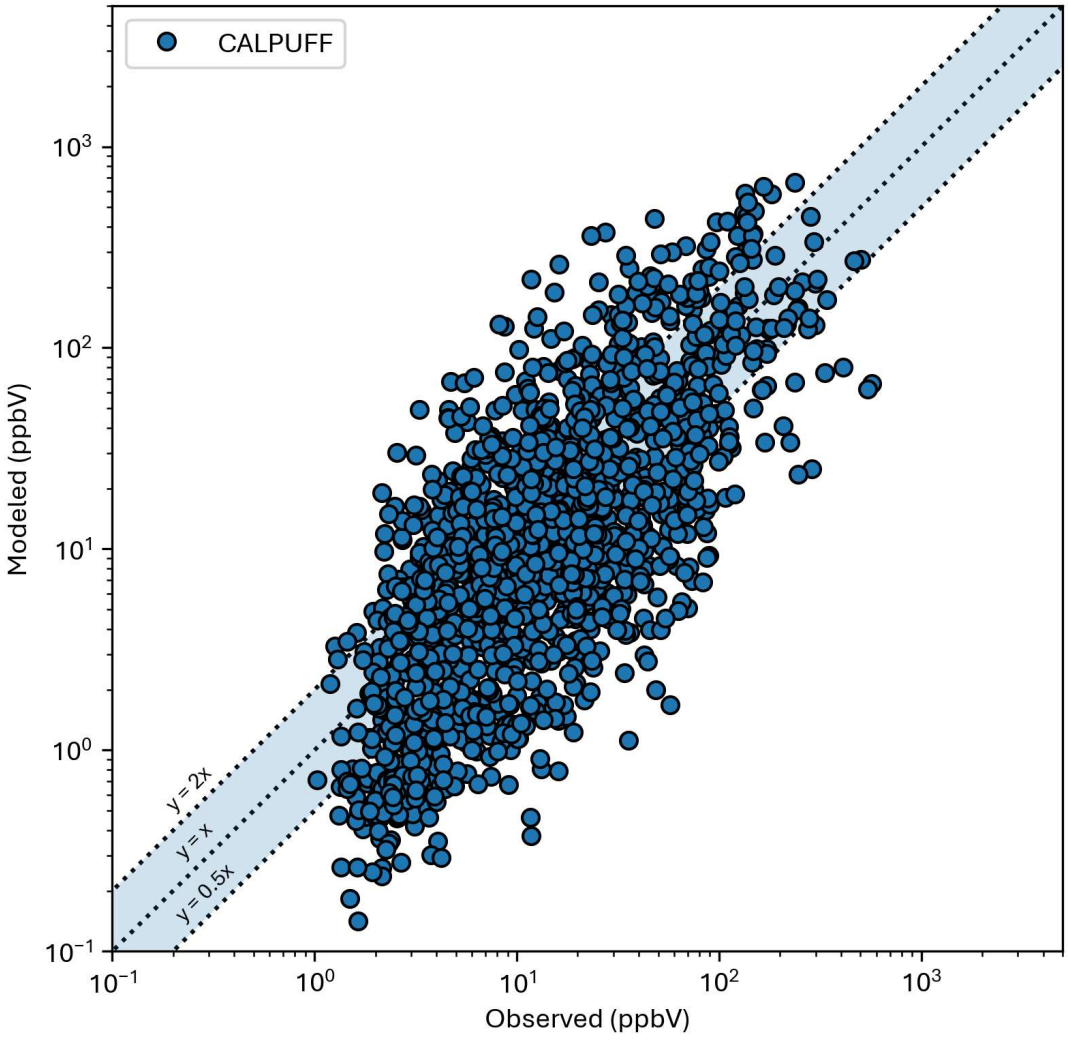


Figure S13: Log-log parity plot showing predicted and observed enhancements paired in time

The distribution of predicted concentrations is compared to the observed distribution in Figure S14. The model generally predicts the range of observed concentrations above 10 ppb well, but it slightly underpredicts the lowest observed enhancements. However, no background corrections were applied to the observations. To estimate these backgrounds, ambient ethane measurements from an Auto-GC in Gregory, TX (Figure S15) were sampled.

Ambient ethane measurements from the Gregory monitor are assumed to be entirely uninfluenced by oil and gas operations in the Eagle Ford Shale**.** However, the monitor is located near several refineries and an ethylene and mono-ethylene glycol production plant that uses ethane as a feedstock (Figure S16)[30]. Background estimates were only calculated for wind directions $\geq$ 300 degrees or $\leq$ 120 degrees to avoid the influence of these sources.

An average background concentration was calculated for both daytime and nighttime periods by averaging all the enhancements occurring during each time-of-day category with measured wind directions $\geq$ 300 degrees or $\leq$ 120 degrees. The average daytime background concentration was 1.85 ppb and the average nighttime background was 2.77 ppb.

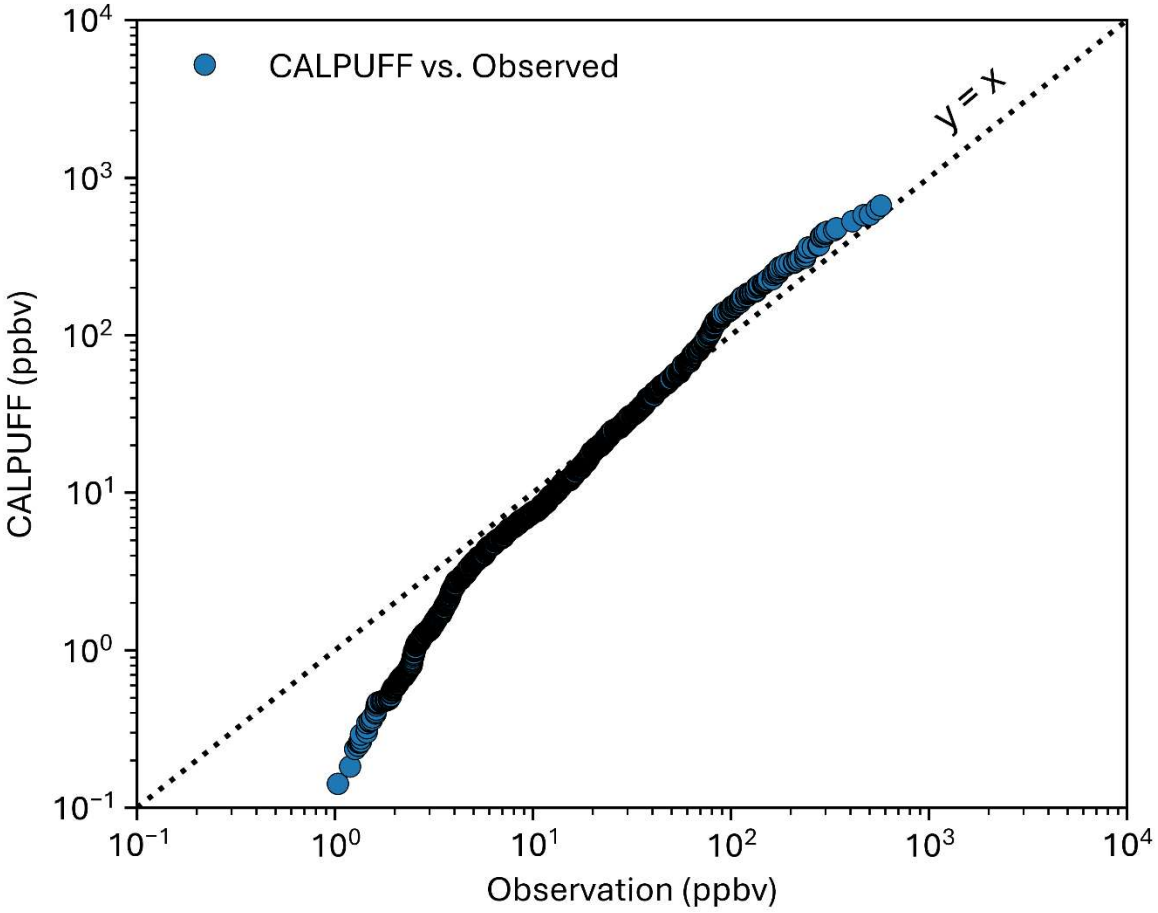


Figure S14: Quantile-quantile plot showing observed and predicted enhancement distributions. Records from each time series (observed and predicted) are sorted and plotted un-paired in time (e.g., the lowest observed enhancement is paired with the lowest predicted enhancement regardless of when they occurred).

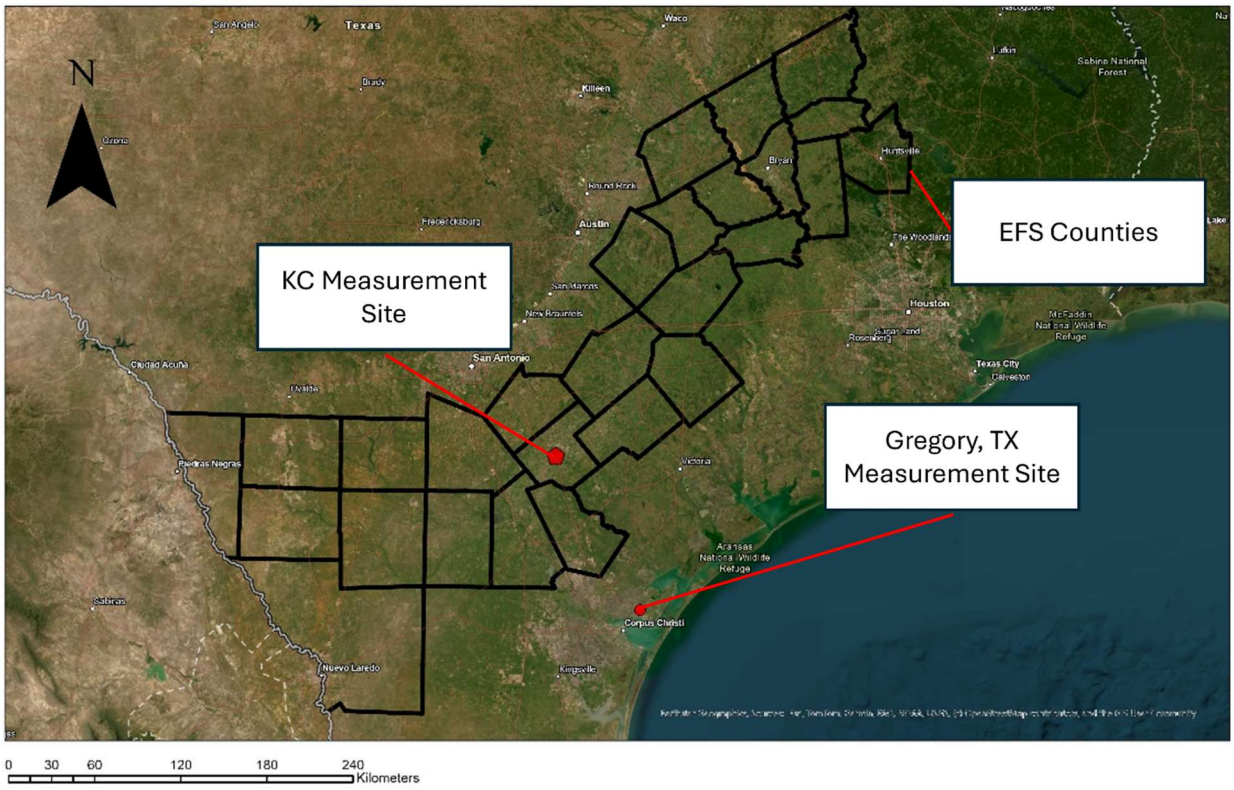


Figure S15: Illustration of the spatial relationship between Corpus Cristi, the Karnes City monitor, and the Eagle Ford Shale. Basemap data provided by Esri.[31]

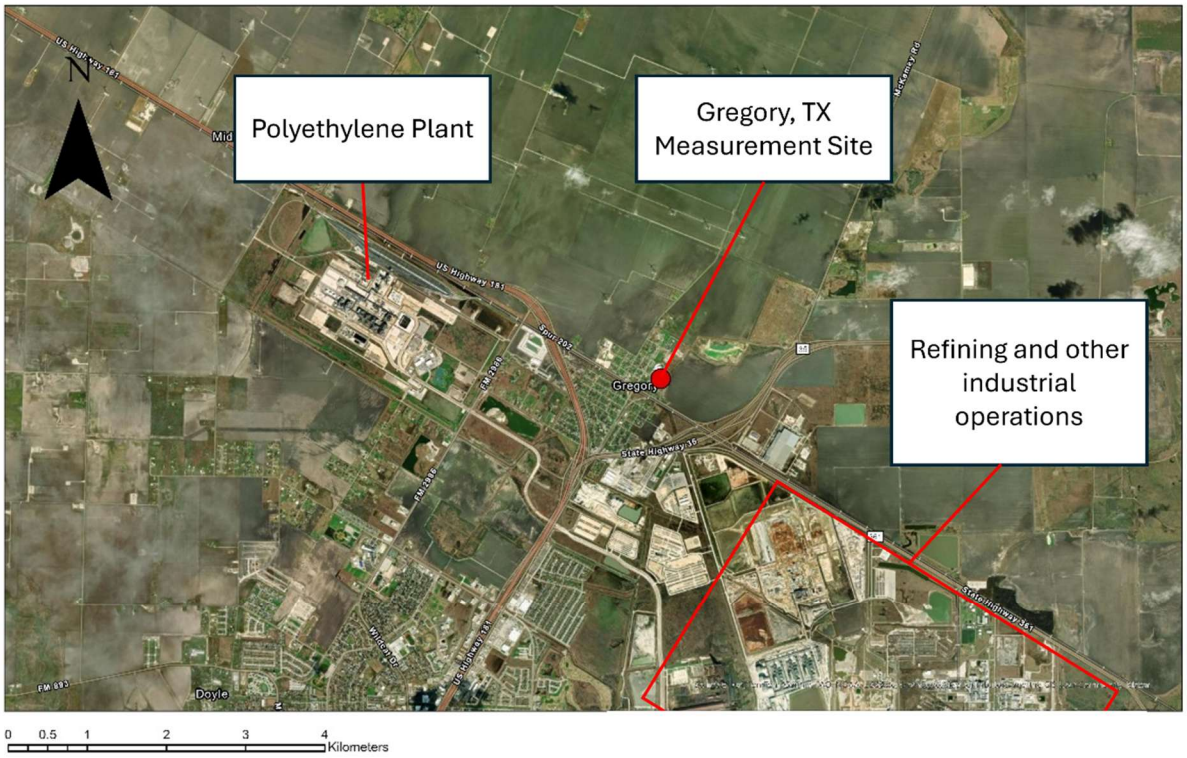


Figure S16: Infrastructure around the Gregory, TX auto-GC measurement location. Basemap provided by Esri.[31]

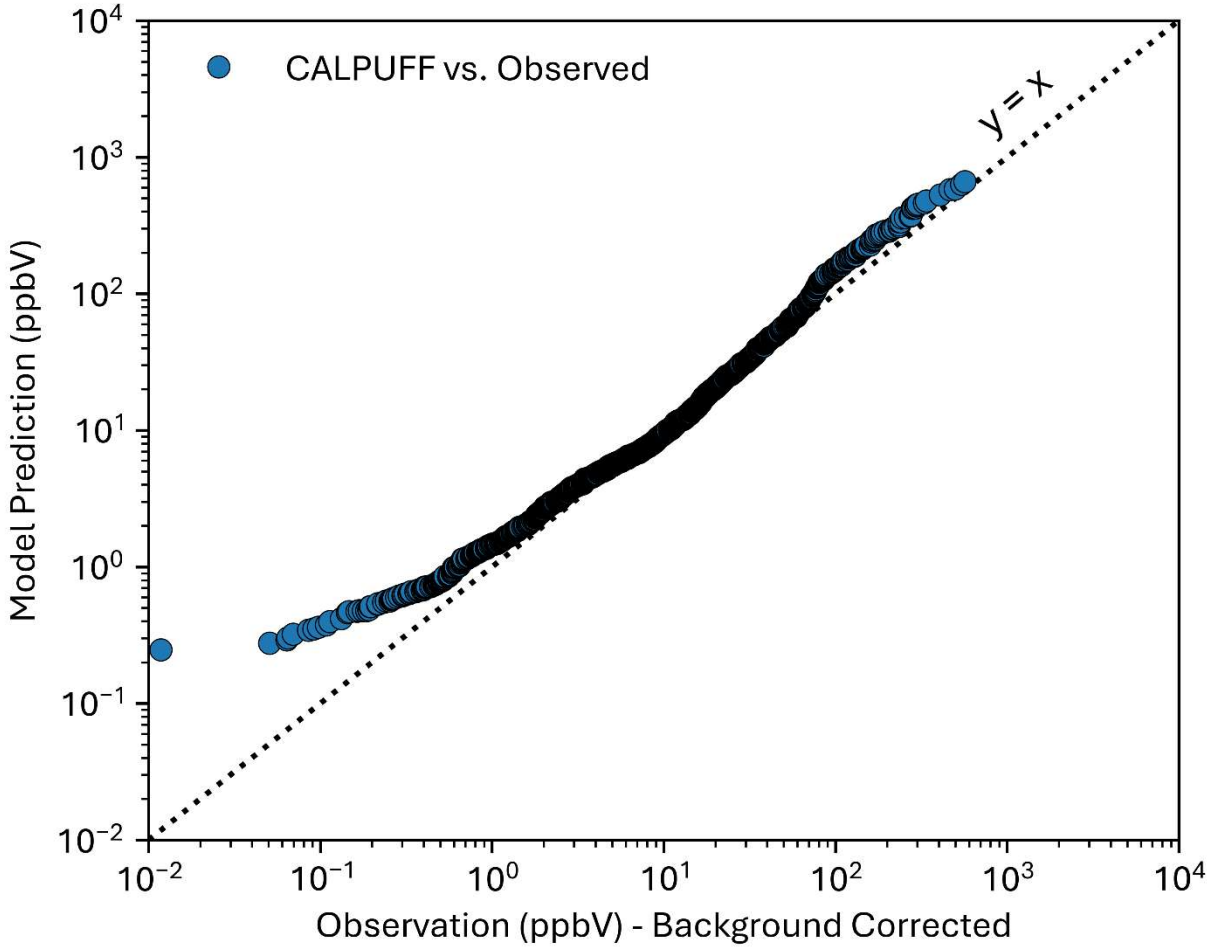


Figure S17: Quantile-quantile plot showing background corrected observations and corresponding model predictions. Records from each time series (observed and predicted) are sorted and plotted un-paired in time (e.g., the lowest observed enhancement is paired with the lowest predicted enhancement regardless of when they occurred).

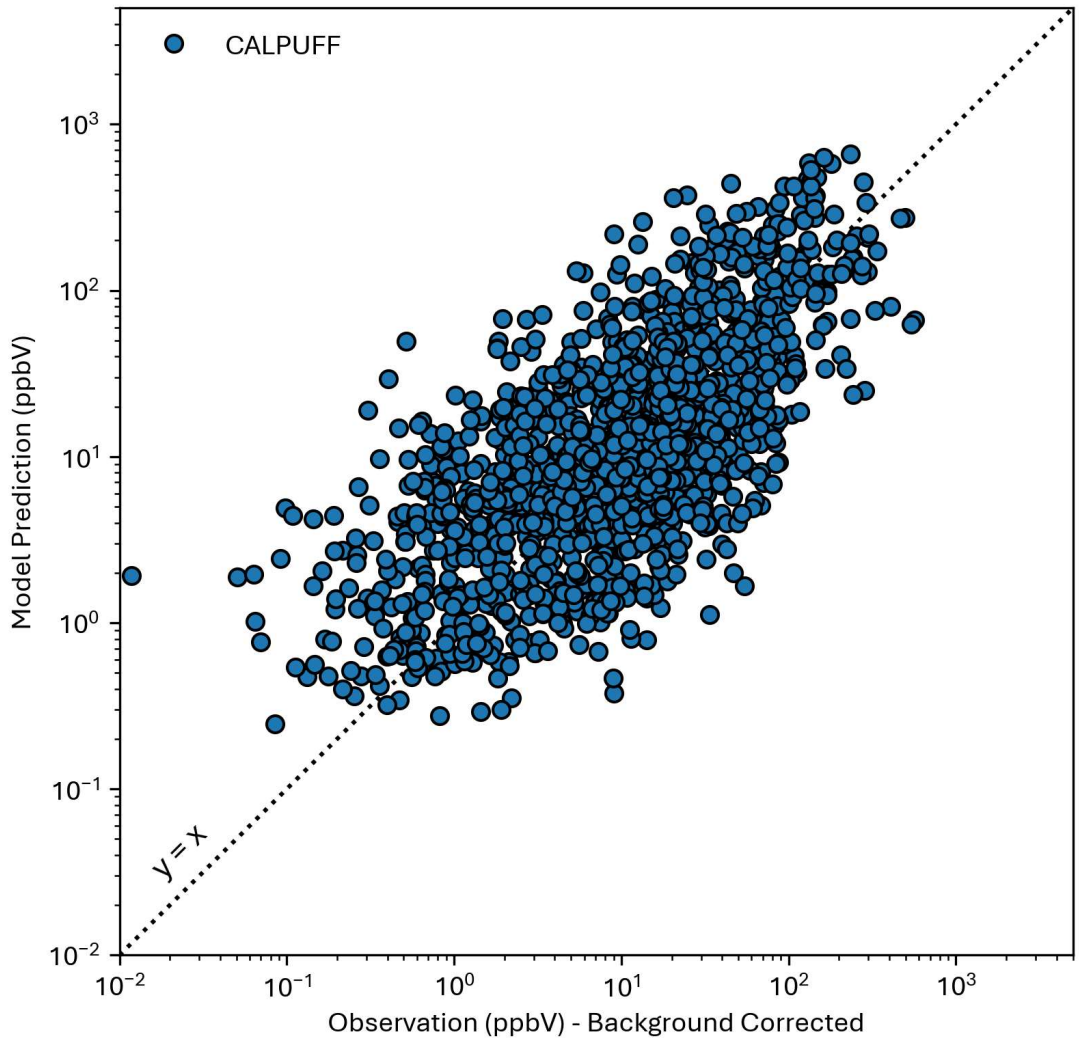


Figure S18: Log-log parity plot of background corrected observations and model predictions paired in time and plotted

S10. Meteorological Sensitivity Analyses

Selecting representative meteorological inputs is an important aspect of dispersion modeling. It is common for both the vertical and horizontal rates of dispersion and the magnitude and direction of advection to be influenced by several of the meteorological parameters input to the model including cloud cover, wind speed, wind direction, and estimates of the planetary boundary layer height. Uncertainties in these parameters, whether they originate in the initial inputs or the extrapolation of these inputs to other locations in the modeling domain, can propagate through the model leading to errors in enhancement predictions. For example, small errors in wind direction (on the order of a few degrees) at a source location can lead to significantly different predictions of plume trajectory than those occurring in the actual wind field.

Several variables have historically been used as the basis of atmospheric stability classes, which qualitatively describe the state of atmospheric turbulence. These variables are wind speed, some measure of incoming solar radiation, and time-of-day. The incoming solar radiation and time-of-day are generally associated, except in the presence of clouds. The planetary boundary layer (PBL) depth, which typically varies reliably with atmospheric stability and acts as an additional barrier to vertical transport, can be used as a more general descriptor, though it is much more challenging to measure directly. Figure S19 shows ground-level plume predictions for a constantly emitting source in the near-field ($\leq$ 200m from the source) over the course of a day. Plume structure is observed to vary considerably based on stability class, mixing height, and wind speed.

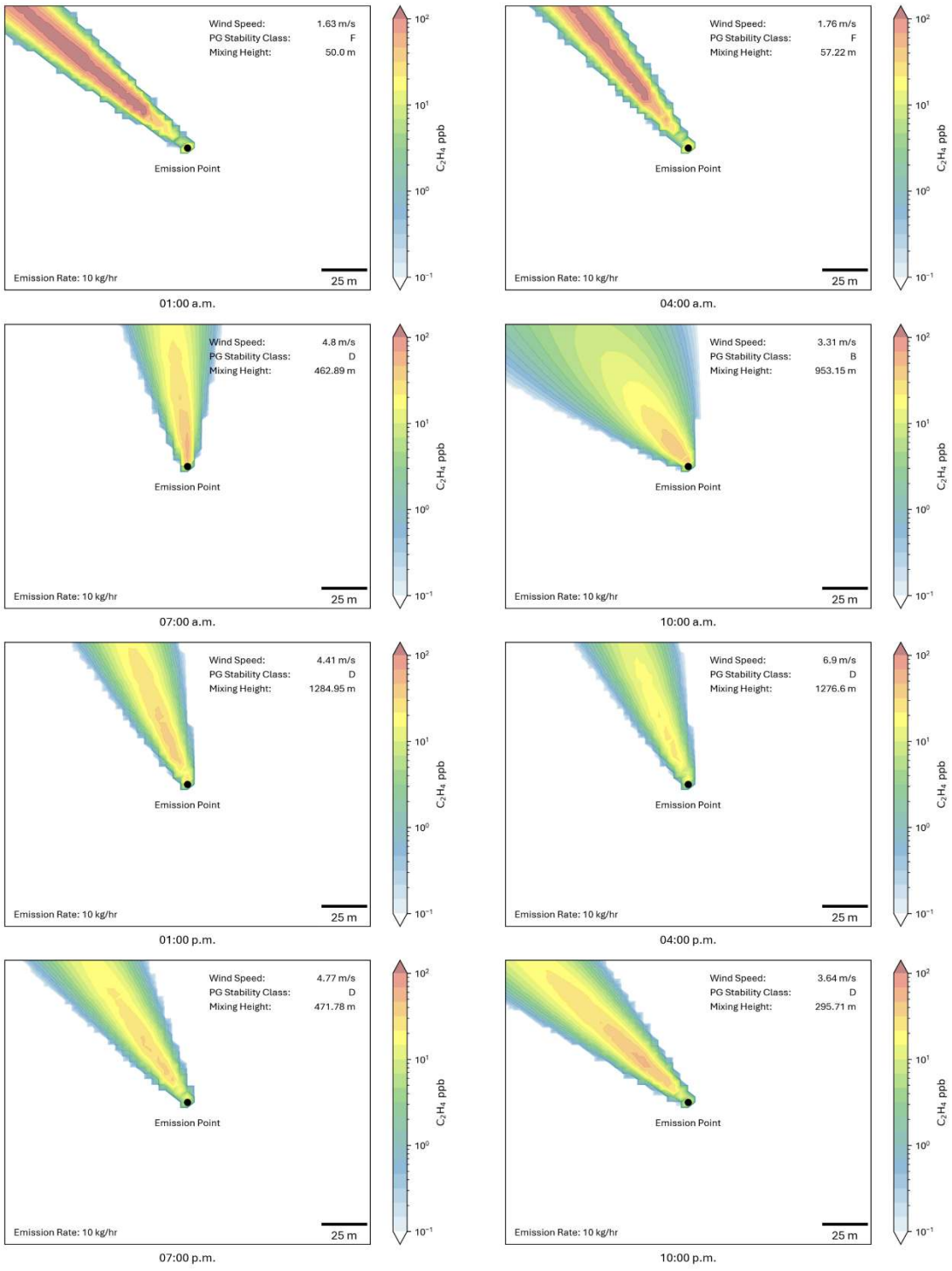
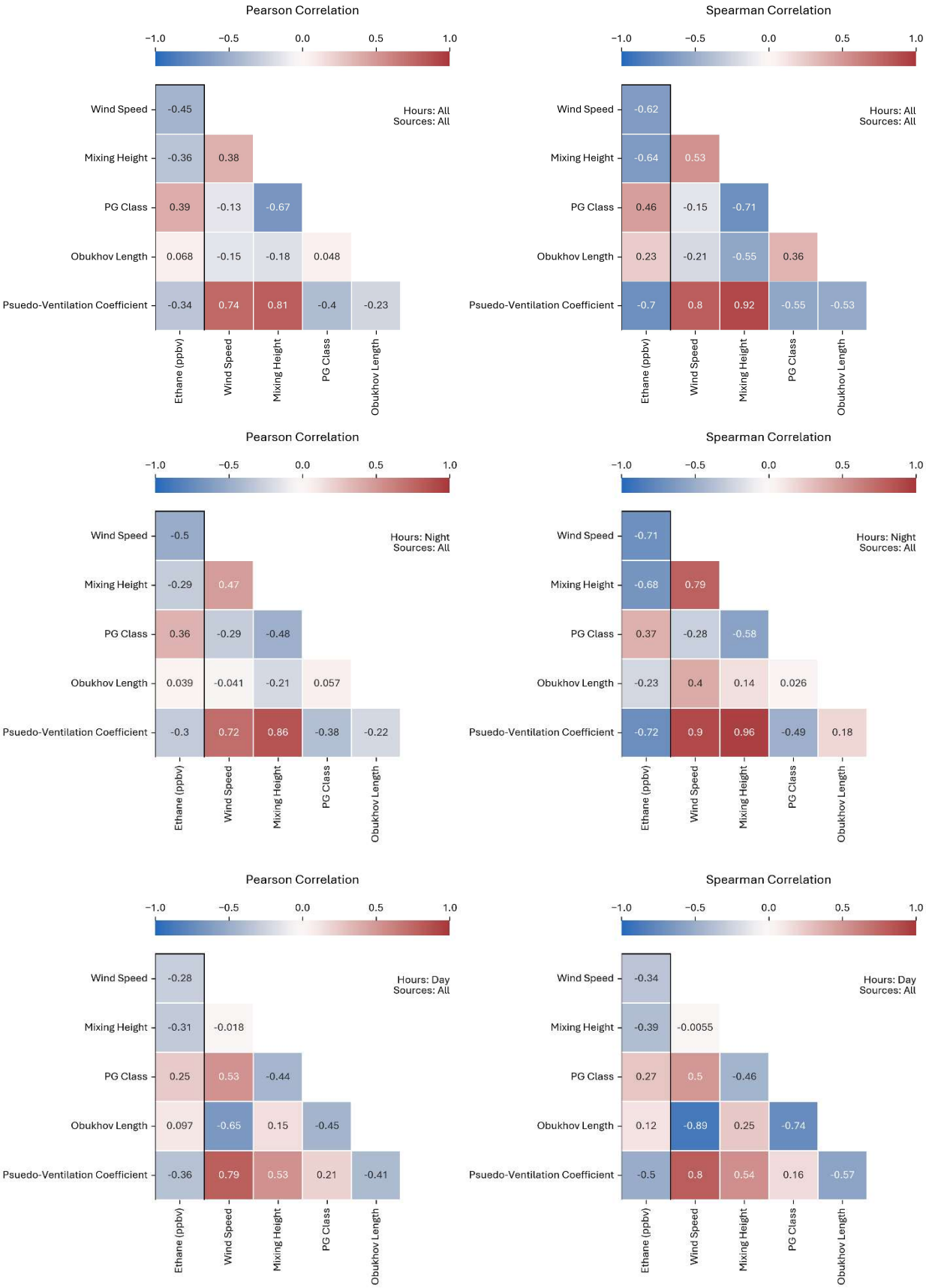


Figure S19: CALPUFF predictions of plume structure for a constantly emitting source over the course of a single day.

To evaluate the sensitivity of ambient alkane concentration predictions from CALPUFF to meteorological inputs, a two-step analysis framework was applied to the simulations. First, predicted ambient ethane concentrations from CALPUFF from every hour of the simulation were regressed onto predictions of several meteorological parameters predicted by CALMET over the measurement site. These regressions were prepared as one column of a correlation matrix (Figure S14) to evaluate associations between meteorological variables and predicted enhancements. Parameters for evaluation were surface-level wind speed, mixing height, PG stability class (1-6 corresponding to A-F), Obhukov length, and a "pseudo-ventilation coefficient" which represents the mixing height multiplied by the surface-level wind speed at a given hour. Ventilation coefficients are traditionally calculated by multiplying PBL depth with the vertically averaged wind speed in the PBL at that time. The pseudo-ventilation coefficient is calculated the same way, but using the surface-level wind speed instead.

In almost all cases, predicted ethane concentrations are only weakly correlated (Pearson's r) with any of the meteorological parameters. Rank ordered correlations (Spearman's ρ) are much stronger, indicating many of the associations between concentration and individual meteorological parameters may be monotonic but not necessarily linear (Figure S14). This is especially true in the case of the pseudo-ventilation coefficient, which tends to be more weakly correlated with predicted ethane concentrations than mixing height or wind speed individually but becomes more highly correlated than in the rank-ordered analyses. Variations in the association between variables are also apparent between day and night and for sources in the near and far field.


Pearson Correlation
−1.0 −0.5 0.0 0.5 1.0
Hours: All
Sources: All
Wind Speed -0.45
Mixing Height -0.36 0.38
PG Class 0.39 -0.13 -0.67
Obukhov Length 0.068 -0.15 -0.18 0.048
Psuedo-Ventilation Coefficient -0.34 0.74 0.81 -0.4 -0.23
Ethane (ppbv)
Wind Speed
Mixing Height
PG Class
Obukhov Length
Spearman Correlation
−1.0 −0.5 0.0 0.5 1.0
Hours: All
Sources: All
Wind Speed -0.62
Mixing Height -0.64 0.53
PG Class 0.46 -0.15 -0.71
Obukhov Length 0.23 -0.21 -0.55 0.36
Psuedo-Ventilation Coefficient -0.7 0.8 0.92 -0.55 -0.53
Ethane (ppbv)
Wind Speed
Mixing Height
PG Class
Obukhov Length
Pearson Correlation
−1.0 −0.5 0.0 0.5 1.0
Hours: Night
Sources: All
Wind Speed -0.5
Mixing Height -0.29 0.47
PG Class 0.36 -0.29 -0.48
Obukhov Length 0.039 -0.041 -0.21 0.057
Psuedo-Ventilation Coefficient -0.3 0.72 0.86 -0.38 -0.22
Ethane (ppbv)
Wind Speed
Mixing Height
PG Class
Obukhov Length
Spearman Correlation
−1.0 −0.5 0.0 0.5 1.0
Hours: Night
Sources: All
Wind Speed -0.71
Mixing Height -0.68 0.79
PG Class 0.37 -0.28 -0.58
Obukhov Length -0.23 0.4 0.14 0.026
Psuedo-Ventilation Coefficient -0.72 0.9 0.96 -0.49 0.18
Ethane (ppbv)
Wind Speed
Mixing Height
PG Class
Obukhov Length
Pearson Correlation
−1.0 −0.5 0.0 0.5 1.0
Hours: Day
Sources: All
Wind Speed -0.28
Mixing Height -0.31 -0.018
PG Class 0.25 0.53 -0.44
Obukhov Length 0.097 -0.65 0.15 -0.45
Psuedo-Ventilation Coefficient -0.36 0.79 0.53 0.21 -0.41
Ethane (ppbv)
Wind Speed
Mixing Height
PG Class
Obukhov Length
Spearman Correlation
−1.0 −0.5 0.0 0.5 1.0
Hours: Day
Sources: All
Wind Speed -0.34
Mixing Height -0.39 -0.0055
PG Class 0.27 0.5 -0.46
Obukhov Length 0.12 -0.89 0.25 -0.74
Psuedo-Ventilation Coefficient -0.5 0.8 0.54 0.16 -0.57
Ethane (ppbv)
Wind Speed
Mixing Height
PG Class
Obukhov Length

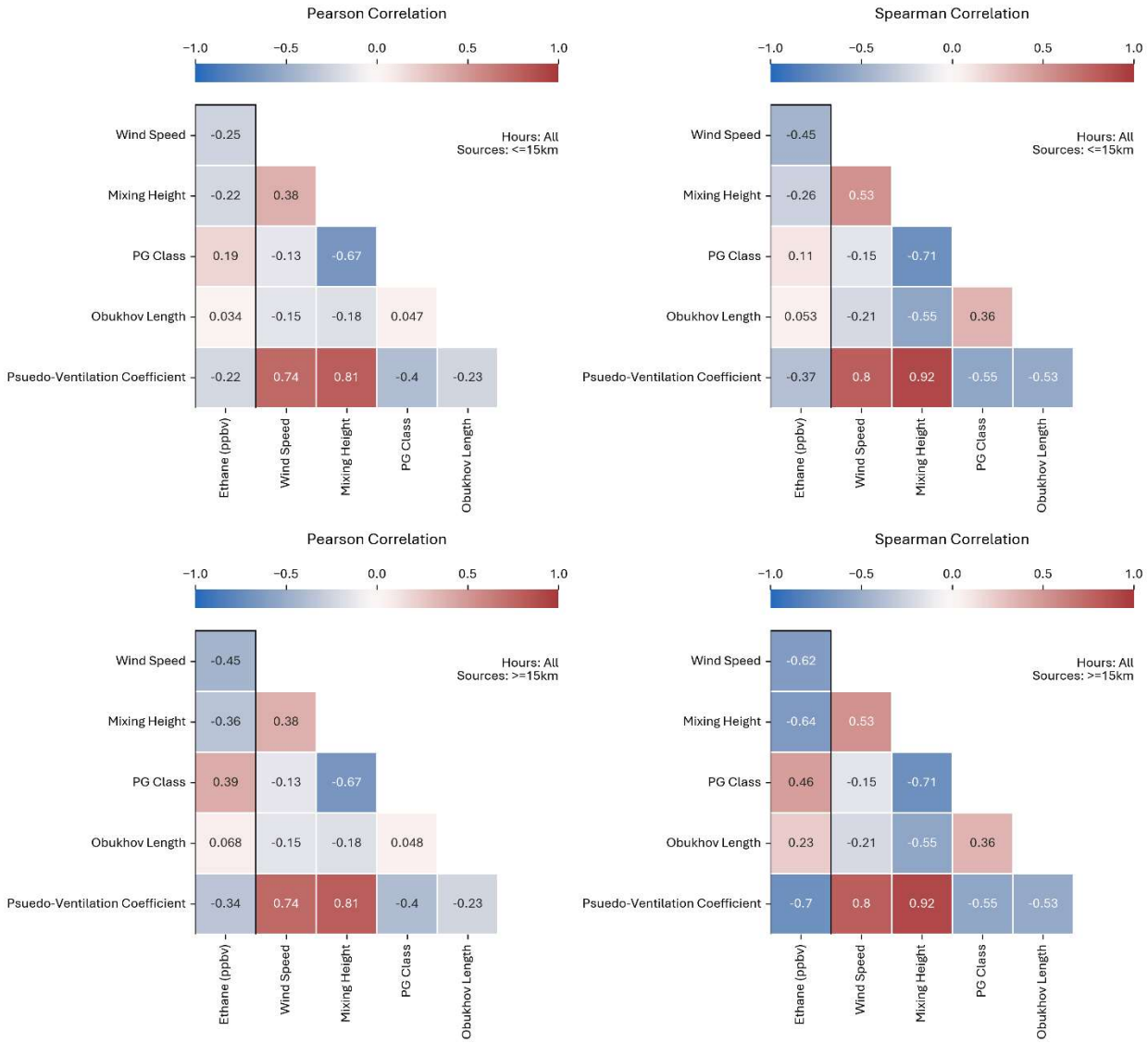


Figure S20: Correlation matrices for predicted ethane concentrations and predictions of meteorological parameters from CALMET.

In the second part of the analysis, simulations were run with modified meteorological inputs. A base-case was generated by using all the same sources from the original modeling, but now with a constant emission rate of 1 kg/hr to remove the influence of emission rates. This serves as a reference case for comparison. The first sensitivity simulation was prepared by increasing the wind speed at every surface meteorological station by 10% at every timestamp in the modeling period. Wind speed was chosen because it is directly input to the model (rather than estimated by the model like Obhukov length and PG class), and because it displays a strong association with observed concentration based on the results shown in Figure S20. The second sensitivity simulation was prepared by shifting the wind direction at every surface meteorological station by +10 degrees. Wind directions from standard ISD measurements are reported to the nearest 10 degrees (e.g., a wind direction measurement of 174° is reported as 170°) so measurements can

be uncertain at least to this magnitude. Figure S21 shows reasonable agreement between the distribution of predicted enhancements between the base-case and sensitivity simulations for both daytime and nighttime enhancements, but the temporal alignment is more variable (Figure S22). In the wind-speed adjusted case, daytime enhancements are much less consistent with the base-case when the base-case enhancements are low. This may be explained by the fact that concentrations are typically lowest when wind speeds are highest, so a 10% increase in these values is a much larger absolute change than for slower wind-speeds. Enhancement predictions are much more sensitive to changes in the wind direction in these comparisons (Figure S22). Notably, the highest nighttime enhancements are relatively insensitive to changes in the wind direction, suggesting that meteorological conditions (e.g., low wind speed, low PBL depth) may have more impact on the observed enhancements than the specific source profile when conditions are favorable for producing high ambient concentrations. The daytime enhancements show a weaker version of this trend indicating that the specific source-profile contributing to enhancements at a given time may be more impactful than the specific meteorological conditions during the daytime.

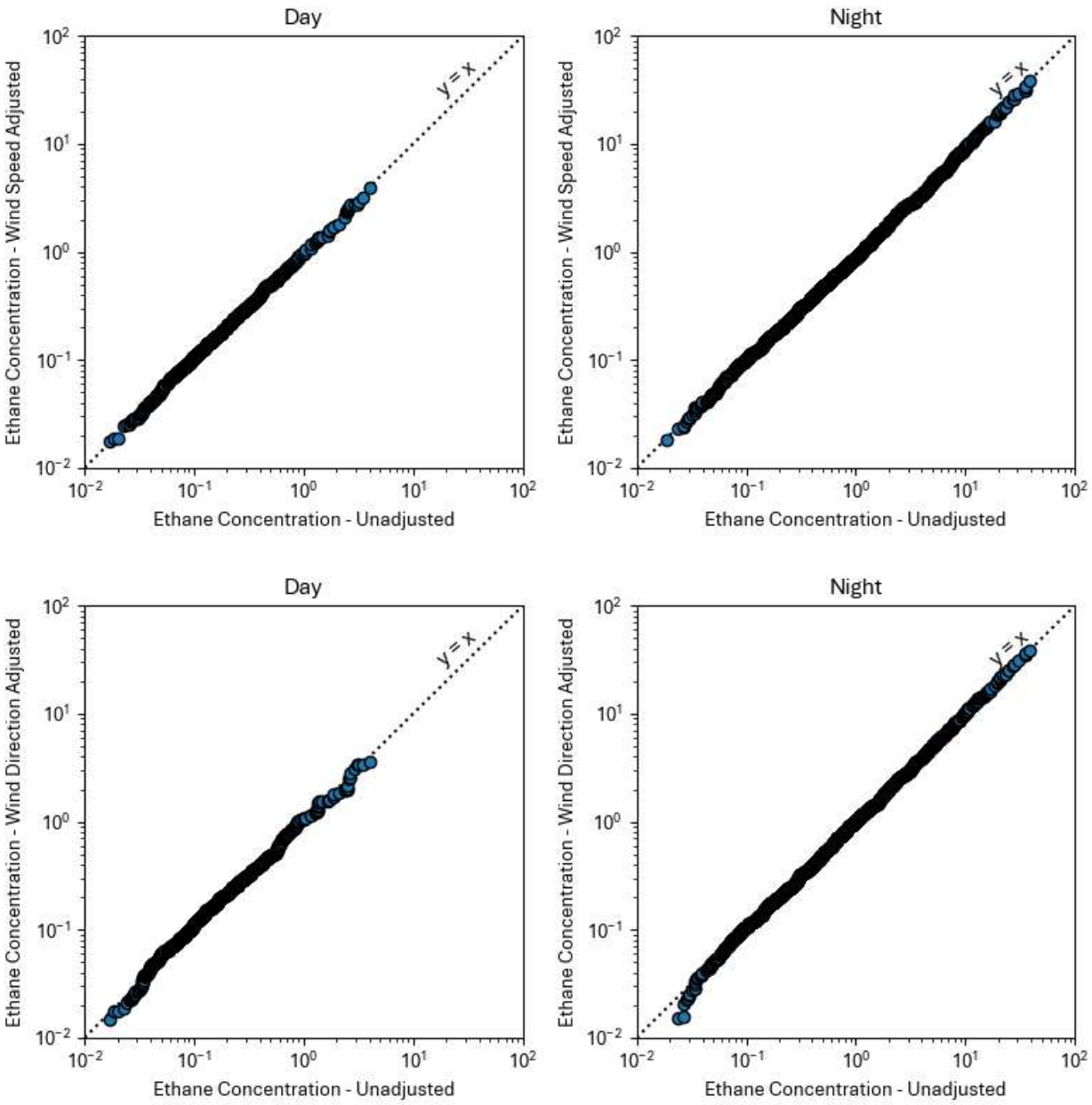


Figure S21: Quantile-quantile plot for wind-speed and wind-direction adjusted simulations compared to the base-case CALPUFF runs

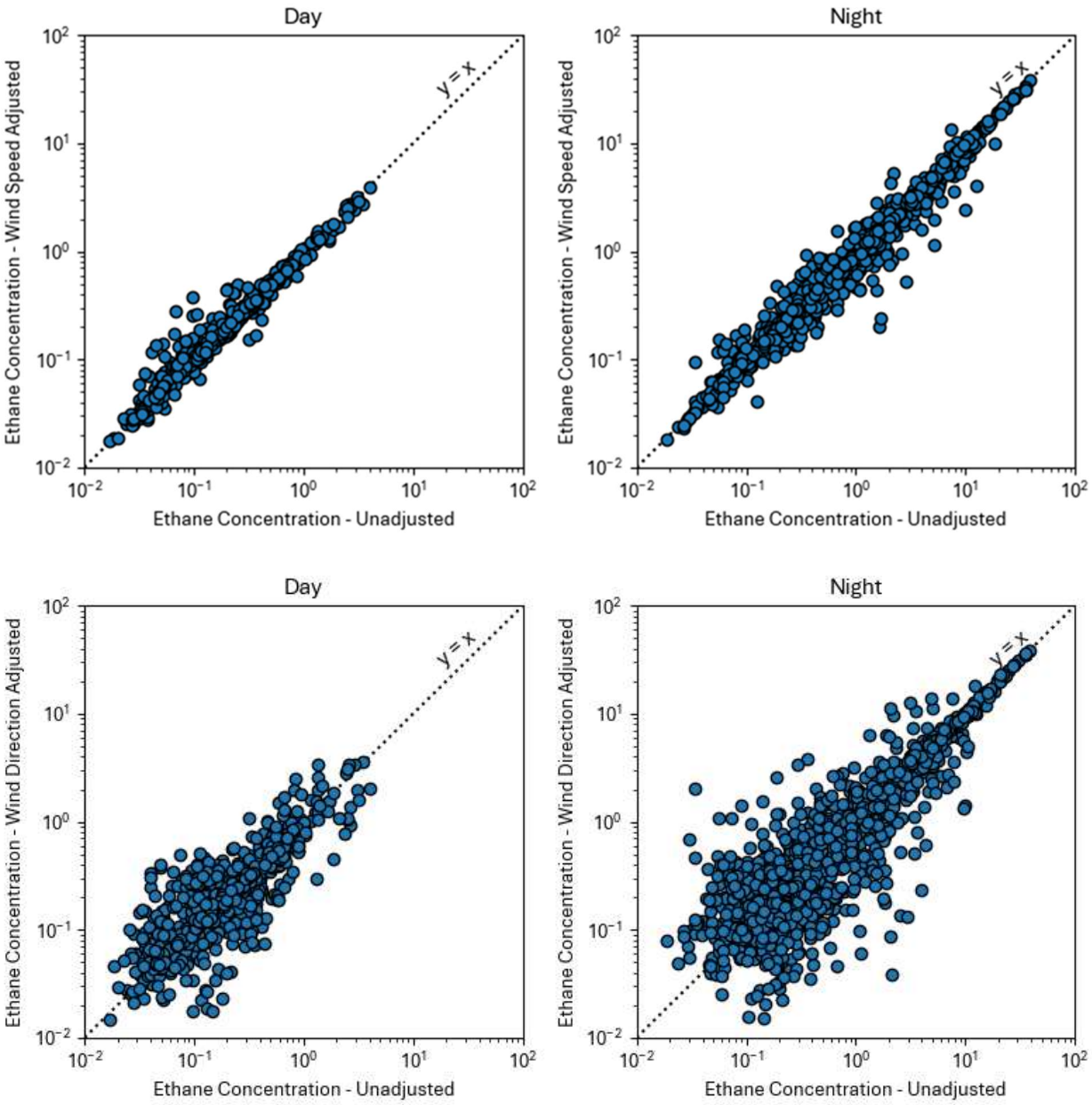


Figure S22: Parity plot for wind-speed and wind-direction adjusted simulations compared to the base-case CALPUFF runs

S11 Back-trajectory analysis

Observed enhancements of ambient ethane concentrations that were not predicted by the model were assumed to be produced by non-routine emission events. To identify potential sources of these unmatched enhancements, a three-step inverse modeling procedure was developed. In this procedure, measurements of several additional hydrocarbons are evaluated at the time of the enhancement to constrain potential source candidates to specific equipment or processes with unique chemical fingerprints. For example, when compared to upstream emissions from tank flashing or other non-combustion processes, flaring emissions will typically have higher ratios of combustion products (e.g. benzene:ethane ratios are greater) while midstream emissions will typically have much lower ratios of heavier alkanes (e.g., propane:ethane will be lower). Back trajectory analyses are then used to further constrain the geographic location of probable source candidates. Finally, an inverse modeling approach is applied that estimates, for each source individually, the lowest emission rate necessary to match the observed enhancement. A demonstration of this technique is provided for the unexplained episode (Figure 7a in the main text)

The chemical fingerprinting analysis used benzene and propane as combustion and non-combustion tracers respectively. Chemical profiles were evaluated for the entire duration of an episode and for the 6 hours immediately preceding and following the episode (Figure S23). The chemical profile within the episode was also compared to historic profiles from the same time of day during the 3-month modeling period (Figure S24) to rule out a shifting background. In both cases, the quantile rank of the propane:ethane and benezne:ethane ratios at the time of the maximum enhancement relative to all other times was evaluated (e.g. if the maximum concentration in an enhancement occurred at 11 a.m., the chemical profile of that enhancement was compared to the chemical profile of all other measurements at 11 a.m. over the modeling period). If a chemical profile of the maximum enhancement was low in propane (e.g., low propane:ethane) compared to historic measurements and the times preceding and following the episode, it was assumed to come from a midstream source. If a chemical profile was enriched in benzene, it was assumed to be from a flare or other combustion source. If no distinctive chemical traits to link the source to midstream or combustion operations could be identified (e.g., high propane:ethane ratios), the source was assumed to be from upstream processes (e.g., tank flashing, unloadings, etc.). Figures S23 and S24 show this approach implemented for the unexplained episode (Figure 7a in main text). Both propane:ethane and benzene:ethane ratios are elevated at the time of the maximum enhancement relative to the hours immediately around the episode. The benzene:ethane ratios are relatively close to the historic median for that time-of-day indicating the source is likely from the upstream sector, and the propane:ethane ratios are high indicating an upstream source enriched in propane relative to other upstream sources, such as tank emissions.

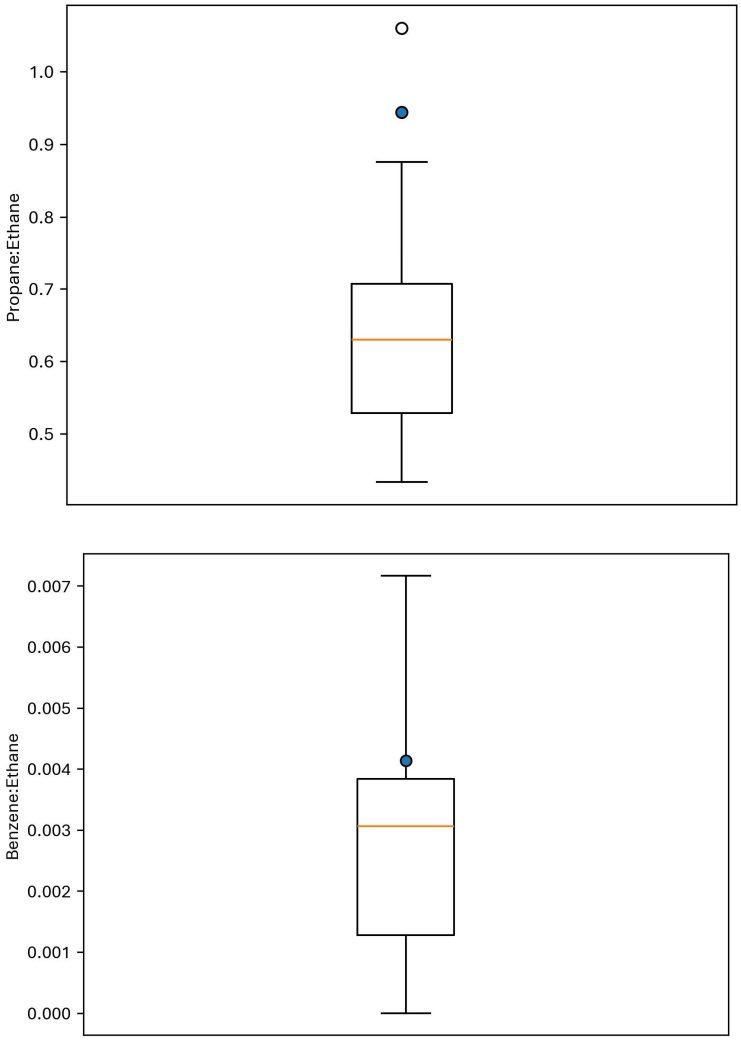


Figure S23: Chemical profiles of the maximum concentration in an unmatched observed enhancement compared to all measurements within a window beginning 6 hours before the start of the episode and ending 6 hours after the episode. Boxplots show the distribution of trace gas ratios for all measurements in the window and the blue dot shows the profile of the maximum enhancement.

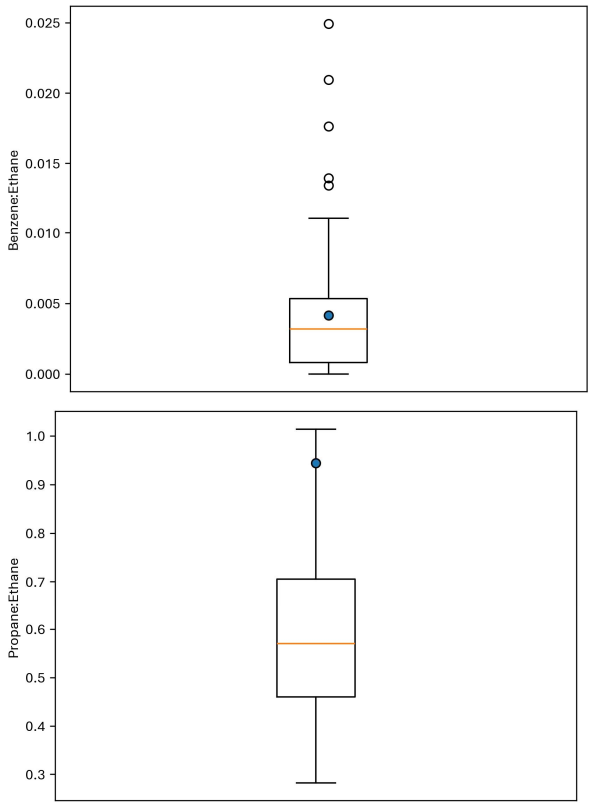


Figure S24: Chemical profiles of the maximum concentration in an unmatched observed enhancement compared to all measurements in the 3-month modeling period that occur during the same hour of the day. Boxplots show the distribution of trace gas ratios for all measurements and the blue dot shows the profile of the maximum enhancement.

In the second step of the analysis, source candidates are identified for an unexplained episode by evaluating parcel back trajectories initialized from the time of the maximum concentration in the observed episode. Back trajectories were simulated using the NOAA Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model through the Real-time Environmental Applications and Display System (READY) interface.[32,33] Data from the High-Resolution Rapid Refresh (HRRR) prognostic weather model were used as the meteorological inputs. HYSPLIT is a hybrid Lagrangian-Eularian transport model that supports both dispersion and trajectory calculations. In the present analysis, only trajectory calculations were used for simplicity.

In the present analysis, particle back trajectories were simulated to generate a candidate source "region" (Figure S25) which defines the geographic boundaries for considering candidate source locations. These regions are the product of several back trajectories run as an ensemble to evaluate uncertainty in the meteorological fields. At the finest scale, the gridded wind fields used in particle back trajectory analyses may have a spatial resolution of a few kilometers and a 1-hour temporal resolution. However, features of the true wind field are continuous in time and space and can range in scale from a few centimeters to a few kilometers. This mismatch in the resolving power of the gridded wind fields can lead to significant variations between the true and predicted particle trajectories when using deterministic approaches (e.g., a single simulated

trajectory). Ensemble methods attempt to characterize the degree of uncertainty associated with model predictions, typically by introducing slight perturbations to the model between runs. A common approach when using gridded meteorological inputs is to shift the meteorological grid by one cell (in either the vertical or horizontal direction or both) around the initialization point (e.g., the particle release point or the or the starting point for the back trajectory).[34] Perturbations around the initialization point are emphasized because small uncertainties in the trajectory very early in the simulation can propagate uncertainty most substantially over time and space. The present work follows the methods outlined by Draxler, et al.[34], which have been implemented as a feature of the HYSPLIT modeling system. Briefly, 27 individual simulations are initialized, with one trajectory being simulated using the meteorological inputs as-is. For each successive simulation, the entire meteorological grid is shifted so that one of the 26 meteorological grid cells surrounding the initialization cell overlaps the initialization point (Figure S26). The trajectory is then calculated using this offset grid. The initialization height is typically selected to be above the first layer of the meteorological grid (typically around 250m AGL) to avoid non-physical behavior like attempting to sample grid cells below ground level. Using this approach, only the bottom two layers are considered representative of the potential trajectory and the trajectories initialized in one of the 9 uppermost cells are discarded. In these simulations, the initialization height was set to 300m AGL to be conservative.

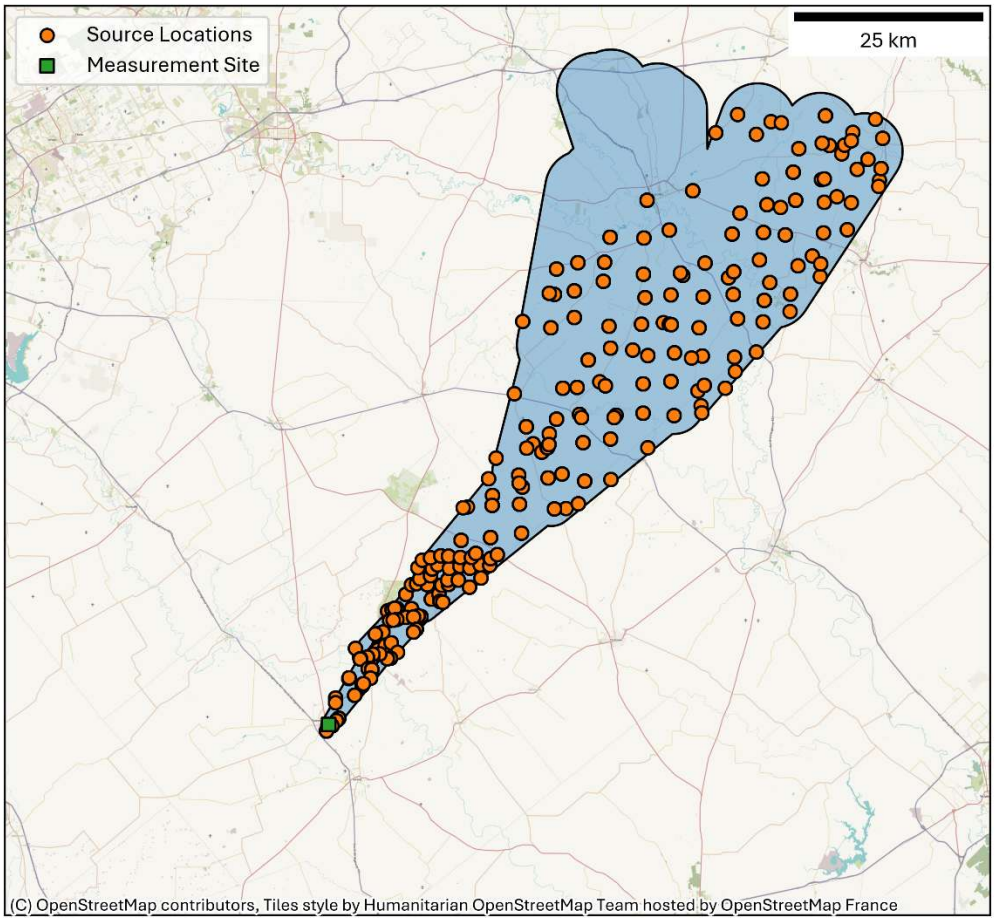


Figure S25: Candidate source region generated using from overlapping ensemble back trajectories. Basemap data provided by OpenStreetMap.[35]

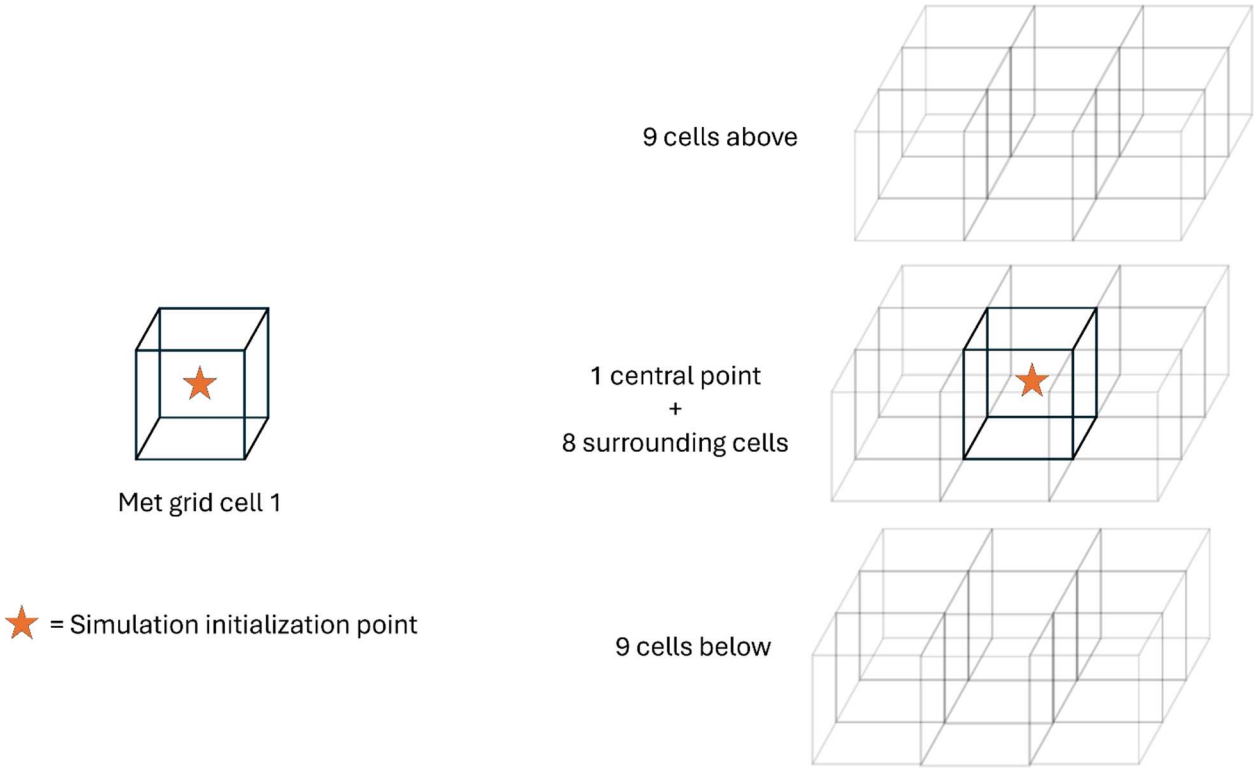


Figure S26: Illustration of the meteorological grid cells sampled during the ensemble simulations**.**

The 18 ensemble members retained for analysis provide particle trajectories with hourly resolution (e.g., the particle location at each hour is simulated, with a line connecting the two points). To account for dispersion, which the trajectory model does not simulate, a buffer of variable width was applied to the straight-line path segments from each trajectory based on the temporal displacement between the segment and the trajectory initialization time. The buffer for each segment was set to 1 km (on either side of the segment) multiplied by the hours since initialization (Figure S27). After the buffers have been created, they are dissolved into a single polygon (i.e., the source candidate region). Sources within this polygon are then retained for further analysis. These sources can be characterized by their distance from the receptor and their equipment type to constrain the candidate search to a specific subset of sources (Figure S28).

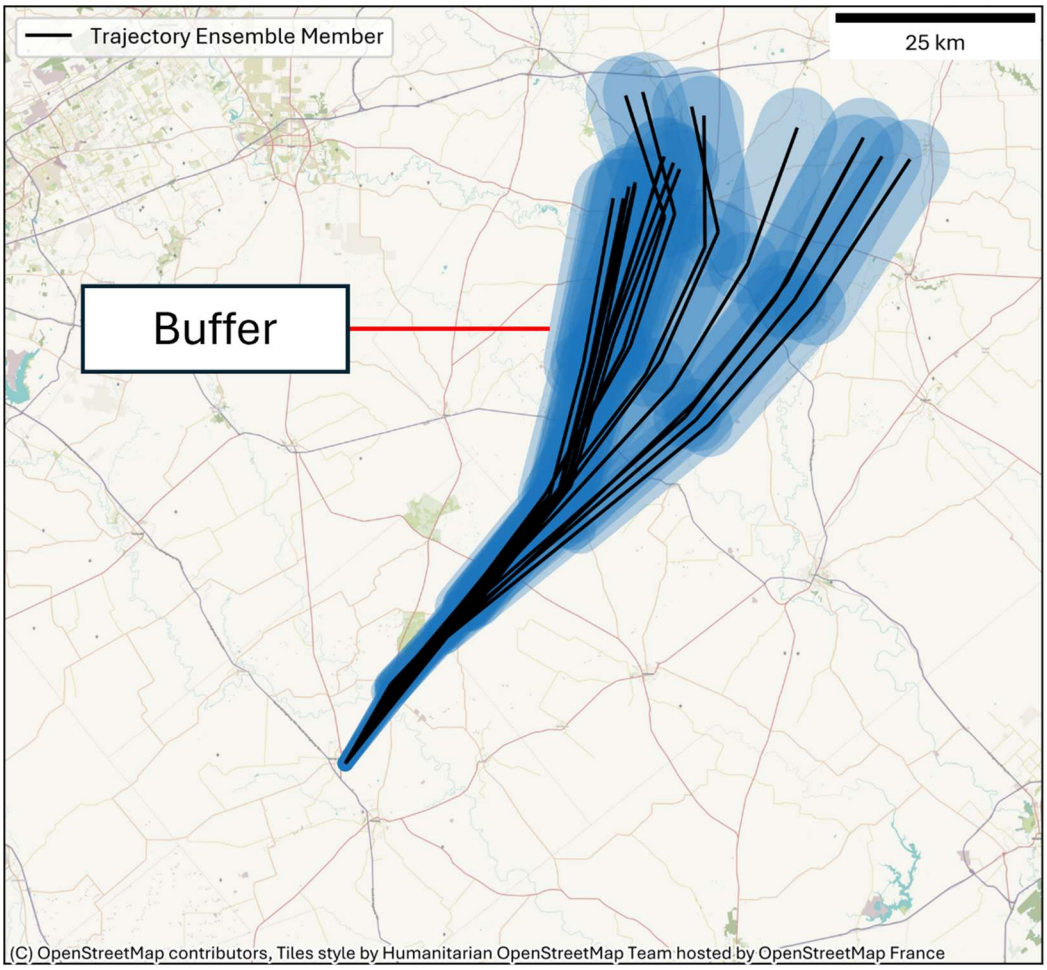


Figure S27: Variable-width buffers for ensemble trajectory members used to establish a candidate source region. Basemap data provided by OpenStreetMap.[35]

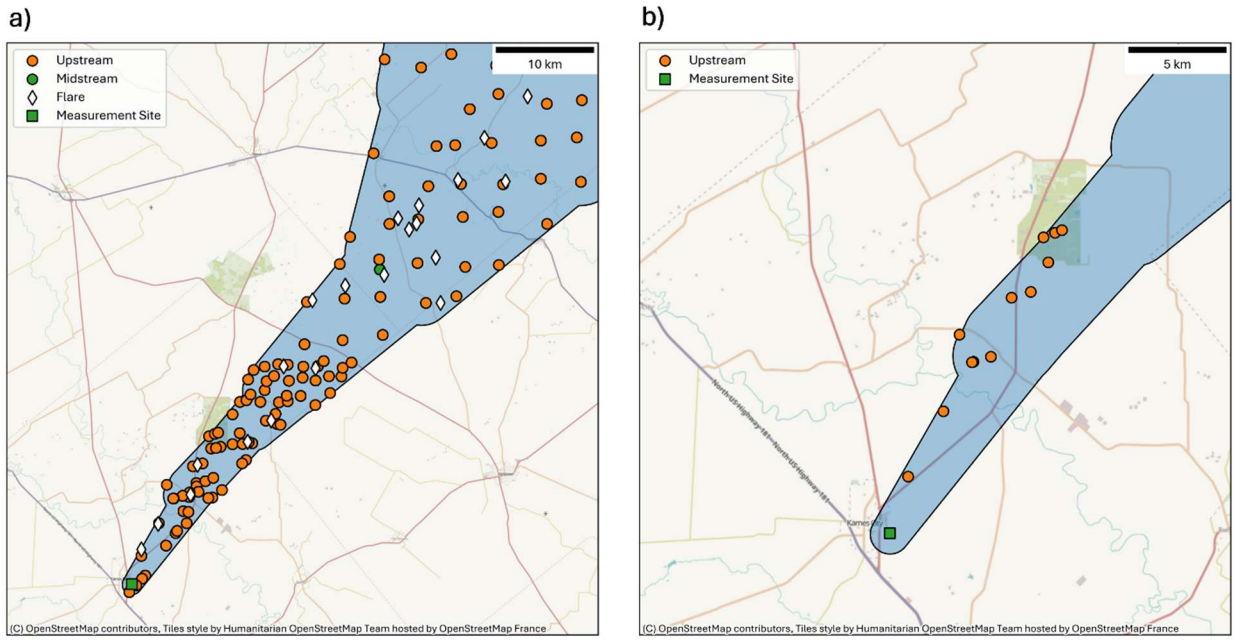


Figure S28: Constraining the distance and source type based on the results of the chemical fingerprinting and hypothetical source approach (a) shows sources along the back trajectory by source type and (b) shows candidate sources whose minimum emission rate is $\leq$ 1 ton/hr and whose chemical fingerprints are consistent with the observed concentration ratios. Flare locations are identified using the Visible Infrared Imaging Radiometer Suite (VIIRS) Nightfire product.[36] Basemap data for both panels provided by OpenStreetMap.[35]

Candidate sources can then be screened using the approach described in the main text. Candidate sources identified using this approach for the unexplained episode in Figure 7a of the main text are shown in Figure S28. The predicted emission rates at these sources that would be necessary to reconcile the observed unexplained enhancement with the model predictions are provided in Table S26.

Source Name	Minimum Emission Rate (kg/hr)
TB0367_LEAK	24.
TB0367_TANK	24.
TB0316_TANK	84.
TB0316_LEAK	87.
TB0316_LEAK	87.
TB0261_TANK	155.
TB0261_LEAK	155.
TB0261_LEAK	155.
TB0260_TANK	159.
TB0260_LEAK	161.
TB0260_LEAK	161.
TB0196_TANK	434.
TB0196_LEAK	437.
TB0196_LEAK	439.
TB0259_TANK	519.
TB0259_LEAK	534.
TB0124_LEAK	539.
TB0259_LEAK	541.
TB0124_LEAK	542.
TB0124_TANK	548.
TB0125_LEAK	707.
TB0125_TANK	709.
TB0125_LEAK	710.
TB0219_LEAK	764.
TB0219_LEAK	774.
TB0219_TANK	795.
TB0179_LEAK	842.
TB0179_TANK	846.
TB0179_LEAK	847.
TB0126_TANK	856.
TB0126_LEAK	868.
TB0126_LEAK	872.
TB0140_LEAK	949.
TB0140_TANK	956.

Table S26: Candidate sources and predicted emission rates necessary to reconcile the unexplained episode in Figure 7a of the main text (TBXXXX refers to the site number assigned to a specific tank battery)

S12 Time Series for Largest Ethane Enhancements

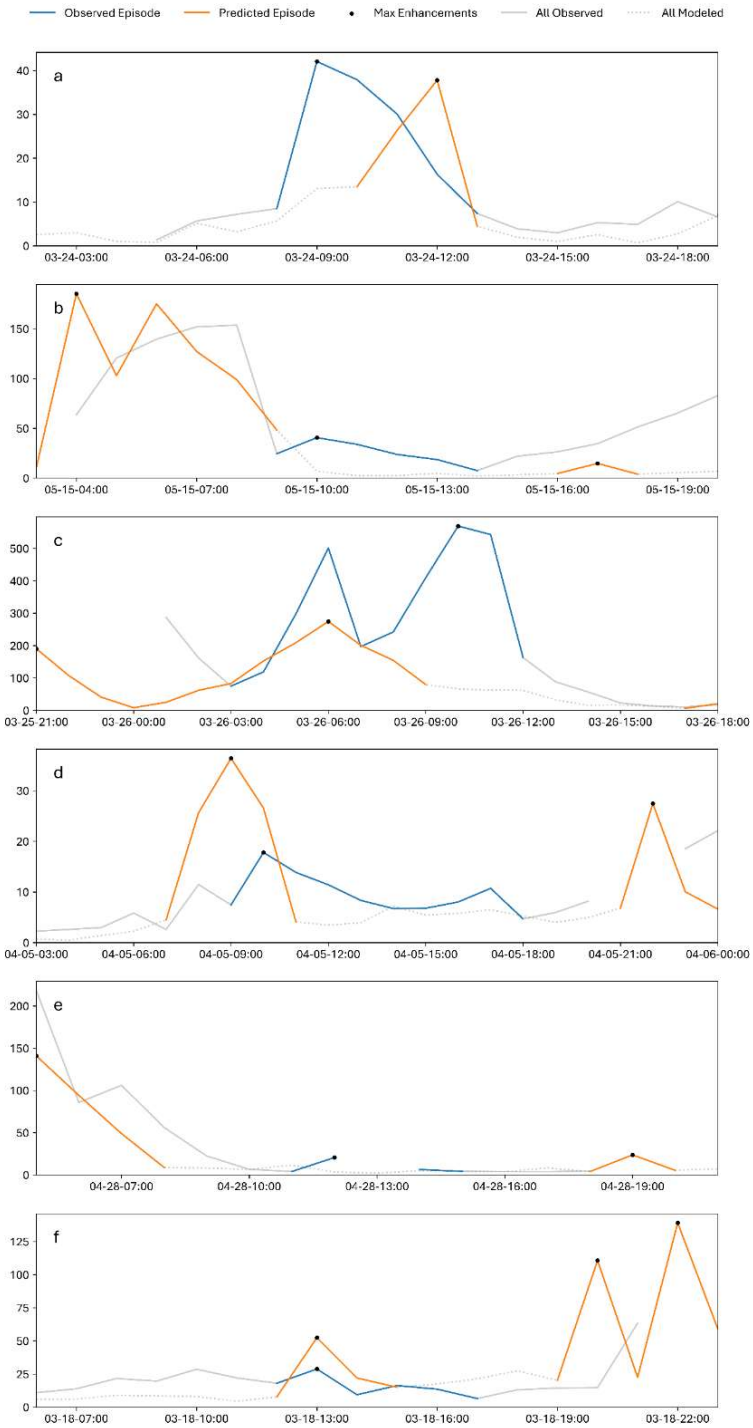


Figure S29 The 6 observed daytime episodes identified by the enhancement identification algorithm with predicted episodes overlaid. The maximum enhancement for each modeled and observed episode is indicated with a black point. Only one observed episode is included in each subfigure, but multiple predicted enhancements may be included. Predicted enhancements are not screened by time-of-day (e.g., predicted episodes that would be categorized as nighttime episodes are included).

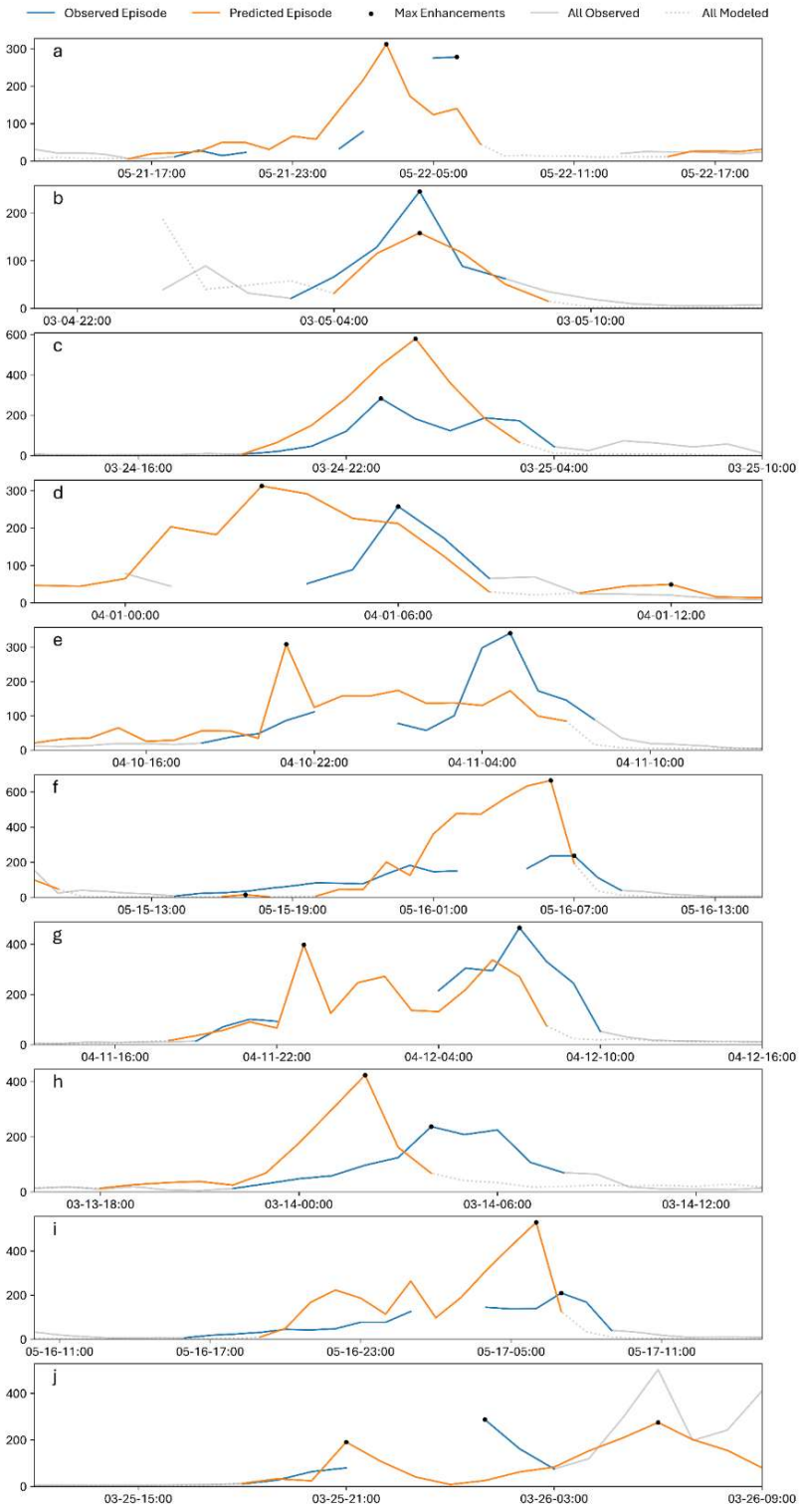


Figure S30 The top 10 observed nighttime episodes identified by the enhancement identification algorithm with predicted episodes overlaid. The maximum enhancement for each modeled and observed episode is indicated with a black point. Only one observed episode is included in each subfigure, but multiple predicted enhancements may be included. Predicted enhancements are not screened by time-of-day (e.g., predicted episodes that would be categorized as daytime episodes are included).

Daytime Enhancement	Matched?	Explanation
A	Y	A majority of the observed episode overlaps with a predicted episode and both episodes primarily occur during the daytime hours before falling off around noon. The maximum concentration in the predicted episode is offset by 3 hours, but the magnitude of the enhancements are similar.
B	N	The observed episode does not overlap with a predicted episode, or any period of enhanced concentrations predicted by the model, but the observed enhancements are below 50 ppb
C	N	The nighttime peak at the beginning of this episode has overlap with a predicted enhancement, but the daytime portion of the peak has no overlap with a predicted episode.
D	N	The first portion of the observed enhancement has overlap with a predicted enhancement, but the behavior of the rest of the episode is not well matched.
E	N	The observed episode does not overlap at all with a predicted episode, or any period of enhanced concentrations predicted by the model, but the observed enhancements are below 50 ppb
F	Y	Almost the entire observed episode overlaps with a predicted episode. The peaks align in time and the broader temporal behavior of the episode.

Table S27 Results of the quantitative matching analysis for the top daytime episodes (6 total)

Nighttime Enhancement	Match	Explanation
A	Y	A majority of the observed episode overlaps with a predicted episode and both episodes occur predominantly at night. The maximum concentration in the observed episode is offset by approximately 4 hours, but the broader temporal behavior of the episodes are consistent. The general magnitude of the enhancements are also relatively consistent. Missing observational data due to automated calibrations complicates this assessment.
B	Y	A majority of the observed episode overlaps with a predicted episode and both episodes occur predominantly at night. The maximum concentrations are well aligned in time and magnitude.
C	Y	A majority of the observed episode overlaps with a predicted episode and both episodes occur predominantly at night. The maximum concentrations are well aligned in time, and relatively well in magnitude (within a factor of 2).
	Y	A majority of the observed episode overlaps with a predicted episode and both episodes occur predominantly at night. The maximum concentrations are offset by a few hours, but the missing data in the observations makes assessing the hours leading up to the observed episode more challenging. Predictions are high for the first half of the episode, but the in latter half predictions are consistent in shape and form, indicating the conditions are appropriate for producing an elevated enhancement.
E	Y	A majority of the observed episode overlaps with a predicted episode and both episodes occur predominantly at night. The maximum concentrations are not well-aligned in time, but the predicted episode has an unusual feature at the beginning of the night that is contributing to the mismatch. Missing data makes it challenging to determine if this peak would match otherwise. At the time of the observed peak, an uptick in enhancement is predicted.
F	N	A majority of the observed episode overlaps with a predicted episode and both episodes occur predominantly at night. The maximum concentrations are well aligned in time, but the average magnitude of the enhancements and the peak enhancements are largely inconsistent. The model overpredicts during this period, indicating that the discrepancy may not be the result of a large emission event.
G	Y	A majority of the observed episode overlaps with a predicted episode and both episodes occur predominantly at night. The maximum concentrations are well aligned in time and magnitude.
H	N	A majority of the observed episode overlaps with a predicted episode but the predicted episode begins and ends much earlier than the observed episode. A majority of the integrated concentration in the predicted episode is mostly confined to the hours around 11 p.m. - 4 a.m. In the observed episode, most of the integrated concentration is observed between the hours of 3 a.m. - 7 a.m.
I	N	A majority of the observed episode overlaps with a predicted episode and both episodes occur predominantly at night. The maximum concentrations are well aligned in time, but not in magnitude. The model predicts much more obvious peaks. However, the model overpredicts, suggesting the source of the discrepancies may not be a large emission source.
J	N	A majority of the observed episode overlaps with a predicted episode and both episodes occur predominantly at night. However, the observed episode includes a relatively high enhancement when the model predicts a near-zero concentration. The broader temporal behavior of the enhancements cannot be commented on as a result of several missing datapoints in the observations.

Table S8: Results of the quantitative matching analysis for the top 10 nighttime episodes

References

(1) IHS Markit. U.S. Well and Production Data [Dataset]. https://ihsmarkit.com/index.html (accessed 2023-10-09).
(2) Allen, D. T.; Cardoso-Saldaña, F. J.; Kimura, Y.; Chen, Q.; Xiang, Z.; Zimmerle, D.; Bell, C.; Lute, C.; Duggan, J.; Harrison, M. A Methane Emission Estimation Tool (MEET) for Predictions of Emissions from Upstream Oil and Gas Well Sites with Fine Scale Temporal and Spatial Resolution: Model Structure and Applications. *Sci. Total Environ.* **2022**, *829*, 154277. https://doi.org/10.1016/j.scitotenv.2022.154277.
(3) U.S. Enivronmental Protection Agency (EPA). Inventory of U.S. Greenhouse Gas Emissions and Sinks, Electronic Tables, Annex 3.6, Table 3.6-7: Activity Data for Natural Gas Systems Sources, for All Years [Dataset]. https://www.epa.gov/sites/default/files/2020-02/2019_ghgi_natural_gas_systems_annex36_tables.xlsx.
(4) U.S. Environmental Protection Agency (EPA). Inventory of U.S. Greenhouse Gas Emissions and Sinks, Electronic Tables, Annex 3.5, Table 3.5-5: Activity Data for Petroleum Systems Sources, for All Years [Dataset]. https://www.epa.gov/sites/default/files/2020-02/2019_ghgi_petroleum_systems_annex35_tables.xlsx.
(5) Code of Federal Regulations. *Title 40, Chapter I, Subchapter C, Part 98, § 98.6 Definitions, Low-Bleed Pneumatic Devices*. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-98 (accessed 2024-12-01).
(6) Allen, D. T.; Pacsi, A. P.; Sullivan, D. W.; Zavala-Araiza, D.; Harrison, M.; Keen, K.; Fraser, M. P.; Daniel Hill, A.; Sawyer, R. F.; Seinfeld, J. H. Methane Emissions from Process Equipment at Natural Gas Production Sites in the United States: Pneumatic Controllers. *Environ. Sci. Technol.* **2015**, *49* (1), 633–640. https://doi.org/10.1021/es5040156.
(7) Allen, D. T.; Torres, V. M.; Thomas, J.; Sullivan, D. W.; Harrison, M.; Hendler, A.; Herndon, S. C.; Kolb, C. E.; Fraser, M. P.; Hill, A. D.; Lamb, B. K.; Miskimins, J.; Sawyer, R. F.; Seinfeld, J. H. Measurements of Methane Emissions at Natural Gas Production Sites in the United States. *Proc. Natl. Acad. Sci.* **2013**, *110* (44), 17768–17773. https://doi.org/10.1073/pnas.1304880110.
(8) United States Environmental Protection Agency. Code of Federal Regulations, Title 40, Chapter I, Subpart C, Part 98, Subpart W, Table W-1B, Default Average Component Counts for Major Onshore Natural Gas Production Equipment and Onshore Petroleum and Natural Gas Gathering and Boosting Equipment. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-98/subpart-W/appendix-Table%20W-1B%20to%20Subpart%20W%20of%20Part%2098 (accessed 2024-11-20).
(9) U.S. Enivronmental Protection Agency (EPA). 2018 Greenhouse Gas Reporting Program (GHGRP), 2019. https://enviro.epa.gov/query-builder/ghg/SUBPART W - PETROLEUM AND NATURAL GAS SYSTEMS (RY 2015-2023)/ (accessed 2024-01-11).
(10) United States Environmental Protection Agency. Code of Federal Regulations, Title 40, Chapter I, Subpart C, Part 98, Subpart W, Table W-1E, Default Whole Gas Leaker Emission Factors for Onshore Petroleum and Natural Gas Production and Onshore Petroleum and Natural Gas Gathering and Boosting. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-98/subpart-W/appendix-Table%20W-1E%20to%20Subpart%20W%20of%20Part%2098 (accessed 2024-11-20).
(11) Allen, D. T.; Sullivan, D. W.; Zavala-Araiza, D.; Pacsi, A. P.; Harrison, M.; Keen, K.; Fraser, M. P.; Daniel Hill, A.; Lamb, B. K.; Sawyer, R. F.; Seinfeld, J. H. Methane Emissions

from Process Equipment at Natural Gas Production Sites in the United States: Liquid Unloadings. *Environ. Sci. Technol.* **2015**, *49* (1), 641–648. https://doi.org/10.1021/es504016r.
(12) Vaughn, T. L.; Bell, C. S.; Pickering, C. K.; Schwietzke, S.; Heath, G. A.; Pétron, G.; Zimmerle, D. J.; Schnell, R. C.; Nummedal, D. Temporal Variability Largely Explains Top-down/Bottom-up Difference in Methane Emission Estimates from a Natural Gas Production Region. *Proc. Natl. Acad. Sci.* **2018**, *115* (46), 11712–11717. https://doi.org/10.1073/pnas.1805687115.
(13) Vaughn, et al. BU Input Data Study Period [Dataset]. https://mountainscholar.org/items/6a216dff-33d4-4437-8acf-01a88609eea0 (accessed 2024-11-10).
(14) Brito, R. .; Pereyra, E. .; Sarica, C. . A Novel Analysis to Detect When and Where Liquid Loading Occurs in Horizontal Gas Wells - Case Studies. In *Day 2 Wed, October 26, 2016*; SPE: The Woodlands, Texas, USA, 2016; p D021S005R003. https://doi.org/10.2118/181223-MS.
(15) U.S. Environmental Protection Agency (EPA). Greenhouse Gas Reporting Program, Natural Gas Processing Plants (2022) [Dataset]. http://ghgdata.epa.gov/ghgp (accessed 2024-09-13).
(16) U.S. Energy Information Administration. Natural Gas Processing Plants (2017) [Dataset]. https://atlas.eia.gov/datasets/eia::natural-gas-processing-plants-2/explore?filters=eyJTdGF0ZSI6WyJUWCJdfQ%3D%3D&location=32.070194%2C-97.332235%2C5.75 (accessed 2024-09-11).
(17) Zimmerle, D.; Vaughn, T.; Luck, B.; Lauderdale, T.; Keen, K.; Harrison, M.; Marchese, A.; Williams, L.; Allen, D. Methane Emissions from Gathering Compressor Stations in the U.S. *Environ. Sci. Technol.* **2020**, *54* (12), 7552–7561. https://doi.org/10.1021/acs.est.0c00516.
(18) Mitchell, A. L.; Tkacik, D. S.; Roscioli, J. R.; Herndon, S. C.; Yacovitch, T. I.; Martinez, D. M.; Vaughn, T. L.; Williams, L. L.; Sullivan, M. R.; Floerchinger, C.; Omara, M.; Subramanian, R.; Zimmerle, D.; Marchese, A. J.; Robinson, A. L. Measurements of Methane Emissions from Natural Gas Gathering Facilities and Processing Plants: Measurement Results. *Environ. Sci. Technol.* **2015**, *49* (5), 3219–3227. https://doi.org/10.1021/es5052809.
(19) Mitchell, A. L.; Tkacik, D. S.; Roscioli, J. R.; Herndon, S. C.; Yacovitch, T. I.; Martinez, D. M.; Vaughn, T. L.; Williams, L.; Sullivan, M.; Floerchinger, C.; Omara, M.; Subramanian, R.; Zimmerle, D.; Marchese, A. J.; Robinson, A. L. Correction to Measurements of Methane Emissions from Natural Gas Gathering Facilities and Processing Plants: Measurement Results. *Environ. Sci. Technol.* **2015**, *49* (20), 12602–12602. https://doi.org/10.1021/acs.est.5b04018.
(20) Texas Railroad Comission. Oil and Gas Well Records. https://rrcsearch3.neubus.com/esd3-rrc/index.php?_module_=esd&_action_=keysearch&profile=17 (accessed 2017-12-12).
(21) Texas Railroad Comission. Oil and Gas Completions. https://webapps.rrc.state.tx.us/CMPL/publicSearchAction.do (accessed 2017-12-12).
(22) Cardoso-Saldaña, F. J.; Pierce, K.; Chen, Q.; Kimura, Y.; Allen, D. T. A Searchable Database for Prediction of Emission Compositions from Upstream Oil and Gas Sources. *Environ. Sci. Technol.* **2021**, *55* (5), 3210–3218. https://doi.org/10.1021/acs.est.0c05925.
(23) U.S. Geological Survey. 3D Elevation Program 1-Meter Resolution Digital Elevation Model (2019) [Dataset]. https://www.usgs.gov/the-national-map-data-delivery (accessed 2024-01-12).

(24) Dewitz, J., and U.S. Geological Survey. National Land Cover Database (NLCD) 2019 Products (Ver. 2.0, June 2021): U.S. Geological Survey Data Release [Dataset]. https://doi.org/10.5066/P9KZCM54 (accessed 2024-01-12).
(25) Texas Commission on Environmental Quality. MeteoStar LEADS® (Leading Environmental Analysis & Display System) 5 Minute Average Surface Meteorology [Dataset].
(26) National Center for Environmental Information (NCEI), National Oceanic and Atmospheric Administration. Integrated Surface Data [Dataset]. https://www.ncei.noaa.gov/data/global-hourly/archive/isd/ (accessed 2024-01-05).
(27) National Weather Service (NWS), Federal Aviation Administration (FAA), Department of Defense (DOD). Automated Surface Observing Systems Program, Page 1 [Dataset]. https://www.ncei.noaa.gov/data/automated-surface-observing-system-one-minute-pg1/ (accessed 2023-05-31).
(28) National Weather Service (NWS), Federal Aviation Administration (FAA), Department of Defense (DOD). Automated Surface Observing Systems Program, Page 2 [Dataset]. https://www.ncei.noaa.gov/data/automated-surface-observing-system-one-minute-pg2/ (accessed 2023-05-31).
(29) National Oceanic and Atmospheric Administration. High Resolution Rapid Refresh (HRRR) Model [Dataset]. https://noaa-hrrr-bdp-pds.s3.amazonaws.com/index.html (accessed 2023-06-03).
(30) *Our products | Gulf Coast Growth Ventures*. GulfCoastGV. https://www.gcgv.com/our-products/our-products (accessed 2025-06-10).
(31) Esri. "World Imagery Hybrid" [Basemap]. Scale Not Given. "Imagery Hybrid." https://www.arcgis.com/home/item.html?id=86265e5a4bbb4187a59719cf134e0018 (accessed 2024-11-25).
(32) Stein, A. F.; Draxler, R. R.; Rolph, G. D.; Stunder, B. J. B.; Cohen, M. D.; Ngan, F. NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. *Bull. Am. Meteorol. Soc.* **2015**, *96* (12), 2059–2077. https://doi.org/10.1175/BAMS-D-14-00110.1.
(33) Rolph, G.; Stein, A.; Stunder, B. Real-Time Environmental Applications and Display sYstem: READY. *Environ. Model. Softw.* **2017**, *95*, 210–228. https://doi.org/10.1016/j.envsoft.2017.06.025.
(34) Draxler, R. R. Evaluation of an Ensemble Dispersion Calculation. *J. Appl. Meteorol.* **2003**, *42* (2), 308–317. https://doi.org/10.1175/1520-0450(2003)042<0308:EOAEDC>2.0.CO;2.
(35) OpenStreetMap. Data available through the Open Database License: https://www.openstreetmap.org/copyright.
(36) Elvidge, C.; Zhizhin, M.; Hsu, F.-C.; Baugh, K. VIIRS Nightfire: Satellite Pyrometry at Night. *Remote Sens.* **2013**, *5* (9), 4423–4449. https://doi.org/10.3390/rs5094423.