Quantifying methane emissions from natural gas production in northeastern Pennsylvania

Zachary R. Barkley[1], Thomas Lauvaux[1], Kenneth J. Davis[1], Aijun Deng[1], Natasha L. Miles[1], Scott J. Richardson[1], Yanni Cao[2], Colm Sweeney[3], Anna Karion[4], MacKenzie Smith[5], Eric A. Kort[6], Stefan Schwietzke[6], Thomas Murphy[7], Guido Cervone[8], Douglas Martins[9], Joannes D. Maasakkers[10]

[1]Department of Meteorology, The Pennsylvania State University, University Park, PA 16802, United States
[2]Department of Geography, The Pennsylvania State University, University Park, PA 16802, United States
[3]NOAA/Earth Systems Research Laboratory, University of Colorado, Boulder, CO, 80305, United States
[4]National Institute of Standards and Technology, Gaithersburg, MD 20899, United States
[5]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, 48109, United States
[6]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA.
[7]Marcellus Center for Outreach and Research, The Pennsylvania State University, University Park, PA 16802, United States
[8]Department of Geography, The Pennsylvania State University, University Park, PA 16802, United States
[9]FLIR Systems, West Lafayette, IN 47906, United States
[10]School of Engineering and Applied Sciences, Harvard University, Pierce Hall, 29 Oxford Street, Cambridge, Massachusetts 02138, United States

Correspondence to: Zachary R. Barkley (zrb5027@psu.edu)

Abstract. Natural gas infrastructure releases methane (CH_4), a potent greenhouse gas, into the atmosphere. The estimated emission rate associated with the production and transportation of natural gas is uncertain, hindering our understanding of its greenhouse footprint. This study presents a new application of inverse methodology for estimating regional emission rates from natural gas production and gathering facilities in northeastern Pennsylvania. An inventory of CH_4 emissions was compiled for major sources in Pennsylvania. This inventory served as input emission data for the Weather Research and Forecasting model with chemistry enabled (WRF-Chem), and atmospheric CH_4 mole fraction fields were generated at 3 km resolution. Simulated atmospheric CH_4 enhancements from WRF-Chem were compared to observations obtained from a three-week flight campaign in May 2015. Modelled enhancements from sources not associated with upstream natural gas processes were assumed constant and known and therefore removed from the optimization procedure, creating a set of observed enhancements from natural gas only. Simulated emission rates from unconventional production were then adjusted to minimize the mismatch between aircraft observations and model-simulated mole fractions for ten flights. To evaluate the method, an aircraft mass balance calculation was performed for four flights where conditions permitted its use. Using the model optimization approach, the weighted mean emission rate from unconventional natural gas production and gathering facilities in northeastern Pennsylvania approach is found to be 0.36% of total gas production, with a 2σ confidence interval

between 0.27-0.45% of production. Similarly, the mean emission estimates using the aircraft mass balance approach is calculated to be 0.40% of regional natural gas production, with a 2σ confidence interval between 0.08-0.72% of production. These emission rates as a percent of production are lower than rates found in any other basin using a top-down methodology, and may be indicative of some characteristics of the basin that makes sources from the northeastern Marcellus region unique.

1 Introduction

The advent of hydraulic fracturing and horizontal drilling technology has opened up the potential to access vast reservoirs of natural gas previously inaccessible, shifting energy trends in the United States away from coal and towards natural gas (EIA, 2016b). From a greenhouse gas (GHG) emissions perspective, natural gas has the potential to be a cleaner energy source than coal. For every unit of energy produced, half as much carbon dioxide (CO_2) is emitted through the stationary combustion of natural gas in comparison to coal (EPA, 2016). However, during the process of extracting and distributing natural gas a percentage of the overall production escapes into the atmosphere through both planned releases and unintended leaks in infrastructure. Though these emissions may be small from an economic perspective, their climatological impacts are not negligible (Alvarez et al., 2012; Schwietzke et al., 2014). Methane (CH_4), the main component of natural gas, is a potent greenhouse gas with a global warming potential over a 20-year period (GWP_{20}) of 84 (Myhre et al., 2013). Over a 100-year period the GWP is reduced to 28 due mostly to interactions with the hydroxyl radical which transform the CH_4 molecule to CO_2. Depending on which timespan is used, the relative climatological impacts of natural gas as an energy source compared to coal can vary. Using the GWP_{20} value, it is estimated that a natural gas emission rate of greater than 3% of total gas production would result in a natural gas power plant having a more negative impact on the climate than a coal-powered plant. Using the GWP_{100} value, this emission rate threshold shifts to 10% of production (Schwietzke et al., 2014; Alvarez et al., 2012). Complicating matters further, the future climate impacts associated with an increased availability of natural gas extends well beyond a simple greenhouse gas footprint comparison against coal. Lower fuel prices linked to this new reservoir of energy can change the course of future energy development globally. With many states and countries attempting to find a suitable balance between their energy policies and greenhouse gas footprint, it is important for the scientific community to be able to quantify and monitor natural gas emission rates.

The drilling and transportation of natural gas can be broken down into five stages: production, processing, storage, transmission, and distribution. The United States Environmental Protection Agency (EPA) uses a bottom-up approach to quantify these emissions, estimating emission rates per facility or component (such as a compressor, unit length of pipeline, pneumatic device) or an average emission per event (such as a well completion or liquids unloading). These "emission factors" are then multiplied by nationwide activity data containing the number of components or events associated with each emission factor, and a total emission rate is produced for the country (EPA, 2015b). This bottom-up approach is a practical methodology for estimating emissions over a large scale but has limitations. A bottom-up inventory depends on the quality and quantity of its emission factors and activity data. Emissions from sources in the natural gas industry can be temporally variable and have a wide range of values depending on a number of factors, such as the quality and age of the device and the gas pressure moving through the component. Furthermore, recent studies have shown that a majority of emissions comes from a small percentage of devices, often referred to as "super-emitters", creating a long-tail distribution of emission sources (Brandt et al., 2014, Omara et al., 2016, Zavala-Araiza et al., 2015, 2017, Frankenberg et al., 2016). These factors make it difficult to sample enough devices and adequately describe the mean emission rate, thus allowing for significant representation errors in the emission factors. Because emission factors are required for hundreds of different components, these errors can accumulate and lead to systematic biases in the total emissions estimate.

One way to compliment results based on inadequate sample sizes in the bottom-up approach is to measure the aggregated enhancement in the atmospheric mole fraction at larger scales through a top-down approach. Instead of measuring emissions from individual devices and scaling up, a top-down approach takes atmospheric greenhouse gas concentrations measured downwind of a continent (e.g. Bousquet et al., 2006), a region (e.g. Lauvaux et al., 2008), a city (e.g. White et al., 1976, Mays et al., 2009, Lamb et al., 2016) or a facility (e.g. Ryerson et al., 2001) and uses inverse methodologies to attribute the enhancements to potential sources upwind. One of these methods, the aircraft mass balance technique, has been performed at many different oil and gas fields to characterize natural gas emissions (Petron et al., 2012, Karion et al., 2013, 2015, Peischl et al., 2015, Conley et al., 2016). While this methodology is able to capture surface fluxes over a large region, it remains difficult to attribute the emissions to any individual source (Cambaliza et al., 2014). Any sources from within the flux region that emit CH_4 will be measured in the downwind observations and be a part of the aggregated regional enhancement. Atmospheric observations may include other sources of CH_4 unrelated to natural gas, such as anaerobic respiration from landfills and wetlands, enteric fermentation from cattle, anaerobic decomposition of manure, CH_4 seepage from coal mining, and many other smaller sources. If the purpose of the study is to solve for the emissions from the natural gas industry, emissions from all sources unrelated to natural gas must be known and removed from the regional flux estimate. Thus, top-down experiments require an accurate CH_4 inventory of the study area and any errors associated with the inventory will propagate into the final emissions estimate. A more advanced technique to separate out non-natural gas sources has been developed using ethane as a tracer for natural gas (Smith et al., 2015). However, such methods may struggle in dry gas basins where smaller ethane to methane ratios within the gas can make the ethane signature more difficult to separate out, or in regions where multiple ethane sources are present. And similar to bottom-up methods, top-down studies fail to address temporal variability, with observations from many of these studies having been collected during a limited number of 2 to 4 hour aircraft flights performed over a period of weeks.

In recent years, both bottom-up and top-down studies have aimed at calculating natural gas emission rates, with bottom-up studies generally finding smaller emission rates than their top-down counterparts (Brandt et al., 2014). The discrepancy between the results from these two methodologies must be better understood if the true emission rate is to be known. Both the bottom-up and top-down approaches have their own inherent sources of error. For the bottom-up approach, a small sample size could result in the omission of any super-emitters, resulting in a low emissions bias. For the top-down approach, difficulty in attributing the measured enhancements to their correct sources can lead to errors when solving for the emissions of a particular sector.

Top-down emission estimates of individual basins have shown variation in the emission rate across the different basins. An aircraft mass balance performed over the Barnett shale in Texas found an emission rate between 1.3-1.9% of production (Karion et al., 2015), yet a similar mass balance study executed over unconventional wells in Uintah County, Utah, calculated an emission rate between 6.2-11.7% of production (Karion et al., 2013). Differences in regional emission rates can perhaps best be illustrated by recent studies in the Marcellus region. The Marcellus shale gas play is part of the Marcellus geological formation running close to the Appalachian mountain chain from West Virginia to southern New York and contains an estimated 140 billion cubic feet of technically recoverable natural gas (EIA, 2012). Reaching peak production by the end of 2015, the Marcellus is the largest producing shale in the U.S., producing 17,000 million standard cubic feet per day (MMSCFD) of natural gas (EIA, 2016a). A bottom-up study measuring emissions from 17 unconventional well-sites in the Marcellus found a median emission rate from the wells of 0.13% of production, but estimated a mean emission rate between 0.38-0.86% of production due to the potential presence of super-emitters which would skew the mean emission rate towards values higher than the median (Omara et al., 2016). An aircraft mass balance study over northeastern Pennsylvania calculated an emission rate between 0.18-0.41%, a number that accounted for emissions from the production, processing, and transmission of the gas (Peischl et al., 2015). Both of these derived estimates fall below emission rates calculated throughout other basins and are below the 3% threshold required for natural gas to be a smaller climate pollutant in comparison to coal over a 20-year timescale. The low rates in the Marcellus compared to other regions could be the result of a systematic difference within the Marcellus that leads to a more efficient extraction of natural gas. However, while useful as a first-guess estimation, current studies performed in the region are based on relatively small sample sizes (1 aircraft mass balance and 88 individual well measurements). A more thorough analysis of the emission rate in the Marcellus would provide insight into regional differences in CH_4 emissions from different shale basins and help improve national estimates of emissions from natural gas.

This study seeks to provide confidence in the emission rate for the northeastern Marcellus by performing the most thorough top-down analysis of the northeastern Marcellus region to date. CH_4 measurements were taken from aircraft observations across 10 flights in northeastern Pennsylvania. A new implementation of modelling CH_4 mole fractions is developed to track complex plume structures associated with different emitters, and an optimal natural gas emission rate is solved for each of the 10 flights. An aircraft mass balance technique is also conducted for 4 of the flights and natural gas emission estimates from this method are compared to those calculated using the modelling technique. Using information on the uncertainty with both methods, a regional emission rate is calculated for the natural gas industry in the northeastern Marcellus region.

2 Methods

The objective of this study is to quantify CH_4 emissions coming from unconventional wells and compressor stations, henceforth referred to as upstream natural gas emissions, in the northeastern Marcellus region (defined as the area contain within 41.1-42.2°N 75.2-77.6°W, see Figure 1) through two different top-down methodologies. CH_4 observations from aircraft data are collected for ten (10) individual flights over a three-week period in May 2015. These data are used to solve for the upstream natural gas emission rate using an aircraft mass balance approach. Additionally, a CH_4 emissions inventory for the region is compiled and input into an atmospheric transport model described below. CH_4 concentrations are modelled for each flight, and the upstream natural gas emission rate within the model is optimized to create the best match between aircraft observations and model projected enhancement, providing another estimate for the upstream natural gas emission rate. The sections below detail the regional CH_4 inventory, the aircraft campaign, the transport model, the model optimization technique, and the mass balance approach used in this study.

2.1 Regional Methane Emission Inventory

In this study we characterize emissions from the natural gas industry into five different sectors: emissions from wells, emissions from compressor facilities, emissions from storage facilities, emissions from pipelines, and emissions in the distribution sector.

To estimate CH_4 emissions from the production sector of the natural gas industry, data were first obtained on the location and production rate of each unconventional well from the Pennsylvania Department of Environmental Protection Oil and Gas Reporting website (PADEP, 2016) and the West Virginia Department of Environmental Protection (WVDEP, 2016). To convert the production rate into an emission rate, we need to assume a first-guess as to the expected leakage from wells in the area. A first-guess natural gas emission rate of 0.13% was applied to the production value of each of the 7000+ producing unconventional wells based on the median rate from Omara et al., (2016). The natural gas emission rate was then converted to a CH_4 emission rate by assuming a CH_4 composition in the natural gas of 95% (Peischl et al., 2015).

In addition to unconventional wells, the domain also contains more than 100,000 shallow conventional wells. Annual conventional production rates for the year 2014 were obtained through the PADEP Oil and Gas Reporting website, the WVDEP, and the New York Department of Environmental Conservation (NYDEC, 2016). Despite the large number of wells, the average conventional well in PA produces 1% of the natural gas compared to its unconventional counterpart. However, it is speculated that the older age of these wells and a lack of maintenance and care for them results in a higher emission rate for these wells as a function of their production (Omara et al., 2016). A first-guess natural gas emission rate of 11% was applied to the production values of the conventional wells based on the median emission rate from the wells sampled in Omara et al., (2016). Similar to the unconventional wells, the natural gas emission rate was then converted to a CH_4 emission rate by assuming a CH_4 composition in the natural gas of 95%.

Compressor stations located within the basin are responsible for collecting natural gas from multiple well locations, removing non-CH_4 hydrocarbons and other liquids from the flow, and regulating pressure to keep gas flowing along gathering and transmission pipelines, and can be a potential source for methane emissions. Data for compressor station locations and emissions comes from a dataset used in Marchese et al., (2015). A total of 489 compressor facilities are listed for Pennsylvania, with 87% of the listed facilities also containing location data. Emissions for each compressor station are calculated through two different methodologies. In the simplest case, a flat emission rate of 32.35 kg hr^{-1} is applied for each station, the mean emission rate of a gathering facility in PA found in Marchese et al., (2015). In the more complex scenario, the same emissions total is used as in the flat rate case, but is distributed among the compressor stations linearly as a function of their energy usage. Wattage between compressors in our dataset can vary greatly, from 10 kW for small compressors to 7000 kW or more at large gathering facilities. Using the wattage as a proxy for emissions allows us to account for the size and throughput of natural gas at each station and assumes larger stations will emit more natural gas compared to smaller stations (Marchese et al., 2015).

Data on locations of underground storage facilities were obtained from the United States Energy Information Administration (EIA, 2015). For each of these locations, a base emission rate of 96.7 kg hr^{-1} was applied according to the average value emitted by a compressor station associated with an underground storage facility (Zimmerle et al., 2016).

To calculate pipeline emissions, data on pipeline locations needed to be collected. Information on transmission pipelines, which connect gathering compressors to distribution networks, is provided by the Natural Gas Pipelines GIS product purchased from Platts, a private organization which collects and creates various infrastructural layers for the natural gas and oil industry (Platts, 2016). Gathering pipeline data, corresponding to the transfer of gas from wellheads to gathering compressors, is nearly non-existent for PA with the exception of Bradford County, which maps out all gathering pipeline infrastructure within the county border. In PA, information on the location of a gathering pipeline elsewhere is only available where a gathering line crosses a stream or river. To account for gathering pipelines in the remainder of the state, a GIS model was created using Bradford County pipelines map in addition to previously generated pipeline maps of Lycoming County (Langlois et al., 2017) as a typical pattern to simulate connecting pipelines between unconventional wells throughout the state (Figure 2). The resulting pattern follows the valley of the Appalachian Mountains, with larger pipelines crossing through the state to connect the different branches of the network. These pipelines were then multiplied by an emission factor of 0.043 kg per mile of pipe, the factor used for gathering pipeline leaks in the Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2013 (EPA, 2015b).

CH_4 emissions from natural gas distribution sources, coal mines, and animal/animal waste were provided from Maasakkers et al., (2016), which takes national scale emissions from the EPA's greenhouse gas inventory for the year 2012 and transforms it into a 0.1° × 0.1° emissions map for the continental U.S. For natural gas distribution emissions, various pipeline data was collected at a state-level and emission factors were accounted for to calculate a total distribution emission for the state. This emissions total was then distributed within the state proportional to the population density. Emission estimates for coal are calculated using information from the Greenhouse Gas Reporting Program (GHGRP) for active mines and the Abandoned Coal Mine Methane Opportunities Database for abandoned mines (EPA, 2008). State-level emissions missions from enteric fermentation and manure management are provided in the EPA's inventory. These emissions were segregated into higher resolutions using county-level data from the 2012 U.S. Census of Agriculture (USDA, 2012) and land-type mapping.

Finally, the EPA's Greenhouse Gas Reporting Program dataset for the year 2014 was used to capture all other major sources of CH_4 in the region otherwise unaccounted for, the majority of which are emissions from landfills and some industrial sources (EPA, 2015a). Sources within the GHGRP that overlap with natural gas sources already accounted for within our inventory were removed to prevent redundancy.

Although our emissions map used for the model runs did not account for potential CH_4 emissions from wetland sources, a series of wetlands emission scenarios was obtained for the region using data from Bloom, et al., (2017). From this dataset, wetland CH_4 emissions make up only 1% of all regional CH_4 emissions in the most extreme scenario, and thus we assume their impact is negligible to this study.

2.2 Aircraft Campaign

Observations for this project were obtained from a 3-week aircraft campaign during the period of May 14th-June 3rd, 2015 and are available for public access (https://doi.org/10.15138/G35K54). The campaign was led by the Global Monitoring Division (GMD) of the National Oceanic and Atmospheric Administration Earth Systems Research and Laboratory (NOAA ESRL), in collaboration with the University of Michigan. During this period, the NOAA Twin Otter aircraft flew throughout the northeast portion of Pennsylvania, providing a total of ten flights across nine days. The aircraft was equipped with a Cavity Ring-Down Spectroscopic analyser (Picarro G2401-m) measuring CH_4, CO_2, CO, and water vapour mole fractions at approximately 0.5Hz with a random error of 1 ppb, 0.1 ppm, 4 ppb, and 50 ppm respectively (Karion et al., 2013). GPS location, horizontal winds, temperature, humidity, and pressure were also recorded at 1 Hz. The majority of observations for each flight occurred during the afternoon hours at heights lower than 1500 m above ground level. Each flight contains at least one vertical profile within and above the boundary layer, with temperature and water vapour observations from these profiles used to estimate the atmospheric boundary layer height and ensure that the aircraft sampled air within the boundary layer throughout the flight. Observations suspected of being located above the boundary layer top are flagged and removed from all calculations.

Flight paths, wind speeds, and CH_4 observations for each of the 10 flights can be seen in Figure 3. For six of the ten flights, a box pattern was flown around a large portion of unconventional natural gas wells in northeastern PA. These flights were performed typically on days with a strong, steady wind, with a clearly defined upwind and downwind transect intended for use in an aircraft mass balance calculation. Five of the six box-pattern flights were composed of two loops circling the gas basin, allowing for two separate calculations of the upstream natural gas emission rate for the flight. On the remaining four flights, raster patterns were performed to help identify spatial complexities of CH_4 emissions within the basin. All ten flights were used in the model optimization calculation of the upstream natural gas emission rate.

2.3 Transport Model

The atmospheric transport model used in this study is the Advanced Weather Research and Forecasting (WRF) model (WRF-ARW, Skamarock et al., 2008) version 3.6.1. The WRF configuration for the model physics used in this research includes the use of: 1) the double-moment scheme (Thompson et al., 2004) for cloud microphysical processes, 2) the Kain-Fritsch scheme (Kain and Fritsch 1990, Kain 2004) for cumulus parameterization on the 9-km grid, 3) the Rapid Radiative Transfer Method for general circulation models (GCMs) (RRTMG, Mlawer et al., 1997, Iacono et al., 2008), 4) the level 2.5 TKE-predicting MYNN planetary boundary layer (PBL) scheme (Nakanishi and Niino 2006), and 5) the Noah 4-layer land-surface model (LSM) that predicts soil temperature and moisture (Chen and Dudhia 2001, Tewari et al., 2004) in addition to sensible and latent heat fluxes between the land surface and atmosphere.

The WRF model grid configuration used in this research contains two grids: 9- and 3-km, each with a mesh of 202x202 grid points. The 9-km grid contains the mid-Atlantic region, the entire northeastern United States east of Indiana, parts of Canada, and a large area of the northern Atlantic Ocean. The 3-km grid contains the entire state of Pennsylvania and most of the state of New York. Fifty vertical terrain-following model layers are used, with the centre point of the lowest model layer located at ~10 m above ground level. The thickness of the layers stays nearly constant with height within the lowest 1 kilometre, with 26 model layers below 850 hPa (~1550 m AGL). One-way nesting is used so that information from the coarse domain translates to the fine domain but no information from the fine domain translates to the coarse domain.

The WRF modelling system used for this study also has four-dimensional data assimilation (FDDA) capabilities to allow meteorological observations to be assimilated into the model (Deng et al., 2009). With WRF FDDA, observations are assimilated through the entire simulation to ensure the optimal model solutions that combine both observation and the dynamic solution, a technique referred to as dynamic analysis. Data assimilation can be accomplished by nudging the model solutions toward gridded analyses based on observations (analysis nudging), or directly toward the individual observations (observation nudging), with a multiscale grid-nesting assimilation framework typically using a combination of these two approaches (Deng et al., 2009; Rogers et al., 2013).

FDDA (Deng et al., 2009) was used in this research, with the same strategy as used in Rogers et al., (2013). Both analysis nudging and observation nudging were applied on the 9-km grid, and only observation nudging was applied on the 3-km grid. In addition to assimilating observations and using the North America Regional Reanalysis model as initial conditions, we reinitialize the WRF model every five days, allowing 12 hours of overlapping period in consideration of model spin-up period to prevent model errors from growing over long periods. The observation data types assimilated include standard WMO surface and upper-air observations distributed by the National Weather Service (NWS), available hourly for surface and 12-hourly for upper air, and the Aircraft Communications Addressing and Reporting System

(ACARS) commercial aircraft observations, available anywhere in space and time with low-level observations near the major airports.

The WRF model used in this study enables the chemical transport option within the model allowing for the projection of CH_4 concentrations throughout the domain. Surface CH_4 emissions used as input for the model come from our CH_4 emissions inventory and are all contained within the 3-km nested grid. Each source of CH_4 within our inventory is defined with its own tracer (Table 1), allowing for the tracking of each individual source's contribution to the overall projected CH_4 enhancement within the model. For this study, CH_4 is treated as an inert gas. The potential for interaction with the hydroxyl radical (OH), the main sink of CH_4, is neglected. A calculation assuming an above-average OH mole fraction over a rural region of 0.5 pptv (Stone et al., 2012) and a reaction rate of 6.5×10^{-15} (Overend et al., 1975) produces a CH_4 sink of 0.5ppb per hour. The duration of a flight can be up to 3 hours, leading to a potential loss of 1.5ppb over the course of a flight. This loss is small but not insignificant. CH_4 plumes associated with natural gas during each flight ranged between 15-70 ppb, and a change of 1.5ppb could theoretically impact observations by as much as 10% of the plume signal. However, this decrease in the CH_4 mole fraction would likely have equal impacts on both the background CH_4 values as well as the enhancement. Because emission calculations are based on the relative difference between the CH_4 background mole fraction and the enhancement downwind, it would take a gradient in the oxidation of OH to impact the results. Considering this relatively low destruction rate, the expected homogeneity of the sink across the region, and the difficulties associated with the simulation of chemical loss processes, we assumed that the CH_4 mass is conserved throughout the afternoon and therefore we ignored the impact of oxidation by OH.

2.4 Model Optimization Technique

2.4.1 Model Optimization Methodology

The objective of the model optimization technique is to solve for an emission rate as a percent of natural gas production that creates the best match between modelled CH_4 concentration maps, provided by the transport model, with actual CH_4 mole fraction observations provided by the aircraft data. The optimization process in this study was originally designed to solve for natural gas emission from unconventional wells and emissions from compressor facilities separately. Because the flow rate of natural gas being processed was not available for each compressor station, emissions at each facility were originally scaled based on the size of the station. However, when running the transport model using this emissions map, enhancements from the compressor stations produced plume structures nearly identical in shape to enhancements from the unconventional wells due to the similar spatial distributions of these two tracers. Without distinct differences between the enhancement patterns from each tracer, it becomes impossible to distinguish which emissions source must be adjusted to obtain the closest match to the observations. For this reason, emissions from compressor facilities are merged with unconventional well emissions in the optimized emission rate. Though the emission rate solved for in this experiment only uses the locations and production for the unconventional wells, this optimized rate represents emissions from both the wells and compressor facilities and are referred to as the modelled upstream natural gas emission rate. Midstream and downstream natural gas processes (such as processing, transmission and distribution of the gas) and emissions from conventional wells are not solved for in this study due to their minimal contribution (less than 5%) to CH_4 emissions in the region encompassed by the aircraft campaign.

Using the transport model WRF-Chem, CH_4 atmospheric enhancements were generated for each flight using different tracers to track different components to the overall CH_4 enhancement (e.g. animal/animal waste, distribution sector, industries). From these concentration fields, the upstream natural gas emission rate was solved for each flight using a three-step model optimization technique. First, a background concentration was determined for each flight and subtracted from the observations to create a set of "observed CH_4 enhancements," using

$$X_{EnhO} = X_{Obs} - X_{bg} \quad , \qquad (1)$$

where X_{Obs} is the CH_4 mole fraction observation from the aircraft, X_{bg} is a chosen background value for the flight, and X_{EnhO} is the calculated CH_4 enhancement at each observation. In this study, the background value is defined as the ambient CH_4 mole fraction over the region not accounted for by any of the sources within the model, with each flight having a unique background value. Box-pattern flights containing 2 loops around the basin may have a different background value assigned for each loop. To determine the background mole fraction, we start with the value of the observed mole fraction in the lowest 2nd percentile of all observations within the boundary layer for a given flight or loop. This chosen background value represents the CH_4 mole fraction across the flight path from sources that are outside of our model domain. Because the background value is meant to represent the CH_4 mole fraction outside the model domain that is otherwise unaccounted for in our model, using the observations with the lowest CH_4 mole fraction is not always a sufficient definition for the background. On certain days, CH_4 enhancements from sources within the model domain can form plumes with wide spatial coverage that cover all observations during a flight. For example, during a flight the lowest CH_4 observations from the aircraft may be 1850 ppb, but the model simulation during that period indicates that all observations within the flight are being impacted by at minimum a 20 ppb enhancement. In this case, we would set our background value for the flight at 1830 ppb, and say that our 1850 ppb observations from the flight are a combination of an 1830 ppb background in addition to a 20 ppb enhancement from sources within the model. By subtracting off this background value from our observations, we create a set of "observed CH_4 enhancements" which can be directly compared to the model projected enhancements

The next step is to remove enhancements from this set that are not associated with emissions from upstream natural gas using

$$X_{GasO} = X_{EnhO} - X_{OtherM} \, , \qquad (2)$$

where X_{OtherM} is the modelled CH_4 enhancement at each observation from sources unrelated to upstream natural gas processes, and X_{GasO} is the observation-derived CH_4 enhancement associated with upstream natural gas emissions for each observation. In this step, each observed CH_4 enhancement has subtracted from it the projected non-natural gas enhancement from the model (i.e. nearest grid point in space) using the corresponding model output time closest to the observation within a 20-minute time interval. This creates a set of observed CH_4 enhancements related only to emissions from upstream gas processes, filtering out potential signals from other CH_4 emitters and providing a set of observed enhancements that can be directly compared to the projected upstream natural gas enhancement within the model. By subtracting these other sources from the observations, we make the assumption that our emissions inventory is accurate for non-natural gas sources and that the transport of these emissions is perfect, both of which are actually uncertain. Because errors exist in both the emissions and transport, it is possible to create a negative observation-derived upstream gas enhancement if model-projected enhancements from other sources are larger than the observation-derived enhancement. From the 10 flights, 16% of the observation-derived enhancements are negative, but only 3% are negative by more than 5 ppb. To avoid solving for unrealistic negative values, these negative observation-based upstream gas enhancements are set to 0. Errors associated with this issue and other uncertainties with our inventory are examined further in the uncertainly analysis section of this paper.

In the final step, the upstream natural gas emission rate within the model is adjusted to create the best match between the modelled upstream gas enhancement and observation-derived upstream gas enhancement using

$$J = X_{GasO} - C * X_{GasM}) \quad (3)$$

where X_{GasO} and X_{GasM} are the observed and modelled enhancement for each observation. In this equation, J is a cost function we are trying to minimize by solving for a scalar multiplier C which, when applied to the modelled natural gas enhancements, creates the smallest sum of the differences between the observation-derived upstream gas enhancement and the modelled upstream enhancement. Because the emission rate within the model is linearly proportional to the model enhancements, we can solve for the upstream natural gas emission rate that minimizes the cost function using

$$E = 0.13\ C \quad (4)$$

where 0.13 was the first guess upstream emission rate (in percent of production) used in the model, and E is the optimized emission rate for the flight as a percentage of the natural gas production at each well. This final value represents an overall emission rate associated with both unconventional wells and compressor stations across the region.

The decision to use a scalar cost function rather than the sum of squares is to account for possible misalignment between any observed CH_4 plume and modelled plumes. There are two potential ways in which misalignment may occur. One possibility is that the modelled wind direction differs from the true wind direction, leading to a plume in the model that is off-centre in relation to the observed plume. The other possibility relates to how the model treats emissions from natural gas as a uniform percent of production. In reality the emissions are more random in nature, and thus the plume may not always develop over the wells with the largest production values. If a cost function is used that minimizes the sum of the squares, any misalignment between the modelled and observed plume will result in the peak of the modelled plume aligning with the height of the tail of the observed plume (Figure 4). Unless the observed plume aligns perfectly with the modelled plume, the optimized emission rate using a sum of squares approach will always bias low. By using a scalar cost function, we solve for an optimized emission rate that results in a plume with the same area under the curve compared to the observed plume (Figure 4). This methodology is not impacted by any misalignment between the modelled vs. observed plumes, preventing the low biases associated with a sum of squares minimization.

2.4.2 Model Optimization Uncertainty Assessment

For each of the ten flights, an uncertainty assessment was performed to obtain a range of likely upstream emission rates for any individual flight. Five different sources of error were considered in this assessment: model wind speed error, model boundary layer height error, CH_4 background error, CH_4 emission inventory error, and model/observation mismatch error. These five sources of error vary substantially from flight to flight depending on conditions, and each can have significant impacts on the total uncertainty (Table 4, 5).

Errors in the modelled wind speed and boundary layer height have impacts on our emission estimates that linearly impact the results. If we assume a constant wind speed, a constant boundary layer height, and no entrainment of air from the top of the boundary layer, we can use the following equation to understand their impacts.

$$\Delta C = \overline{F_0}\left(\frac{\Delta x}{U * D}\right) \quad (5)$$

where ΔC is the total CH_4 enhancement of the column of air contained within the boundary layer, $\overline{F_0}$ is the average emission rate over the path the parcel travelled, Δx is the distance the column of air travelled, U is the wind speed and D is the boundary layer height. Using this equation, we can see the linear relationship between the model wind speed, model boundary layer height, and the calculated emission rate. As an example, if wind speeds in the model are biased low, natural gas enhancements projected by the model would increase inversely. To compensate for this effect, the optimized emission rate would decrease proportionally. A similar case can be made for bias in the boundary layer height. Both errors in the wind speed and boundary layer height have known impacts on the optimized emission rate which can be corrected for, as long as the errors of each are known.

To calculate the error in the model wind speed, we assume aircraft observations are truth and use

$$U_e = \frac{\bar{U}_m - \bar{U}_{obs}}{\bar{U}_{obs}} \quad (6)$$

where $\bar{U}_{obs}$ is the mean observed wind speed by the aircraft across all points within the boundary layer, $\bar{U}_m$ is the mean modelled wind speed by the model across all points closest in time and space to each observation, and U_e is the wind speed error percentage.

To compute the error in the modelled boundary layer height, the observed boundary layer height for each flight is assumed to be the true boundary layer height and the boundary layer height percentage error, H_e, is estimated using:

$$H_e = \frac{\bar{H}_m - \bar{H}_{obs}}{\bar{H}_{obs}} \quad (7)$$

where $\bar{H}_{obs}$ is the average observed boundary layer height across each of the aircraft profiles for a given flight, $\bar{H}_m$ is the model boundary layer height closest in time and space to the location of the observed profiles averaged over all profiles. For both the observation and the model, boundary layer heights were determined by locating height of the potential temperature inversion associated with the top of the boundary layer. On the May 22 flight where a potential temperature inversion could not easily be identified in the observations, changes in water vapour, CO_2 and CH_4 mixing ratios were used to identify the boundary layer top.

Errors in the model wind speed and boundary layer height are calculated for each of the ten flights. From these errors, a corrected optimized emission rate is calculated for each flight using Eq. (8):

$$E_{new} = \frac{E}{(1+U_e)(1+H_e)} \quad (8)$$

where E is the original emission rate and E_{new} is the corrected optimized upstream natural gas emission rate as a percent of production.

In addition to errors related to wind speed and boundary layer height, we quantify three other sources of error in each flight: errors in the selected CH_4 background value, errors in the CH_4 inventory, and errors associated with the overall model performance (Table 5). Unlike the wind speed and boundary layer errors which have easily computable impacts on the emission estimates, these other three sources of error and their impact on the optimized emission rate are more difficult to quantify.

The background error relates to the value chosen for each flight which represents the ambient CH_4 concentration in the boundary layer unrelated to emission sources within the model. In this study background values ranged from 1897-1923ppb. Though background values should not have high variability during a 2-3 hour mid-afternoon flight, entrainment from the boundary layer top can lead to the mixing in of tropospheric air that has different CH_4 mole fraction values from those within the boundary layer, resulting in a change in the afternoon background value with time. Furthermore, for days on which all aircraft observations (including those upwind of the unconventional wells) are impacted by various CH_4 plumes predicted within the model, it is difficult to determine the background CH_4 concentration accurately. Additionally, observations corresponding to locations with no modelled enhancement may in fact have been impacted by missing sources in our inventory, highlighting the difficult nature of knowing with certainty where and what the background is for any given flight. Understanding this uncertainty is crucial; any error in subtracting off the background value directly impacts each observation's observed natural gas enhancement. For example, a background value of 1 ppb below the true background for a given flight would add 1 ppb to each observed natural gas enhancement for all observations, creating a high bias with the optimized upstream emission rate. To account for this error, each flight's optimization processes was rerun iterating the background value by ±5 ppb, and the ratio of the percent change in the emission rate compared to the original case was defined as the resulting error in the emission rate due to background uncertainty. This ±5 ppb background error range is an estimate at the range of possible error in the background based on changes observed in the upwind measurements from each of the flights and is meant to be a conservative estimate of the error. The impact this error can have in the emission rate varies depending on the magnitude of the observed downwind enhancements during a flight. A plume containing a CH_4 enhancement of 50 ppb will have a smaller relative error from a 5 ppb change compared to one with an enhancement of only 10 ppb. Thus, days with high wind speeds and a high boundary layer height (and thus enhancements of a smaller magnitude) tend to be affected the most by background errors.

Similar to background errors, errors from the CH_4 emissions inventory are difficult to quantify. In the model optimization technique, we subtract out enhancements from sources unrelated to unconventional natural gas before solving for the upstream gas emission rate. In doing so, we are making the assumption that our emissions inventory for sources unrelated to upstream natural gas processes are accurate. In truth, each emission source in our inventory comes from a different dataset and has its own unique error bounds, many of which are unknown. To simulate the potential errors associated with unknowns in our inventory, we use a Monte Carlo approach and iterate the unconventional emissions optimization approach for each flight 10,000 times, applying a random multiplier between 0-2 for each of the different sources not associated with unconventional natural gas production. The resulting range of optimized natural gas emission rates was fit to a Gaussian distribution and the 2σ emission range was calculated. Despite varying the emissions used in the error analysis by 0 to 200% their original value, their impacts on the optimized natural gas emission rate are minimal on most days due to the northeastern Marcellus region having very few emission sources not related to upstream natural gas processes. Only for the flights on May 24th do we see errors from the inventory contribute significantly to the overall daily error, when the coal plume in southwestern PA enters the centre of the study region and has a large role in the upstream natural gas emission rate calculation for that day (Table 5).

The final source of error attempts to quantify the similarity of the pattern of modelled and observed natural gas enhancements, referred to here as the model performance error. Figure 5 shows an example of two days, one of which the model appears to recreate the observations, and the other of which the model poorly matches the shape of the observed enhancements. Comparing these two simulations with no other information, we hypothesize that one should put more trust in the upstream natural gas emission rate calculated for the flight whose modelled upstream enhancements match structurally compared to the emission rate from the flight whose modelled enhancement bares little semblance to the observed enhancement. The model performance error is designed to account for the trustworthiness of the optimized upstream emission rate based on how well the model simulates a given day. The model performance error is calculated using a modified normalized root mean squared error formula given in Eq. (9):

$$e_{Perf} = \frac{\bar{\sigma}_{\Delta X}}{\Delta X_{gas}} \quad (9)$$

In this equation, $\bar{\sigma}_{\Delta X}$ is the standard deviation of the difference between the modelled and observation-derived upstream natural gas CH_4 enhancement using the optimized emission rate, and ΔX_{gas} is the observed magnitude of enhancement from the major natural gas plume observed in each flight. Here, ΔX_{gas} serves as a normalization factor to account for the varying strength of the enhancement from flight to flight, and ensures that days with increased enhancements due to meteorological conditions or true daily fluctuations in the upstream natural gas emissions do not proportionally impact the performance error percentage. For example, a day with high winds and a deep boundary layer would produce smaller enhancements, leading to a small $\bar{\sigma}_{\Delta X}$ regardless of model performance unless normalized by ΔX_{gas}.

2.5 Aircraft Mass Balance Method and Uncertainty Assessment

An aircraft mass balance calculation was performed for four applicable flights from the aircraft campaign as an alternative method to calculate upstream natural gas emission rates independent of the transport model. The aircraft mass balance approach uses the CH_4 enhancement between a downwind and upwind transect to calculate the total CH_4 flux of the area contained between the two transects. We use the mass balance equation from Karion et al., (2013):

$$E = \bar{U}cos(\bar{\theta}) \int_{-b}^{b} \Delta X \int_{z=0}^{z_{top}} n_{air} dzdx \qquad (10)$$

where E is the total flux (in mol s^{-1}) coming from the enclosed flight track, $\bar{U}$ is the mean wind speed (in m s^{-1}), $\bar{\theta}$ is the mean angle of the wind perpendicular to the flight track, ΔX is the CH_4 enhancement measured along the downwind flight track from –b to b (expressed as a mole fraction), n_{air} is the molar density of air within the boundary layer (in mol m^{-3}), and each of the integrals represents the summing over all air being measured within our transect in both the horizontal (x) and the vertical (z). By simplifying further and using the mean enhancement along each downwind transect as the enhancement and choosing z_{top} to be the top of the boundary layer, we can transform the previous equation into the following:

$$E = 37.3LDU\Delta\bar{X}cos(\bar{\theta}) \qquad (11)$$

where L is the length of the transect (in meters), D is the depth of the boundary layer (in meters) found using observations from vertical ascents during each flight, $\Delta\bar{X}$ is the mean enhancement across the transect (expressed as a mole fraction), $\bar{U}$ and $\bar{\theta}$ are the mean wind speed (in m s^{-1}) and wind direction relative to the angle of the transect, and 37.3 is the average molar density of dry air within the boundary layer (in mol m^{-3}) assuming an average temperature and pressure of 290K and 900hPa.

Of the 6 days from the aircraft campaign with a clearly defined upwind and downwind transect, one day (May 14th) contained a surface high-pressure centre in the middle of the flight resulting in erratic wind patterns, and another day (May 25th) had CH_4 plumes from southwestern PA affecting portions the flight observations. These days were not used for a mass balance, and calculations were performed for the remaining four box-pattern flights (May 22nd, May 23rd, May 28th, May 29th). From this list of remaining flights, three of them contained two loops around a portion of the Marcellus basin. A mass balance was performed on each loop, resulting in a total of 7 mass balance calculations for the region across 4 days. Table 6 summarizes the results from the mass balance flights.

For each flight, a total flux within the box encompassed was calculated using Eq. (11). Using this flux, a natural gas emission rate based on production from within the box was calculated using Eq. (12)

$$E_{\%} = \frac{E - E_{other}}{P} \quad (12)$$

where E is the total flux from Eq. (11) (in kg hr^{-1}), E_{other} are the emissions enclosed in the box from sources not related to upstream natural gas processes (in kg hr^{-1}), P is the total CH_4 from natural gas being produced within the box (in kg hr^{-1}), and $E_{\%}$ is the resulting natural gas emission rate as a percent of total production within the box.

As an error analysis for the mass balance flight, we look at four potential sources of error (Table 7). One source of uncertainty comes from the observed wind speed used in Eq. (11). For our experiment, we take the mean observed wind speed from the aircraft and assume this value represents the mean wind speed within the entire box during the 2-4 hour period it would take for air to travel from the upwind transect to the downwind transect. To understand the uncertainty and biases associated with this assumption, we recreate wind observations along the flight path using values from WRF-Chem, and compare the mean wind speed from the simulated observations to the mean model winds contained within the box integrated throughout the boundary layer during the 3 hour period closest to the flight time. By making this comparison, we are able to understand the representation error associated with treating the wind speed observations from the aircraft as the wind speed within the entire box during the period it would take for air to cross from the upwind transect to the downwind transect. On average, modelled wind speeds following the flight were 7% faster than integrated wind speeds within the box, due to the inability for aircraft observations to account for slower wind speeds closer to the surface. This bias was removed from each day's calculated wind speed. After accounting for the wind speed bias, the average error of the modelled wind speed following the flight path compared to the modelled winds within the box was 3%. This 3% uncertainty was applied to each flight and used as the potential uncertainty in the mean wind speed. Errors in the wind direction were neglected, as each flight used in the mass balance completely surrounded the basin using downwind transects at multiple angles, and thus small errors in the wind angle would result in a negligible net change on the total flux calculated.

Another source of uncertainty is error in the boundary layer height. For each flight, between 2-3 vertical profiles were performed, and the mean height was used in Eq. (11). The standard deviation of different heights from each transect was used as the uncertainty. On May 22nd, a boundary layer height could be interpreted from only one vertical transect. For this day, we assume an uncertainty of ±200 m (±9%).

Uncertainty in the CH_4 background mole fraction was estimated similar to the boundary layer height. On three of the four flights, two upwind transects were performed. The mean observed CH_4 mole fraction between the two transects was used as the background value for the entire flight, and the standard deviation between the loops was used as the uncertainty. On both the May 23rd and May 28th flights, background differences between the two transects were less than the instrument error of 1 ppb. On these days, we use the instrument error as the background error. On May 22nd, only one upwind transect was usable for the calculation. For this day, we assume a conservative estimate in the uncertainty of the background of ±5 ppb.

Finally, we assess uncertainty in the emissions inventory. After a CH_4 flux is calculated for each loop, emissions from sources contained within the box that are not associated with upstream natural gas processes must be subtracted out to solve for the upstream natural gas emission rate. Any errors associated with our inventory will result in a CH_4 source attribution error. To account for the potentially large uncertainty with the emission sources in our inventory, we vary these non-natural gas emissions by a factor of 2 to test the impact on the solved upstream natural gas emission rate. Because northeastern Pennsylvania contains few sources of CH_4 emissions outside of natural gas production, the impact of this uncertainty is typically less than 20% of the total emissions calculated within the box.

3 Results

3.1 Methane Inventory

From the first-guess CH_4 inventory created in this study, a total anthropogenic CH_4 emission rate of 2.76 Tg CH_4 year^{-1} is projected within our inner model domain (Figure 6) with values for individual source contributions shown in Table 2. This total emissions estimate assumes a leak rate of 0.13% of gas production for unconventional wells, and does not account for emissions from natural gas transmission and storage facilities outside of PA due to a lack of information available from other states. Within the model domain, the area encompassing southwestern PA and northeastern WV stands out as the largest contributor to CH_4 emissions, with emissions from conventional gas, unconventional gas, and coal mines all having significant contributions to the total. In particular, the large emissions from coal make this region unique in comparison to other shales. The EPA's Greenhouse Gas Reporting Program dataset for the year 2014 lists individual coal mines in the southwestern portion of our domain as 8 of the top 10 CH_4 emitting facilities across the entire United States. This large area source of CH_4 can have an impact on CH_4 concentrations hundreds of kilometres downwind and must be taken into account when winds are from the southwest (Figure 7). Examples of this plume and its impacts on the aircraft campaign are discussed in Section 3.2.1.

3.2 Model Optimization Results

3.2.1 Case Studies

From the aircraft campaign, a total of 10 flights across 9 days were used in the model optimization technique. For each one of these flights, CH_4 concentration fields were produced using WRF-Chem, and the emission rate from upstream gas processes was adjusted as outlined in the methods section to find the rate that best matches the total observed CH_4 enhancement. For box flights with two loops completed around the basin, emission rates were calculated for each loop independent from one another and then averaged for the flight. Table 3 provides the general meteorology for the 10 flights.

During each of the observational periods, we use the transport model to project the mole fraction enhancement across the region for each of the different CH_4 tracers (Figure 8). From these projections, we see three common sources of CH_4 which can significantly influence the observed mole fractions in our study region of northeastern PA. The first is emissions from unconventional gas in northeastern PA. Although the first-guess total emissions from upstream production in the Marcellus are small compared to the overall contributions from other sources within the domain, their proximity to the aircraft track results in unconventional gas having the largest contribution to observed enhancements throughout the domain covered by most of the flights, often producing signals downwind of about 20-80 ppb above background levels. The second most influential source of enhancements in our study region comes from various sources of CH_4 emissions located in southwestern PA. Despite being more than 400 km away from our study region, large plumes from coal and other sources in the southwestern corner of the state can contribute enhancements as high as 50 ppb across portions of the flight when winds are from the southwest, affecting background measurements and masking signals from the unconventional gas. One final, but less influential source of CH_4 enhancement is animal agriculture in southeastern PA. Lancaster County is home to roughly 20% of all cattle in the state, with more than 200,000 cattle and calves as of 2012. A southerly wind can result in a 5-15 ppb enhancement across the flight path due to enteric fermentation and manure management from these cattle. Because of coal, conventional gas, and cattle sources located south of the basin, signals from flights with a southerly component to the wind can be difficult to interpret without modelling the projected plumes associated with these sources. Observations on these days contrast to days with a northerly wind component, where a lack of CH_4 sources north of the study region results in observations with a more clearly defined background and unconventional natural gas enhancement.

For each of the ten flights, variability in the model-observation offset was observed. The first loop of the May 29th flight is the best example of a case where comparisons between the modelled and observed enhancements match closely after optimization. For this flight, a box pattern was flown encompassing a majority of the unconventional wells in northeastern PA, and enhancements were observed along the western and northern transects of the flight. Modelled enhancements from sources unrelated to upstream gas emissions showed a broad CH_4 plume associated mostly with animal agriculture along the western edge of the flight, and a smaller enhancement on the eastern edge associated with two landfills in the Scranton/Wilkes-Barre urban corridor (Figure 9). Both of these enhancements are subtracted off from the observations to produce a set of observation-derived enhancements due to upstream natural gas production and gathering facilities. Any enhancements in this new observational dataset are located almost entirely along the northern transect of the flight, directly downwind of the natural gas activity in the region. The observation-derived upstream gas enhancement is then directly compared to the modelled upstream enhancement using its first guess emission rate, and an optimized upstream emission rate of 0.26% of production (i.e. a doubling of the first guess) is calculated by minimizing the difference between the two datasets (Figure 10).

The match between observed and modelled CH_4 enhancements on the first loop of the May 29th flight is closer than any other flight in the campaign. The success of the model on this day is likely due to a number of ideal conditions. In general, inconsistencies between the modelled and observed mean wind speeds and boundary layer heights can have a linear bias on the projected enhancements, but for this flight differences between the observed and modelled wind speed and boundary layer height were near 0 for both loops (Figure 11, 12). Observed wind directions throughout the course of the flight had little directional spread and the averaged observed wind direction was only 9° different compared to modelled values, resulting in a transport of the CH_4 plumes that the model was able to match well. Furthermore, the observed mean wind speed was 4.6 m s^{-1}, a moderate wind which allows for a steady transport of any enhancements towards the downwind transect, but not strong enough to dilute their magnitude, resulting in an easily observable enhancement downwind of the basin. Finally, intrusions from sources unrelated to upstream gas were small on this day due to favourable wind conditions, reducing the probability of incorrectly attributing the observed enhancements to the wrong source. Enhancements from upstream natural gas processes were between 15-40 ppb along our downwind transect. By comparison, enhancements from other sources were lower than 15 ppb along a majority of the flight, and most of these enhancements were located west of the downwind transect, making them easier to identify and remove without unintentionally impacting enhancements from the natural gas plume. All of these different factors likely contributed to producing a situation where the model was successfully able to match CH_4 observations during the May 29th flight.

Flights that occurred on days with a southwest wind had a tendency to produce CH_4 observations that were intuitively difficult to interpret due to convolved CH_4 sources in southwestern Pennsylvania. One of these complex observation sets occurred during the late afternoon flight on May 24th, 2015 (Figure 13). Observations on this day show a CH_4 enhancement that decreased with latitude, with higher CH_4 mole fractions observed farther south. Given the location of the wells in the middle of the flight path and the WSW wind pattern in the region, this north/south CH_4 gradient is unexpected and counterintuitive compared to where one would expect the enhancements to be based solely on the presence of the gas industry in northeastern PA. However, through modelling each of the many contributors of CH_4 within our inventory, we are able to recreate this latitudinal CH_4 gradient and better understand the observed patterns (Figure 13). Throughout an 18-hour period leading up to the May 24th flight, winds from the SSW transport emissions from coal in southwestern PA northeastward until they reach the centre of the state, where a westerly wind then shifts the plume across the study region such that it only intersects the southern half of the flight path. Because of both the magnitude of the coal emissions and an accumulation that occurred in the southwestern portion of the state during the previous night, the modelled enhancement from the coal plume is substantial (>20 ppb) as it crosses over the flight path and covers up much of the signal from upstream gas emissions. Nonetheless, the transport model is able to account for these far-reaching sources and attempt to separate out their contribution to the observed enhancements. We are able to recreate the May 24th flight observations more accurately than most other flights, with a correlation coefficient of 0.71 between the observations and model CH_4 values. Although the model successfully recreates the overall observed CH_4 pattern on this flight, attempting to match model vs. observation-derived enhancements specifically from upstream natural gas contributions is much more difficult. Contributions from non-natural natural gas sources are large such that they overwhelm much of the signal from local natural gas sources. After subtracting out non-natural gas sources from the observations, the correlation specifically between modelled and observation-derived upstream natural gas enhancements is only 0.11.

Despite the model's success at recreating observations from the May 24th late-afternoon flight, there is reason to be careful when interpreting results on day with observations influenced by distant sources. In particular, some transport error is unavoidable in atmospheric reanalyses, and the longer the time and distance a plume takes to reach the observations, the more its position and magnitude will be susceptible to these errors. During the early May 24th flight, a small 50 km shift in the location of the coal plume across the study region would change projected enhancements at some observations by as much as 20 ppb. Furthermore, errors in the transport speed could create scenarios where the coal plume either arrives in the study region too early or exits too late, creating a projected enhancement pattern that does not agree with the observations (Figure 14). Additionally, inaccuracies with the emission estimates of non-unconventional gas sources in the inventory will impact the magnitude of their CH_4 enhancements, creating additional errors in the optimization process when subtracting out these enhancements from the observations. The early-afternoon May 24th flight and May 25th flight are both examples where influences from CH_4 sources in southwest PA create complex structures in the enhancements, which the model is not able to match as well as the late-afternoon flight on May 24th (Figure 15). And although observations and modelled enhancements closely match throughout portions of these two flights, a slight shift in the modelled wind direction can lead to vastly differing results due to the large offset small changes in the wind field can have on an emission source hundreds of kilometres away. Thus, results from the flights on May 24th and May 25th should be taken with caution. A deeper analysis of these errors can be found in Section 3.2.2.

3.2.2 Emission Rates and Uncertainty Assessment

Table 4 shows the wind speed and boundary layer height errors for each flight as well as the optimized and corrected natural gas emission rates. On days where model performance was poor in regards to the wind speed and boundary layer height, we can see changes in the corrected emission rate. For most days, this change is less than 20% different than the original optimized emission rate. However, both May 14th and May 25th have corrected emission rates which are around a factor of 2 different from their original value. Whether these corrected emission rates are more accurate than the original optimized rates is debatable. To calculate these alternative emission rates, we must assume that the wind speeds and boundary layer heights from our limited number of observations are the true values in the atmosphere, which may not be the case. Regardless of which rate is more accurate for each flight, the overall 16% high bias in the model wind speed and the -12% low bias in the model boundary layer result in compensating errors that cancel out, and the mean emission rates across all flights end up similar. Thus, any errors associated with these two meteorological variables has a trivial impact on the overall calculated emission rate for the region and the uncorrected emission rates are used for the final mean and uncertainty calculations.

Table 5 summarizes the background error, inventory error, and model performance error, and assumes independence between the three error sources to calculate the total uncertainty for each flight. The largest uncertainty exists for the May 22nd flight, where an unexplained enhancement along the northern transect led to a poor match between the modelled enhancements and the observed enhancements. This may explain the anomalously high optimized emission rate for that day. Other flights with large uncertainty are those that occurred on May 24th, where enhancements from southwestern PA are believed to be influencing large portions of the observations.

Based on the conservative methodology used to calculate these uncertainties, we assume the total uncertainty for each flight represents a 2σ range of possible emission rates and calculate a weighted mean and a 2σ confidence interval for the overall upstream emission rate across the ten flights. From this approach, we find a mean upstream emission rate of 0.36% of production and a 2σ confidence interval from 0.27-0.45% of production.

3.3 Aircraft Mass Balance Results

Calculated emission rates varied extensively between flights used for the mass balance analysis, ranging from 0.11% to 1.04% of natural gas production (Table 6). Comparing emission rates between loops on the same day, we see more consistency in the values. This result is not surprising, as on each of the days with multiple loops, upwind and downwind CH_4 concentrations patterns tended to be similar between loops. Thus, differences in the total emission rate are likely due to either errors specific to each day (such as background variability, errors in meteorology) or real daily variability in the upstream natural gas emission rate.

From Table 7, we can see the largest error with regards to the absolute uncertainty in the emission rate occurs on the May 22nd flight. It is on this day where we have the largest uncertainty in the background value, with observations towards the end of the flight becoming unusable due to a rapid and unexplained decrease in the CH_4 mole fraction of 8 ppb over a 30 minute period (Figure 16). This day also features the highest boundary layer height and fastest winds of all flights done in this study, reducing the magnitude of the enhancement associated with the natural gas plume and thus amplifying the effects an uncertain background has on the overall uncertainty of the calculated CH_4 flux. Uncertainty across the other three flights is smaller, and results between individual loops on the May 23rd and May 28th flight provide more confidence in the calculated flux for those days.

Using the mean estimated CH_4 emissions and uncertainty for each loop, we calculate a daily mean emission rate and uncertainty for each of the four days. We then solve for an unweighted mean across the four flights to derive our overall emissions estimate from the aircraft mass balance approach, and use the standard error of the flights to estimate the uncertainty. In doing so, we derive a natural gas emission rate from upstream processes of 0.40% of production, with a 2σ confidence interval from 0.08-0.72% of production. Here, we use the arithmetic mean rather than a weighted mean due to the linear relationship between the size of the emission rate and the size of the errors. Because errors associated with ABL height and wind speed have a proportional impact on the calculated emissions within the box, days with a high emissions estimate produce large uncertainties relative to days with a small emission rate. Using a weighted mean approach assigns more weight to the days with low estimated emissions, and produces an overall emission estimate too low and certain to have confidence in (0.12±0.02 percent of gas production).

4 Discussion

4.1 Upstream Emission Rate

From this study, we estimate with a 2σ confidence interval an emission rate between 0.27-0.45% of gas production using the model optimization method and 0.08-0.72% of gas production using the aircraft mass balance. Figure 17 provides the emission range estimates from upstream natural gas processes using both the model optimization technique and mass balance technique when applicable. Top-down studies of other basins in the U.S. have all found emission rates greater than 1% of production, and thus the rates calculated for the northeastern Marcellus basin are the lowest observed yet, raising questions as to why the values in this region appear to be low. One possibility may be related to the well efficiency of the northeastern Marcellus region compared to other major shale plays (Table 8). In terms of gas production per unconventional well, the Marcellus is the highest of all major basins in the U.S. Furthermore, the gas production per well increases by nearly a factor of two when focusing specifically on Susquehanna and Bradford Counties in northeastern Pennsylvania where the majority of the wells from this study are located (Figure 1). The large difference in production per well between the northeastern Marcellus and other shales may partly explain the low emission rates as a percentage of production. Throughout this study, we normalize natural gas emissions as a percentage of total production under the assumption that higher throughput of natural gas in a system should lead to higher emissions in the system. However, if leaks are more influenced by the number of components in operation rather than the throughput passing through the wells, a high production-per-well system such as the unconventional wells in the northeastern Marcellus could end up having a very low emission rate as a percentage of production, but a similar emission rate compared to other basins based on the number of wells, compressors, etc. A thorough bottom-up study of the Marcellus region measuring emissions on a device level could provide an answer to this hypothesis.

Although we calculate a low emission rate for this region, rates calculated for May 22 and May 25 stand out as outliers where emissions fall well-above our uncertainty bounds. It is possible that emissions from natural gas sources were higher on these days compared to others. Releases of natural gas into the atmosphere from short timeframe events such as liquids unloading and venting can add a temporal component to the emission rate. Such events occurring at an increased frequency during the May 22 and May 25 flights could be responsible for the higher emission rates. However, these two days both have issues that could have affected the optimized emission rate. On May 22, we observe a sudden drop in the observed CH_4 values that is nearly as large as the main plume on that day, creating concerns about background concentrations. On May 25, a southwesterly wind was present, and while the model showed the coal plume to be west of the flight path, a small shift in the model wind direction would shift the coal plume over the region. For these reasons we are sceptical but not dismissive of the high emission rates found during these two flights.

4.2 Advantages of Combining Observations with Model Output

One of the major advantages of using a chemical transport model to solve for natural gas emission rates compared to a standard mass-balance approach is that the transport model is able to account for the complex and oftentimes non-uniform plume structures originating from sources outside the flight path that can affect observations. When performing a mass balance over a basin, it is assumed that the upwind transect is representative of the air exiting the downwind transect after subtracting out all sources within the box. However, this assumption is only true if winds contained within the flight path are in perfect steady state during the time it take for air to move from the upwind transect to the downwind transect, and that measurements from the downwind transect occurred at a much later time so that the air being measured is the same air measured from the upwind transect. These conditions are not easily achieved for regional scale mass balances due to the long times needed for the air from the upwind transect to reach the downwind transect. As an example, from the four mass balance flights performed for this study the average time for air to move from the upwind transect to the downwind transect was 4 hours whereas the average time between the aircraft's upwind and downwind measurements was ~40 minutes. The aircraft observations can be thought of as a snapshot in time, which can be problematic if large scale plumes from outside the domain are moving through the region and impacting only certain portions of the observations during the flight's short timeframe. By using a transport model for a domain much larger than that of the flight paths, we are able to track these far-reaching plumes and identify situations where the background CH_4 concentrations may be spatially heterogeneous.

The potential usefulness of using a transport model alongside a mass balance calculation can best be demonstrated from observations taken over the Marcellus during a 2013 aircraft campaign (Peischl et. al 2015). During this flight the prevailing winds were from the WSW, and the largest CH_4 enhancements were observed along the western edge of the flight path, upwind of the unconventional wells. Using our transport model, we are able to recreate the day of flight and attempt to use our inventory and explain this feature (Figure 18). Comparisons between modelled output and observations show a 60 ppb CH_4 enhancement from coal and conventional wells in southwest PA stretching close to the western edge of the aircraft observations, a plume structure similar to the one observed during the May 24th flight from our own study. Though this plume does not initially align with the observed transect with the largest enhancements, we recognize that the coal and gas plume travels for more than 20 hours (a distance of 400 km) from its source before reaching the flight path. If we allow for a 10% error in the transport speed and therefore advance the transport model by an additional two hours past the time in which the aircraft observed these high values, we are able to line up the centre of the plume with the largest observed CH_4 mole fractions along the western edge of the flight. In addition to the 60 ppb enhancement along the centre of the plume, the model projects 20 ppb enhancements along the edges and in front of the plume centre. These smaller enhancements have an influence along different portions of the flight which varies in magnitude, making it difficult to assess a proper background CH_4 value upwind of the wells and potentially masking natural gas enhancements downwind of them. But by using a transport model, we are able to see the potential impact of these far-reaching sources which would otherwise not be considered in a regional mass balance and better understand the complex CH_4 plume structures which can occur in a given region under specific wind conditions.

5 Conclusion

Using the model optimization technique presented in this study, we find a weighted mean natural gas emission rate from unconventional production and gathering facilities of 0.36% of production with a 2σ confidence interval from 0.27-0.45% of production. This emission rate is supported by four mass balance calculations, which produce a mean of 0.40% and a 2σ confidence interval from of 0.08-0.72% of production. Applied to all the wells in our study region, this mean rate results in a leakage rate of 20 Mg CH_4 hr^{-1} for the year 2015. The emission rate found in this top-down study quantified as a percent of production is significantly lower than rates found using top-down methodology at any other basin, and indicates the presence of some fundamental difference in the northeastern Marcellus gas industry that is resulting in more efficient extraction and processing of the natural gas.

The ten flights that took place in this study reveal large regional variations in the CH_4 enhancement patterns depending on the prevailing wind direction. On days with a northwest wind, observed enhancements come primarily from natural gas sources, and a small plume associated with it can be seen on the downwind leg of each flight with few enhancements upwind of the wells. Flights tjat took place with winds conditions predominantly from the southwest were more difficult to interpret. Plumes associated with coal and other potential sources of CH_4 in the southwestern Pennsylvania create complex enhancement patterns affecting both the upwind and downwind portions of the flight, making both the background CH_4 mole fraction and enhancements from the gas industry difficult to interpret. The stark difference between observations that occurred with a northwest wind compared to a southwest wind illustrates the importance of having multiple flights across days with various wind conditions to better understand the major influences on CH_4 concentrations throughout a region. The regional influences in Pennsylvania also demonstrate the utility of deriving an emissions inventory that provides input data to drive a transport model, allowing one to forecast CH_4 mole fractions on difficult days and better understand the daily uncertainties associated with heterogeneous background conditions.

Though this study presented observations from ten flights over a three-week period, it is not able to account for the potential of long term temporal variability in the emission rates. In May 2015 when the flights took place, the entire Marcellus basin was nearing peak production and active drilling and hydraulic fracturing was still ongoing in the region. By mid-2016, the rate of drilling of new wells in the northeast Marcellus had decreased and natural gas production had begun to decline in the area. A snapshot of the emission rate during one month of a basin in its peak production is insufficient to characterize emissions from an area that is likely to be producing and transporting gas at various intensities for decades. We need to quantify the long-term climatological impacts of gas production. Future work examining the temporal variability of CH_4 emissions within natural gas basins would complement short-term, high-intensity studies such as this one, and aid with understanding how well the calculated emission rates represent the gas basin over the course of time.

Acknowledgements

This work has been funded by the U.S. Department of Energy National Energy Technology Laboratory (project DE-FE0013590). We thank in-kind contributions from the Global Monitoring Division of the National Oceanic and Atmospheric Administration, and from the Earth and Environmental Systems Institute, the Department of Meteorology and Atmospheric Science, and the College of Earth and Mineral Science of The Pennsylvania State University. We also want to thank the Pennsylvania College of Technology in Williamsport, PA for access to their Technology Aviation Center facilities during the aircraft campaign. We also want to thank Lillie Langlois from the Department of Ecosystem Science and Management for sharing pipeline information, Anthony J. Marchese and Dan Zimmerle (Colorado State University) for information on compressor stations, Jeff Peischl for sharing data from his 2013 flight campaign, and Bernd Haupt (Penn State University) for data processing and management during the project. We thank the two anonymous reviewers who volunteered their time to improve the contents of this paper. Finally, we would like to thank Dennis and Joan Thomson for their creation and continued support of the Thomson Distinguished Graduate Fellowship.

References

Alvarez, R. A., Pacala, S. W., Winebrake, J. J., Chameides, W. L., and Hamburg, S. P.: Greater focus needed on methane leakage from natural gas infrastructure, Proceedings of the National Academy of Sciences, 109, 6435-6440, 10.1073/pnas.1202407109, 2012.

Bloom, A. A., Bowman, K., Lee, M., Turner, A. J., Schroeder, R., Worden, J. R., Weidner, R., McDonald, K. C., and Jacob, D. J.: A global wetland methane emissions and uncertainty dataset for atmospheric chemical transport models, Geosci. Model Dev., 10.5194/gmd-10-2141-2017, 2017.

Bradford County: Maps of Natural Gas Development in Bradford County, available at: http://www.bradfordcountypa.org/index.php/natural-gas-information.

Bousquet, P., Ciais, P., Miller, J. B., Dlugokencky, E. J., Hauglustaine, D. A., Prigent, C., Van der Werf, G. R., Peylin, P., Brunke, E. G., Carouge, C., Langenfelds, R. L., Lathiere, J., Papa, F., Ramonet, M., Schmidt, M., Steele, L. P., Tyler, S. C., and White, J.: Contribution of anthropogenic and natural sources to atmospheric methane variability, Nature, 443, 439-443, http://www.nature.com/nature/journal/v443/n7110/suppinfo/nature05132_S1.html, 2006.

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., and Harriss, R.: Methane Leaks from North American Natural Gas Systems, Science, 343, 733-735, 10.1126/science.1247045, 2014.

Cambaliza, M. O. L., Shepson, P. B., Caulton, D. R., Stirm, B., Samarov, D., Gurney, K. R., Turnbull, J., Davis, K. J., Possolo, A., Karion, A., Sweeney, C., Moser, B., Hendricks, A., Lauvaux, T., Mays, K., Whetstone, J., Huang, J., Razlivanov, I., Miles, N. L., and Richardson, S. J.: Assessment of uncertainties of an aircraft-based mass balance approach for quantifying urban greenhouse gas emissions, Atmos. Chem. Phys., 14, 9029-9050, 10.5194/acp-14-9029-2014, 2014.

Chen, F., and Dudhia, J.: Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity, Monthly Weather Review, 129, 569-585, 10.1175/1520-0493(2001)129<0569:caalsh>2.0.co;2, 2001.

Conley, S., Franco, G., Faloona, I., Blake, D. R., Peischl, J., and Ryerson, T. B.: Methane emissions from the 2015 Aliso Canyon blowout in Los Angeles, CA, Science, 10.1126/science.aaf2348, 2016.

Deng, A., Stauffer, D., Gaudet, B., Dudhia, J., Hacker, J., Bruyere, C., Wu, W., Vandenberghe, F., Liu, Y., and Bourgeois, A.: Update on WRF-ARW end-to-end multi-scale FDDA system, 10th Annual WRF Users' Workshop, Boulder, CO, June 23, 2009.

Deng, A., Gaudet, B., Dudhia, J., and Alapaty, K.: Implementation and Evaluation of a New Shallow Convection Scheme in WRF, 94th American Meteorological Society Annual Meeting, Atlanta, GA, February 2-6, 2014.

EIA: Annual Energy Outlook 2012 with Projections to 2035, available at: http://www.eia.gov/forecasts/aeo/pdf/0383(2012).pdf, 2012.

EIA: Layer Information for Interactive State Maps, available at: https://www.eia.gov/maps/layer_info-m.cfm, 2015

EIA: Shale in the United States, available at: https://www.eia.gov/energyexplained/index.cfm?page=natural_gas_where, 2016a

EIA: Monthly Energy Review: June 2016. [Available online at http://www.eia.gov/totalenergy/data/monthly/archive/00351607.pdf , 2016b.

EPA: Abandoned Coal Mine Methane Opportunities Database, available at: https://www.epa.gov/cmop/abandoned-underground-mines, 2008.

EPA: Greenhouse Gas Reporting Program 2014, available at: https://www.epa.gov/ghgreporting/ghg-reporting-program-data-sets, 2015a.

EPA: Inventory of US Greenhouse Gas Emissions and Sinks: 1990-2013 Annex 3.6, available at: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2013, 2015b.

EPA: Inventory of US Greenhouse Gas Emissions and Sinks: 1990-2014 Annex 2.2, available at: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2014, 2016.

Frankenberg C., Thorpe, A. K., Thompson, D. R., Hulley, G., Kort, E. A., Vance, N., Borchardt, J., Krings, T., Gerilowski, K., Sweeney, C., Conley, S., Bue, B. D., Aubrey, A. D., Hook, S., and Green, R. O.: Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region, Proceedings of the National Academy of Sciences, 113, 9734-9739, 2016.

Hughes, J. D.: Drilling Deeper: A Reality Check on US Government Forecasts for a Lasting Tight Oil & Shale Gas Boom, Post Carbon Institute, Santa Rosa, California, 2014.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, Journal of Geophysical Research: Atmospheres, 113, 10.1029/2008jd009944, 2008.

Jimenez, P. A., Hacker, J. P., Dudhia, J., Haupt, S. E., Ruiz-Arias, J. A., Gueymard, C. A., Thompson, G., Eidhammer, T., and Deng, A.: WRF-Solar: Description and Clear-Sky Assessment of an Augmented NWP Model for Solar Power Prediction, Bulletin of the American Meteorological Society, 97, 1249-1264, 10.1175/bams-d-14-00279.1, 2016.

Jiménez, P. A., Alessandrini, S., Haupt, S. E., Deng, A., Kosovic, B., Lee, J. A., and Monache, L. D.: The Role of Unresolved Clouds on Short-Range Global Horizontal Irradiance Predictability, Monthly Weather Review, 144, 3099-3107, 10.1175/mwr-d-16-0104.1, 2016.

Kain, J. S., and Fritsch, J. M.: A One-Dimensional Entraining/Detraining Plume Model and Its Application in Convective Parameterization, Journal of the Atmospheric Sciences, 47, 2784-2802, 10.1175/1520-0469(1990)047<2784:aodepm>2.0.co;2, 1990.

Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update, Journal of Applied Meteorology, 43, 170-181, 10.1175/1520-0450(2004)043<0170:tkcpau>2.0.co;2, 2004.

Karion, A., Sweeney, C., Pétron, G., Frost, G., Michael Hardesty, R., Kofler, J., Miller, B. R., Newberger, T., Wolter, S., Banta, R., Brewer, A., Dlugokencky, E., Lang, P., Montzka, S. A., Schnell, R., Tans, P., Trainer, M., Zamora, R., and Conley, S.: Methane emissions estimate from airborne measurements over a western United States natural gas field, Geophysical Research Letters, 40, 4393-4397, 10.1002/grl.50811, 2013.

Karion, A., Sweeney, C., Kort, E. A., Shepson, P. B., Brewer, A., Cambaliza, M., Conley, S. A., Davis, K., Deng, A., Hardesty, M., Herndon, S. C., Lauvaux, T., Lavoie, T., Lyon, D., Newberger, T., Pétron, G., Rella, C., Smith, M., Wolter, S., Yacovitch, T. I., and Tans, P.: Aircraft-Based Estimate of Total Methane Emissions from the Barnett Shale Region, Environmental Science & Technology, 49, 8124-8131, 10.1021/acs.est.5b00217, 2015.

Lamb, B. K., Cambaliza, M. O. L., Davis, K. J., Edburg, S. L., Ferrara, T. W., Floerchinger, C., Heimburger, A. M. F., Herndon, S., Lauvaux, T., Lavoie, T., Lyon, D. R., Miles, N., Prasad, K. R., Richardson, S., Roscioli, J. R., Salmon, O. E., Shepson, P. B., Stirm, B. H., and Whetstone, J.: Direct and Indirect Measurements and Modeling of Methane Emissions in Indianapolis, Indiana, Environmental Science & Technology, 50, 8910-8917, 10.1021/acs.est.6b01198, 2016.

Langlois, L. A., Drohan, P. J., and Brittingham, M. C.: Linear infrastructure drives habitat conversion and forest fragmentation associated with Marcellus shale gas development in a forested landscape. Journal of Environmental Management. 197, 167-176, 10.1016/j.jenvman.2017.03.045. 2017.

Lauvaux, T., Uliasz, M., Sarrat, C., Chevallier, F., Bousquet, P., Lac, C., Davis, K. J., Ciais, P., Denning, A. S., and Rayner, P. J.: Mesoscale inversion: first results from the CERES campaign with synthetic data, Atmos. Chem. Phys., 8, 3459-3471, 10.5194/acp-8-3459-2008, 2008.

Maasakkers, J. D., Jacob, D. J., Sulprizio, M. P., Turner, A. J., Weitz, M., Wirth, T., Hight, C., DeFigueiredo, M., Desai, M., Schmeltz, R., Hockstad, L., Bloom, A. A., Bowman, K. W., Jeong, S., and Fischer, M. L.: Gridded National Inventory of U.S. Methane Emissions, Environmental Science & Technology, 50, 13123-13133, 10.1021/acs.est.6b02878, 2016.

Marchese, A. J., Vaughn, T. L., Zimmerle, D. J., Martinez, D. M., Williams, L. L., Robinson, A. L., Mitchell, A. L., Subramanian, R., Tkacik, D. S., Roscioli, J. R., and Herndon, S. C.: Methane Emissions from United States Natural Gas Gathering and Processing, Environmental Science & Technology, 49, 10718-10727, 10.1021/acs.est.5b02275, 2015.

Mays, K. L., Shepson, P. B., Stirm, B. H., Karion, A., Sweeney, C., and Gurney, K. R.: Aircraft-Based Measurements of the Carbon Footprint of Indianapolis, Environmental Science & Technology, 43, 7816-7823, 10.1021/es901326b, 2009.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, Journal of Geophysical Research: Atmospheres, 102, 16663-16682, 10.1029/97jd00237, 1997.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., and Mendoza, B.: Anthropogenic and natural radiative forcing, Climate change, 423, 2013.

Nakanishi, M., and Niino, H.: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog, Boundary-Layer Meteorology, 119, 397-407, 2006.

NYDEC: NY 2014 Oil & Gas Production Data: available at: http://www.dec.ny.gov/energy/36159.html, 2016.

Omara, M., Sullivan, M. R., Li, X., Subramanian, R., Robinson, A. L., and Presto, A. A.: Methane Emissions from Conventional and Unconventional Natural Gas Production Sites in the Marcellus Shale Basin, Environmental Science & Technology, 50, 2099-2107, 10.1021/acs.est.5b05503, 2016.

Overend, R. P., Paraskevopoulos, G., and Cvetanović, R. J.: Rates of OH Radical Reactions. I. Reactions with H2, CH4, C2H6, and C3H8 at 295 K, Canadian Journal of Chemistry, 53, 3374-3382, 10.1139/v75-482, 1975.

PADEP: PA Oil and Gas Well Historical Production Report, available at: http://www.depreportingservices.state.pa.us/ReportServer/Pages/ReportViewer.aspx?%2fOil_Gas%2fOil_Gas_Well_Historical_Production_Report, 2016.

Peischl, J., Ryerson, T. B., Aikin, K. C., de Gouw, J. A., Gilman, J. B., Holloway, J. S., Lerner, B. M., Nadkarni, R., Neuman, J. A., Nowak, J. B., Trainer, M., Warneke, C., and Parrish, D. D.: Quantifying atmospheric methane emissions from the Haynesville, Fayetteville, and northeastern Marcellus shale gas production regions, Journal of Geophysical Research: Atmospheres, 120, 2119-2139, 10.1002/2014jd022697, 2015.

Pétron, G., Frost, G., Miller, B. R., Hirsch, A. I., Montzka, S. A., Karion, A., Trainer, M., Sweeney, C., Andrews, A. E., Miller, L., Kofler, J., Bar-Ilan, A., Dlugokencky, E. J., Patrick, L., Moore, C. T., Ryerson, T. B., Siso, C., Kolodzey, W., Lang, P. M., Conway, T., Novelli, P., Masarie, K., Hall, B., Guenther, D., Kitzis, D., Miller, J., Welsh, D., Wolfe, D., Neff, W., and Tans, P.: Hydrocarbon emissions characterization in the Colorado Front Range: A pilot study, Journal of Geophysical Research: Atmospheres, 117, 10.1029/2011jd016360, 2012.

Platts: Maps and Geospatial Data: available at: http://www.platts.com/maps-geospatial, 2016.

Rogers, R., Deng, A., Stauffer, D., Jia, Y., Soong, S., Tanrikulu, S., Beaver, S., and Tran, C.: Fine particulate matter modeling in Central California. Part I: Application of the Weather Research and Forecasting model, 91th Annual Meeting, 2011.

Ryerson, T. B., Trainer, M., Holloway, J. S., Parrish, D. D., Huey, L. G., Sueper, D. T., Frost, G. J., Donnelly, S. G., Schauffler, S., Atlas, E. L., Kuster, W. C., Goldan, P. D., Hübler, G., Meagher, J. F., and Fehsenfeld, F. C.: Observations of Ozone Formation in Power Plant Plumes and Implications for Ozone Control Strategies, Science, 292, 719-723, 10.1126/science.1058113, 2001.

Schwietzke, S., Griffin, W. M., Matthews, H. S., and Bruhwiler, L. M. P.: Natural Gas Fugitive Emissions Rates Constrained by Global Atmospheric Methane and Ethane, Environmental Science & Technology, 48, 7714-7722, 10.1021/es501204c, 2014.

Smith, M. L., Kort, E. A., Karion, A., Sweeney, C., Herndon, S. C., and Yacovitch, T. I.: Airborne Ethane Observations in the Barnett Shale: Quantification of Ethane Flux and Attribution of Methane Emissions, Environmental Science & Technology, 49, 8158-8166, 10.1021/acs.est.5b00219, 2015.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 2, DTIC Document, 2005.

Stone, D., Whalley, L. K., and Heard, D. E.: Tropospheric OH and HO2 radicals: field measurements and model comparisons, Chemical Society Reviews, 41, 6348-6404, 10.1039/c2cs35140d, 2012.

Sweeney, C., Karion, A., Kort, E. A., Smith, M. L., Newberger T., Schwietzke, S., Wolter, S., and Lauvaux, T.: Aircraft Campaign Data over the Northeastern Marcellus Shale, May-June 2015, Version: 2017-03-29, Path: https://doi.org/10.15138/G35K54, 2015

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R.: Implementation and verification of the unified NOAH land surface model in the WRF model, 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, 2004.

Thompson, G., Field, P. R., Rasmussen, R. M., and Hall, W. D.: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization, Monthly Weather Review, 136, 5095-5115, 10.1175/2008mwr2387.1, 2008.

USDA: Census Ag Atlas Maps, available at: https://www.agcensus.usda.gov/Publications/2012/Online_Resources/Ag_Atlas_Maps/Livestock_and_Animals/, 2012.

White, W. H., Anderson, J. A., Blumenthal, D. L., and Wilson, W. E., Formation and transport of secondary air-pollutants: Ozone and aerosols in St. Louis urban plume: Science 194, 187-189, 10.1126/science.959846, 1976.

WVDEP: WV Oil and Gas Database and Map Information, available at: http://www.dep.wv.gov/oil-and-gas/databaseinfo/Pages/default.aspx, 2016.

Zavala-Araiza, D., Lyon., D., Alvarez., R. A., Palacios., V., Harriss., R., Lan., X., Talbot., R., and Hamburg., S. P.: Toward a Functional Definition of Methane Super-Emitters: Application to Natural Gas Production Sites, Enviornmental Science & Technology, 49, 8167-8174, 10.1021/acs.est.5b00133, 2015

Zavala-Araiza, D., Alvarez, R. A., Lyon, D. R., Allen, D. T., Marchese, A. J., Zimmerle, D. J., and Hamburg, S. P.: Super-emitters in natural gas infrastructure are caused by abnormal process conditions, Nature Communications, 8, 14012, 10.1038/ncomms14012, 2017.

Zimmerle, D. J., Williams, L. L., Vaughn, T. L., Quinn, C., Subramanian, R., Duggan, G. P., Willson, B., Opsomer, J. D., Marchese, A. J., Martinez, D. M., and Robinson, A. L.: Methane Emissions from the Natural Gas Transmission and Storage System in the United States, Environmental Science & Technology, 49, 9374-9383, 10.1021/acs.est.5b01669, 2015.

Table 1: List of tracers used in the transport model.

Tracer #	Name	Description of source
1	Unconventional Wells	Emissions from unconventional wells.
2	Storage Facilities	Emissions from compressors associated with natural gas storage.
3	Pipelines	Emissions from gathering and transmission pipelines
4	Distribution	Emissions from the distribution sector of the natural gas industry.
5	Conventional Wells	Emissions from conventional wells.
6	Landfills/Other	Emissions from landfills and uncharacterized industrial sources.
7	Coal	Emissions from active and abandoned coal mining.
8	Animals/Waste	Emissions from enteric fermentation and manure management
9	Production Compressors (HP)	Emissions from compressor stations characterized as "production". Emissions scaled linearly with wattage.
10	Gathering Compressors (HP)	Emissions from compressor stations characterized as "gathering". Emissions scaled linearly with wattage.
11	Other Compressors (HP)	Emissions from all other compressor stations. Emissions scaled linearly with wattage.
12	Production Compressors (C)	Emissions from compressor stations characterized as "production". Emissions constant among compressors.
13	Gathering Compressors (C)	Emissions from compressor stations characterized as "gathering". Emissions constant among compressors.
14	Other Compressors (C)	Emissions from all other compressor stations. Emissions constant among compressors.

Table 2: Annual emission rate totals from anthropogenic sources within the innermost model domain based on values from the inventory within this study

Source	Total Emission Rate (Gg CH_4 $year^{-1}$)
Unconventional Wells	125
Conventional Wells	607
Gathering Compressor Facilities	118
Storage Facilities	69
Gathering/Transmission Pipelines	8
Natural Gas Distribution	213
Underground, Surface, and Abandoned Coal Mines	831
Enteric Fermentation/Manure Management	371
Landfills	420
Total	**2762**

Table 3: Meteorological statistics from the May 2015 flight campaign.

Day	Flight Pattern	# of Loops	# of Vertical Profiles	ABL Depth (m)	Mean Observed Wind Speed (m/s)	Mean Observed Wind Direction	Model Background Value (ppm)
May 14	Box	1	2	1300	2.9	30°	1.908
May 21	Raster	N/A	2	1300	3.9	231°	1.905
May 22	Box	2	2	2300	10.1	300°	1.910
May 23	Box	2	2	1400	4.4	276°	1.906
May 24[1]	Other	N/A	2	1500	4.4	270°	1.923
May 24[2]	Raster	N/A	2	2050	4.8	272°	1.907
May 25	Box	1	2	1800	9.0	217°	1.920
May 28	Box	2	3	1400	7.1	322°	1.897
May 29	Box	2	2	1000	4.6	195°	1.899
June 3	Raster	N/A	1	1250	2.7	149°	1.898

Table 4: Optimized natural gas emission rates for each flight as well as corrected emission rates adjusting for errors in the model wind speed and boundary layer height For wind speed and boundary layer height error, a negative value represents a model value less than the observations.

Day	Optimized NG Emission Rate (% of production)	Wind Speed Error (6)	Boundary Layer Height Error (7)	Corrected NG Emission Rate (% of production)
May 14	0.37	-31%	-33%	0.80
May 21	0.53	3%	39%	0.37
May 22	1.15	37%	-18%	1.02
May 23	0.45	34%	-9%	0.37
May 24	0.68	48%	-21%	0.58
May 24	0.36	48%	-21%	0.30
May 25	0.99	3%	-43%	1.69
May 28	0.33	-4%	-8%	0.37
May 29	0.35	4%	1%	0.33
June 3	0.26	19%	-8%	0.24
Average	**0.55**	**16%**	**-12%**	**0.61**

Table 5: Emission rates and potential errors associated with the model optimization technique. r-values represent the correlation between the model and observation-derived upstream natural gas enhancements.

Day	Optimized Upstream Emission Rate (% of production)	r-value Model vs Obs NG Sources	Background Error	Non-Upstream Gas Inventory Error	Model Performance Error	Total Error	2σ Confidence Interval (% of Production)
May 14	0.37	0.20	±24%	±19%	±17%	±35%	±0.13
May 21	0.53	0.31	±24%	±13%	±30%	±41%	±0.22
May 22	1.15	0.47	±38%	±5%	±37%	±53%	±0.61
May 23	0.45	0.10	±39%	±13%	±42%	±59%	±0.26
May 24[1]	0.68	0.31	±24%	±81%	±17%	±86%	±0.58
May 24[2]	0.36	0.11	±51%	±150%	±31%	±161%	±0.57
May 25	0.99	0.43	±29%	±15%	±30%	±44%	±0.44
May 28	0.33	0.33	±76%	±12%	±20%	±79%	±0.26
May 29	0.35	0.58	±24%	±11%	±19%	±33%	±0.12
June 3	0.26	0.37	±31%	±12%	±24%	±41%	±0.11

Table 6: Emission rates from mass balance calculations on applicable days, with emission ranges associated with a ±5 ppb error in the background value.

Flight	CH_4 Production within box (Gg hr^{-1})	Mass Balance CH_4 Flux (kg hr^{-1})	Non-Upstream CH_4 Emissions (kg hr^{-1})	Calculated Upstream Emission Rate (% of production)	2σ Confidence Interval (% of Production)
May 22_1	4.96	53800	2250	1.04	±1.09
May 22_2	4.96	27400	2250	0.51	±1.08
May 23_1	4.05	5600	934	0.11	±0.07
May 23_2	4.05	5500	934	0.11	±0.07
May 28_1	3.73	7100	706	0.17	±0.11
May 28_2	3.73	6000	843	0.14	±0.10
May 29_1	4.63	27900	1622	0.57	±0.30

Table 7: Relative error associated with the different sources of uncertainty in the aircraft mass balance.

Flight	Wind Speed Error	Background Error	ABL Error	Inventory Error	Total Error (1σ)	Upstream Emission Rate (% of Production) w/ 2σ Confidence Interval
May 22_1	±3%	±56%	±9%	±5%	±57%	1.04 ±1.09
May 22_2	±3%	±121%	±9%	±8%	±121%	0.51 ±1.08
May 23_1	±3%	±24%	±7%	±20%	±32%	0.11 ±0.07
May 23_2	±3%	±26%	±7%	±21%	±34%	0.11 ±0.07
May 28_1	±3%	±31%	±7%	±11%	±34%	0.17 ±0.11
May 28_2	±3%	±33%	±7%	±16%	±38%	0.14 ±0.10
May 29_1	±3%	±28%	±20%	±8%	±36%	0.57 ±0.30

Table 8: Production statistics from mid-2014 for various shales across the United States (Hughes 2014).

	Barnett	Fayetteville	Haynesville	Marcellus	Bradford/ Susquehanna County, PA
# of Producing Wells	16100	4500	3100	7000	1558
Total Production (Bcf day^{-1})	5.0	2.8	4.5	12	5.01
Production per well (MMcf day^{-1})	0.31	0.56	1.25	1.71	3.22

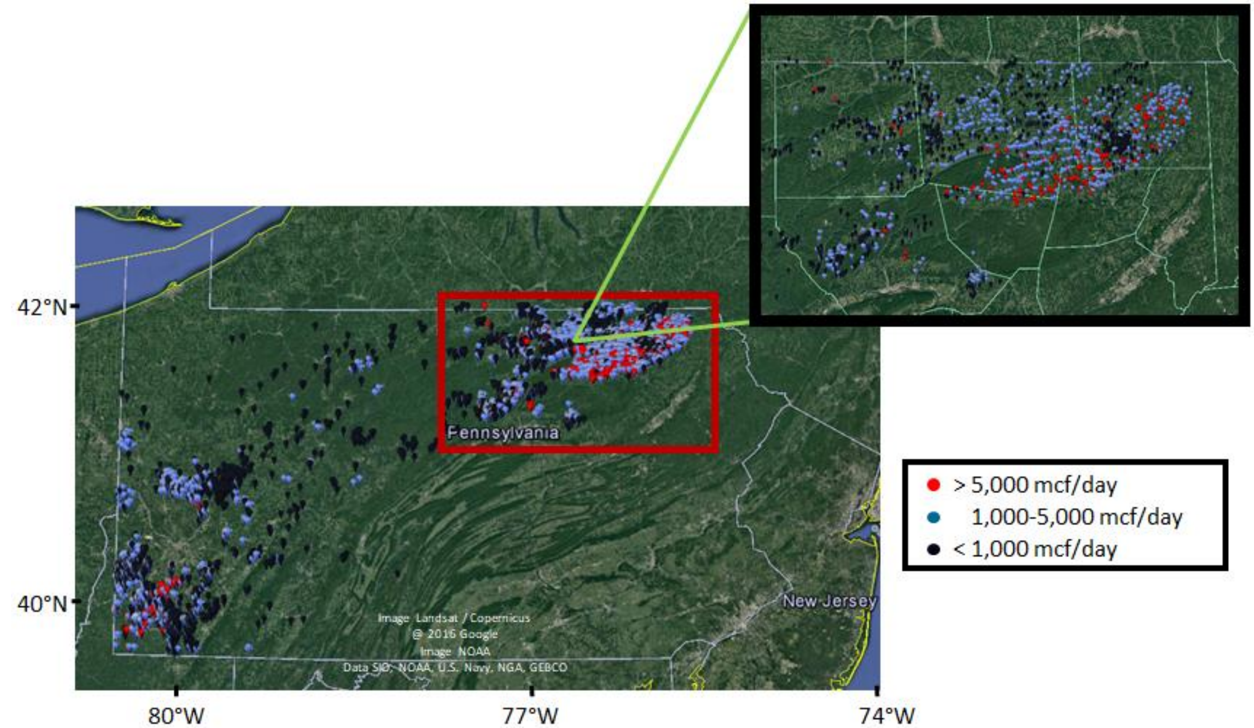

Figure 1: A map of the unconventional wells in Pennsylvania dotted in purple. Production values of wells for May 2015 are indicated by the marker colour. Red rectangle and zoom-in show the region of focus for this study, 41.1-42.2°N 75.2-77.6°W.

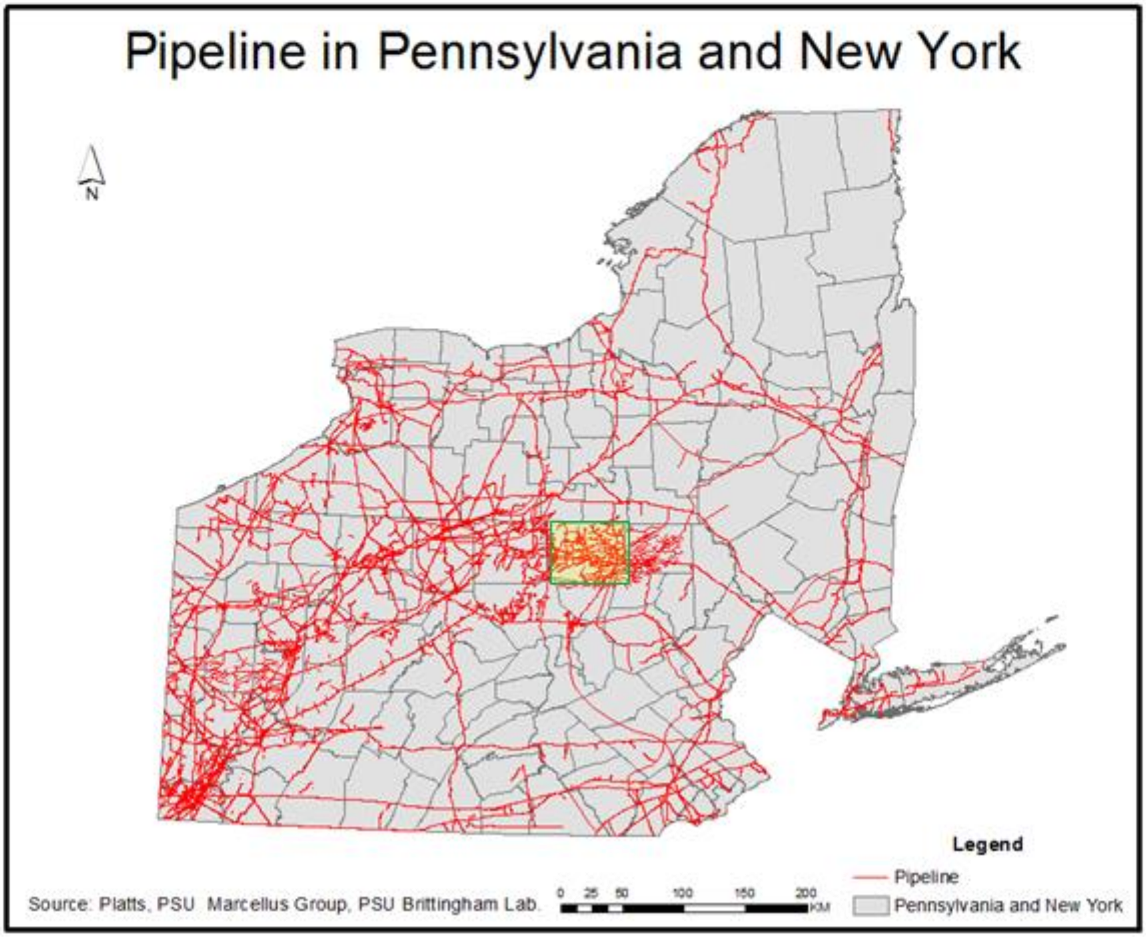

Figure 2: A map of transmission and gathering pipelines for the state of PA and NY. Transmission pipelines are provided by Platts Natural Gas Pipelines product. Gathering pipelines associated with unconventional wells in PA are extrapolated using information on existing gathering pipelines provided by Bradford County, PA (highlighted in yellow).

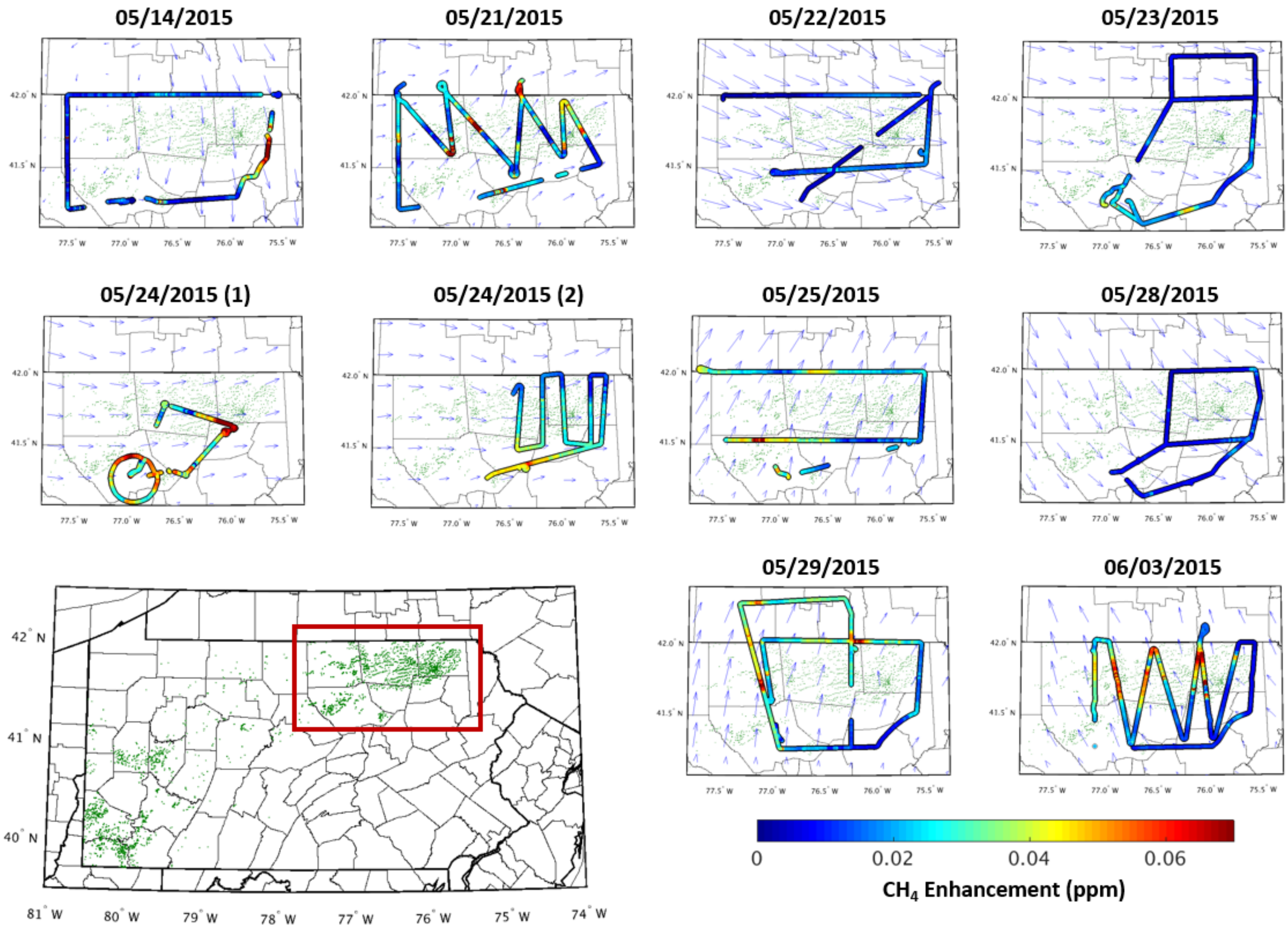

Figure 3: Observed CH_4 enhancements within the boundary layer from each of the 10 afternoon flights used in this study, with green dots showing the location of unconventional wells in PA and blue arrows showing the modelled wind direction during the time of the flight. CH_4 enhancements are calculated by taking the observed CH_4 mole fraction values and subtracting off the flight's background CH_4 value shown in Table 3.

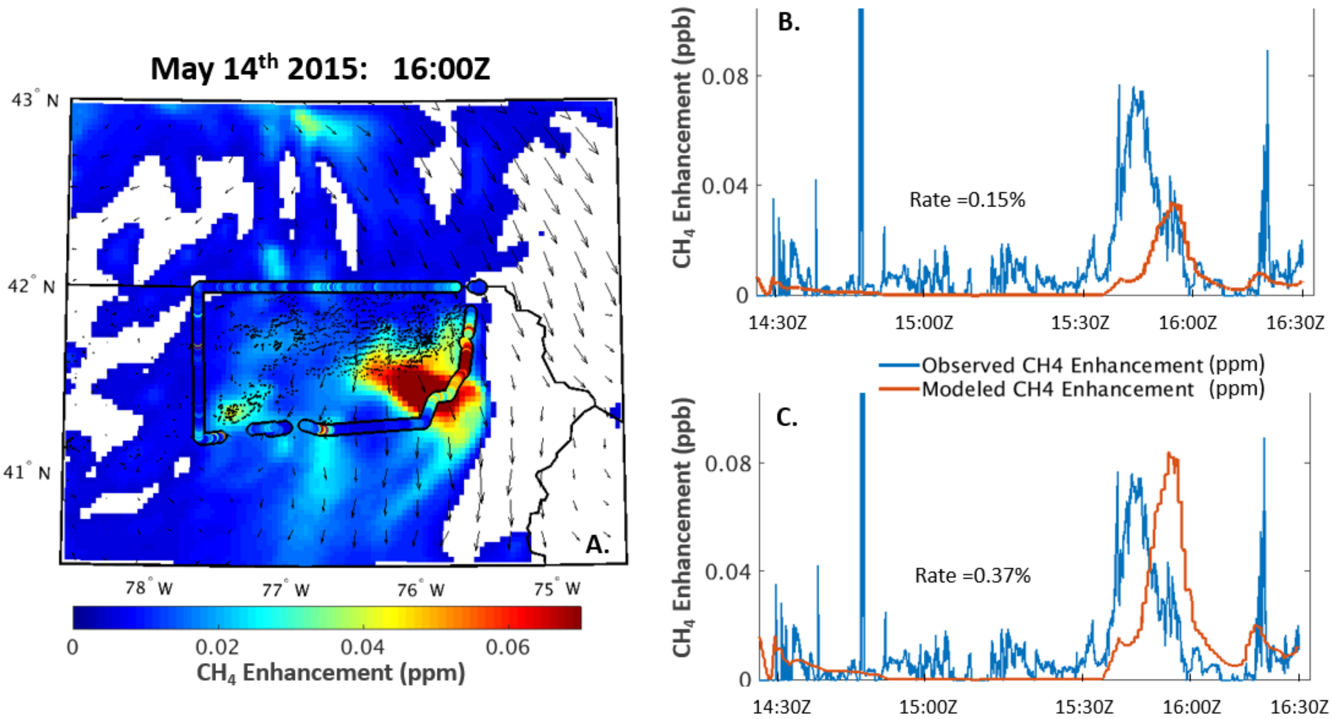

Figure 4: (a.) Observed vs model projected CH_4 enhancements during the May 14th, 2015 at 16Z. (b.) Comparison of observed natural gas enhancement to modelled natural gas enhancement along flight path, with upstream emission rate optimized by minimizing the absolute error between the datasets. (c.) Same as previous, but optimized by minimizing the sum of the error between the datasets.

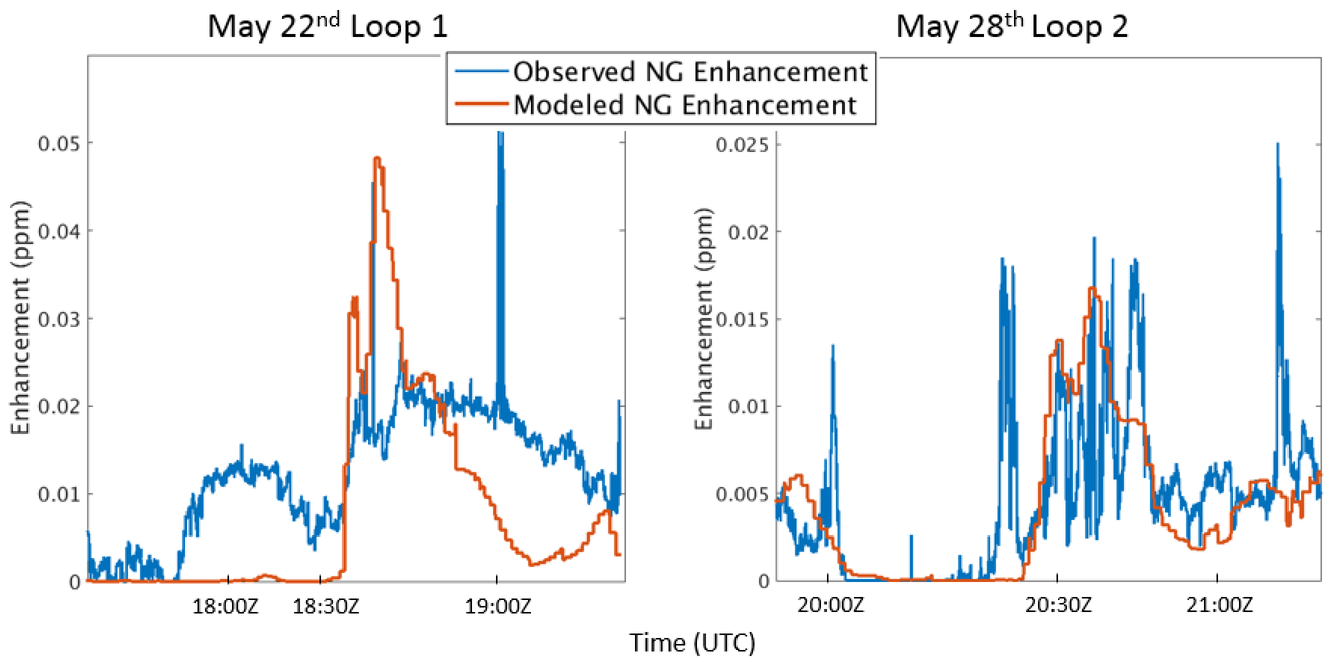

Figure 5: Comparison of observed natural gas enhancement to modelled natural gas enhancement for segments along the (left) May 22nd flight and (right) May 28th flight. A distinct lack of representativeness of the observations in the modelled enhancement can be seen in the May 22nd flight compared to the May 28th flight.

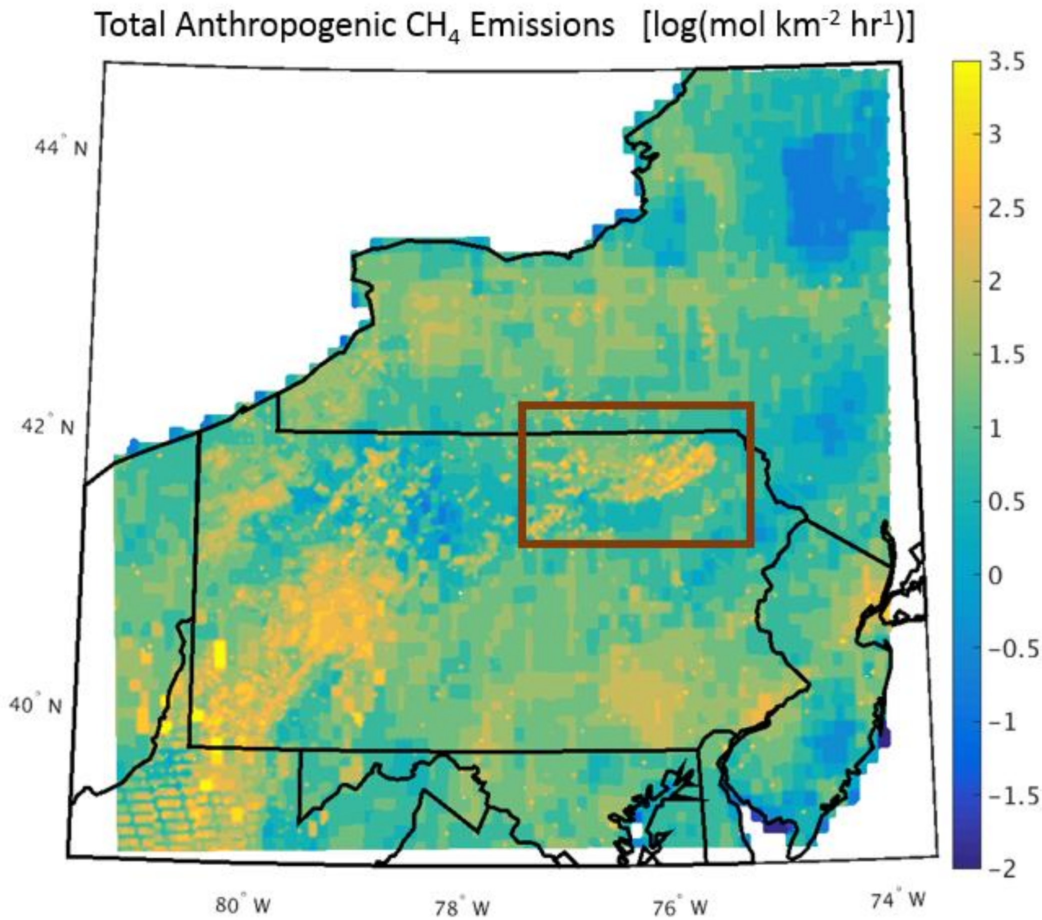

Figure 6: A log scale contour of the anthropogenic CH_4 emissions inventory from this study used within the transport model. The red rectangle surrounds the study region where the aircraft campaign took place.

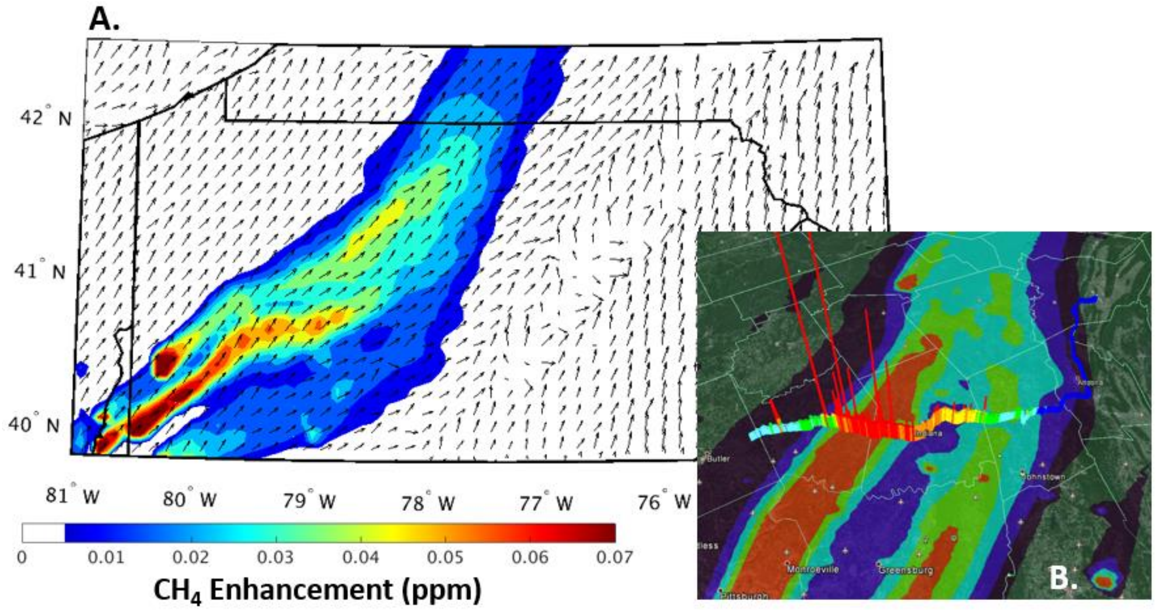

Figure 7: (a.) Model projected CH_4 enhancement at the surface associated with underground, surface. and abandoned coal mines on May 27th, 2015 at 19Z, with the shaded regions showing the CH_4 enhancement and the arrows representing the wind direction. (b.) Projected enhancement from a. mapped over measured CH_4 enhancement from a driving campaign. The height and colour of the bars represents the scale of the CH_4 enhancement.

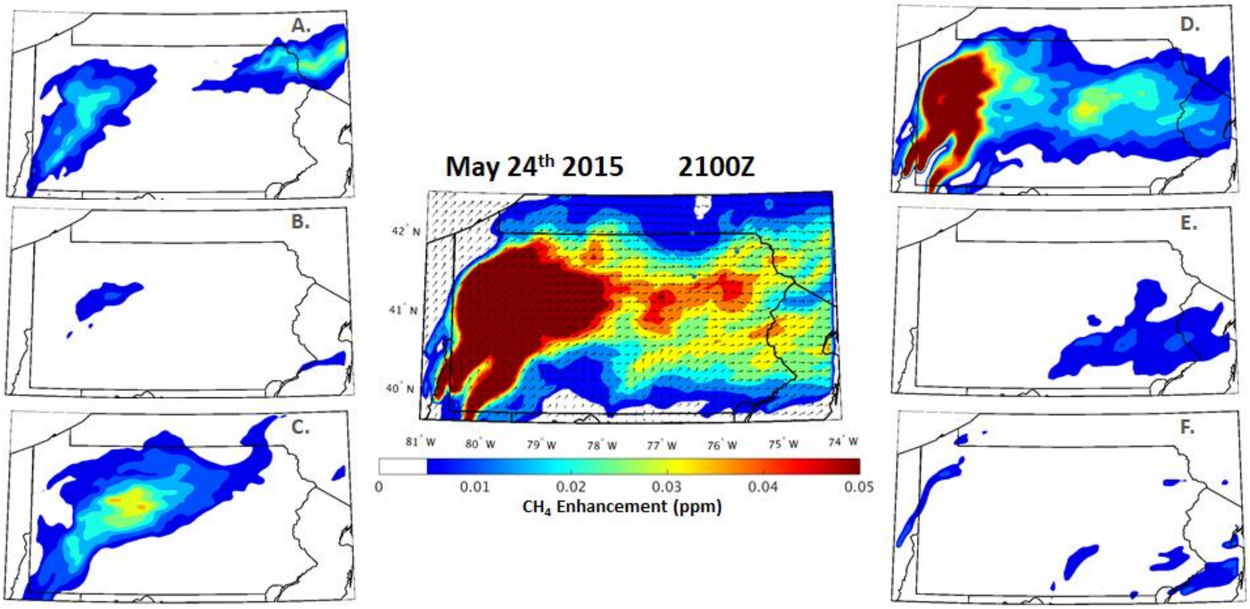

Figure 8: Projected CH_4 enhancements during the late afternoon flight of May 24th, 2015 at 2100Z, 700m above ground level from (A) upstream unconventional gas processes (B) downstream unconventional gas processes (C) conventional production (D) coal mines (E) animal emissions and (F) landfills and other sources within the EPA GHG Inventory Report. The centre figure is a map of the combined enhancement from sources A-F.

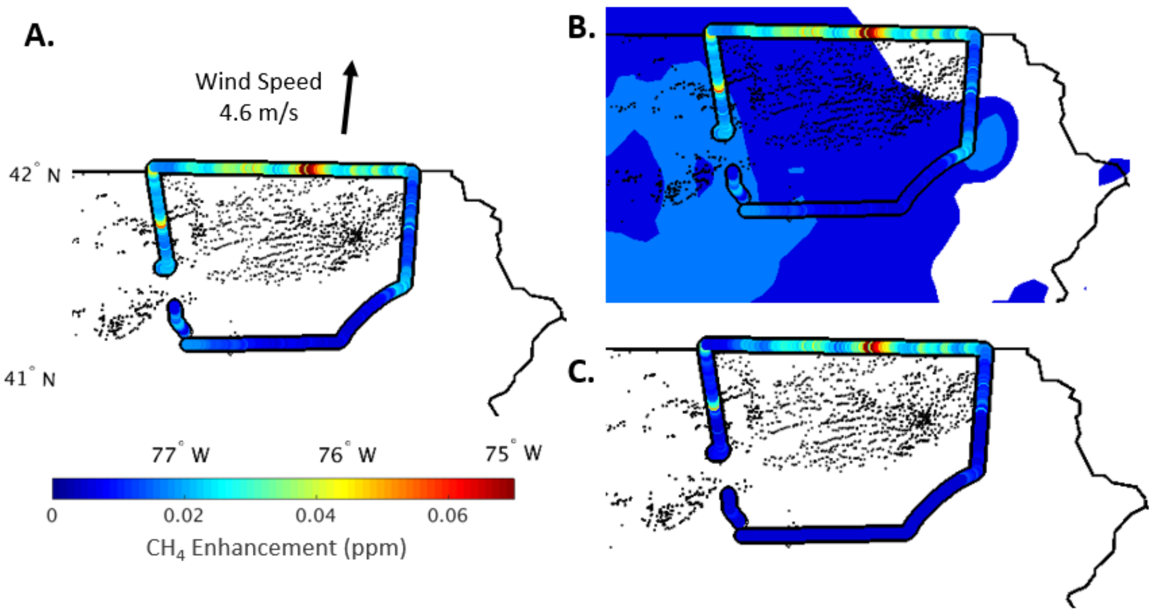

Figure 9: (a.) Observed CH_4 enhancements from within the boundary layer during the first loop of the May 29th aircraft campaign. (b.) Aircraft observations laid overtop modelled CH_4 concentrations at 700 m from sources unrelated to emissions from upstream gas production. (c.) Observed CH_4 enhancements from the May 29th flight after subtracting off modelled sources in b. The new set of observations represent the observation-derived upstream gas enhancement during the flight.

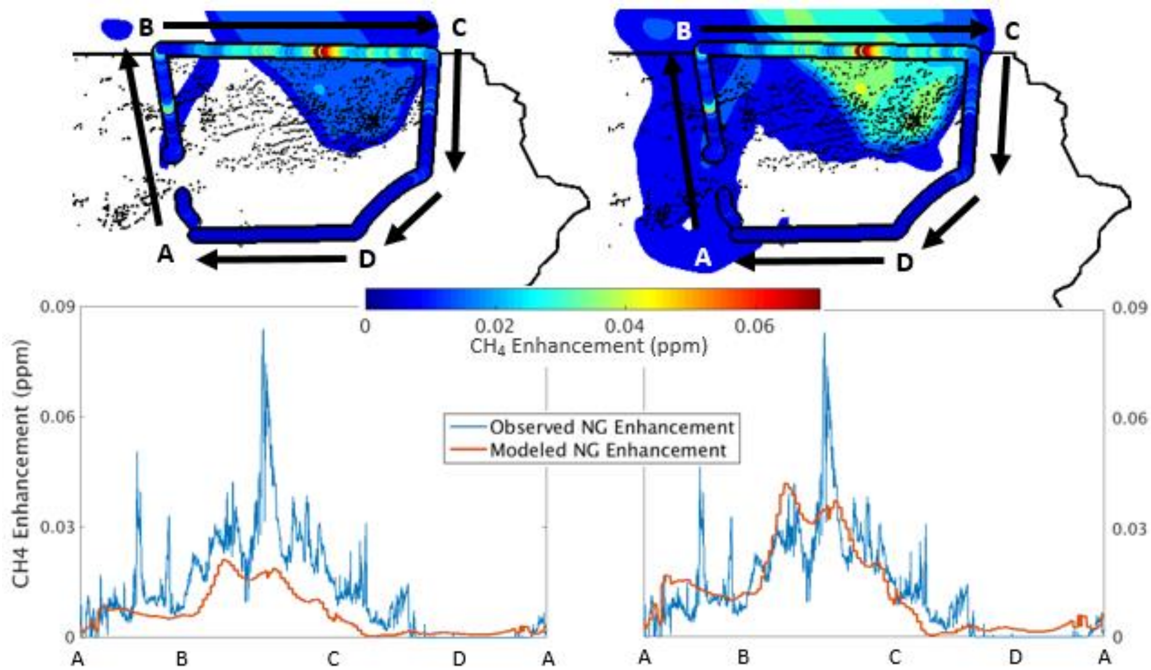

Figure 10: (top-left) Observed enhancement from unconventional natural gas production overtop projected upstream natural gas enhancements at 700 m from the first loop of the May 29th flight, using an upstream gas emission rate of 0.13% of production. (bottom-left) Direct comparison of the observed natural gas enhancement vs. the modelled enhancement following the path from A-D using an unconventional emission rate of 0.13%. (top-right, bottom-right). Same as left figures, except using the optimized upstream emission rate of 0.26%

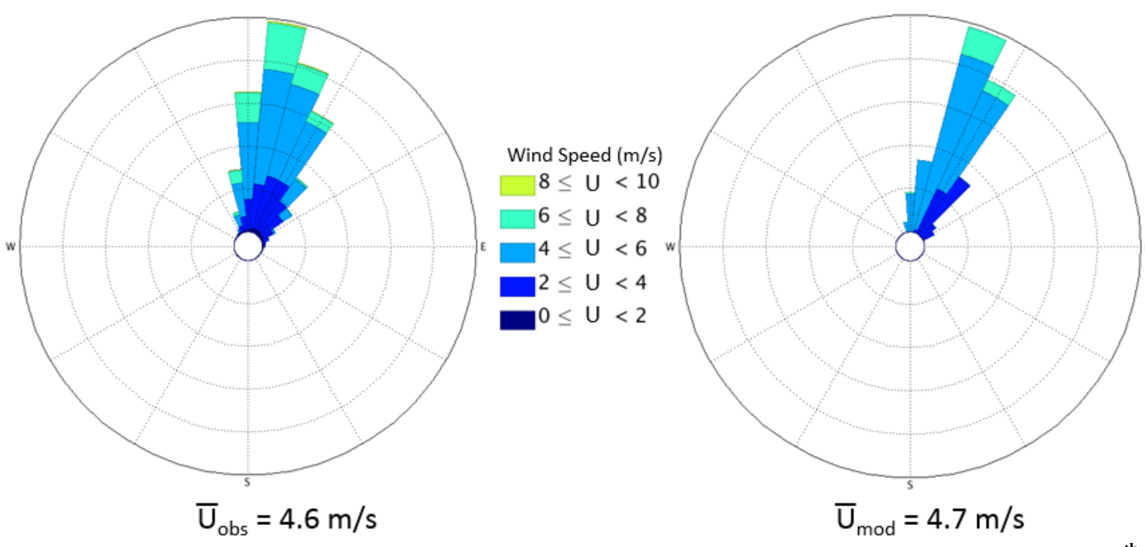

Figure 11: Wind rose of aircraft observations (left) within the boundary from the first loop of the May 29th flight compared to modelled winds following the path (right).

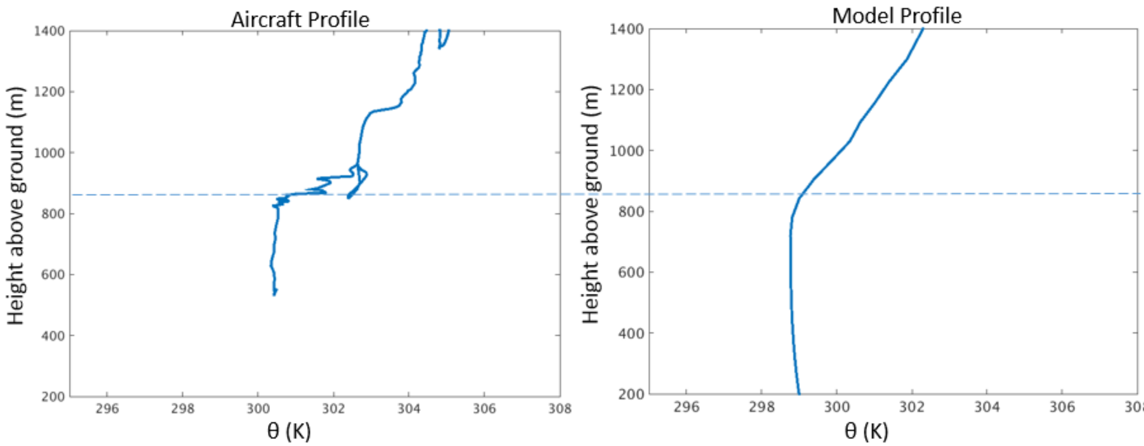

Figure 12: (left) Observed potential temperature profile with height from the first aircraft spiral on the May 29th flight at 17Z. (right) Modelled potential temperature at the location and time at which the aircraft spiral occurred. In both cases, an inversion in the potential temperature profile begins to occur around 850m.

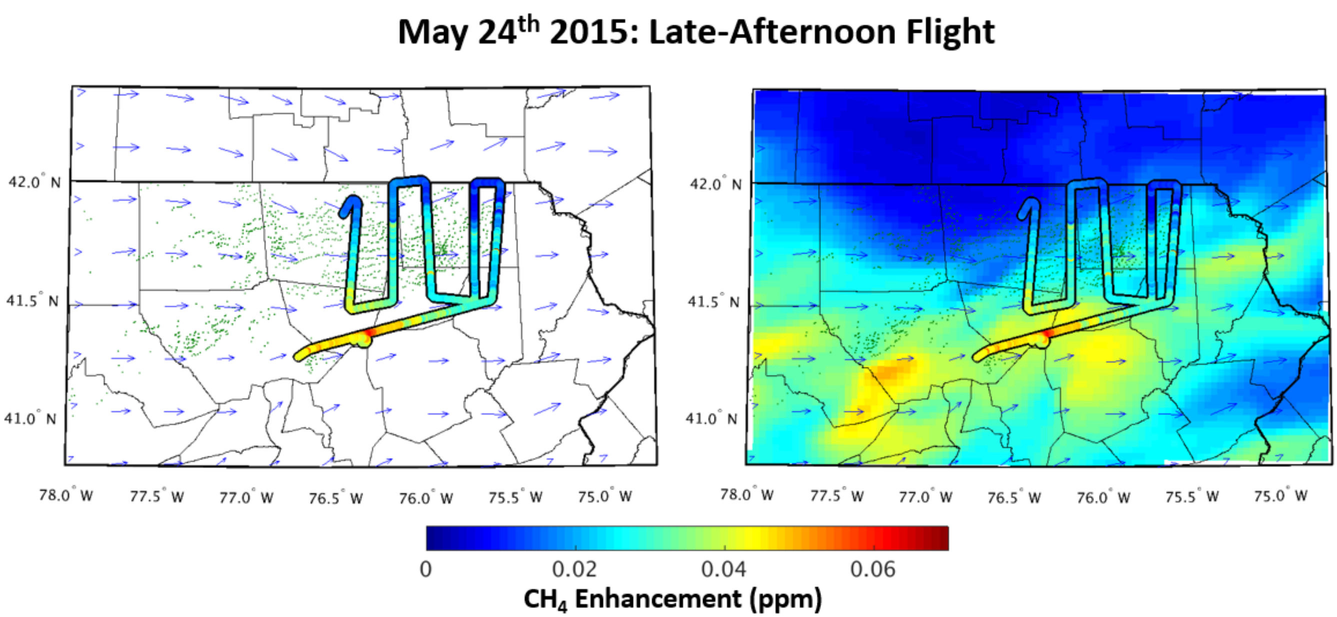

Figure 13: (left) Observed CH_4 enhancement from the late-afternoon flight on May 24th, 2015. (right) Observed CH_4 enhancement compared to the model projected CH_4 enhancement from the sum of all sources in the region. The colour scale of observed and projected enhancements is scaled 1:1, with matching colours indicating matching values. Modelled wind vectors and CH_4 concentrations are from 700 m model height level.

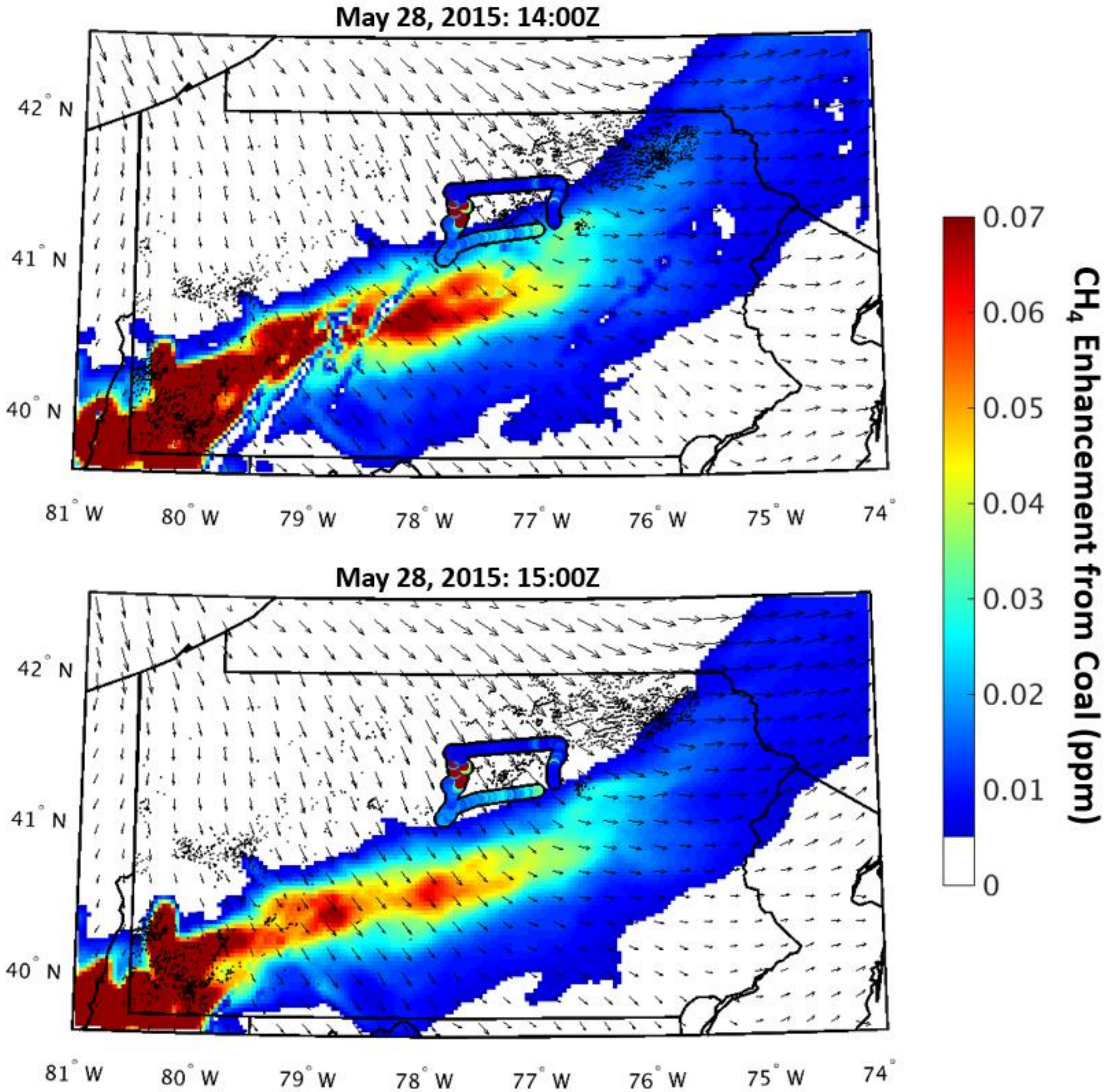

Figure 14: Observed CH4 enhancements from an early flight on May 28th, 2015 compared to projected CH4 enhancements from coal emissions modelled at (top) 14:00Z and (bottom) 15:00Z. The one hour time difference results in vastly different projected enhancements across the southern portion of observations. Modelled wind vectors and CH_4 concentrations are from the 700 m model height level.

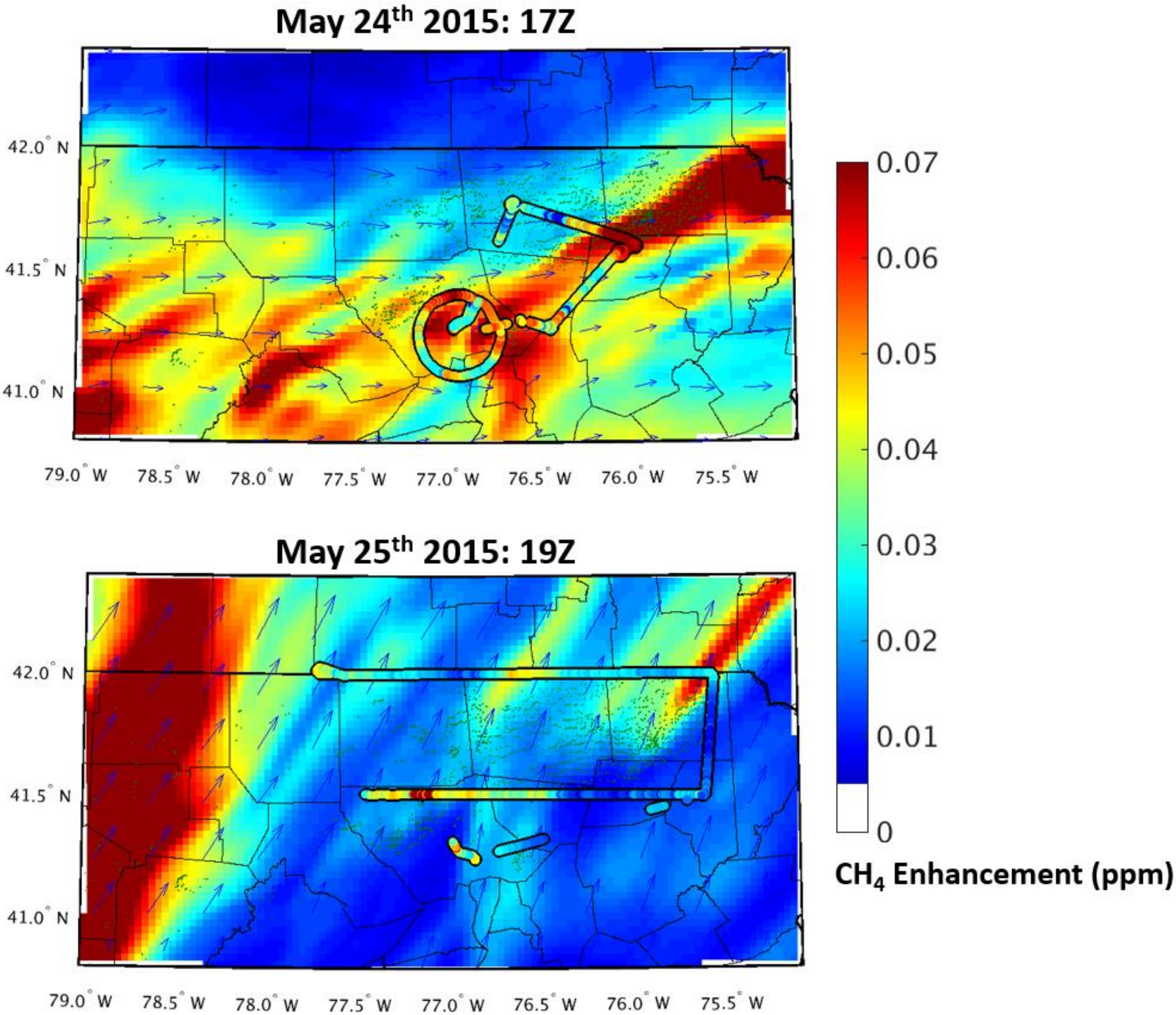

Figure 15: Observed vs model projected CH_4 enhancements during (top) the early afternoon flight of May 24th, 2015 at 17Z and (bottom) the flight of May 25th, 2016 at 19Z. Modelled wind vectors and CH_4 concentrations are from 700 m model height level.

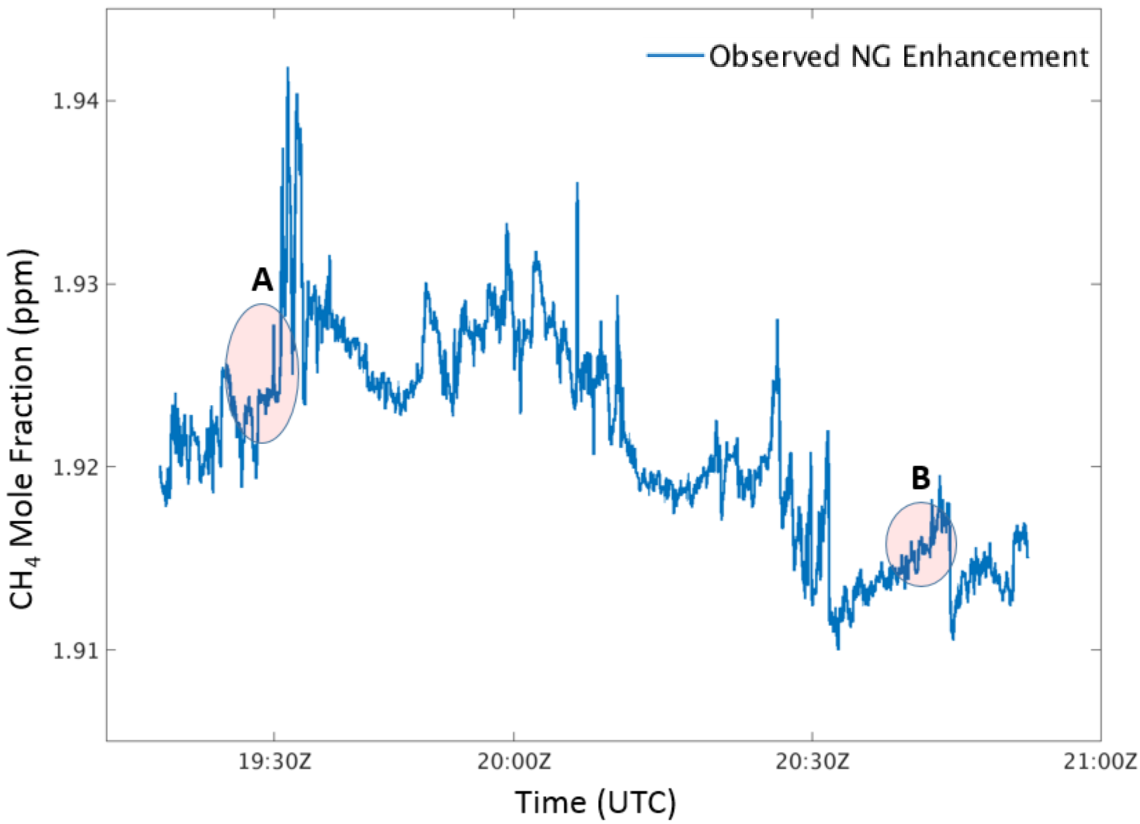

Figure 16: Time series of CH_4 mole fractions from the second loop of the May 22nd flight. Observations at the shaded areas below A and B were taken at similar locations in space, showing the change in the background mole fraction across time.

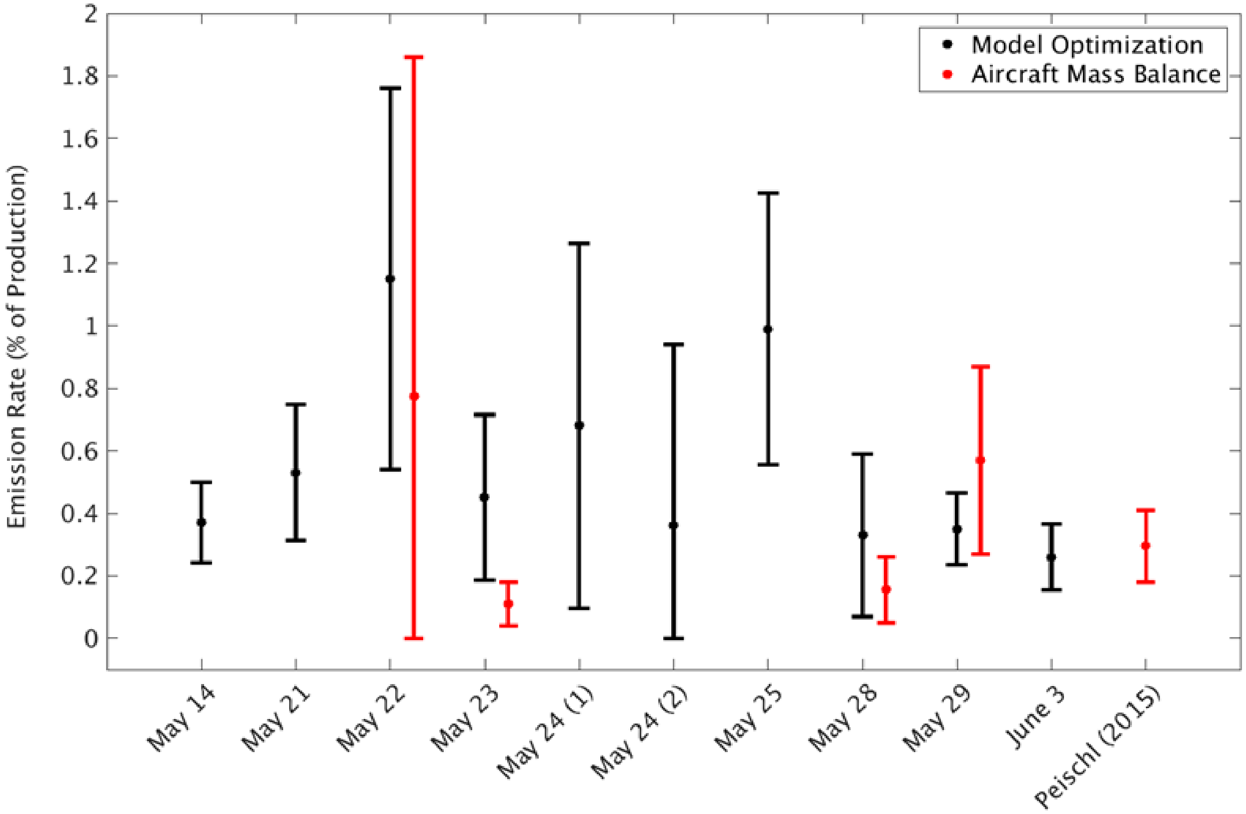

Figure 17: Calculated upstream natural gas emission rates using (black) model optimization technique and (red) aircraft mass balance technique. Error bars represent the 2σ confidence interval for each flight. Mass balance performed in Peischl et al (2015) included for comparison.

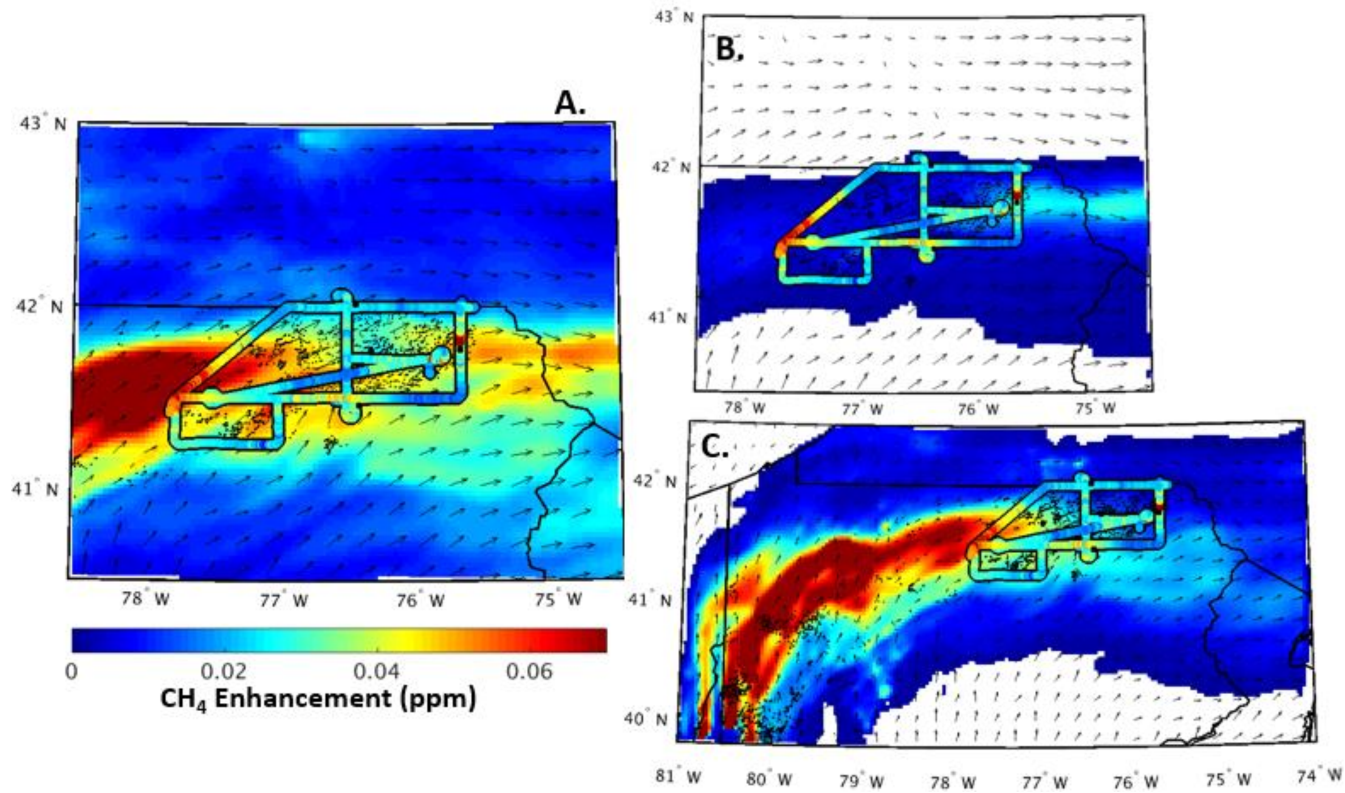

Figure 18: Observations vs modelled enhancements of the flight from Peischl et. al (2015) for July 6th, 2013. (a.) Observed enhancements from the flight over model projected enhancements from all sources at 21Z. (b.) Projected enhancement from upstream gas processes using a 0.4% emission rate. (c.) Projected enhancement from coal sources in southwestern PA. Modelled wind vectors and CH_4 concentrations are from 700 m model height level.