

# Source apportionment of methane and nitrous oxide in California's San Joaquin Valley at CalNex 2010 via positive matrix factorization

A. Guha<sup>1\*</sup>, D. R. Gentner<sup>2†</sup>, R. J. Weber<sup>1</sup>, R. Provencal<sup>3</sup>, and A. H. Goldstein<sup>1,2</sup>

<sup>1</sup> Department of Environmental Science, Policy and Management, University of California, Berkeley, USA

<sup>2</sup> Department of Civil and Environmental Engineering, University of California, Berkeley, USA

<sup>3</sup> Los Gatos Research Inc., Mountain View, California, USA

\*now at: Planning and Climate Protection Division, Bay Area Air Quality Management District, San Francisco, CA, USA

†now at: Department of Chemical & Environmental Engineering, Yale University, New Haven, CT, USA

Correspondence to: A. Guha (aguha@baaqmd.gov)

## Abstract

Sources of methane (CH<sub>4</sub>) and nitrous oxide (N<sub>2</sub>O) were investigated using measurements from a site in southeast Bakersfield as part of the CalNex (California at the Nexus of Air Quality and Climate Change) experiment from mid-May to June-end, 2010. Typical daily minimum mixing ratios of CH<sub>4</sub> and N<sub>2</sub>O were higher than daily minima that were simultaneously observed at a mid-oceanic background station (NOAA, Mauna Loa) by approximately 70 ppb and 0.5 ppb, respectively. Substantial enhancements of CH<sub>4</sub> and N<sub>2</sub>O (hourly averages > 500 ppb and > 7 ppb, respectively) were routinely observed suggesting the presence of large regional sources. Collocated measurements of carbon monoxide (CO) and a range of volatile organic compounds (VOCs) (e.g. straight-chain and branched alkanes, cycloalkanes, chlorinated alkanes, aromatics, alcohols, isoprene, terpenes and ketones) were used with a Positive Matrix Factorization (PMF) source apportionment method to estimate the contribution of regional sources to observed enhancements of CH<sub>4</sub> and N<sub>2</sub>O.

The PMF technique provided a “top-down” deconstruction of ambient gas-phase observations into broad source categories, yielding a 7-factor solution. We identified these source factors as emissions from evaporative and fugitive; motor vehicles; livestock and dairy; agricultural and soil management; daytime light and temperature driven; non-vehicular urban; and nighttime terpene biogenics and anthropogenics. The dairy and livestock factor accounted for a majority of the CH<sub>4</sub> (70 -

90%) enhancements during the duration of experiments. The dairy and livestock factor was also a principal contributor to the daily enhancements of  $\text{N}_2\text{O}$  (60–70%). Agriculture and soil management accounted for ~20–25% of  $\text{N}_2\text{O}$  enhancements over a 24-hour cycle, not surprisingly given that organic and synthetic fertilizers are known to be a major source of  $\text{N}_2\text{O}$ . The  $\text{N}_2\text{O}$  attribution to the agriculture and soil management factor had a high uncertainty in the conducted bootstrapping analysis. This is most likely due to an asynchronous pattern of soil-mediated  $\text{N}_2\text{O}$  emissions from fertilizer usage and collocated biogenic emissions from crops from the surrounding agricultural operations that is difficult to apportion statistically when using PMF. The evaporative / fugitive source profile, which resembled a mix of petroleum operation and non-tailpipe evaporative gasoline sources, did not include a PMF resolved- $\text{CH}_4$  contribution that was significant (<2%) compared to the uncertainty in the livestock-associated  $\text{CH}_4$  emissions. The uncertainty of the  $\text{CH}_4$  estimates in this source factor, derived from the bootstrapping analysis, is consistent with the ~3% contribution of fugitive oil and gas emissions to the statewide  $\text{CH}_4$  inventory. The vehicle emission source factor broadly matched VOC profiles of on-road exhaust sources. This source factor had no statistically significant detected contribution to the  $\text{N}_2\text{O}$  signals (confidence interval of 3% of livestock  $\text{N}_2\text{O}$  enhancements) and negligible  $\text{CH}_4$  (confidence interval of 4% of livestock  $\text{CH}_4$  enhancements) in the presence of a dominant dairy and livestock factor. The CalNex PMF study provides a measurement-based assessment of the state  $\text{CH}_4$  and  $\text{N}_2\text{O}$  inventories for the southern San Joaquin valley. The state inventory attributes ~18% of total  $\text{N}_2\text{O}$  emissions to the transportation sector. Our PMF analysis directly contradicts the state inventory and demonstrates there were no discernible  $\text{N}_2\text{O}$  emissions from the transportation sector in the southern SJV region.

## 1. Introduction

Methane ( $\text{CH}_4$ ) and nitrous oxide ( $\text{N}_2\text{O}$ ) are the two most significant non- $\text{CO}_2$  greenhouse gases (GHGs) contributing about 50% and 17% of the total direct non- $\text{CO}_2$  GHG radiative forcing ( $\sim 1 \text{ W m}^{-2}$ ), respectively (Figure SPM.5; IPCC, 2013).  $\text{CH}_4$ , with a lifetime of ~10 years and Global Warming Potential (GWP) of 34 on a 100-year basis, accounting for climate-carbon feedbacks (Table 8.7, Myhre et al., 2013; Montzka et al., 2011), is emitted by both anthropogenic and natural sources (e.g. wetlands, oceans, termites etc.). Anthropogenic global  $\text{CH}_4$  emissions are due to agricultural activities (enteric fermentation in livestock, manure management and rice cultivation) (McMillan et al., 2007; Owen and Silver, 2014), energy sector (oil and gas operations and coal mining), waste management (landfills and waste water treatment), and biomass burning (some of which is natural) (Smith et al., 2007; NRC, 2010).  $\text{N}_2\text{O}$  has a higher persistence in the atmosphere (lifetime of ~120 years) and stronger infrared radiation

absorption characteristics than CH<sub>4</sub> giving it a GWP of 298 (Table 8.7, Myhre et al., 2013; Montzka et al., 2011). Agriculture is the biggest source of anthropogenic N<sub>2</sub>O emissions since the use of synthetic fertilizers and manure leads to microbial N<sub>2</sub>O emissions from soil (Crutzen et al., 2007; Galloway et al., 2008). Management of livestock and animal waste is another important biological source of N<sub>2</sub>O, while industrial processes including fossil fuel combustion have been estimated to account for ~10% of total global anthropogenic N<sub>2</sub>O emissions (Denman et al., 2007).

In 2006, the state of California adopted Assembly Bill 32 (AB32) into a law known as the Global Warming Solutions Act, which committed the state to cap and reduce anthropogenic GHG emissions to 1990 levels by 2020. A statewide GHG emission inventory (CARB, 2013) maintained by the Air Resources Board of California (CARB) is used to report, verify and regulate emissions from GHG sources. In 2011, CH<sub>4</sub> accounted for 32.5 million metric tonnes (MMT) CO<sub>2</sub>-eq representing 6.2% of the statewide GHG emissions, while N<sub>2</sub>O emissions totaled 13.4 MMT CO<sub>2</sub>-eq representing about 3% of the GHG emissions inventory (Figure 1). CARB's accurate knowledge of GHG sources and statewide emissions is a key component to the success of any climate change mitigation strategy under AB32. CARB's GHG inventory is a "bottom-up" summation of emissions derived from emission factors and activity data. The bottom-up approach is reasonably accurate for estimation and verification of emissions from mobile and point sources (vehicle tailpipes, power plant stacks etc.) where the input variables are well-understood and well-quantified. The main anthropogenic sources of CH<sub>4</sub> in the CARB inventory are ruminant livestock and manure management, landfills, wastewater treatment, fugitive and process losses from oil and gas production and transmission, and rice cultivation while the major N<sub>2</sub>O sources are agricultural soil management, livestock manure management and vehicle fuel combustion (CARB, 2013). The emission factors for many of these sources have large uncertainties as they are biological in nature and their production and release mechanisms are inadequately understood thus making these sources unsuitable for direct measurements (e.g. emissions of N<sub>2</sub>O from farmlands). Many of these sources (e.g. CH<sub>4</sub> from landfills) are susceptible to spatial heterogeneity and seasonal variability. Unfortunately, a more detailed understanding of source characteristics is made difficult because CH<sub>4</sub> and N<sub>2</sub>O are often emitted from a mix of point and area sources within the same source facility (e.g. dairies in the agricultural sector) making bottom-up estimation uncertain. There is a lack of direct measurement data or "top-down" measurement-based approaches to independently validate seasonal trends and inventory estimates of CH<sub>4</sub> and N<sub>2</sub>O in California's Central Valley, which has a mix of several agricultural sources and oil and gas operations, both of which are known major sources of GHGs.

In the recent past, regional emission estimates derived from measurements from a tall tower at Walnut Grove in Central California coupled with inverse dispersion techniques (Fischer et al., 2009) reported underestimation of CH<sub>4</sub> and N<sub>2</sub>O emissions especially in the Central Valley. Comparison of regional surface footprints determined from WRF-STILT algorithm between Oct-Dec 2007 indicate posterior CH<sub>4</sub> emissions to be higher than California-specific inventory estimates by 37 ± 21% (Zhao et al., 2009). Predicted livestock CH<sub>4</sub> emissions are 63 ± 22% higher than *a priori* estimates. A study over a longer period (Dec 2007 - Nov 2008) at the same tower (Jeong et al., 2012a) generated posterior CH<sub>4</sub> estimates that were 55 - 84% larger than California-specific prior emissions for a region within 150 km from the tower. For N<sub>2</sub>O, inverse estimates for the same sub-regions (using either EDGAR32 and EDGAR42 *a priori* maps) were about twice as much as *a priori* EDGAR inventories (Jeong et al., 2012b). Recent studies have incorporated WRF-STILT inverse analysis on airborne observations across California (Santoni et al., 2012). The authors conclude that CARB CH<sub>4</sub> budget is being underestimated by a factor of 1.64 with aircraft-derived emissions from cattle and manure management, landfills, rice, and natural gas infrastructure being around 75%, 22%, 460%, and 430% more than CARB's current estimates for these categories, respectively. Statistical source footprints of CH<sub>4</sub> emissions generated using FLEXPART-WRF modeling and CalNex-Bakersfield CH<sub>4</sub> concentration data are consistent with locations of dairies in the region (Gentner et al., 2014a). The authors conclude that the majority of CH<sub>4</sub> emissions in the region originate from dairy operations. Scaled-up CH<sub>4</sub> rice cultivation estimates derived from aircraft CH<sub>4</sub>/CO<sub>2</sub> flux ratio observations over rice paddies in the Sacramento valley during the growing season when emissions are at their strongest (Peischl et al., 2012) are around three times larger than inventory estimates. CH<sub>4</sub> budgets derived for the Los Angeles (LA) basin from aircraft observations (Peischl et al., 2013) and studies involving comparison with CO enhancements and inventory at Mt. Wilson (Hsu et al., 2010; Wunch et al., 2009) indicate higher atmospheric CH<sub>4</sub> emissions in the LA basin than expected from bottom-up accounting.

Recent literature seems to suggest that the CARB bottom-up inventory is underestimating CH<sub>4</sub> and N<sub>2</sub>O sources, especially from the livestock sector and perhaps from the oil and gas industry as well. Source apportionment studies of non-CO<sub>2</sub> GHGs over the Central Valley can provide critical information about under-inventoried or unknown sources that seek to bridge the gap between 'bottom-up' and 'top-down' methods. GHG emission inventories can potentially be constrained through simultaneous measurements of GHGs and multiple gas species (VOCs) that are tracers of various source categories. This study provides CH<sub>4</sub> and N<sub>2</sub>O source attribution during a six-week study involving a complete suite of continuous GHG and VOC tracer measurements during the CalNex 2010 campaign in Bakersfield which is

located in the southern part of the Central Valley (May - June 2010). The objective of this study is to partition the measured  $\text{CH}_4$ ,  $\text{N}_2\text{O}$  and VOC enhancements into statistically unique combinations using Positive Matrix Factorization (PMF) apportionment technique. We classify these combinations as plausible source factors based on our prior knowledge of the chemical origin of mutually co-varying groups of VOC tracers found in each statistical combination. We examine the source categorization using observations from source-specific, ground site and airborne measurements and results from other source apportionment studies. We also compare the relative abundance of  $\text{CH}_4$  and  $\text{N}_2\text{O}$  enhancements in each source factor with the CARB inventory estimates in order to assess the accuracy of the inventory. We hypothesize that the PMF analysis will be able to parse the atmospheric observations into unique statistical source combinations that, as analysts, we would be able to distinguish from each other on the basis of unique VOC source markers. We should, thus, be able to appropriately attribute the  $\text{CH}_4$  and  $\text{N}_2\text{O}$  apportioned to each of these factor profiles to a major source category. We then proceed to answer the scientific question if our top down assessment of the  $\text{CH}_4$  and  $\text{N}_2\text{O}$  inventory can improve our understanding of the bottom-up CARB inventory in the region.

## **2. Experimental Setup**

### **2.1. Field Site and Meteorology**

Measurements were conducted from 19 May to 25 June 2010 at the Bakersfield CalNex supersite (35.3463°N, 118.9654°W) (Figure 2) in the southern San Joaquin Valley (SJV) (Ryerson et al., 2013). The SJV represents the southern half of California's Central Valley. It is 60 to 100 km wide, surrounded on three sides by mountains, with the Coastal Ranges to the west, the Sierra Nevada Mountains to the east, and the Tehachapi Mountains to the southeast.

The measurement site was located to the southeast of the Bakersfield urban core in Kern County (Figure 2). The east-west Highway 58 is located ~ 0.8 km north; the north-south Highway 99 is ~ 7 km west. The city's main waste water treatment plant (WWTP) and its settling ponds are located east and south (< 2.5 km), respectively. Numerous dairy and livestock operations are located south-southwest at 10 km distance and onwards. The metropolitan region has three major oil refineries located within 10 km (two to the northwest; one to the southeast). A majority of Kern County's high-production active oil fields (> 10000 barrels (bbl) per day) (CDC, 2013) are located to the west/northwest and are distant (~ 40 - 100 km). Kern River oilfield (~ 60000 bbl day<sup>-1</sup>), one of the largest in the country, and Kern Front (~ 11000 bbl day<sup>-1</sup>) are located about 10 - 15 km to the north. There are several other oil

fields dotted within the urban core (5 - 20 km) which are less productive ( $< 2000 \text{ bbl day}^{-1}$ ) or not active ( $< 100 \text{ bbl day}^{-1}$ ). The whole region is covered with agricultural farmlands with almonds, grapes, citrus, carrots and pistachios amongst the top commodities by value and acreage (KernAg, 2010).

The meteorology and transport of air masses in the southern SJV is complex and has been addressed previously (Bao et al., 2007; Beaver and Palazoglu, 2009). The wind rose plots (Figure 3) shown here present a simplified distribution of microscale wind speed and direction at the site for the campaign duration, the latter often being non-linear over larger spatial scales. The plots depict broad differences in meteorology during daytime and nighttime. A mesoscale representation of the site meteorology during this study period was evaluated through back-trajectory footprints generated from each hourly sample using FLEXPART Lagrangian transport model with WRF meteorological modeling (Gentner et al., 2014a). The 6-h and 12-h back trajectory footprints are generated on a  $4 \times 4 \text{ km}$  resolution with simulations originating from top of the 18-m tall tower. The site experiences persistent up-valley flows from the north and northwest during afternoons and evenings, usually at high wind speeds. The direction and speed of the flow during nights is quite variable (Figure 3). On some nights, the up-valley flows diminish as night-time inversion forms a stable layer near the ground, and eventually downslope flows off the nearby mountain ranges bring winds from the east and south during late night and early morning periods. On other nights, fast moving northwesterly flows extend in to middle of the night leading to unstable conditions through the night. The daytime flows bring plumes from the up wind metropolitan region (Figure 3), and regional emissions from sources like dairies and farmlands located further upwind. The slow nighttime flows and stagnant conditions cause local source contributions to be more significant than during daytime, including those from nearby petroleum operations and dairies (Gentner et al., 2014a), and agriculture (Gentner et al., 2014b).

### **3. Methods**

#### **3.1. Trace gas measurements and instrumentation**

Ambient air was sampled from the top of a tower (18.7 m a.g.l) through Teflon inlet sampling lines with Teflon filters to remove particulate matter from the gas stream.  $\text{CH}_4$ ,  $\text{CO}_2$  and  $\text{H}_2\text{O}$  were measured using a Los Gatos Research (LGR Inc., Mountain View, CA) Fast Greenhouse Gas Analyzer (FGGA, Model 907-0010).  $\text{N}_2\text{O}$  and  $\text{CO}$  were measured by another LGR analyzer (Model 907-0015) with time response of  $\sim 0.1$  to  $0.2 \text{ Hz}$ . These instruments use off-axis Integrated Cavity Output Spectroscopy (ICOS) (O'Keefe, 1998; Paul et al., 2002; Hendriks et al., 2008; Parameswaran et al., 2009). The FGGA

instrument internally calculates and automatically applies a water vapor correction to counter the dilution effect of water on a target molecule and calculates CH<sub>4</sub> and CO<sub>2</sub> on a dry (and wet) mole fraction basis. We report dry mole fraction mixing ratios in this study. The FGGA instrument had a 1 $\sigma$ -precision of 1 ppb (for CH<sub>4</sub>) and 0.15 ppm (for CO<sub>2</sub>) while the N<sub>2</sub>O/CO instrument had a 1 $\sigma$ -precision of 0.3 ppb, respectively over short time periods (< 10 s). Prior to the campaign, the precision of measurements of each instrument used in this study were determined as the 1-sigma standard deviation of a data set over a given length of time measuring a fixed standard (scuba tank) and found to conform to the manufacturer specifications. The instruments were housed at ground level in a thermally insulated temperature controlled 7-foot wide cargo wagon trailer developed by the instrument manufacturers (Los Gatos Research Inc.). CO was coincidentally measured using another instrument (Teledyne API, USA, Model # M300EU2) with a precision of 0.5% of reading and output as 1-minute averages. The mixing ratios from the two collocated CO instruments correlated well ( $r \sim 0.99$ ) and provided a good stability check for the LGR instrumentation. Scaled Teledyne CO data was used to gap-fill the LGR CO data. The coincident gas-phase VOC measurements were made using a gas chromatograph (GC) with a quadrapole mass selective detector and a flame ionization detector (Gentner et al., 2012).

Hourly calibration checks of the three GHGs and CO were performed using near-ambient level scuba tank standards through the entire campaign. The scuba tanks were secondary references and were calibrated before and after the experiments using primary standards conforming to the WMO mole fraction scale obtained from the Global Monitoring Division (GMD) at the NOAA Earth System Research Lab. The calibration tests confirmed that there was no issue in short term stability of these species. During data processing, final concentrations were generated from the raw data values using scaling factors obtained from comparison of measured and target concentrations during secondary calibration checks. Diurnal plots of measured species are generated from 1-min averages. PMF analyses in the following sections are based on 30-minute averages to match the time resolution of VOC measurements. The meteorological data measured at the top of the tower included relative humidity (RH), temperature (T), and wind speed (WS) and direction (WD).

### 3.2. Positive Matrix Factorization (PMF)

Source apportionment techniques like PMF have been used in the past to apportion ambient concentration datasets into mutually co-varying groups of species. PMF is especially suitable for studies where *a priori* knowledge of contributing number of sources impacting the measurements, chemical

nature of source profiles and relative contribution of each source to the concentration time series of a measured compound are unknown or cannot be assumed. PMF has been applied to ambient particulate matter studies (Lee et al., 1999; Kim et al., 2004); in determining sources of atmospheric organic aerosols (OA) (Ulbrich et al., 2009; Slowik et al., 2010; Williams et al., 2010); and in gas phase measurements of VOCs in major metropolitan cities (Brown et al., 2007; Bon et al., 2011). PMF is a receptor-only unmixing model which breaks down a measured data set containing time series of a number of compounds into a mass balance of an arbitrary number of constant source factor profiles (FP) with varying concentrations over the time of the data set (time series or TS) (Ulbrich et al., 2009).

In real world ambient scenarios, emissions sources are often not known or well-understood. PMF technique requires no a priori information about the number or composition of factor profiles or time trends of those profiles. The constraint of non-negativity in PMF ensures that all values in the derived factor profiles and their contributions are constrained to be positive leading to physically meaningful solutions. PMF requires the user to attribute a measure of experimental uncertainty (or weight) to each input measurement. Data point weights allow the level of influence to be related to the level of confidence the analyst has in the measured data (Hopke, 2000). In this way, problematic data such as outliers, below-detection-limit (BDL), or altogether missing data can still be substituted into the model with appropriated weight adjustment (Comero et al., 2009) allowing for a larger input data set, and hence a more robust analysis. PMF results are quantitative; it is possible to obtain chemical composition of sources determined by the model (Comero et al., 2009). PMF is not data-sensitive and can be applied to data sets that are not homogenous and/or require normalization without introducing artifacts.

### 3.3. Mathematical Framework of PMF

The PMF model is described in greater detail elsewhere (Paatero and Tapper, 1994; Paatero 1997; Comero et al., 2009; Ulbrich et al., 2009) and we will briefly mention some concepts relevant to the understanding of the analysis carried out in this study. The PMF input parameters involve a  $m \times n$  data matrix  $\mathbf{X}$  with  $i$  rows containing mixing ratios at sampling time  $t_i$  and  $j$  columns containing time series of each tracer  $j$ . A corresponding uncertainty matrix  $\mathbf{S}$  reports measurement precision (uncertainty) of the signal of each tracer  $j$  at every  $t_i$  ( $s_{ij}$ ). The PMF model can then be resolved as:

$$X_{ij} = \sum_p g_{ip} f_{pj} + e_{ij} \quad (1)$$



where  $p$  refers to the number of contributing factors in the solution as determined by the analyst (discussed below),  $g_{ij}$  (mass concentration) are elements of a  $m \times p$  matrix  $\mathbf{G}$  whose columns represent the factor time series while  $f_{ij}$  (mass fraction) are elements of a  $p \times n$  matrix  $\mathbf{F}$  whose rows represent the factor chemical profiles.  $e_{ij}$  are the elements of a  $m \times n$  matrix  $\mathbf{E}$  containing residuals not fit by the model matrix at each data point.

The PMF algorithm uses a least-squares algorithm to iteratively fit the values of  $\mathbf{G}$  and  $\mathbf{F}$  by minimizing a “quality of fit” parameter  $Q$  (Bon et al., 2011), defined as:

$$Q = \sum_{i=1}^m \sum_{j=1}^n (e_{ij}/s_{ij})^2 \quad (2)$$

In this way, PMF minimizes the sum of squares of error-weighted model-measurement deviations. The theoretical value of  $Q$ , denoted by  $Q$ -expected ( $Q_{exp}$ ) can be estimated as:

$$Q_{exp} \equiv (m \times n) - p \times (m + n) \quad (3)$$

If all the errors have been estimated within the uncertainty of the data points (i.e.  $e_{ij} s_{ij}^{-1} \sim 1$ ) and the model fits the data perfectly, then  $Q$  should be approximately equal to  $Q_{exp}$ .

### 3.4. Data preparation for PMF analysis

For this study, measurements from the FGGA, LGR  $\text{N}_2\text{O}/\text{CO}$  analyzer and the GC were combined into a unified data set to create matrices  $\mathbf{X}$  and  $\mathbf{S}$ . Only VOCs that are a part of broad chemical composition of nearby sources (like dairies and vehicle emissions) or could potentially serve as source specific tracers (e.g. iso-octane as a tailpipe emissions tracer; isoprene as a biogenic tracer) were included. Isomers were limited (e.g. 2,3-Dimethylbutane over 2,2-Dimethylbutane) and VOCs with large number of missing values were not included. The input data set represented major chemical families like straight-chain and branched alkanes, cycloalkanes, alkenes, aromatics, alcohols, aldehydes, ketones and chlorinated as well as organosulfur compounds. In spite of best efforts by the authors, it was not possible to quantify the magnitude of observed concentrations of benzene relative to the positive artifacts coming from the Tenax TA adsorbent (previously documented elsewhere). Hence, benzene was not included in the PMF analysis. In all, there were a total of 653 half-hour samples of data collected from 22 May to 25 June. In the days prior to and after this period, there was no  $\text{N}_2\text{O}$  and/or VOC data collected and hence the PMF analysis is limited to this period. Table 1 lists all the compounds included in the PMF analysis along with a spectrum of observed and background concentrations.

PMF analysis resolves the covariance of mixing ratio enhancements and thus characterizes the chemical composition of emissions from various sources. Hence, for this analysis, only enhancements representing local emissions were included in the data set after subtracting local background concentrations from the original signals. Background concentrations were derived as the minima in the time series (0th percentile) for each of the 50 tracers included in the PMF analysis ( $\text{CH}_4$ ,  $\text{N}_2\text{O}$ ,  $\text{CO}$  and 46 VOCs). For VOCs, tracers with a minimum value less than two times the limit of detection (LOD, in ppt) and a maximum value larger than hundred times the LOD were assumed to have a negligible background (0 ppt) (Table 1). The 99th percentile for each tracer was treated as the effective-maximum mixing ratio and the upper limit of the range for the “normalization” of time series. Enhancements above the 99th percentile are often extreme values. Such outliers, even if true enhancements, represent isolated and short-duration footprints of high-emission events that are difficult for PMF to reconstruct. In order to maintain the robustness of PMF analysis, outliers were selectively down-weighted by increasing their uncertainty in proportion to the uncertainty of other data points (described below). Finally, the enhancements in each time series were “normalized” by dividing every sample by the difference in the 99th percentile and background (the range) as seen in Equation 4. This process scaled the enhancements in each time series (final data points in  $\mathbf{X}$ ) within a range of 0 to 1. This allowed for a consistent scheme to represent tracers with vastly different concentrations (e.g. ppm level of  $\text{CH}_4$  vs ppt level of propene) and improve the visual attributes of PMF output plots to follow. Data points denoting zero enhancement (lower limit) were replaced by a very small positive number (i.e.  $\exp(-5)$ ) to avoid ‘zeros’ in the data matrix  $\mathbf{X}$ .

$$x_{ij} = \frac{(\text{Mixing ratio}_{ij} - \text{Background}_j)}{(\text{Maximum mixing ratio}_j - \text{Background}_j)} \quad (4)$$

For the VOCs, guidelines set forth by Williams et al., (2010) were adopted to calculate the uncertainty estimates. An analytical uncertainty (AU) of 10% was used; a limit of detection (LOD) of 1 ppt and a limit of quantification (LOQ) of 2 ppt (Gentner et al., 2012) was used to calculate the total uncertainty for each  $x_{ij}$ :

$$s_{ij} \equiv 2 \times \text{LOD}, \quad \text{if } x_{ij} \leq \text{LOD}, \quad (5a)$$

$$s_{ij} \equiv \text{LOQ}, \quad \text{if } \text{LOD} < x_{ij} \leq \text{LOQ}, \quad (5b)$$

$$s_{ij} \equiv \left( (AU \times x_{ij})^2 + (\text{LOD})^2 \right)^{0.5}, \quad \text{if } x_{ij} > \text{LOQ} \quad (5c)$$

Using this approach, detection limit dictates the errors for low enhancements (near LOD) while errors for larger enhancements of VOCs are tied more to the magnitude of the data value ( $x_{ij}$ ) itself.

The GHG and CO measurements have high precision and significantly lower detection limits than ambient levels. The relatively low values of GHGs in the uncertainty matrix, compared to VOCs, is substituted with those calculated using a custom approach. The GHG and CO uncertainties are assumed to be proportional to the square root of the data value and an arbitrary scaling factor determined through trial and error in order to produce lower values of  $Q Q_{exp}^{-1}$ :

$$s_{ij} \equiv A \times (x_{ij})^{0.5}, \text{ where } A = 1 \text{ (for CH}_4\text{)}, 0.25 \text{ (for CO}_2\text{)}, 0.5 \text{ (for CO)}, 0.1 \text{ (for N}_2\text{O)} \quad (6)$$

This method attributes larger percentage uncertainties to smaller enhancements and hence lesser weight in the final solution and vice versa. This approach leads to an uncertainty matrix that attributes relatively similar percentage errors to both GHGs and VOCs, which should lead to a better fitting of the data through PMF.

Missing values are replaced by geometric mean of the tracer time series and their accompanying uncertainties are set at four times this geometric mean (Polissar et al., 1998) to decrease their weight in the solution. Based on the *a priori* treatment of the entire input data (scaling) and the corresponding outputs of the PMF analysis, a weighting -approach (for measurements from different instruments) as used in (Slowik et al., 2010) is not found to be necessary.

### 3.5. PMF source analysis

We use the customized software tool (PMF Evaluation Tool v2.04, PET) developed by Ulbrich et al. (2009) in Igor Pro (Wavemetrics Inc., Portland, Oregon) to run PMF, evaluate the outputs and generate statistics. The PET calls the PMF2 algorithm (described in detail in Ulbrich et al., 2009) to solve the bilinear model for a given set of matrices **X** and **S** for different numbers of factors  $p$  and for different values of FPEAK or SEED (defined and described later). The tool also stores the results for each of these combinations in a user friendly interface that allows simultaneous display of factor profiles (FP) and time series (TS) of a chosen solution along with residual plots for individual tracers. A detailed explanation of PMF analysis performed in this study is provided in the Supplement (see Section S). The supplement describes the PMF methodology of how the final number of user-defined factors was chosen (Section S1), the outcomes of linear transformations (rotations) of various PMF solutions (Section S2) and how uncertainties in the chosen solution were derived (Section S3). The standard deviations in the mass

fractions of individual tracers in each factor profile and time series of each factor mass is evaluated using a bootstrapping analysis (Norris et al., 2008; Ulbrich et al., 2009) and described in Section S3. The uncertainty of a tracer contribution to a source factor is derived from the 1-sigma deviation of the averaged mass fraction of that tracer to that factor from 100 bootstrapping runs. This is the quantity we report and refer to as ‘uncertainty’ throughout the text in Section 4. The percentage ranges reported in the abstract and in Section 4 are derived from the relative apportionment of CH<sub>4</sub> and N<sub>2</sub>O to different source factors over the 653 half-hourly samples collected during the experiment period. This range represents the mean diurnal range observed and as seen in Figure 7. This diurnal range combined with bootstrapping based uncertainty estimates is used to understand better the contribution of each source factor to the observed enhancements of a target GHG and the analyst’s confidence in those estimates.

#### 4. Results and Discussion

In Bakersfield, there are a multitude of pollutant sources, ranging from local to regional, from biogenic to anthropogenic, and from primary to secondary. We recognize that PMF analysis is not capable of precise separation of all sources. In PMF analysis, the analyst chooses the number of factor profiles to include in the solution and assigns a source category interpretation for each identified factor. The PMF factors are not unique sources but really statistical combinations of coincident sources. The chemical profile of each factor may contain some contributions from multiple sources that are collocated, or have a similar diurnal pattern of emissions. The cycle of daytime dilution of the boundary layer and nighttime inversion can also result in a covariance that can lead to emissions from unrelated sources being apportioned to a single source factor. Such limitations have been observed previously by Williams et al. (2010) while applying PMF in an urban-industrial setting like Riverside, California. The user must infer the dominant source contributions to these individual factors. Our factor profile (FP) nomenclature is based on the closest explanation of the nature and distribution of emission sources in the region. The source factor names should be treated with caution bearing in mind the physical constraints of the solution and not used to over-explain our interpretation of the region’s CH<sub>4</sub> and N<sub>2</sub>O inventories.

A seven factor solution has been chosen to optimally explain the variability of the included trace gases. The factors have been named based on our interpretation of the emission “source” categories they represent, with corresponding colors which remain consistent in the discussion across the rest of the paper: evaporative and fugitive (black), dairy and livestock (orange), motor vehicles (red), agricultural + soil management (purple), daytime biogenics + secondary organics (light blue), non-

vehicular urban (green) and nighttime anthropogenic + terpene biogenics (navy blue). Figure 4 presents the Factor Profile (FP) plots of each factor. The sum of the normalized contributions of the 50 species in each “source” is equal to 1 in the FP plots. Figures 5a through 5g present the diurnal profiles based on mean hourly concentrations (in normalized units) of each PMF factor with standard deviations explaining the variability. The interpretation of the individual FPs is discussed below (in Section 4.2-4.8). Molar emission factor (EF) of tracers with respect to (w.r.t) one another can be derived for each FP. These EFs can then be compared to those from previous source-specific and apportionment studies (Table 2 through 5). The ratio of PMF-derived total CH<sub>4</sub> enhancement to the input measured CH<sub>4</sub> enhancement ranges from 0.90 to 0.95 (mol mol<sup>-1</sup>) through the whole time series except outliers with really high values (>500 ppb). For N<sub>2</sub>O, the ratio is somewhat lower (0.82-0.92 mol mol<sup>-1</sup>) and this is reflected in the higher PMF-derived uncertainties. The apportionment of some N<sub>2</sub>O mass into a statistically weak and time-varying factor is discussed in Section 4.5. The general assessment is that PMF analysis is able to reconstruct a majority of the measured enhancements for both CH<sub>4</sub> and N<sub>2</sub>O.

#### **4.1. Time trends of measured CH<sub>4</sub>, CO<sub>2</sub>, CO, and N<sub>2</sub>O**

The time series of CH<sub>4</sub>, CO<sub>2</sub>, CO, and N<sub>2</sub>O mixing ratios have been plotted in Figures 6a through 6d while the diurnal variations have been plotted in Figures 6e-6h, respectively. The color markers in each plot indicate the median wind direction. The daily minima for the three GHGs and CO occur during the late afternoon period when daytime heating, mixing and subsequent dilution occurs rapidly. The daily minimum values of CH<sub>4</sub> and N<sub>2</sub>O were larger than that observed at National Oceanic and Atmospheric Administration’s (NOAA) Mauna Loa station at 19.48°N latitude in Hawaii (Dlugokencky et al., 2014) by at least 70 ppb and 0.5 ppb, respectively, for this period. We also compare Bakersfield (at 35.36°N latitude) observations to that from NOAA’s Trinidad Head station which is located on the coast in Northern California and is more representative of mid-latitudes at 40.97°N latitude. Although there was no N<sub>2</sub>O data collected at Trinidad Head, the CH<sub>4</sub> concentrations observed in discrete flask samples collected every few days during summer of 2010 (not necessarily a daily low background) were consistently lower than the daily minimum CH<sub>4</sub> concentration curve at Bakersfield by 10 - 15 ppb. This indicates that there are significant GHG emissions from regional sources around Bakersfield that get added to the already higher local background concentrations, thus keep the local mixing ratio levels quite high. Winds during the highest temperature period between noon and evening (12:00 - 20:00 hour local time) almost always arrive through the urban core in the northwest. Any PMF factor whose dominant source direction is northwest is likely to contain contributions from VOCs emitted from urban

sources, regional sources further upwind or contain contributions from secondary tracers generated from photochemical processing during the day. The three GHGs show a sharp increase during nighttime when the inversion layer builds up and traps primary emissions close to the ground. For CO, measured concentrations show two distinct peaks in the diurnal plot (Figure 6g). The observed early morning peak in the concentration is a combination of decreased dilution and fresh emissions from the morning motor vehicle traffic. The late evening peak in CO concentrations is not coincident with rush hour and is a result of build-up of evening emissions in the boundary layer that is getting shallower as the night progresses. Figure 6a indicates CH<sub>4</sub> enhancements of 500 ppb or more on almost every night with peak mixing ratios exceeding 3000 ppb on several occasions indicating an active methane source(s) in the region. Figure 6d shows that peak N<sub>2</sub>O mixing ratios rise above 330 ppb on almost every night suggesting large sources in the region. Huge enhancements of CH<sub>4</sub>, CO<sub>2</sub> and N<sub>2</sub>O (on DOY 157, 164, and 165) (in Figures 6a, 6b and 6d, respectively) may appear well-correlated to each other due to regional sources emitting into the inversion layer. However, the shapes of the diurnal cycles differ indicating different emission distributions, with the early morning maximum in CH<sub>4</sub> occurring before the maxima for CO<sub>2</sub> and N<sub>2</sub>O, and the morning maximum for CO occurring slightly later. These differences in timing allow PMF analysis to differentiate their contributions into separate factors.

#### **4.2. Factor 1: Evaporative and fugitive emissions**

Factor 1 has a chemical signature indicative of evaporative and fugitive losses of VOCs. The FP of this source is dominated by C<sub>3</sub> to C<sub>6</sub> straight-chain and branched alkanes and some cycloalkanes (Figure 4). The average diurnal cycle of Factor 1 (Figure 5a) shows a broad peak during late night and early morning hours after which the concentrations begin to decrease as the day proceeds reaching a minimum at sunset before beginning to rise again. This is strong indication of a source containing primary emissions that build up in the shallow pronounced nighttime inversions of southern SJV. The subsequent dilution of primary emissions as the mixed layer expands leads to low concentrations during the daytime.

Most of the propane, n-butane and pentanes signal is apportioned to this factor, but not the typical vehicle emission tracers like isooctane or CO or any of the alkenes or aromatics. Absence of these tracers in the FP suggests this factor is not related to vehicular exhaust and is a combination of non-tailpipe emissions and fugitive losses from petroleum operations. None of the CH<sub>4</sub> signal at the SJV site is apportioned to this factor, but almost all of the small straight-chain alkanes, exclusively apportion to this factor. This is in agreement with Gentner et al. (2014a) where the authors show that VOC emissions

from petroleum operations are due to fugitive losses of associated gas from condensate tanks following separation from CH<sub>4</sub>. Table 2 compares EFs derived from this PMF study for the non-tailpipe (evaporative) and fugitive petroleum operation source factor with those from the Gentner et al. (2014a) study done on the same CalNex dataset using an independent source receptor model with chemical mass balancing and effective variance weighting method, and also to, reports of fugitive emissions from the oil and natural gas sources (Pétron et al., 2012; Gilman et al., 2013) and similar factors produced by other PMF studies (Buzcu and Fraser, 2006; Leuchner and Rappenglück, 2010; Bon et al., 2011). Good agreement of Factor 1 VOC EFs with those from the mentioned studies confirms petroleum operations in Kern County as the major source contributing to this factor. The PMF apportionment indicates that this source factor does not contribute to CH<sub>4</sub> enhancements observed at the SJV site (Figure 7a) and thus most of the ‘associated’ CH<sub>4</sub> is likely separated from the condensate prior to emission. As mentioned before, a tiny fraction (~ 5%; Section 4) of the total input CH<sub>4</sub> enhancement is not resolved into source-apportioned contributions. There could be a minor contribution to CH<sub>4</sub> signal from this source, which is unresolved within the framework of uncertainties in the PMF analysis.

#### **4.3. Factor 2: Motor vehicle emissions**

Factor 2 has a chemical signature consistent with the tailpipe exhausts of gasoline and diesel motor vehicles. This source factor includes the combustion tracer CO, and other vehicular emissions tracers, such as isooctane (Figure 4). Alkenes are a product of incomplete fuel combustion in motor vehicles, and almost all of the propene and a significant portion of the isobutene signal are attributed to this source factor. The diurnal variation of Factor 2 shows two distinctive peaks (Figure 5b). The first peak occurs in the morning between 06:00 and 07:00 local time and is influenced by morning rush hour traffic, with suppressed mixing allowing vehicle emissions to build up. As the day proceeds, accelerated mixing and dilution (and perhaps chemical processing of reactive VOCs) reduce the enhancements to a minimum by late afternoon. The evening peak mainly occurs as the dilution process slows down after sunset and emissions build up. The increased motor vehicle traffic in the evening adds more emissions to the shrinking boundary layer. This build-up reaches a peak around 22:00. The occasional high wind events from the northwest (unstable conditions) and fewer vehicles operating on the roads during late nighttime hours contribute to the relatively lower levels of enhancements as compared to the peaks on either side of this nighttime period.

Table 3 compares selective PMF derived EFs from vehicle emissions with the measured gasoline composition collected during CalNex in Bakersfield (Gentner et al., 2012), analysis of gasoline samples

from Riverside in Los Angeles basin (Gentner et al., 2009) and ambient VOC emission ratios measured during CalNex at the Pasadena supersite (Borbon et al., 2013). Although, the two Bakersfield studies employ different source apportionment techniques (and so do the studies conducted in the Los Angeles basin), we observe a broad agreement of relative emission rates of vehicular emission tracers. This agreement validates our assertion that Factor 2 represents a broad suite of vehicular tailpipe emissions.

The PMF derived  $\text{CH}_4/\text{CO}$  EF in Factor 2 is 0.58 ( $\text{mol mol}^{-1}$ ) and is significantly higher than the range of 0.03 - 0.08 ( $\text{mol mol}^{-1}$ ) calculated from results of a vehicle dynamometer study of 30 different cars and trucks (Nam et al., 2004) and an EF of 0.014 ( $\text{mol mol}^{-1}$ ) calculated for SJV district during summer of 2010 using EMFAC, which is CARB's model for estimating emissions from on-road vehicles operating in California (EMFAC, 2011). While it is certainly a possibility that current in-use  $\text{CH}_4$  emission factor in the inventory may be an underestimation, it seems more logical that the relatively high proportion of  $\text{CH}_4$  signal in the vehicle source factor profile is due to contributions from coincident urban sources (e.g. natural gas leaks) mixed into the vehicle gasoline exhausts resulting in a 'mixing' phenomena as discussed in the supplement. In spite of the non-negligible proportion of  $\text{CH}_4$  in the Factor 2 source profile, the contribution of the factor to  $\text{CH}_4$  enhancements (Figure 7a) at Bakersfield is negligible relative to the dairy and livestock factor.

The state GHG inventory attributes about 18% of the 2010 statewide  $\text{N}_2\text{O}$  emissions to the on-road transportation sector (CARB 2012). Our PMF analysis shows essentially a negligible enhancement of  $\text{N}_2\text{O}$  associated with the vehicle emission Factor 2 with a PMF derived  $\text{N}_2\text{O}/\text{CO}$  EF of 0.00015 ( $\text{mol mol}^{-1}$ ). The EMFAC generated  $\text{N}_2\text{O}/\text{CO}$  EF in SJV during summer of 2010 is more than 20 times higher at 0.0034 ( $\text{mol mol}^{-1}$ ). The PMF derived 'vehicle emissions' contribution to  $\text{N}_2\text{O}$  is in stark contrast to the inventory and is an important outcome suggesting a significant error in EFs used to derive the statewide inventory for  $\text{N}_2\text{O}$ .

#### **4.4. Factor 3: Dairy and livestock emissions**

Factor 3 has a chemical signature indicative of emissions from dairy operations. This source factor is the largest contributor to  $\text{CH}_4$  enhancements (Figure 7a) and a significant portion of the  $\text{N}_2\text{O}$  signal (Figure 7c). The FP also has major contributions from methanol ( $\text{MeOH}$ ) and ethanol ( $\text{EtOH}$ ), with minor contributions from aldehydes and ketones (Figure 4). A separate PMF analysis with a broader set of VOC measurements at the same site showed that most of the acetic acid ( $\text{CH}_3\text{COOH}$ ) and some formaldehyde ( $\text{HCHO}$ ) signal attributed to this factor as well (Allen Goldstein, personal communication,



2014). All the above-mentioned VOCs are emitted in significant quantities from dairy operations and cattle feedlots (Filipy et al., 2006; Shaw et al., 2007; Ngwabie et al., 2008; Chung et al., 2010). About 70-90% of the diurnal CH<sub>4</sub> signal is attributed to this factor (Figure 7a) depending on the time of day. The CH<sub>4</sub> dairy and livestock mass fraction which is used to calculate this diurnal range has an uncertainty of 29% derived using the bootstrapping method. This source factor contributes about 60 - 70% of the total N<sub>2</sub>O daily enhancements as seen in Figure 7c. The bootstrapping uncertainty in the N<sub>2</sub>O dairy and livestock mass fraction is 33%.

Comparing the Factor 3 profile to dairy source profiles from various studies is challenging. A dairy is, in essence, a collection of area sources with distinct emission pathways and chemical characteristics. Hence, a lot of dairy studies do not look at facility-wide emissions instead focusing on specific area sources within the facility. In contrast, PMF captures the covariance of CH<sub>4</sub>, N<sub>2</sub>O, and VOCs emitted from the ensemble source as downwind plumes from dairies arrive at the site. Table 4 compares the PMF derived EFs of CH<sub>4</sub> w.r.t MeOH and EtOH with those from other studies. Previously, cow chamber experiments (Shaw et al., 2007; Sun et al., 2008) have measured emissions from ruminants and their fresh manure; emissions have also been studied in a German cowshed (Ngwabie et al., 2008) and EFs have been derived from SJV dairy plumes sampled from aircraft (Gentner et al., 2014a; Guha et al., in prep). Since enteric fermentation and waste manure is the predominant CH<sub>4</sub> source in dairies, CH<sub>4</sub> emission rates calculated by Shaw et al. (2007) are representative of a whole facility. However, their MeOH/CH<sub>4</sub> ratios are lower than those determined by PMF and aircraft studies. Animal feed and silage are the dominant sources of many VOCs including MeOH and EtOH (Alanis et al., 2010; Howard et al., 2010) and the ratios in (Shaw et al., 2007) do not reflect these emissions. In Ngwabie et al. (2008), experiments were performed in cold winter conditions (-2 to 8°C) when temperature dependent VOC emissions from silage and feed are at a minimum. The authors comment that MeOH emissions from California dairies are likely higher, as the alfalfa-based feed is a big source of MeOH owing to its high pectin content (Galbally and Kirstine, 2002). These observations explain why MeOH/CH<sub>4</sub> ratios in these studies are lower than PMF derived ratios. The PMF range for EtOH /CH<sub>4</sub> EF for Factor 3 agrees with the slope derived from ground-site data (Gentner et al., 2014a) and is similar, but somewhat larger than the German dairy study (Ngwabie et al., 2008) . Miller and Varel (2001) and Filipy et al. (2006) did not measure CH<sub>4</sub> emission rates so a direct derivation of EF w.r.t CH<sub>4</sub> is not possible. These studies, however, reported EtOH emission rates (from dairies and feedlots in United States) which are used to derive EFs w.r.t to CH<sub>4</sub> using an averaged CH<sub>4</sub> emission rate from (Shaw et al., 2007). Using this method, we get EFs that are comparable to PMF derived EF of CH<sub>4</sub>/EtOH (Table 4).

Hence, we demonstrate within reasonable terms that the relative fractions of masses in Factor 3 are consistent with CH<sub>4</sub> and VOC emissions from dairies.

Enteric fermentation is a part of the normal digestive process of livestock such as cows, and is a large source of CH<sub>4</sub> while the storage and management of animal manure in lagoons or holding tanks is also a major source of CH<sub>4</sub>. According to the state GHG inventory (CARB, 2013), ~ 58% of the statewide CH<sub>4</sub> emissions results from a combination of these two processes. N<sub>2</sub>O is also emitted during the breakdown of nitrogen in livestock manure and urine and accounts for about 10% of the statewide N<sub>2</sub>O emission inventory. Kern County has a big dairy industry with about 160,000 milk cows representing 10% of the dairy livestock of the state in 2012 and another 330,000 heads of cattle for beef (KernAg, 2011; CASR, 2013). The dominant contributions to CH<sub>4</sub> and N<sub>2</sub>O signal and the general agreement of dairy EFs with PMF EFs from Factor 3 indicate that the extensive cattle operations in the county are a big source of these emissions. We do observe that the proportion of regional N<sub>2</sub>O enhancements attributed to this sector is a significantly larger proportion of the total N<sub>2</sub>O emissions as compared to the state inventory.

#### **4.5. Factor 4: Agricultural and soil management emissions**

The chemical profile of Factor 4 is a mix of emissions from agricultural activities around the site. Factor 4 includes a major portion of the N<sub>2</sub>O signal along with a number of VOCs that have crop/plant signatures like methacrolein, methyl ethyl ketone (Jordan et al., 2009; McKinney et al., 2011), methanol and acetone (Goldstein and Schade, 2000; Hu et al., 2013; Gentner et al., 2014b) (Figure 4). While many of these oxygenated VOCs have several prominent sources, studies have reported substantial simultaneous emissions from natural vegetation and agricultural crops. At a rural site in the Northeast, Jordan et al. (2009) reported high concentrations of oxygenated VOCs and correlations between the diurnal concentrations of acetone, methanol, and methyl ethyl ketone. Kern County is one of the most prolific agricultural counties in California. The four main crops grown (by value as well as acreage) in 2010 were almonds, grapes, citrus and pistachios (KernAg, 2011). Table 5 compares the PMF derived EFs for acetone/MeOH from Factor 4 with ratios of basal emission factors (BEFs) from crop-specific greenhouse and field measurements (Fares et al., 2011, 2012; Gentner et al., 2014b). The good agreement of the ratios confirms that the FP of Factor 4 is an aggregate of biogenic VOC emissions from the agricultural sector. Nitrous oxide is emitted when nitrogen is added to soil through use of synthetic fertilizers and animal manure, while crops and plants are responsible for the VOC emissions. Hence this source factor is a combination of collocated sources (soils and crops). The PMF solution to this factor has

uncertainties greater than those for other factors (Figure S4). This is potentially because not all crops emit the same combination of VOCs nor are all agricultural fields fertilized at the same time. The existence of this statistically weak factor is confirmed by bootstrapping runs (Section S3) and numerous PMF trials all of which produce a distinct factor with N<sub>2</sub>O as a dominant contributor along with certain biogenic VOCs, though often in varying proportions. CO<sub>2</sub> is not included in the PMF analysis reported in the paper, most importantly because negative CO<sub>2</sub> fluxes during daytime can introduce artifacts in PMF analysis and result in erroneous apportionment. But PMF runs involving CO<sub>2</sub> indicate that most of the CO<sub>2</sub> is apportioned to this factor. Plant and soil respiration (especially during the night) is a major source of CO<sub>2</sub> and the apportionment of CO<sub>2</sub> to Factor 4 confirms the nature of this source. The temporal correlation between CO<sub>2</sub> and N<sub>2</sub>O is also evident in their average diurnal cycles (Figures 6f and 6h), which have a coincident early morning peak. The absence of monoterpenes from the FP of this factor can be explained by their shorter atmospheric lifetimes compared to VOCs like acetone and MeOH and the rapid daytime mixing which dilutes the terpenoid emissions arriving at the site during the day. At night, when atmospheric dilution is low, monoterpene emissions from agriculture are more likely to get apportioned into a separate source factor dominant during nighttime, when temperature-sensitive biogenic emissions of MeOH and acetone can be expected to be a minor constituent in the FP (see Section 4.8).

Factor 4 is a significant source of GHGs contributing about 20 - 25% of the total N<sub>2</sub>O enhancements in the diurnal cycle (Fig. 7c) but with a relatively large 1 $\sigma$  confidence interval of 70% in the agriculture and soil N<sub>2</sub>O mass fraction. Kern County is one of the premier agricultural counties of California accounting for \$4.2 billion (about 18%) of the total agricultural revenue from fruits and nuts, vegetables and field crops (KernAg, 2011; CASR, 2013) and is also the biggest consumer of synthetic fertilizers. Agricultural soil management accounts for about 60% of the statewide N<sub>2</sub>O emission inventory (CARB, 2013). Our assessment of diurnal source distribution of N<sub>2</sub>O emissions from the agriculture source factor (Figure 7c) in presence of another dominating source (dairy and livestock) is consistent with the inventory estimates from agricultural and soil management sector.

#### **4.6. Factor 5: Daytime biogenics and secondary organics**

The chemical composition and diurnal profile of Factor 5 points to a source whose emissions are either primary biogenic VOCs with temperature-dependent emissions (e.g. isoprene), or products of photochemical oxidation of primary VOCs (e.g. acetone) (Figure 4). Isoprene is a dominant component of the source FP and is mostly apportioned to Factor 5. Figure 5e shows a steady increase in the PMF

factor mass concentration during the daytime hours that hits a peak during afternoons indicating that this source is dependent on sunlight and temperature. Potential source contributions come from oak forests on the foothills of the western edge of the SJV or scattered isoprene producing plants in the SJV (note that most crops do not emit significant amounts of isoprene). Factor 5 includes contribution from VOCs that have primary light and temperature driven (crops), as well as secondary sources in the Central Valley e.g. acetone (Goldstein and Schade, 2000), methanol (Gentner et al., 2014b) and aldehydes. A similar PMF analysis with a different objective (Goldstein et al., in prep) shows that secondary organics like glyoxal, formaldehyde and formic acid mostly apportion to Factor 5. The CO apportioned to this factor could potentially be a product of mobile and/or stationary combustion co-located or up/downwind of the biogenic VOC source. CO can also come from coincident isoprene oxidation (Hudman et al., 2008). This daytime source is not responsible for any of the observed CH<sub>4</sub> and N<sub>2</sub>O enhancements.

#### **4.7. Factor 6: Non-vehicular/miscellaneous urban emissions**

The chemical signature of Factor 6 is composed of VOCs associated with an array of applications and processes, including solvents, fumigants, industrial-byproducts, etc. The diurnal profile of Factor 6 (Figure 5f) is somewhat different from that of evaporative and fugitive source (Figure 5a) and dairies (Figure 5c) in that even during the middle of the day when vertical mixing is at its strongest, the enhancements contributing to the factor are substantial. This suggests that the source(s) is in close proximity to the site and hence most likely located within the urban core. The FP has CO as an important component but relative absence of fugitive source markers (e.g. light alkanes) and vehicle emissions tracers (e.g. isooctane, cycloalkanes etc.) indicate that the origin of this source factor is potentially non-mobile combustion. Also present in a major proportion is carbon disulfide (CS<sub>2</sub>), chlorinated alkanes like 1,2-dichloroethane and 1,2-dichloropropane, isobutene (product of incomplete combustion), and minor contributions from aromatics and aldehydes (Figure 4). There are a myriad of potential sources that could be contributing to this factor, and we don't have specific tracers or other information to ascribe it to a single source or group of sources. Hence we call Factor 6 an 'urban emissions source'. There is a very minor CH<sub>4</sub> contribution from this factor which results in a tiny and negligible contribution to the PMF source apportionment of CH<sub>4</sub> (Figure 7a). The source factor does not contribute to the N<sub>2</sub>O enhancements.

#### **4.8. Factor 7: Nighttime anthropogenic and terpene biogenic emissions**

Factor 7 is primarily composed of biogenic compounds belonging to the terpene family and p-cymene (Figure 4). Factor 7 mostly influences the site during late night and early morning hours (Figure 5g) when nighttime downslope flows usually dominate bringing winds from the east and south to the site. The entire flow path from the base of the foothills to the site is covered with agricultural crops emitting into a shallow nighttime boundary layer. These crops include grapes, almonds, citrus and pistachios, which are the top four agricultural commodities grown in the county (KernAg, 2011; CASR, 2013), and these produce considerable monoterpene emissions (Fares et al., 2012; Gentner et al., 2014b). The spatial distribution of terpene compounds from statistical source footprint derived from FLEXPART back-trajectories is consistent with the location of croplands in southern SJV (Gentner et al., 2014b). Biogenic VOCs emitted from forests in the foothills are likely minor contributors to the downslope flows arriving at the site owing to their lifetime and distance (> 50 km) (Tanner and Zielinska, 1994).

Following the rapid rise in enhancements in the early morning hours, contributions of Factor 7 to total signal decrease rapidly when the flow moves to more typical daytime wind directions (Figure 5g). A nearby source (e.g. the WWTP), that is upwind of the site for only a certain part of the diurnal cycle, is expected to be more directionally constrained and emissions profile from such a source will look similar to the diurnal profile of Factor 7. Among source factors which contain non-negligible fractional contribution of both CH<sub>4</sub> and N<sub>2</sub>O (i.e. dairies, agriculture and soil management, and Factor 7), the PMF derived CH<sub>4</sub>/N<sub>2</sub>O EF of  $42 \pm 20$  (gC gN<sup>-1</sup>) from Factor 7 is most similar to the bottom-up inventory EF of 56 (gC gN<sup>-1</sup>) for waste water treatment in Kern County (KernGHG, 2012). Given the proximity of the WWTP and previous observations of GHGs from them, it is possible that there is a minor but noticeable contribution (~5%) to CH<sub>4</sub> and N<sub>2</sub>O enhancements from this nighttime source (Figures 7a and 7c).

## 5. Implications

This study demonstrates the potential of PMF technique to apportion atmospheric gas-phase observations of CH<sub>4</sub> and N<sub>2</sub>O into source categories using a broad array of tracers. PMF is not commonly employed to perform for source attribution of these GHGs because studies generally lack simultaneous measurements of specific source-markers. Applying this statistical technique on a GHG-VOC unified data set, well-represented by a broad suite of VOC classes, allows a set of compounds acting as source markers to be partitioned into separate profiles leading to easier identification of their sources.

We provide clear analysis that dairy and livestock operations are the largest sources of emissions in the Bakersfield region accounting for a majority of the CH<sub>4</sub> (70 - 90%) and N<sub>2</sub>O ( 60 -70%) emissions. As per the CARB inventory (Figure 1), dairy operations are the dominant source of non-CO<sub>2</sub> GHGs in the state and our analysis agrees with that broad trend. However, in the recent past, a number of top-down CH<sub>4</sub> and N<sub>2</sub>O emission studies in the Central Valley have reported underestimation of the non-CO<sub>2</sub> GHG inventory (Zhao et al., 2009; Santoni et al., 2012; Jeong et al., 2012a, 2012b; Miller et al., 2013). These studies attribute a majority of this underestimation to the dairy sector. Our results emphasize the significance of this sector in the SJV although we do not derive total emission estimates to compare directly with the inventory.

The contribution of fugitive emissions from the oil and gas industry in Bakersfield to CH<sub>4</sub> emissions is found to be negligible especially in the presence of the much larger dairy source. The PMF analysis, though, clearly establishes an evaporative and fugitive source that contributes to emissions of lighter hydrocarbons. This supports the conclusion that the majority of the CH<sub>4</sub> is being separated at the point of extraction from the ‘associated gas’ and is not released with fugitive emissions (Gentner et al., 2014a). Kern County produces 75% of all the oil produced in California (~6% of US production) and has 81% of the state’s 60000+ active oil wells (CDC, 2013). There is, however, a surprising scarcity of measured data to quantify the estimates of fugitive CH<sub>4</sub> from the prolific oil fields in the county and validate the bottom-up, activity data-based inventory. Currently, fugitive emissions from fossil fuel extraction and distribution contribute ~5% to the county’s CH<sub>4</sub> emissions inventory (KernGHG, 2012). Nationwide, a number of recent studies have reported significantly higher emissions of fugitive CH<sub>4</sub> from oil and gas production operations in other regions (Pétron et al., 2012; Karion et al., 2013; Miller et al., 2013; Kort et al., 2014) . The PMF apportionment in this study (~2%) is consistent with the fraction of fugitive CH<sub>4</sub> emissions in the regional and state inventories but the PMF method, by itself, cannot accurately constrain emissions from minor sources owing to the uncertainties in the dominant sources.

We find that the vehicle emissions source factor identified in this study makes no detectable contribution to observed N<sub>2</sub>O enhancements. Our findings do not agree with the significant contribution (~18%) of the transportation sector to the state’s N<sub>2</sub>O emission inventory (CARB, 2013). Vehicle dynamometer studies have indicated rapidly declining N<sub>2</sub>O EFs with advancement in catalyst technologies, declining sulfur content in fuel and newer technology vehicles (Huai et al., 2004). N<sub>2</sub>O emissions from California vehicles, required to meet progressively stringent emission standards, are expected to decline and should have a minimal contribution to the CARB inventory in this decade.

However, it seems the updates to the mobile N<sub>2</sub>O emissions inventory is not keeping in pace with the improvements in vehicle catalyst technologies and corresponding decline in tailpipe N<sub>2</sub>O emissions. Bakersfield is a fairly large population urban region (~500,000) and the essentially non-existent contribution of the PMF vehicle emissions source to the N<sub>2</sub>O apportionment and large divergence of the PMF derived N<sub>2</sub>O/CO EF from the state inventory EF for motor vehicles is a significant outcome pointing to overestimation of N<sub>2</sub>O from motor vehicles in the inventory.

## Acknowledgements

This work was supported by a contract from the California Air Resources Board (CARB) #08-316 and in kind support from Los Gatos Research (LGR) Inc. We thank Dr. John Karlik, the staff at UC Cooperative Extension at Bakersfield, Aaron Gardner and Dr. Doug Baer at LGR Inc., and Dr. Ying Hsu at CARB for their logistical support. We also appreciate the subject matter guidance from Dr. Marc Fischer at Lawrence Berkeley National Laboratory.

## References

- Alanis, P., Ashkan, S., Krauter, C., Campbell, S. and Hasson, A. S.: Emissions of volatile fatty acids from feed at dairy facilities, *Atmos. Environ.*, 44(39), 5084–5092, doi:10.1016/j.atmosenv.2010.09.017, 2010.
- Bao, J.-W., Michelson, S. a., Persson, P. O. G., Djalalova, I. V. and Wilczak, J. M.: Observed and WRF-Simulated Low-Level Winds in a High-Ozone Episode during the Central California Ozone Study, *J. Appl. Meteorol. Climatol.*, 47(9), 2372–2394, doi:10.1175/2008JAMC1822.1, 2007.
- Beaver, S. and Palazoglu, A.: Influence of synoptic and mesoscale meteorology on ozone pollution potential for San Joaquin Valley of California, *Atmos. Environ.*, 43(10), 1779–1788, doi:10.1016/j.atmosenv.2008.12.034, 2009.
- Bon, D. M., Ulbrich, I. M., de Gouw, J. a., Warneke, C., Kuster, W. C., Alexander, M. L., Baker, a., Beyersdorf, a. J., Blake, D., Fall, R., Jimenez, J. L., Herndon, S. C., Huey, L. G., Knighton, W. B., Ortega, J., Springston, S. and Vargas, O.: Measurements of volatile organic compounds at a suburban ground site (T1) in Mexico City during the MILAGRO 2006 campaign: measurement comparison, emission ratios, and source attribution, *Atmos. Chem. Phys.*, 11(6), 2399–2421, doi:10.5194/acp-11-2399-2011, 2011.
- Borbon, A., Gilman, J. B., Kuster, W. C., Grand, N., Chevaillier, S., Colomb, a., Dolgorouky, C., Gros, V., Lopez, M., Sarda-Estevé, R., Holloway, J., Stutz, J., Petetin, H., McKeen, S., Beekmann, M., Warneke, C., Parrish, D. D. and de Gouw, J. a.: Emission ratios of anthropogenic volatile organic compounds in northern mid-latitude megacities: Observations versus emission inventories in Los Angeles and Paris, *J. Geophys. Res. Atmos.*, 118(4), 2041–2057, doi:10.1002/jgrd.50059, 2013.

688 Brown, S. G., Frankel, A. and Hafner, H. R.: Source apportionment of VOCs in the Los Angeles area using  
689 positive matrix factorization, *Atmos. Environ.*, 41(2), 227–237, doi:10.1016/j.atmosenv.2006.08.021,  
690 2007.

691 Buzcu, B., and M. P. Fraser (2006), Source identification and apportionment of volatile organic  
692 compounds in Houston, TX, *Atmos. Environ.*, 40(13), 2385–2400, doi:10.1016/j.atmosenv.2005.12.020.

693 CARB (2013), California Greenhouse Gas Inventory for 2000-2012 - by IPCC Category.  
694 <http://www.arb.ca.gov/cc/inventory/data/data.htm>. accessed on September 23, 2014.

695 CASR (2011), California Agricultural Statistics Review, 2013-14. California Department of Food and  
696 Agriculture, Sacramento, CA.

697 CDC (2013), 2012 Preliminary Report Of California Oil And Gas Production Statistics . California  
698 Department of Conservation, Division of Oil, Gas and Geothermal Resources.  
699 [ftp://ftp.consrv.ca.gov/pub/oil/annual\\_reports/2012/PR03\\_PreAnnual\\_2012.pdf](ftp://ftp.consrv.ca.gov/pub/oil/annual_reports/2012/PR03_PreAnnual_2012.pdf). accessed on  
700 September 23, 2014.

701 Chung, M. Y., M. Beene, S. Ashkan, C. Krauter, and A. S. Hasson (2010), Evaluation of non-enteric sources  
702 of non-methane volatile organic compound (NMVOC) emissions from dairies, *Atmos. Environ.*, 44(6),  
703 786–794, doi:10.1016/j.atmosenv.2009.11.033.

704 Comero, S., Capitani, L., and Gawlik, B. M.: Positive Matrix Factorization - An introduction to the  
705 chemometric evaluation of environmental monitoring data using PMF, JRC Scientific and Technical  
706 Reports, EUR 23946 EN-2009.

707 Crutzen, P. J., Mosier, a. R., Smith, K. a. and Winiwarter, W.: N<sub>2</sub>O release from agro-biofuel production  
708 negates global warming reduction by replacing fossil fuels, *Atmos. Chem. Phys. Discuss.*, 7(4), 11191–  
709 11205, doi:10.5194/acpd-7-11191-2007, 2007.

710 Denman, K.L., G. Brasseur, A. Chidthaisong, P. Ciais, P.M. Cox, R.E. Dickinson, D. Hauglustaine, C. Heinze,  
711 E. Holland, D. Jacob, U. Lohmann, S Ramachandran, P.L. da Silva Dias, S.C. Wofsy and X. Zhang, 2007:  
712 Couplings Between Changes in the Climate System and Biogeochemistry. In: *Climate Change 2007: The*  
713 *Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the*  
714 *Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B.  
715 Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and  
716 New York, NY, USA.

717 Dlugokencky, E.J., A.M. Crotwell, P.M. Lang, K.A. Masarie (2014), Atmospheric Methane Dry Air Mole  
718 Fractions from quasi-continuous measurements at Barrow, Alaska and Mauna Loa, Hawaii, 1986-2013,  
719 Version: 2014-08-12, Path: <ftp://ftp.cmdl.noaa.gov/ccg/ch4/in-situ/>.

720 EMFAC (2011). 2011 Mobile Source Emission Inventory - Current methods and data. California Air  
721 Resources Board, Sacramento, CA. <http://www.arb.ca.gov/msei/modeling.htm>. accessed on September  
722 20, 2014.



723 Fares, S., Gentner, D. R., Park, J.-H., Ormeno, E., Karlik, J. and Goldstein, A. H.: Biogenic emissions from  
 724 Citrus species in California, *Atmos. Environ.*, 45(27), 4557–4568, doi:10.1016/j.atmosenv.2011.05.066,  
 725 2011.

726 Fares, S., Park, J.-H., Gentner, D. R., Weber, R., Ormeño, E., Karlik, J. and Goldstein, a. H.: Seasonal cycles  
 727 of biogenic volatile organic compound fluxes and concentrations in a California citrus orchard, *Atmos.*  
 728 *Chem. Phys.*, 12(20), 9865–9880, doi:10.5194/acp-12-9865-2012, 2012.

729 Filipy, J., B. Rumburg, G. Mount, H. Westberg, and B. Lamb (2006), Identification and quan tification of  
 730 volatile organic compounds from a dairy, *Atmos. Environ.*, 40(8), 1480–1494,  
 731 doi:10.1016/j.atmosenv.2005.10.048.

732 Fischer, M.L., C. Zhao., W.J. Riley, and A.C. Andrews. 2009. Observation of CH<sub>4</sub> and other Non-CO<sub>2</sub>  
 733 Green House Gas Emissions from California. California Energy Commission, PIER Energy-Related  
 734 Environmental Research. (Report # 500-2009-096).

735 Galbally, I. E., and Kirstine, W.: The Production of Methanol by Flowering Plants and the Global Cycle of  
 736 Methanol, *Journal of Atmospheric Chemistry*, 43, 195–229, 10.1023/A:1020684815474, 2002. Galloway,  
 737 J. N., Townsend, A. R., Erisman, J. W., Bekunda, M., Cai, Z., Freney, J. R., Martinelli, L. a, Seitzinger, S. P.  
 738 and Sutton, M. a: Transformation of the nitrogen cycle: recent trends, questions, and potential  
 739 solutions., *Science*, 320(5878), 889–92, doi:10.1126/science.1136674, 2008.

740 Galloway, J. N., A. R. Townsend, J. W. Erisman, M. Bekunda, Z. Cai, J. R. Freney, L. a Martinelli, S. P.  
 741 Seitzinger, and M. a Sutton (2008), Transformation of the nitroge n cycle: recent trends, questions, and  
 742 potential solutions., *Science*, 320(5878), 889–92, doi:10.1126/science.1136674.

743 Gentner, D. R., Ford, T. B., Guha, a., Boulanger, K., Brioude, J., Angevine, W. M., de Gouw, J. a., Warneke,  
 744 C., Gilman, J. B., Ryerson, T. B., Peischl, J., Meinardi, S., Blake, D. R., Atlas, E., Lonneman, W. a.,  
 745 Kleindienst, T. E., Beaver, M. R., Clair, J. M. St., Wennberg, P. O., VandenBoer, T. C., Markovic, M. Z.,  
 746 Murphy, J. G., Harley, R. a. and Goldstein, a. H.: Emissions of organic carbon and methane from  
 747 petroleum and dairy operations in California’s San Joaquin Valley, *Atmos. Chem. Phys.*, 14(10), 4955–  
 748 4978, doi:10.5194/acp-14-4955-2014, 2014a.

749 Gentner, D. R., Harley, R. a, Miller, A. M. and Goldstein, A. H.: Diurnal and seasonal variability of  
 750 gasoline-related volatile organic compound emissions in Riverside, California., *Environ. Sci. Technol.*,  
 751 43(12), 4247–52 [online] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19603630>, 2009.

752 Gentner, D. R., Isaacman, G., Worton, D. R., Chan, A. W. H., Dallmann, T. R., Davis, L., Liu, S., Day, D. A.,  
 753 Russell, L. M., Wilson, K. R., Weber, R., Guha, A. and Harley, R. A.: Elucidating secondary organic aerosol  
 754 from diesel and gasoline vehicles through detailed characterization of organic carbon e missions, ,  
 755 doi:10.1073/pnas.1212272109/, 2012.

756 Gentner, D. R., Ormeño, E., Fares, S., Ford, T. B., Weber, R., Park, J.-H., Brioude, J., Angevine, W. M.,  
 757 Karlik, J. F. and Goldstein, a. H.: Emissions of terpenoids, benzenoids, and other biogenic gas-phase  
 758 organic compounds from agricultural crops and their potential implications for air quality, *Atmos. Chem.*  
 759 *Phys.*, 14(11), 5393–5413, doi:10.5194/acp-14-5393-2014, 2014b.

760 Gilman, J. B., B. M. Lerner, W. C. Kuster, and J. a de Gouw (2013), Source signature of volatile organic  
 761 compounds from oil and natural gas operations in northeastern colorado., *Environ. Sci. Technol.*, 47(3),  
 762 1297–305, doi:10.1021/es304119a.

763 Goldstein, A. H. and Schade, G. W.: Quantifying biogenic and anthropogenic contributions to acetone  
 764 mixing ratios in a rural environment, *Atmos. Environ.*, 34(29–30), 4997–5006, doi:10.1016/S1352-  
 765 2310(00)00321-6, 2000.

766 Goldstein, A. H. et al.: Source apportionment of OH reactivity at Bakersfield Positive Matrix Factorization  
 767 (PMF) during CalNex 2010, *in preparation*.

768 Guha, A., Misztal, P. K., Karl, T., Peischl, J., Ryerson, T. B., Jonsson, H.H., and Goldstein, A.H.: Identifying  
 769 and mapping methane (CH<sub>4</sub>) sources over California from airborne mixing ratio, eddy covariance flux and  
 770 VOC source tracer measurements, *in preparation*, 2015.

771 Hendriks, D. M. D., a. J. Dolman, M. K. van der Molen, and J. van Huissteden (2008), A compact and  
 772 stable eddy covariance set-up for methane measurements using off-axis integrated cavity output  
 773 spectroscopy, *Atmos. Chem. Phys.*, 8(2), 431–443, doi:10.5194/acp-8-431-2008.

774 Hopke, P.: A guide to positive matrix factorization, Department of Chemistry, Calrkson University,  
 775 Potsdam, NY pp 1–16, <http://www.epa.gov/ttnamti1/files/ambient/pm25/workshop/laymen.pdf>, 2000

776 Howard, C. J., Kumar, A., Malkina, I., Mitloehner, F., Green, P. G., Flocchini, R. G. and Kleeman, M. J.:  
 777 Reactive organic gas emissions from livestock feed contribute significantly to ozone production in  
 778 central California., *Environ. Sci. Technol.*, 44(7), 2309–14, doi:10.1021/es902864u, 2010.

779 Hsu, Y.-K., VanCuren, T., Park, S., Jakober, C., Herner, J., FitzGibbon, M., Blake, D. R. and Parrish, D. D.:  
 780 Methane emissions inventory verification in southern California, *Atmos. Environ.*, 44(1), 1–7,  
 781 doi:10.1016/j.atmosenv.2009.10.002, 2010.

782 Hu, L., D. B. Millet, S. Y. Kim, K. C. Wells, T. J. Griffis, E. V. Fischer, D. Helmig, J. Hueber, and a. J. Curtis  
 783 (2013), North American acetone sources determined from tall tower measurements and inverse  
 784 modeling, *Atmos. Chem. Phys.*, 13(6), 3379–3392, doi:10.5194/acp-13-3379-2013.

785 Huai, T., Durbin, T. D., Wayne Miller, J. and Norbeck, J. M.: Estimates of the emission rates of nitrous  
 786 oxide from light-duty vehicles using different chassis dynamometer test cycles, *Atmos. Environ.*, 38(38),  
 787 6621–6629, doi:10.1016/j.atmosenv.2004.07.007, 2004.

788 Hudman, R. C., Murray, L. T., Jacob, D. J., Millet, D. B., Turquety, S., Wu, S., Blake, D. R., Goldstein, a. H.,  
 789 Holloway, J. and Sachse, G. W.: Biogenic versus anthropogenic sources of CO in the United States,  
 790 *Geophys. Res. Lett.*, 35(4), L04801, doi:10.1029/2007GL032393, 2008.

791 IPCC, 2013: Summary for Policymakers. In: *Climate Change 2013: The Physical Science Basis*.  
 792 Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on  
 793 Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia,  
 794 V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York,  
 795 NY, USA.

796 Jeong, S., C. Zhao, A. E. Andrews, L. Bianco, J. M. Wilczak, and M. L. Fischer (2012a), Seasonal variation of  
797 CH<sub>4</sub> emissions from central California, *J. Geophys. Res.*, 117(D11), D11306, doi:10.1029/2011JD016896.

798 Jeong, S., C. Zhao, A. E. Andrews, E. J. Dlugokencky, C. Sweeney, L. Bianco, J. M. Wilczak, and M. L.  
799 Fischer (2012b), Seasonal variations in N<sub>2</sub>O emissions from central California, *Geophys. Res. Lett.*,  
800 39(16), n/a–n/a, doi:10.1029/2012GL052307.

801 Jordan, C., Fitz, E., Hagan, T., Sive, B., Frinak, E., Haase, K., Cottrell, L., Buckley, S., and Talbot, R.: Long-  
802 term study of VOCs measured with PTR-MS at a rural site in New Hampshire with urban influences,  
803 *Atmos. Chem. Phys.*, 9, 4677–4697, doi:10.5194/acp-9-4677-2009, 2009.

804 KernGHG (2012), Kern County Communitywide Greenhouse Gas Emission Inventory, Methodology  
805 Documents Volume 2; Final Report - May 2012. Prepared by San Joaquin Air Pollution Control District for  
806 Kern County Planning and Community Development. <http://www.kerncog.org/climate-change>.

807 KernAg (2010). 2010 Kern County Agricultural Crop Report. Department of Agriculture and  
808 Measurement Standards. [http://www.kernag.com/caap/crop-reports/crop10\\_19/crop2010.pdf](http://www.kernag.com/caap/crop-reports/crop10_19/crop2010.pdf).  
809 accessed on September 22 2014.

810 Karion, A., Sweeney, C., Pétron, G., Frost, G., Michael Hardesty, R., Kofler, J., Miller, B. R., Newberger, T.,  
811 Wolter, S., Banta, R., Brewer, A., Dlugokencky, E., Lang, P., Montzka, S. a., Schnell, R., Tans, P., Trainer,  
812 M., Zamora, R. and Conley, S.: Methane emissions estimate from airborne measurements over a  
813 western United States natural gas field, *Geophys. Res. Lett.*, 40(16), 4393–4397, doi:10.1002/grl.50811,  
814 2013.

815 Kim, E., Hopke, P. K. and Edgerton, E. S.: Improving source identification of Atlanta aerosol using  
816 temperature resolved carbon fractions in positive matrix factorization, *Atmos. Environ.*, 38(20), 3349–  
817 3362, doi:10.1016/j.atmosenv.2004.03.012, 2004.

818 Kort, E. A., C. Frankenberg, K. R. Costigan, R. Lindenmaier, M. K. Dubey, and D. Wunch (2014), Four  
819 corners: The largest US methane anomaly viewed from space, *Geophys. Res. Lett.*, 41, 6898–6903,  
820 doi:10.1002/2014GL061503.

821 Lee, E., Chan, C. K. and Paatero, P.: Application of positive matrix factorization in source apportionment  
822 of particulate pollutants in Hong Kong, *Atmos. Environ.*, 33(19), 3201–3212, doi:10.1016/S1352-  
823 2310(99)00113-2, 1999.

824 Leuchner, M., and B. Rappenglück (2010), VOC source–receptor relationships in Houston during TexAQS-  
825 II, *Atmos. Environ.*, 44(33), 4056–4067, doi:10.1016/j.atmosenv.2009.02.029.  
826

827 McKinney, K. a., Lee, B. H., Vasta, a., Pho, T. V. and Munger, J. W.: Emissions of isoprenoids and  
828 oxygenated biogenic volatile organic compounds from a New England mixed forest, *Atmos. Chem. Phys.*,  
829 11(10), 4807–4831, doi:10.5194/acp-11-4807-2011, 2011.

830 McMillan, A. M. S., Goulden, M. L. and Tyler, S. C.: Stoichiometry of CH<sub>4</sub> and CO<sub>2</sub> flux in a California rice  
831 paddy, *J. Geophys. Res.*, 112(G1), G01008, doi:10.1029/2006JG000198, 2007.

832 Miller, D. N., and Varel, V. H.: In vitro study of the biochemical origin and production limits of odorous  
833 compounds in cattle feedlots, *Journal of Animal Science*, 79, 2949-2956, 2001.

834 Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. a, Andrews, A. E., Biraud, S. C., Dlugokencky, E. J.,  
835 Eluszkiewicz, J., Fischer, M. L., Janssens-Maenhout, G., Miller, B. R., Miller, J. B., Montzka, S. a, Nehrkorn,  
836 T. and Sweeney, C.: Anthropogenic emissions of methane in the United States., *Proc. Natl. Acad. Sci. U.*  
837 *S. A.*, 110(50), 20018–22, doi:10.1073/pnas.1314392110, 2013.

838 Montzka, S. a, Dlugokencky, E. J. and Butler, J. H.: Non-CO<sub>2</sub> greenhouse gases and climate change.,  
839 *Nature*, 476(7358), 43–50, doi:10.1038/nature10322, 2011.

840 Myhre, G., D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee,  
841 B. Mendoza, T. Nakajima, A. Robock, G. Stephens, T. Takemura and H. Zhang, 2013: Anthropogenic and  
842 Natural Radiative Forcing. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working*  
843 *Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F.,  
844 D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley  
845 (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 659–740,  
846 doi:10.1017/CBO9781107415324.018.

847 Nam, E., Jensen, T. and Wallington, T.: Methane emissions from vehicles, *Environ. Sci. ...*, 38(7), 2005–  
848 2010 [online] Available from: <http://pubs.acs.org/doi/abs/10.1021/es034837g> (Accessed 16 April 2013),  
849 2004.

850 NRC: National Research Council Report - Verifying Greenhouse Gas Emissions: Methods to Support  
851 International Climate Agreements, The National Academies Press, Washington, DC, 124 pp., 2010.

852 Norris, G., Vedantham, R., Wade, K., Brown, S., Prouty, J., and Foley, C.: EPA Positive Matrix  
853 Factorization (PMF) 3.0 Fundamentals and User Guide, Washington DC, USA, 2008.

854 Ngwabie, N. M., Schade, G. W., Custer, T. G., Linke, S. and Hinz, T.: Abundances and Flux Estimates of  
855 Volatile Organic Compounds from a Dairy Cowshed in Germany, *J. Environ. Qual.*, 37(2), 565,  
856 doi:10.2134/jeq2006.0417, 2008.

857 O’Keefe, A.: 1998: integrated cavity output analysis of ultra-weak absorption, *Chem. Phys. Lett.*, 293,  
858 331–336, doi:10.1016/S0009-2614(98)00785-4, 1998.

859 Owen, J. J. and Silver, W. L.: Greenhouse gas emissions from dairy manure management: a review of  
860 field-based studies., *Glob. Chang. Biol.*, doi:10.1111/gcb.12687, 2014.

861 Paatero, P. (1997), Least squares formulation of robust non-negative factor analysis, *Chemom. Intell.*  
862 *Lab. Syst.*, 37(1), 23–35, doi:10.1016/S0169-7439(96)00044-5.

863 Paatero, P., and U. Tapper (1994), Positive matrix factorization: A non-negative factor model with  
864 optimal utilization of error estimates of data values, *Environmetrics*, 5(April 1993), 111–126.

865 Parameswaran, K. R., Rosen, D. I., Allen, M. G., Ganz, A. M., and Risby, T. H.: O<sub>2</sub>-axis integrated

866 cavity output spectroscopy with a mid-infrared interband cascade laser for real-time breath ethane  
867 measurements., *Appl. Optics*, 48, B73–79, doi: [10.1364/AO.48.000B73](https://doi.org/10.1364/AO.48.000B73), 2009.

868 Paul, J. B., Scherer, J. J., O'Keefe, A., Lapson, L., Anderson, J. R., Gmachl, C. F., Capasso, F., and Cho, A. Y.:  
869 Infrared cavity ringdown and integrated cavity output spectroscopy for trace species monitoring, 2002,  
870 1-11, *Proc. SPIE 4577, Vibrational Spectroscopy-based Sensor Systems*, 1 (February 13, 2002);  
871 doi:10.1117/12.455722

872 Peischl, J., Ryerson, T. B., Brioude, J., Aikin, K. C., Andrews, a. E., Atlas, E., Blake, D., Daube, B. C., de  
873 Gouw, J. a., Dlugokencky, E., Frost, G. J., Gentner, D. R., Gilman, J. B., Goldstein, a. H., Harley, R. a.,  
874 Holloway, J. S., Kofler, J., Kuster, W. C., Lang, P. M., Novelli, P. C., Santoni, G. W., Trainer, M., Wofsy, S. C.  
875 and Parrish, D. D.: Quantifying sources of methane using light alkanes in the Los Angeles basin,  
876 *California, J. Geophys. Res. Atmos.*, 118(10), 4974–4990, doi:10.1002/jgrd.50413, 2013.

877 Peischl, J., Ryerson, T. B., Holloway, J. S., Trainer, M., Andrews, a. E., Atlas, E. L., Blake, D. R., Daube, B.  
878 C., Dlugokencky, E. J., Fischer, M. L., Goldstein, a. H., Guha, a., Karl, T., Kofler, J., Kosciuch, E., Misztal, P.  
879 K., Perring, a. E., Pollack, I. B., Santoni, G. W., Schwarz, J. P., Spackman, J. R., Wofsy, S. C. and Parrish, D.  
880 D.: Airborne observations of methane emissions from rice cultivation in the Sacramento Valley of  
881 California, *J. Geophys. Res.*, 117, D00V25, doi:10.1029/2012JD017994, 2012.

882 Pétron, G., Frost, G., Miller, B. R., Hirsch, A. I., Montzka, S. a., Karion, A., Trainer, M., Sweeney, C.,  
883 Andrews, A. E., Miller, L., Kofler, J., Bar-Ilan, A., Dlugokencky, E. J., Patrick, L., Moore, C. T., Ryerson, T.  
884 B., Siso, C., Kolodzey, W., Lang, P. M., Conway, T., Novelli, P., Masarie, K., Hall, B., Guenther, D., Kitzis, D.,  
885 Miller, J., Welsh, D., Wolfe, D., Neff, W. and Tans, P.: Hydrocarbon emissions characterization in the  
886 Colorado Front Range: A pilot study, *J. Geophys. Res.*, 117(D4), D04304, doi:10.1029/2011JD016360,  
887 2012.

888 Polissar, A. V., P. K. Hopke, P. Paatero, W. C. Malm, and J. F. Sisler (1998), *Atmospheric aerosol over*  
889 *Alaska: 2. Elemental composition and sources*, *J. Geophys. Res.*, 103(D15), 19045–19057,  
890 doi:10.1029/98JD01212.

891 Ryerson, T. B., Andrews, a. E., Angevine, W. M., Bates, T. S., Brock, C. a., Cairns, B., Cohen, R. C., Cooper,  
892 O. R., de Gouw, J. a., Fehsenfeld, F. C., Ferrare, R. a., Fischer, M. L., Flagan, R. C., Goldstein, a. H., Hair, J.  
893 W., Hardesty, R. M., Hostetler, C. a., Jimenez, J. L., Langford, a. O., McCauley, E., McKeen, S. a., Molina,  
894 L. T., Nenes, a., Oltmans, S. J., Parrish, D. D., Pederson, J. R., Pierce, R. B., Prather, K., Quinn, P. K.,  
895 Seinfeld, J. H., Senff, C. J., Sorooshian, a., Stutz, J., Surratt, J. D., Trainer, M., Volkamer, R., Williams, E. J.  
896 and Wofsy, S. C.: The 2010 California Research at the Nexus of Air Quality and Climate Change (CalNex)  
897 field study, *J. Geophys. Res. Atmos.*, 118(11), 5830–5866, doi:10.1002/jgrd.50331, 2013.

898 Santoni, G. W.; Xiang, B.; Kort, E. A.; Daube, B.; Andrews, A. E.; Sweeney, C.; Wecht, K.; Peischl, J.;  
899 Ryerson, T. B.; Angevine, W. M.; Trainer, M.; Nehrkorn, T.; Eluszkiewicz, J.; Wofsy, S. C. (2012)  
900 California's Methane Budget derived from CalNex P-3 Aircraft Observations and the WRF-STILT  
901 Lagrangian Transport Model. AGU Fall Meeting 2012, San Francisco, CA.

902 Shaw, S. L., Mitloehner, F. M., Jackson, W., Depeters, E. J., Fadel, J. G., Robinson, P. H.,  
903 Holzinger, R., and Goldstein, A. H.: Volatile organic compound emissions from dairy cows

904 5 and their waste as measured by proton-transfer-reaction mass spectrometry., *Environ. Sci. Technol.*,  
905 41, 1310–1316, doi:10.1021/es061475e, 2007.

906 Slowik, J. G., Vlasenko, a., McGuire, M., Evans, G. J. and Abbatt, J. P. D.: Simultaneous factor analysis of  
907 organic particle and gas mass spectra: AMS and PTR-MS measurements at an urban site, *Atmos. Chem.*  
908 *Phys.*, 10(4), 1969–1988, doi:10.5194/acp-10-1969-2010, 2010.

909 Smith, P., D. Martino, Z. Cai, D. Gwary, H. Janzen, P. Kumar, B. McCarl, S. Ogle, F. O’Mara, C. Rice, B.  
910 Scholes, O. Sirotenko, 2007: Agriculture. In *Climate Change 2007: Mitigation. Contribution of Working*  
911 *Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [B. Metz,  
912 O.R. Davidson, P.R. Bosch, R. Dave, L.A. Meyer (eds)], Cambridge University Press, Cambridge, United  
913 Kingdom and New York, NY, USA.

914 Sun, H., Trabue, S. L., Scoggin, K., Jackson, W. a., Pan, Y., Zhao, Y., Malkina, I. L., Koziel, J. a. and  
915 Mitloehner, F. M.: Alcohol, Volatile Fatty Acid, Phenol, and Methane Emissions from Dairy Cows and  
916 Fresh Manure, *J. Environ. Qual.*, 37(2), 615, doi:10.2134/jeq2007.0357, 2008.

917 Tanner, R. L. and Zielinska, B.: Determination of the biogenic emission rates of species contributing to  
918 VOC in the San Joaquin Valley OF California, *Atmos. Environ.*, 28(6), 1113–1120, doi:10.1016/1352-  
919 2310(94)90288-7, 1994.

920 Ulbrich, I. M. et al. (2009), Interpretation of organic components from Positive Matrix Factorization of  
921 aerosol mass spectrometric data, *Atmos. Chem. Phys.*, 9(9), 2891–2918, doi:10.5194/acp-9-2891-2009.

922 Williams, B. J., Goldstein, a. H., Kreisberg, N. M., Hering, S. V., Worsnop, D. R., Ulbrich, I. M., Docherty, K.  
923 S. and Jimenez, J. L.: Major components of atmospheric organic aerosol in southern California as  
924 determined by hourly measurements of source marker compounds, *Atmos. Chem. Phys.*, 10(23), 11577–  
925 11603, doi:10.5194/acp-10-11577-2010, 2010.

926 Wunch, D., Wennberg, P. O., Toon, G. C., Keppel-Aleks, G. and Yavin, Y. G.: Emissions of greenhouse  
927 gases from a North American megacity, *Geophys. Res. Lett.*, 36(15), L15810,  
928 doi:10.1029/2009GL039825, 2009.

929 Zhao, C., Andrews, A. E., Bianco, L., Eluszkiewicz, J., Hirsch, A., MacDonald, C., Nehrkorn, T. and Fischer,  
930 M. L.: Atmospheric inverse estimates of methane emissions from Central California, *J. Geophys. Res.*,  
931 114(D16), D16302, doi:10.1029/2008JD011671, 2009.

932

933

934

935

936

937

938 **Table 1.** PMF dataset with total samples (N) and mixing ratio range (in ppt).

Class	Compound	N	1st percentile	99th percentile	Background
GHG	CH <sub>4</sub> <sup>a,c</sup>	619	1855.0	3400.8	1813.6
	CO <sub>2</sub> <sup>b,c</sup>	619	390.8	468.3	390.0
	N <sub>2</sub> O <sup>a,d</sup>	490	323.3	339.5	323.2
combustion tracer	CO <sup>a,d</sup>	653	118.9	330.6	102.1
straight chain alkanes	propane	592	580.8	30839.0	455.5
	n-butane	587	96.4	12649.0	73.6
	n-pentane	647	93.2	3805.4	64.4
	n-hexane	647	23.1	960.5	17.2
	dodecane	643	1.56	54.3	0
branched alkanes	isopentane	646	165.4	7490.5	100.4
	2,3-dimethylbutane	650	52.5	1747.7	41.1
	2,5-dimethylhexane	651	2.37	145.8	0
	isooctane	647	16.6	476.9	12.3
	4-ethylheptane	651	1.45	52.6	0
	dimethyl undecane	643	0.46	24.9	0
cyclo alkanes	methylcyclopentane	647	23.3	1329.6	20.3
	methylcyclohexane	649	8.10	813.9	0
	ethylcyclohexane	651	1.78	169.1	0
alkenes	propene	592	34.7	3299.9	28.6
	isobutene	595	16.7	422.1	10.7
aromatics	toluene	647	48.8	1749.5	33.1
	ethylbenzene	647	5.83	282.0	0
	m,p-xylene	647	21.8	1127.1	21.8
	o-xylene	647	4.31	405.0	0
	cumene	640	0.55	22.8	0
	1-ethyl-3,4-methylbenzene	651	2.22	358.6	0
	p-cymene	649	0.84	93.9	0
	indane	647	0.45	27.9	0
	1,3-dimethyl-4-ethylbenzene	635	0.46	23.9	0
unsaturated aldehyde	naphthalene	654	0.44	19.9	0
	methacrolein	573	14.2	337.0	0
alcohol	methanol	429	2636.81	88691.8	1085.2
	ethanol	598	1021.93	65759.8	1021.9
	isopropyl alcohol	583	25.7	2001.0	25.7
ketone	acetone	663	142.9	3505.8	142.9
	methyl ethyl ketone	605	8.55	1111.2	0
	methyl isobutyl ketone	629	2.03	71.9	0
aldehyde	propanal	636	3.68	140.8	0
	butanal	589	1.72	35.1	0
biogenics	isoprene	651	9.70	310.0	0
	alpha-pinene	740	1.67	525.8	0
	d-limonene	641	1.10	357.1	0
	nopinone	614	0.78	89.5	0
	alpha-thujene	591	0.52	23.8	0
	camphene	645	0.72	100.3	0
chloroalkanes	chloroform	647	34.1	209.3	31.6
	tetrachloroethylene	641	3.41	120.9	0
	1,2-dichloroethane	640	20.6	103.8	20.6
	1,2-dichloropropane	627	2.40	28.4	0
sulfides	carbon disulfide	610	7.84	133.7	0
thiol	ethanethiol	491	4.54	685.8	0

<sup>a</sup> parts per billion volume (ppb)

<sup>b</sup> parts per million (ppm)

<sup>c</sup> measured using LGR Fast Green House Gas Analyzer

<sup>d</sup> measured using LGR N<sub>2</sub>O/CO analyzer

**Table 2.** Comparison of light alkane ratios to propane ( $\text{gC gC}^{-1}$ ) from PMF fugitive and evaporative factor with those from other PMF studies and oil and gas operations.

Study	Source	propane	n-butane	n-pentane	n-hexane	isopentane
Bakersfield PMF evaporative and fugitive factor <sup>a</sup>	This study	1	$0.52 \pm 0.02$	$0.18 \pm 0.01$	$0.06 \pm 0.003$	$0.33 \pm 0.02$
Bakersfield petroleum operations source profile <sup>b</sup>	Gentner et al. (2014a)	1	$0.53 \pm 0.1$	$0.09 \pm 0.02$	$0.04 \pm 0.01$	$0.08 \pm 0.02$
Mexico city PMF LPG factor <sup>c</sup>	Bon et al. (2011)	1	0.5 (0.4 - 0.7)	0.05 (0.04 - 0.07)	0.02 (0.02 - 0.03)	0.07 (0.06 - 0.1)
Wattenberg field BAO, Colorado <sup>d</sup>	Gilman et al. (2013)	1	$0.75 \pm 1.37$	$0.32 \pm 0.6$	$0.08 \pm 0.13$	$0.28 \pm 0.52$
Wattenberg field BAO, Colorado <sup>e</sup>	Petron et al. (2012)	1	0.58 - 0.65	0.22 - 0.31	NA	0.22 - 0.31
PMF natural gas and evaporation factor, Houston Ship Channel <sup>f</sup>	Leuchner and Rappengluck (2010)	1	0.33	0.27	0.12	0.37
PMF natural gas factor, Houston Ship Channel <sup>h</sup>	Buzcu and Fraser (2006)	1	$0.67 \pm 0.16$	$0.07 \pm 0.18$	NA	NA

<sup>a</sup> Uncertainties calculated from propagation of errors (standard deviations) over FPEAK range of -1.6 to 0.4.

<sup>b</sup> Ratios calculated from Table 4, Gentner et al., 2014a; uncertainties defined as  $\pm 20\%$  to account for variability in oil well data.

<sup>c</sup> Uncertainties calculated from propagation of uncertainties over FPEAK range of -3 to 3.

<sup>d</sup> Emission ratios derived from multivariate regression analysis; error bars derived from propagation of uncertainty using mean and standard deviation of samples.

<sup>e</sup> Range over 5 regressions conducted over data collected in different seasons and from mobile lab samples.

<sup>f</sup> Ratios derived from mean and standard deviations, with propagation of uncertainty.

<sup>g</sup> Estimated from Figure 2, Leuchner and Rappengluck, 2010.

<sup>h</sup> Estimated from Figure 2, Buzcu and Fraser, 2006.



**Table 3.** Comparison of hydrocarbon ratios to toluene ( $\text{gC gC}^{-1}$ ) from PMF vehicle emission factor with similar ratios from other California specific studies.

Study	Bakersfield PMF vehicle emissions factor <sup>a</sup>	Bakersfield gasoline source profile <sup>b,c</sup>	Riverside liquid gasoline profile <sup>e</sup>	CalNex Los Angeles ambient emission ratios <sup>g</sup>
Source	This study	Gentner et al. (2012)	Gentner et al. (2009)	Borbon et al. (2013)
CH <sub>4</sub>	$8.1 \pm 2.1$	NA	NA	NA
CO	$14.0 \pm 0.4$	NA	NA	45
toluene	1	1	1	1
isopentane	$0.69 \pm 0.01$	$0.77 \pm 0.04$	0.64-0.84	1.95
isooctane	$0.29 \pm 0.03$	$0.34 \pm 0.02$	0.64-0.80	NA
n-dodecane	$0.03 \pm 0.001$	$(0.02 \pm 0.007)^d$	NA	NA
methylcyclopentane	$0.24 \pm 0.01$	$0.32 \pm 0.02$	NA	NA
ethyl benzene	$0.17 \pm 0.01$	$0.14 \pm 0.01$	NA	0.2
m/p - xylene	$0.65 \pm 0.01$	$0.65 \pm 0.03$	$(0.45-0.52)^f$	0.64
o - xylene	$0.22 \pm 0.01$	$0.23 \pm 0.01$	NA	0.24

<sup>a</sup> errors are standard deviation of 12 unique PMF solutions between FPEAK = -1.6 to +0.4; see section S2.

<sup>b</sup> derived from liquid gasoline fuel speciation profile (Table S9; Gentner et al., 2012).

<sup>c</sup> errors bars derived from propagation of uncertainties.

<sup>d</sup> derived by combining diesel fuel and gasoline speciation profile (Table S9 and S10; Gentner et al., 2012) and gasoline and diesel fuel sale data in Kern County (Table S1, Gentner et al., 2012).

<sup>e</sup> summer data.

<sup>f</sup> only m-xylene.

<sup>g</sup> derived from Linear Regression Fit slope of scatterplot from CalNex Pasadena supersite samples.

**Table 4.** Comparison of PMF dairy and livestock emission factors ( $\text{mmol mol}^{-1}$ ) with previous studies.

Study	Source	Cow/manure type (if applicable)	methanol / methane EF avg. (range)	ethanol / methane EF avg. (range)
PMF analysis of regional measurements	This study		15 - 47	9 - 32.2
Environmental chamber with cows and/or manure	Shaw et al. (2008)	Dry	3.2 (0.6 - 7.4)	NA
		Lactating	1.9 (0.8 - 3.6)	NA
Environmental chamber with cows and/or manure	Sun et al. (2008)	Dry	13.4 (4 - 25)	14.4 (11 - 19)
		Lactating	19.2 (15 - 25)	24.2 (18 - 32)
Cowshed with regular dairy operations (winter)	Ngwabie et al. (2008)		2.0 (1.6 - 2.4)	9.3 (4 - 16)
Cow stall area with regular dairy operations (summer)	Filipy et al. (2006)		NA	(42 - 127) <sup>a</sup>
Manure from cattle feedlot	Miller and Varel (2001)	Fresh (< 24 hr)	NA	14 <sup>b</sup>
		Aged (> 24 hr)		118 <sup>b</sup>
Measured slope of regression (CalNex 2010)	Gentner et al. (2014a)		7.4 (7 - 16) <sup>c</sup>	18 <sup>d</sup>
Sampling of dairy plumes from aircraft (CABERNET 2011)	Guha et al. ( <i>in prep</i> )		9.6 (9 - 30) <sup>c</sup>	NA

<sup>a</sup> calculated based on  $\text{CH}_4$  emission rate of  $4160 \mu\text{g cow}^{-1} \text{s}^{-1}$  for mid-lactating cows (Shaw et al., 2007).

<sup>b</sup> calculated based on  $\text{CH}_4$  emission rate of  $4160 \mu\text{g cow}^{-1} \text{s}^{-1}$  for mid-lactating cows (Shaw et al., 2007) and ethanol emission rate for fresh and aged manure of 175 and  $1223 \mu\text{g cow}^{-1} \text{s}^{-1}$ , respectively, derived by Filipy et al. (2006).

<sup>c</sup> slope of regression with range of measured slopes (in parentheses) from sampling of dairy plumes by aircraft.

<sup>d</sup> ground site data; lower limit of slope of non-vehicular ethanol versus methane

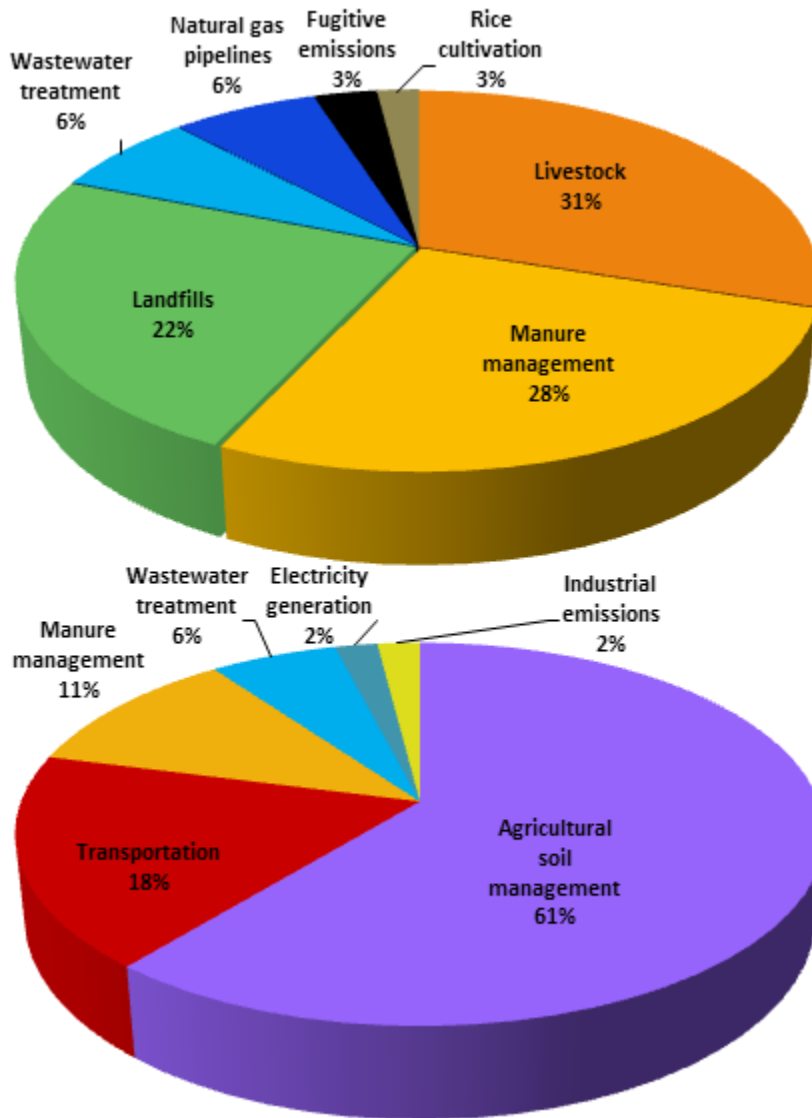
**Table 5.** Comparison of PMF agricultural and soil management emission factor for acetone versus methanol ( $\text{gC gC}^{-1}$ ) with ratios of basal emission factors generated for major crops grown in the Kern County. Errors denote standard deviations computed by propagation of uncertainty.

Bakersfield PMF agricultural and soil management factor	Almond greenhouse summer 2008	Table grape greenhouse summer 2008	Pistachio greenhouse summer 2008	Navel oranges greenhouse summer 2008 <sup>a</sup>	Valencia oranges greenhouse summer 2008
This study	Gentner et al. (2014b)	Gentner et al. (2014b)	Gentner et al. (2014b)	Fares et al. (2011)	Fares et al. (2012)
$0.58 \pm 0.37$	$0.14 \pm 0.2$	$0.04 \pm 0.02$	$0.5 \pm 0.6$	$0.57 \pm 0.1$	$0.5 \pm 0.3$

<sup>a</sup> branch with flowers not removed.

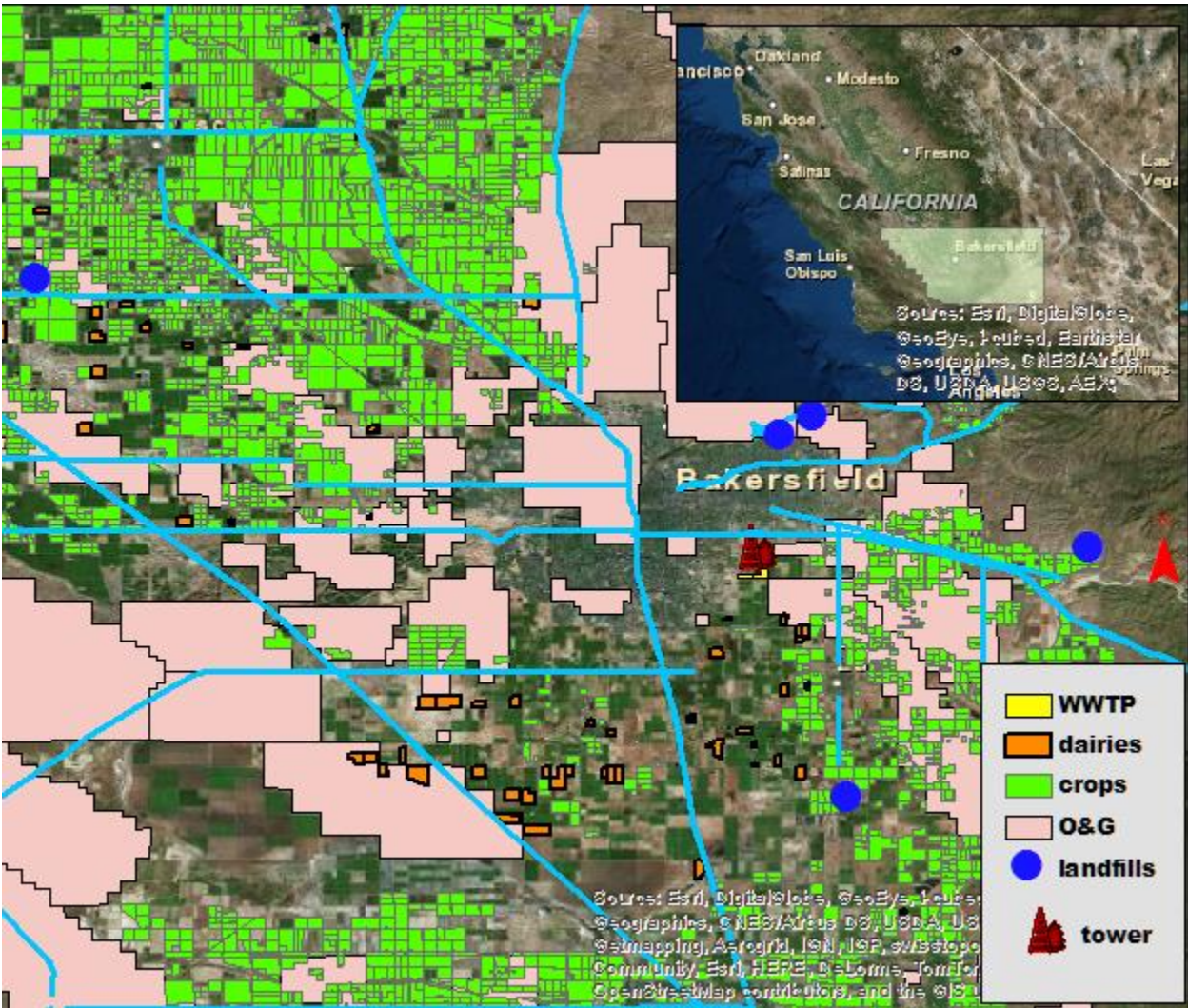
Figures

Figure 1



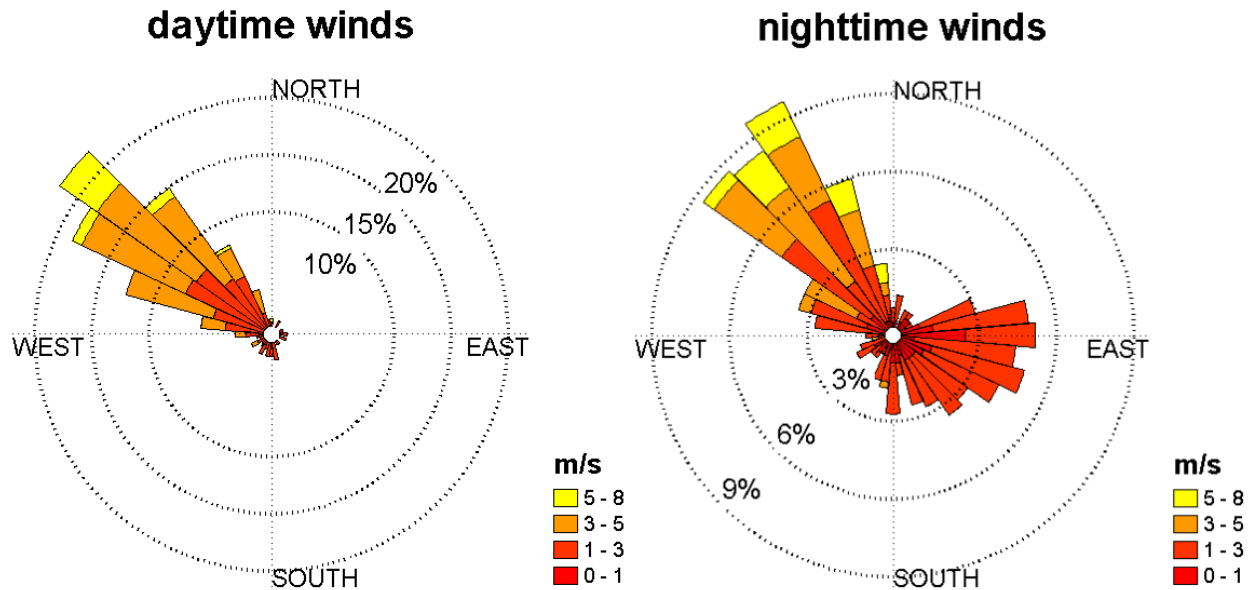
**Figure 1.** 2011 California emission inventory for (a) methane (CH<sub>4</sub>) - 32.5 million ton CO<sub>2</sub>eq at GWP = 25; and (b) nitrous oxide (N<sub>2</sub>O) - 13.4 million ton CO<sub>2</sub>eq at GWP = 298. (Source: CARB GHG Inventory Tool, Aug 2013)

**Figure 2**

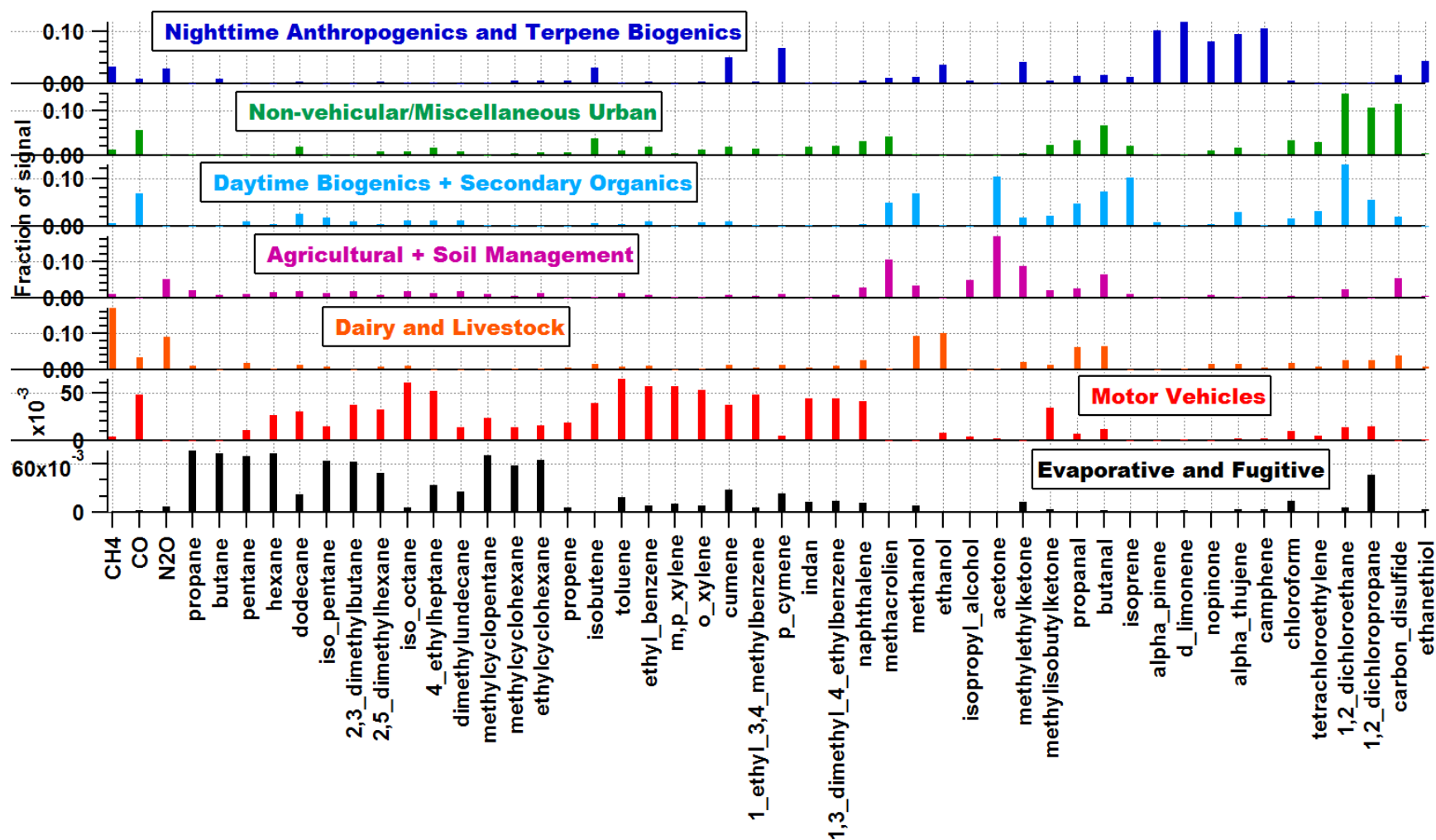


**Figure 2.** Map of potential sources of methane and nitrous oxide in/around the city of Bakersfield and the surrounding parts of the valley. The inset map is a zoomed out image of the southern part of San Joaquin Valley (SJV) with location of Kern County superimposed. The light blue lines mark the highways, WWTP stands for waste water treatment plant, and O&G stands for oil and gas fields. The location of the CalNex experiment site is marked by the ‘tower’ symbol.

968 **Figure 3**

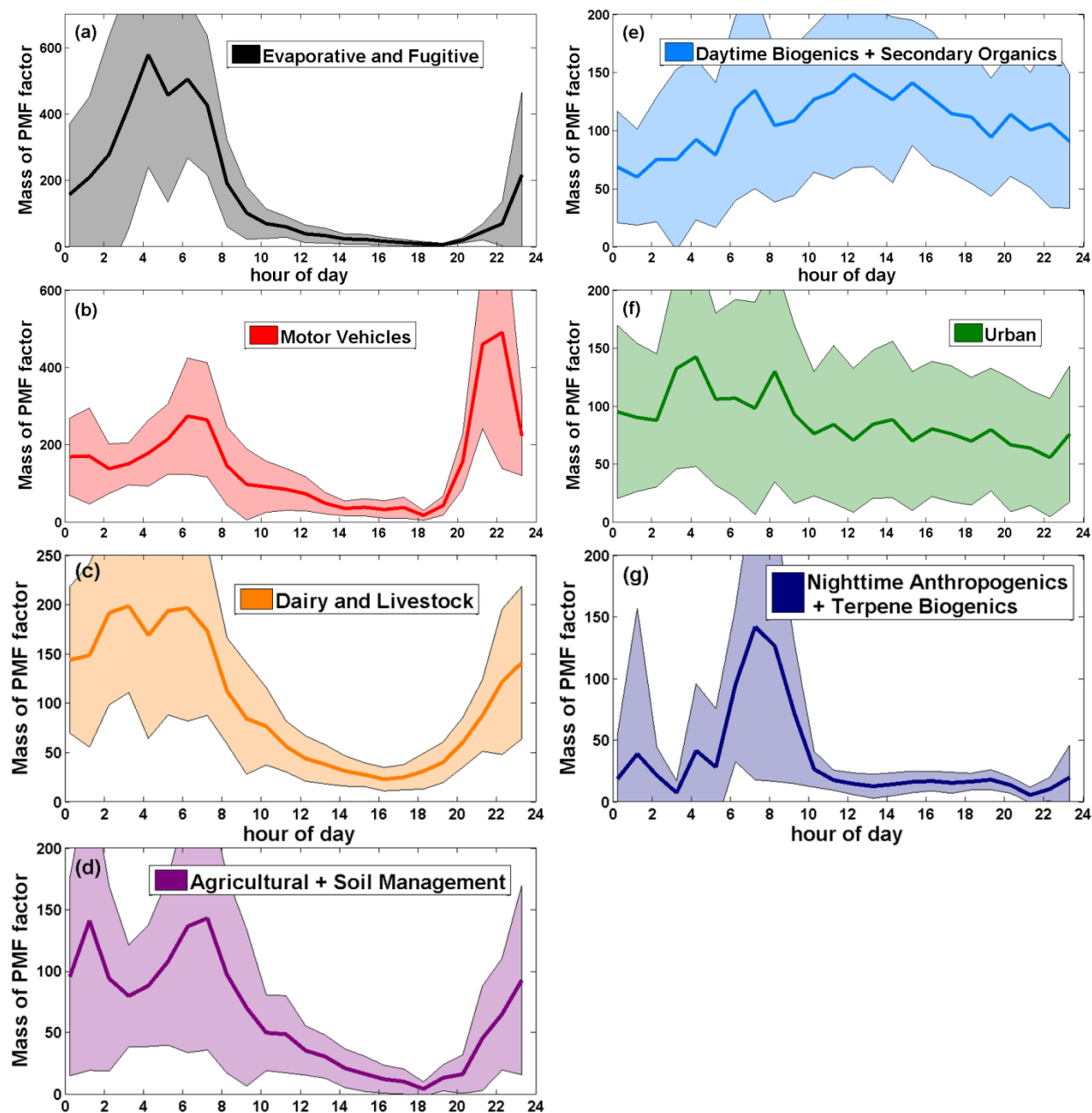


970 **Figure 3.** Wind rose plots showing mean wind direction measured at the site during **(left)** day time  
 971 (07:00-16:00 hour), and **(right)** nighttime (17:00-06:00 hour) during the experiment period in summer  
 972 2010. The concentric circles represent the percentage of total observations; each colored pie represents a  
 973 range of 10° while the colors denote different wind speed ranges.



975  
 976 **Figure 4.** Source profile of the seven factors derived using PMF. The source factors are evaporative and fugitive, motor vehicles, dairy and  
 977 livestock, agricultural + soil management, daytime biogenics + secondary organics, urban, and nighttime anthropogenics + terpene biogenics. The  
 978 y-axis represents the normalized fraction of mass in each source factor, while the x-axis lists all the chemical species included in the PMF analysis.

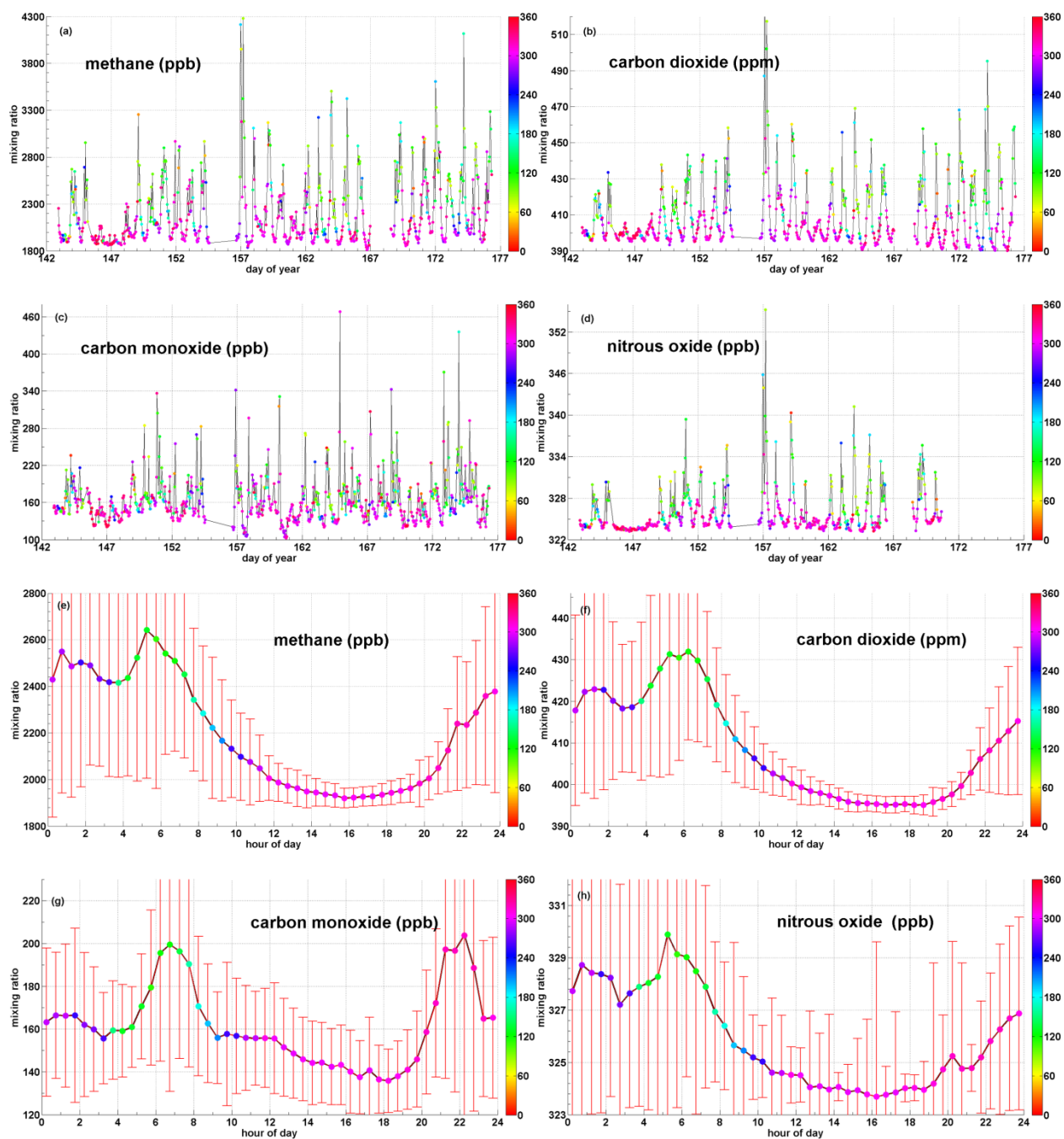
979 **Figure 5**



980  
981 **Figure 5.** Mean hourly diurnal plots of PMF source factor concentration enhancements for (a)  
982 evaporative and fugitive, (b) motor vehicles, (c) dairy and livestock, (d) agricultural + soil management,  
983 (e) daytime biogenics and secondary organics, (f) non-vehicular/miscellaneous urban and (g) nighttime  
984 anthropogenics + terpene biogenics. The y-axis represents sum of normalized mass concentrations from  
985 all tracers contributing to the factor. The x-axis is hour of day (local time). The solid lines represent the  
986 mean and the shaded area represents the standard deviation (variability) at each hour.

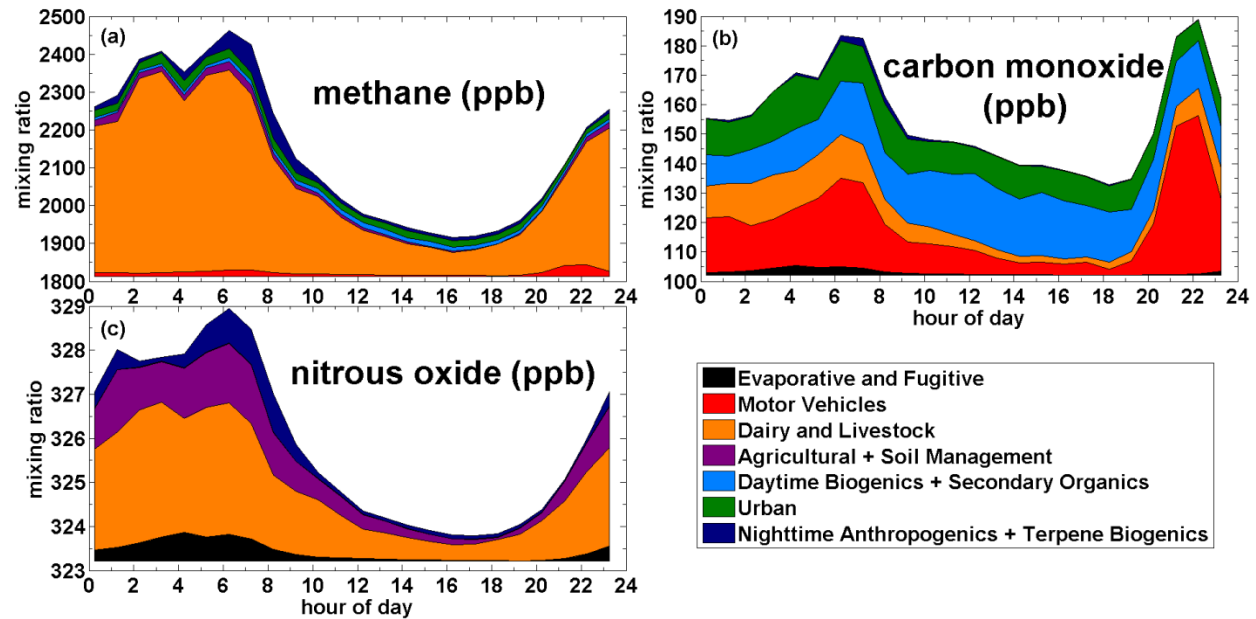


987 **Figure 6**



988  
 989 **Figure 6.** Time series of (a) CH<sub>4</sub>, (b) CO<sub>2</sub>, (c) CO, and (d) N<sub>2</sub>O obtained from 30-min averages from  
 990 May 15- June 30, 2010. The color bar indicates the average wind direction during each 30-min period.  
 991 Mixing ratios plotted as average diurnal cycles for (e) CH<sub>4</sub>, (f) CO<sub>2</sub>, (g) CO and (h) N<sub>2</sub>O along with wind  
 992 direction. The curve and the red whiskers represent the mean and the standard deviations about the mean,  
 993 respectively.

**Figure 7**



**Figure 7.** Diurnal plot of PMF derived (a) CH<sub>4</sub>, (b) CO, and (c) N<sub>2</sub>O concentrations sorted by PMF source category. The legend on the bottom right shows the names of the PMF source factor which each color represents. The PMF derived enhancements from each source have been added to the background concentrations.