1 **Emissions of organic carbon and methane from petroleum and dairy operations in**
2 **California's San Joaquin Valley**
3

4 Drew R. Gentner[1]*, Trevor B. Ford[2], Abhinav Guha[3], Kelsey Boulanger[1,i], Jerome Brioude[4,5],
5 Wayne M. Angevine[4,5], Joost A. de Gouw[4,5], Carsten Warneke[4,5], Jessica B. Gilman[4,5], Tom B.
6 Ryerson[4,5], Jeff Peischl[4,5], Simone Meinardi[6], Donald R. Blake[6], Elliot Atlas[7], William A.
7 Lonneman[8], Tadeusz E. Kleindienst[8], Melinda R. Beaver[9,ii], Jason M. St. Clair[9], Paul O.
8 Wennberg[9], Trevor C. VandenBoer[10,iii], Milos Z. Markovic[10,iv], Jennifer G. Murphy[10], Robert A.
9 Harley[1], and Allen H. Goldstein[1,3]
10

11 [1] Department of Civil and Environmental Engineering, University of California, Berkeley, CA
12 94720, USA.
13 [2] Department of Chemistry, University of California, Berkeley, CA, 94720, USA.
14 [3] Department of Environmental Science, Policy and Management, University of California,
15 Berkeley, CA, 94720, USA.
16 [4] Cooperative Institute for Research in Environmental Sciences, University of Colorado,
17 Boulder, CO 80309, USA.
18 [5] Chemical Sciences Division, Earth System Research Laboratory, National Oceanic and
19 Atmospheric Administration, Boulder, CO 80305, USA.
20 [6] Department of Chemistry, University of California, Irvine, CA, 92697, USA.
21 [7] Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL,
22 33149, USA
23 [8] National Exposure Research Laboratory, Environmental Protection Agency, Research Triangle
24 Park, NC, USA
25 [9] California Institute of Technology, Pasadena, CA, USA.
26 [10] Department of Chemistry, University of Toronto, Toronto, ON, CA.
27 [i] now at: Department of Civil and Environmental Engineering, Massachusetts Institute of
28 Technology, Cambridge, MA 02139, USA.
29 [ii] now at: National Exposure Research Laboratory, Environmental Protection Agency, Research
30 Triangle Park, NC, USA
31 [iii] now at: Department of Chemistry, Memorial University of Newfoundland, NL, CA.
32 [iv] now at: Chemical Sciences Division, Earth System Research Laboratory, National Oceanic and
33 Atmospheric Administration, Boulder, CO 80305, USA.; Cooperative Institute for Research in
34 Environmental Sciences, University of Colorado, Boulder, CO 80309, USA.
35

36 *Corresponding author: Drew Gentner (drew.gentner@yale.edu), now at: Department of
37 Chemical & Environmental Engineering, Yale University, New Haven, CT, 06511, USA
38

41

42 **Abstract**

43       Petroleum and dairy operations are prominent sources of gas-phase organic compounds

44 in California's San Joaquin Valley. It is essential to understand the emissions and air quality

45 impacts of these relatively understudied sources, especially for oil/gas operations in light of

46  increasing U.S. production. Ground site measurements in Bakersfield and regional aircraft

47  measurements of reactive gas-phase organic compounds and methane were part of the CalNex

48  (California Research at the Nexus of Air Quality and Climate Change) project to determine the

49  sources contributing to regional gas-phase organic carbon emissions. Using a combination of

50  near-source and downwind data, we assess the composition and magnitude of emissions, and

51  provide average source profiles. To examine the spatial distribution of emissions in the San

52  Joaquin Valley, we developed a statistical modeling method with the FLEXPART-WRF

53  transport and meteorological model using ground-based data. We present evidence for large

54  sources of paraffinic hydrocarbons from petroleum operations and oxygenated compounds from

55  dairy (and other cattle) operations. In addition to the small straight-chain alkanes typically

56  associated with petroleum operations, we observed a wide range of branched and cyclic alkanes,

57  most of which have limited previous in situ measurements or characterization in petroleum

58  operation emissions. Observed dairy emissions were dominated by ethanol, methanol, acetic

59  acid, and methane. Dairy operations were responsible for the vast majority of methane emissions

60  in the San Joaquin Valley; observations of methane were well-correlated with non-vehicular

61  ethanol, and multiple assessments of the spatial distribution of emissions in the San Joaquin

62  Valley highlight the dominance of dairy operations for methane emissions. The petroleum

63  operations source profile was developed using the composition of non-methane hydrocarbons in

64  unrefined natural gas associated with crude oil. The observed source profile is consistent with

65  fugitive emissions of condensate during storage or processing of associated gas following

66  extraction and methane separation. Aircraft observations of concentration hotspots near oil wells

67  and dairies are consistent with the statistical source footprint determined via our FLEXPART-

68  WRF-based modeling method and ground-based data. We quantitatively compared our

69  observations at Bakersfield to the California Air Resources Board emission inventory and find

70  consistency for relative emission rates of reactive organic gases between the aforementioned

71  sources and motor vehicles in the region. We estimate that petroleum and dairy operations each

72  comprised 22% of anthropogenic non-methane organic carbon at Bakersfield and were each

73  responsible for 8-13% of potential precursors to ozone. Yet, their direct impacts as potential

74  secondary organic aerosol (SOA) precursors were estimated to be minor for the source profiles

75  observed in the San Joaquin Valley.

76

## 1. Introduction

California's San Joaquin Valley contains a large density of dairy farms and is an important region for oil and natural gas production in the United States. Both sources are prominent in the California Air Resources Board (CARB) emission inventory of reactive organic gases (ROG) in the San Joaquin Valley (California Air Resources Board, 2010). Recent work has described large emissions and impacts from new oil/gas operations with increased U.S. production (Petron et al., 2012; Gilman et al., 2013; Carter and Seinfeld, 2012; Schnell et al., 2009; Kemball-Cook et al., 2010; Pacsi et al., 2013). Petroleum operations include extraction, storage, transport, and processing; all of which can have varying degrees of fugitive emissions of methane and other gas-phase organic carbon, such as Volatile Organic Compounds (VOCs) (Leuchner and Rappengluck, 2010; Buzcu and Fraser, 2006, Katzenstein et al., 2003; Petron et al., 2012; Gilman et al., 2013). Crude oil and unrefined natural gas are composed of a suite of organic compounds that span a range of vapor pressures, and are either produced by thermogenic or biogenic processes in the reservoirs (USGS, 2007; Ryerson et al., 2011). Thermogenic gas is geochemically produced via the cracking of larger compounds in oil and can either be termed associated or non-associated depending on the presence of oil (USGS, 2007). The vast majority of wells in the San Joaquin Valley are oil wells and most have associated gas, also known as wet thermogenic gas (USGS, 2007). Thermogenic wet gas is predominately found in oil wells and contains substantial amounts of non-methane hydrocarbons ranging 3-40% $C_2$ and greater content (Table 1) (USGS, 2007). The San Joaquin Valley has historically been an active region for oil/gas production. Currently, crude oil production in Kern County, located at the Southern end of the San Joaquin Valley, is 450,000 barrels day$^{-1}$, which represents 69% of production within California and 8% of national production (U.S. EIA, 2010; California Energy Commission, 2010).

There have been several studies on fugitive emissions from oil and gas operations, including emissions from isolated facilities at oil or gas fields, extraction facilities using advanced recovery methods (i.e. hydraulic fracturing), and urban areas with industrial storage and processing facilities (Leuchner and Rappengluck, 2010; Buzcu and Fraser, 2006, Katzenstein et al., 2003; Petron et al., 2012; Gilman et al., 2013). These studies all provide important advances in the characterization of emissions from petroleum operations, but there is considerable variability between regions due to differences in reservoirs and production methods.

108  The specific equipment/processes, state/county regulations, and regional composition of crude

109  oil and natural gas are critical for determining the potential emission pathways and composition

110  of fugitive emissions. So, regional studies remain important to effectively characterize petroleum

111  operation sources.

112      Previous research on dairy farms and livestock operations has reported emissions of

113  methane, alcohols, carbonyls, esters, acids, and other organic hydrocarbons. Among these,

114  emissions are dominated by methane, methanol, ethanol, and acetic acid (Alanis et al., 2010;

115  Chung et al., 2010; Hafner et al., 2013; Howard et al., 2010a; Howard et al., 2010b; Malkina et

116  al., 2011; Sun et al., 2008; Shaw et al., 2007). Howard et al. (2010b) recently concluded that

117  emissions from dairy operations are dominant contributors to ozone production in California's

118  central valley (comprised of the San Joaquin Valley and the Sacramento Valley to the North), but

119  modeling studies suggest a larger role for VOC emissions from motor vehicles (Hu et al., 2012).

120  Methane and oxygenated organic compounds are emitted via several pathways and sources, all

121  co-located at dairies (and their farms). Silage processing/fermentation, bovine enteric

122  fermentation, and animal waste are among the most dominant sources (Alanis et al., 2010;

123  Chung et al., 2010; Hafner et al., 2013; Malkina et al., 2011; Sun et al., 2008; Shaw et al., 2007).

124  The composition of emissions from each of these sources is different and varies widely

125  depending on factors such as feed composition. The animal feed, also known as total mixed

126  rations, is typically comprised of corn and other grains (i.e. silage), with corn being the abundant

127  type in the U.S. (Hafner et al., 2013). The silage is fermented on-site in large piles and mixed

128  with various adjuncts (e.g. almond shells, fruit, fat). The site-by-site heterogeneity in feed

129  composition and the processing of both animal feed and waste leads to variability in the source

130  profile and emission ratios of organic compounds from dairy operations. This work aims to

131  reduce this uncertainty by estimating the average source profile for dairy operation emissions in

132  the San Joaquin Valley.

133      The objectives of this work are to examine the magnitude, chemical composition, and

134  spatial distribution of organic carbon emissions from petroleum and dairy operations in the San

135  Joaquin Valley. This is accomplished using multiple gas-phase organic carbon data sets from

136  stationary ground sites and aircraft platforms. Our approach includes the development of a

137  method to assess the spatial distribution of sources (i.e. a statistical source footprint) via ground

138  site measurements and meteorological modeling. We examine the relative abundance of

139  emissions from petroleum and dairy operations against other prominent anthropogenic sources in

140  the San Joaquin Valley, and evaluate their potential to impact air quality. We also provide a

141  quantitative assessment of petroleum and dairy operations emissions relative to motor vehicle

142  emissions in the CARB emission inventory.

143

144  **2. Materials & Methods**

145  **2.1 Measurement Sites and Instrumentation**

146      Gas-phase organics and other gases were measured May 18 - June 30, 2010 in

147  Bakersfield, CA during the CalNex (California Research at the Nexus of Air Quality and Climate

148  Change) project. The ground supersite (35.3463° N, 118.9654° W) was located in southeast

149  Bakersfield, a city in the southern San Joaquin Valley. With the exception of gas-sampling

150  canisters and ion chromatography to measure acids, measurements were made from the top of an

151  18 m tower. Measurements of a few light VOCs are included from canister measurements at

152  ground-level to further characterize the observed sources. Canisters were taken as 3-hour

153  averages in the morning (5-8 PST) and analyzed via U.S. Environmental Protection Agency

154  (EPA) methods for an array of organic compounds (Klouda et al., 2002). Supporting methane

155  measurements were made using integrated cavity output spectroscopy (Los Gatos Research, Fast

156  Greenhouse Gas Analyzer) with 1-min time resolution. Acetic acid and other acids were

157  measured using both Chemical Ionization Mass Spectrometry (CIMS) and Ambient Ion Monitor

158  - Ion Chromatography (AIM-IC). These two instruments were located at different heights on the

159  sampling tower in Bakersfield and had different measurement frequencies. With both sets of data

160  averaged to hourly resolution, the acetic acid data were well correlated to each other (r=0.84)

161  with a slope near unity. Details on their sampling and measurement methods have been

162  published previously (Crounse et al., 2006; Markovic et al., 2012).

163      As part of the CalNex project, measurements were also made from the National Oceanic

164  and Atmospheric Administration (NOAA) WD-P3 research aircraft. VOC canister samples were

165  collected on the aircraft and analyzed offline (Barletta et al., 2013). High time resolution data on

166  selected organic compounds and methane were collected on the aircraft using proton transfer

167  reaction mass spectrometry (PTR-MS) and a Picarro flight-ready greenhouse gas analyzer

168  (model 1301-m), respectively (de Gouw and Warneke, 2007; Peischl et al., 2012). High-

169  resolution data was averaged to 1-minute intervals and select flights in the central valley were

170 used to evaluate the spatial distribution of methane concentrations (flight dates: 5/7, 5/11, 5/12,

171 6/14, 6/16, 6/18, 2010).

172

173 **2.2 Source Apportionment Methods**

174 **2.2.1 Petroleum Operations**

175      Using six weeks of in situ VOC data from the Bakersfield ground site, we assessed

176 emissions from petroleum operations during spring and summer 2010.  Contributions to

177 observed VOC concentrations at the site from petroleum operations were determined (along with

178 other motor vehicle sources) using a source receptor model with chemical mass balancing and

179 effective variance weighting focused on hydrocarbon emissions from petroleum-related sources

180 (Gentner et al., 2012). The model used ten compounds emitted from the sources of interest

181 (petroleum operations, non-tailpipe gasoline emissions, gasoline exhaust, and diesel exhaust)

182 along with reliable information on the fractional composition of the ten compounds from each of

183 the sources (i.e. source profiles). The ten compounds used were dependent species, but the model

184 also calculated the predicted concentrations of all the independent compounds not included in the

185 model, but emitted by the petroleum-related sources and measured at the site.

186      The compounds used in the over-constrained (i.e. more tracer compounds than sources)

187 source receptor model were propane, n-butane, n-pentane, isopentane, m/p-xylene, o-xylene,

188 isooctane, n-nonane, n-undecane, and n-dodecane to model motor vehicle and petroleum

189 operation sources. Due to high background concentrations, measurements of propane and n-

190 butane were corrected by local background values of 500 and 100 pptv, respectively. The ten

191 tracer compounds were carefully selected because together they captured the dynamics of all

192 four petroleum-related sources. The atmospheric lifetimes of the most reactive species did not

193 bias the model since the vast majority of contributions (i.e. emissions) were within short

194 transport times to the site. The petroleum operations source had the longest transport times (up to

195 6 hours) from source to field site, which did not present a problem because that source was

196 represented and modeled by the least reactive species that had negligible degradation during

197 transport. Extensive details on these methods and model validation are described in detail in

198 Gentner et al. (2012).

199      A priori source profile information for the model was constructed using U.S. Geological

200 Survey data on associated thermogenic natural gas composition from wells in the San Joaquin

201  Valley (Table 1) (USGS, 2007) and regional gasoline/diesel fuel composition data (Gentner et al.,

202  2012). There was substantial variability between wells and sampling methods in the data

203  compiled by the USGS, so standard deviations for the petroleum operations source profile were

204  ±80-300%. Due to this large uncertainty, we represented the uncertainty for all the source

205  profiles in the model by standard errors (similar to the U.S. EPA CMB 8.2 model), defined as the

206  standard deviation divided by the square root of the sample size (N=49).

207  The source receptor model effectively modeled the compounds included in the initial

208  petroleum operations source profile (Table 1), but there were an array of hydrocarbons (not

209  among the compounds used in the model) that episodically exceed predicted concentrations

210  based on emissions from motor vehicles. Many of the excess hydrocarbon concentrations were

211  well correlated with each other and the petroleum operations source factor, likely indicating

212  emissions from the petroleum operations source. Emissions of additional compounds from

213  petroleum operations (not present in the initial limited petroleum gas profile) are derived from

214  the residual mass that is well-correlated with the petroleum operations source. The residuals, or

215  excess concentrations beyond contributions from motor vehicles, were filtered for values that

216  exceeded the uncertainties of model calculations, which are determined in part by the 10-20%

217  variability in gasoline and diesel fuel.

218  Similarly, we calculated the expected ethanol emissions from gasoline vehicles for hourly

219  data. Taking the difference between these predicted concentrations and total observed ethanol

220  results in non-vehicular ethanol concentrations that must be attributed to other ethanol sources,

221  but were not correlated with the petroleum operations source.

222

223  **2.2.2 Dairy Operations**

224  A reliable source profile for dairy operations in the San Joaquin Valley was not available

225  in the literature for all the compounds of interest in this study, so the source profile was

226  established using a mix of aircraft and ground measurements. The emission ratios of organic

227  compounds to methane were calculated using flight and ground data for compounds that had

228  evident, quantifiable emissions from dairy operations to construct the source profile. The ratio of

229  methanol to methane in dairy operation emissions was determined using 1-min aircraft data

230  points sampled in the plumes from farms and facilities in the San Joaquin Valley. Acetic acid

231  and ethanol ratios could not be determined using the flight data due to a lack of measurements

232 and spatial incongruence of canister to methane data, respectively. Ratios of these two
233 compounds to methane were determined using ground site data from Bakersfield. Dairies have
234 been shown in previous studies to be major sources of methane, methanol, ethanol, acetic acid,
235 and other oxygenated species; and there is a large concentration of dairies in the San Joaquin
236 Valley (Alanis et al., 2010; Chung et al., 2010; Howard et al., 2010a; Howard et al., 2010b;
237 Malkina et al., 2011; Sun et al., 2008; Shaw et al., 2007). So each compound is compared to
238 methane to methane via regression with close attention to enhancements form other sources that
239 may skew the observed dairy operations emission ratio.

240 Predicted concentrations in Bakersfield of methanol, ethanol, and acetic acid from dairy
241 operations were estimated using the determined emission ratios to methane and measurements of
242 methane at the Bakersfield ground site. A local background methane concentration of 1.87 ppmv
243 was subtracted prior to multiplication by the emission ratio. These predicted concentrations were
244 compared with observed concentrations to determine the fraction of each compound emitted
245 from dairy operations.

246 OH reactivities and ozone formation potentials reported in this paper are from literature
247 on OH reaction constants and maximum incremental reactivities (MIRs), respectively (Carter,
248 2007; Atkinson and Arey, 2003).

249

250 **2.3 Methods to Determine Spatial Distribution of Emissions**

251 Several methods are used in this work to assess the spatial distribution of organic carbon
252 sources. In addition to the use of aircraft data collected from the NOAA WD-P3 mobile platform
253 during the CalNex campaign, we developed a method that uses a Lagrangian transport and
254 meteorological model (FLEXPART-WRF) to calculate the distribution of air parcels (i.e. back
255 trajectory footprints) for each hourly sample prior to measurement at a ground site. We combine
256 these footprints with ambient compound data from the CalNex site to assess the spatial
257 distribution of emissions for a given compound in a region. Our method builds upon previous
258 techniques (i.e. TrajStat) to estimate source location(s) using ground site data and the Hybrid
259 Single Particle Lagrangian Integrated Trajectory Model (HYSPLIT) (Wang et al., 2009).

260 We generated 6- and 12-hour back-trajectory footprints with $4 \times 4$ km resolution for each
261 hourly sample using the FLEXPART Lagrangian transport model with WRF meteorological
262 modeling (Figure 1). Simulations were initiated from the top of the 18 m tower using WRF runs

263  EM4N in Angevine et al. (2012); further details on FLEXPART and WRF modeling can also be

264  found in Brioude et al. (2012) and Metcalf et al. (2012). Here, we integrate this

265  transport/meteorological model with statistical back-trajectory analysis to explore the

266  distribution and relative magnitude of gas-phase organic carbon sources at ground level.

267  The back trajectory footprint produced by FLEXPART-WRF represents the area where

268  the air parcel(s) of interest (i.e. a 30-min VOC sample) contacts the surface layer. The statistical

269  source footprint (the final output) represents the calculated distribution of ground-level emissions.

270  Utilizing this concentration-weighted trajectory analysis allows us to find the emissions potential

271  of every point in a region, which is represented by the average concentration of a compound in

272  each cell ($\overline{C_{ij}}$) on a grided map with i and j representing the axes:

$$\overline{C_{ij}} = \frac{1}{\sum_0^t (\tau_{ijt})} \sum_0^t (c_t \cdot \tau_{ijt}) \qquad (1)$$

273

274  where $\tau_{ijt}$ is the time each back-trajectory footprint spends at ground level (<100 m) in the ijth

275  cell for the VOC sample at time t, and $c_t$ is the measured concentration of a compound at the

276  ground site.  Each cell has a corresponding $n_{ij}$ value, representing the number of individual

277  footprints included in each cell, which was determined as the number of samples contributing to

278  a cell's average ($\overline{C_{ij}}$) (Seibert et al., 1994). To reduce bias from cells with few samples (i.e. low

279  $n_{ij}$ values), a weighting function multiplies the ($\overline{C_{ij}}$) result by a factor of 1, 0.7, 0.4, or 0.05 for

280  cells with $n_{ij}$ values above the $Q_{90}$, $Q_{75}$, $Q_{50}$ or below the $Q_{50}$ percentiles, respectively (Polissar et

281  al., 2001).  Contour maps were then plotted using these final values and shown with a 1 arc

282  second elevation map obtained from the USGS National Map Seamless Server

283  (http://viewer.nationalmap.gov/viewer/). It is insufficient to only consider the distribution of

284  wind directions against compound concentrations when complex meteorology affects the

285  transport of air masses. This is the case in California's central valley. Similarly, basic single

286  HYSPLIT back-trajectory analysis can oversimplify the footprint of measurements into one

287  single path and not accurately represent the distribution of ground-level residence times for an air

288  parcel (Figure 2).

289

290  **3. Results and Discussion**

291  Figure 3 shows measurements of a selection of compounds plotted against carbon

292  monoxide, a common technique to assess contributions from anthropogenic emissions (after

293     filtering for biomass burning events). Some compounds have ratios to CO consistent with

294     measurements from the Los Angeles air basin during the same time period (Borbon et al., 2013).

295     However, there are several compounds with frequent enhancements above the Los Angeles

296     slope, indicating additional sources of these compounds that are not abundant in LA. Most of the

297     compounds shown in Figure 3 have been previously linked to petroleum and dairy operations

298     (e.g. Gilman et al., 2013 and Shaw et al., 2007), and their enhancements here are evidence for

299     substantial emissions in the San Joaquin Valley.

300

301     **3.1 Emissions from Petroleum Operations**

302        Petroleum operations emit a substantial amount of hydrocarbons, with a smaller

303     distribution of molecular weights than emissions from gasoline sources. The 25th percentiles for

304     propane and n-butane are similar to other urban ground sites during the summer, but higher

305     concentrations were observed for the 50th and 75th percentiles, by up to a factor of 2 compared to

306     Pittsburgh, PA (2002) (Millet et al., 2005). The 75th percentiles in the San Joaquin Valley are

307     also higher by 25-50% compared to measurements from 2005 in Riverside, CA, a much more

308     populated region (Gentner et al., 2009). Between the CalNex field sites at Bakersfield and

309     Pasadena, median and smaller values (10th and 25th percentiles) were similar and lower at

310     Bakersfield, respectively. Yet, 75th percentile concentrations were greater at Bakersfield by 53%

311     for propane (5.6 vs. 3.7 ppbv) and 16% for n-butane (5.6 vs. 3.7 ppbv). Previous work in the

312     South Coast air basin has also reported emissions of light alkanes from oil/gas operations, but

313     there is a lesser prevalence of oil/gas fields in that air basin compared to the San Joaquin Valley

314     (Peischl et al., 2013).

315        The source receptor model with chemical mass balancing used in Gentner et al. (2012)

316     effectively modeled emissions of most compounds in a motor vehicle emissions study at the

317     Caldecott tunnel and many of the compounds that are most prevalent in gasoline and diesel

318     emissions at Bakersfield. We used the non-methane composition of thermogenic wet gas

319     reported by the USGS (Table 1) to construct the initial petroleum operations source profile in our

320     source receptor model. The composition of unrefined natural gas has substantial variability

321     among all the wells sampled, but the average composition was very effective for modeling the in

322     situ data from Bakersfield. In many cases, ratios in ambient data can be impacted by differences

323     in the rates of chemical reaction in the atmosphere; as is the case in Los Angeles (Borbon et al.,

324  2013). At Bakersfield, the timescales for transport from source to measurement site are much

325  shorter than the timescales of reaction for the species considered here. So, variability due to

326  chemical processing is negligible for all but the most reactive primary emitted compounds in our

327  Bakersfield data.

328  In addition to the compounds known to be in thermogenic wet gas (Table 1), the model

329  under-predicted the observed concentrations of numerous alkanes. These compounds are

330  summarized in Table 2 and Figure 4, which show their average unexplained concentrations and

331  the percent of total mass that is unexplained (determined by the residuals in the chemical mass

332  balance source receptor model). Most of the unexplained concentrations of these alkanes were

333  well correlated ($r \geq 0.75$) with the petroleum operations source contribution from the model and

334  are attributed to this source. The presence of the branched and cyclic alkanes in unrefined

335  petroleum gas is not surprising as there are significant amounts of $C_{5\text{-}7}$ straight chain alkanes in

336  the reported composition (Table 1) and a select few have been measured in other studies (Gilman

337  et al., 2013; Ryerson et al., 2011). Yet, there are limited previous in situ measurements for many

338  of the compounds reported here, especially many of the cyclic alkanes. Concentrations of

339  aromatics observed at Bakersfield matched predicted concentrations from motor vehicle sources

340  in our model, but other studies have observed aromatic emissions from petroleum operations

341  (e.g. Gilman et al., 2013).

342  The additional compounds attributed here to the petroleum operations source profile

343  increase the mass of emissions by 10.6% as shown by the regression of the correlated

344  "unexplained" compounds with the petroleum gas source ($r=0.95$) (Figure 5). The weight

345  fraction of each correlated compound in the "unexplained" mass is given in Table 2 with similar

346  fractions in the overall source profile as the known $C_{5\text{-}7}$ compounds in thermogenic wet gas

347  (Table 1). In all, the interquartile range of the unrefined petroleum gas source contribution was

348  7.6-89 ppbC, with a diurnal pattern that was strongly dependent on meteorological dilution

349  (Figure S3). This source represented a substantial fraction of anthropogenic emissions. For

350  comparison, the mass concentration of compounds emitted by the observed petroleum operations

351  source ranged from 30-40% to 100-150% of the sum of compounds from motor vehicles during

352  the afternoon and nighttime, respectively (Figure S4).

353  The remaining branched and cyclic compounds that were not highly correlated with the

354  petroleum gas source represent a relatively small amount of mass and we could not confidently

355  infer a specific source for these compounds. The excess $C_{13-16}$ branched alkanes were well-
356  correlated (r≥0.80) with each other, but not with any other compounds. The excess
357  concentrations of $C_{10-11}$ branched alkanes were correlated with each other, and one of the
358  compounds, 2,6-dimethyloctane, was well-correlated (r≥0.80) with the three $C_9$ cycloalkanes that
359  do not correlate well with the petroleum operations source. These remaining compounds have
360  ozone formation potentials similar to other observed compounds, ranging from 0.6 to 1.6 $gO_3$ $g^{-1}$,
361  but their excess concentrations after modeling were minimal—average values from 0 to 0.15
362  ppbC each (Figure 4). Work by Liu et al. (2012) and Chan et al. (2013) at CalNex-Bakersfield
363  inferred a source of higher molecular weight organic carbon, potentially from petroleum
364  operations, but we did not observe any significant correlation.

365      Unrefined thermogenic wet gas is largely comprised of methane when extracted at the
366  wells. Yet, at the Bakersfield ground site observations of methane and contributions from the
367  petroleum operations source were not well correlated (Figure S5). Additionally, the potential
368  methane emissions expected based on the thermogenic wet gas source profile (Table 1) would
369  exceed all of the observed methane enhancements above background concentrations by over
370  30%. Despite the absent methane emissions, the large source of hydrocarbons is well modeled by
371  the source profile from unrefined thermogenic wet gas in the San Joaquin Valley when using
372  propane and larger compounds.

373      We compared the relative ratios of hydrocarbons in the thermogenic wet gas profile data
374  to regression slopes of in situ data and canister data to further explore emissions from petroleum
375  operations using ethane and isobutane, which were not available in our in situ data. The light
376  alkanes discussed here were very well correlated in measurements from Bakersfield. Regressions
377  with $C_5$ and larger compounds have more scatter due to emissions from gasoline-related sources,
378  so they excluded here and addressed using the source receptor model (example in Figure S2). For
379  the light alkanes, which have relatively minimal contributions from motor vehicles at the site, we
380  compare ratios between atmospheric data and the source profile expected for petroleum
381  operations (Table 1) with the results summarized in Table 3. Ratios of n-butane to isobutane
382  strongly support the conclusion of a petroleum operations source as they are identical with 1.7 ±
383  0.4 and 1.7 ± 0.04 (r=0.99) in the oil well data and in canister measurements from Bakersfield,
384  respectively. The process of methane separation from the associated petroleum gas can remove a
385  fraction of very light alkanes (i.e. $C_{2-3}$) and affect their relative composition to other

386   hydrocarbons in the condensate (Armendariz, 2009; Hendler et al., 2006). This is consistent with

387   our observations of ratios involving $C_{2\text{-}3}$ alkanes. The ethane to propane ratio ($gC\ gC^{-1}$) observed

388   via canister measurements at the Bakersfield site ($0.6 \pm 0.06$, $r=0.93$) (Figure S1) is significantly

389   lower than expected based on the thermogenic wet well composition in the San Joaquin Valley

390   ($1.2 \pm 0.2$). Similarly, the ethane to n-butane ratio is significantly lower in the canister data ($1.1 \pm$

391   $0.1$) relative to the unrefined gas data ($3.4 \pm 0.8$). The propane to n-butane ratio in the in situ and

392   canister data ($1.9 \pm 0.01$ ($r=0.98$) & $1.8 \pm 0.1$ ($r=0.98$)) was slightly lower than in the oil well

393   data ($2.9 \pm 0.7$). The selective removal of ethane and propane along with methane changes the

394   overall petroleum operations source profile observed at Bakersfield, primarily for ethane, which

395   was not used in our model. This also results in a 33% decrease in the propane weight fraction of

396   the source profile. A revised source profile is shown in Table 4 with the addition of the

397   previously "unexplained" compounds. We modified the propane content of the source profile to

398   reflect this slight change in the propane composition relative to n-butane, and it resulted in very

399   minor changes to the source receptor model outputs and maintained the same robust model

400   diagnostics. The results reported in the paper reflect this minor change. The new source profile

401   (Table 4) does affect the overall ozone formation potential. Including these "new" compounds

402   increases the ozone forming potential of the reported petroleum operations source profile to 0.82

403   $gO_3\ g^{-1}$, due to the addition of more reactive cycloalkanes and branched alkanes to the initial

404   source profile (Table 1).

405       The successful modeling of these emissions using the source profile constructed from

406   well data and the consistency of hydrocarbon ratios between wells and Bakersfield

407   measurements (canisters and in situ data) contributes to the strong evidence of emissions from

408   petroleum operations. Overall, our results infer that the VOC source characterized and classified

409   as petroleum operations in this analysis is not a major source of methane in this region. In many

410   cases, methane emissions are coincident with emissions of non-methane hydrocarbons at

411   petroleum extraction or processing sites due to either co-emission from the same

412   equipment/reservoir or co-located emission pathways at the same facility (Katzenstein et al.,

413   2003; Petron et al., 2012; Gilman et al., 2013). For comparison, we include light hydrocarbon

414   ratios from other relevant studies in Table 3. Given regional variability in oil/gas deposit

415   composition, direct extrapolation between regions should only be done with careful attention to

416   compositional differences in wells and other fuels, especially in urban areas where there are

numerous sources of light hydrocarbons. Despite this expected heterogeneity, ratios are similar between most of the studies within the calculated uncertainties. The consistency between ratios of ethane to propane and n-butane between our ambient measurements and condensate tank samples in (Hendler et al., 2009) supports the case for emissions from condensate storage tanks or associated equipment. Our observation of a major petroleum operations source with minimal coincident methane is consistent with composition measurements of condensate storage tank emissions, which contain the separated non-methane liquids and have been shown in two Texas-based studies to be dominated by non-methane hydrocarbons (Armendariz, 2009; Hendler et al., 2006). The studies demonstrated that condensate tanks emit 4-6 times more VOCs than methane whereas all other emission pathways emit 3-15 times more methane than VOCs, and methane was on average only 15±11 wt% of 20 vent gas samples from condensate tanks (Armendariz, 2009; Hendler et al., 2009).

Similar results can also be found in positive matrix factorization (PMF) studies in the urban area of Houston, a prominent region for petroleum imports and refining. They reported considerable emissions attributed to oil/gas operations and petrochemical production of other chemicals (Leuchner and Rappengluck, 2010; Buzcu and Fraser, 2006). One evident source, termed oil/natural gas evaporation from refineries, was comprised of $C_{2-7}$ straight and branched alkanes, as well as cyclopentane, cyclohexane, and methylcyclopentane. In Leuchner and Rappengluck (2010), this source accounted for 27% of observed VOC mass at the urban site outside of the Houston shipping channel, and resulted in atmospheric concentrations ranging from 10-40 ppbC diurnally.

The good agreement of the observed non-methane hydrocarbon source profile with the measured composition of associated gas in oil wells (accounting for the selective reduction of $C_{2-3}$ alkanes) suggests that emissions occurs via a pathway involving non-methane volatile components separated from thermogenic wet gas. This is very likely a fugitive emission pathway(s), occurring predominantly after methane separation, during the extraction, storage, or processing of crude oil, associated gas, or condensate. In 2012 and 2013, California issued targeted standards to reduce emissions of VOCs and methane from oil and natural gas operations. These efforts to control VOCs are primarily directed at storage tanks and other relevant equipment, with a focus on emissions during production and transmission from equipment that stores and handles crude oil or condensate, and effective control technologies (California Air

448  Resources Board, 2012, 2013). Spatial mapping of emissions in Section 3.3 suggests an area

449  source with a similar distribution to oil wells in the San Joaquin Valley.

450      The results of this section along with the following sections form and augment the

451  conclusion that the vast majority of methane enhancements observed in the San Joaquin Valley

452  are due to emissions from dairy operations. In particular, Section 3.3 shows the statistical source

453  footprint of emissions from petroleum operations in stark contrast to both the statistical source

454  footprint of methane emissions and the spatial distribution of methane concentrations measured

455  via aircraft in California's central valley with large spikes over areas with high concentrations of

456  dairies. It is very possible that there are emissions of methane in the San Joaquin Valley from

457  other petroleum operations that are downstream from our observed source, perhaps related to

458  natural gas marketing. The results of this study infer that these emissions are minor compared to

459  dairy operations, and are predominantly not co-located with our characterized petroleum

460  operations source.

461

462  **3.2 Emissions from Dairy Operations**

463      We observed evidence for substantial emissions from dairy operations in the San Joaquin

464  Valley. These emissions, unlike the petroleum operations source, were dominated by small

465  alcohols, acetic acid, and methane. Concentrations of the major non-methane organic

466  compounds—methanol, ethanol, and acetic acid (average and interquartile range concentrations

467  in Table 5)—are higher than previous measurements at other locations. Compared to another

468  urban ground site in Pittsburgh during summer 2002 (Millet et al., 2005), the ethanol and

469  methanol interquartile ranges and geometric means were greater in Bakersfield, by

470  approximately 300% and 50%, respectively. Despite the larger human population of the South

471  Coast air basin, nighttime geometric means were 70% and 240% greater in Bakersfield compared

472  to coincident measurements at Pasadena, CA (CalNex) for ethanol and methanol, respectively.

473  The mean and median ethanol concentrations at the urban Bakersfield site were 12.8 and 7.6

474  ppbv, respectively. These values are several times greater than observations of urban and

475  continental ethanol mixing ratios globally, as reported by Kirstine et al. (2012). However, a

476  comparison of methanol concentrations is within the typical range of observed values globally

477  (Heikes et al., 2002).

478      The methanol to methane emission ratio in dairy operation plumes measured on the

479  aircraft was $7.4 \pm 0.6$ mmol mol$^{-1}$ (aka ppbv ppmv-1); this slope of the regression (r=0.89) is

480  nearer to the lower limit of the 7-16 mmol mol$^{-1}$ range in the plumes (Figure 6). This ratio was

481  constructed from multiple transects and shows a range of ratios indicating some near-source

482  variability in emissions from the different pathways of emissions. This ratio could be improved

483  by collecting a larger sample size of data from more locations in future source characterization

484  studies.

485       Ground site ethanol and acetic acid data were compared to methane to determine their

486  emission ratios with close attention to enhancements from other sources. For ethanol and

487  somewhat for acetic acid, there is a clear slope that emerges (Figures 7-8) against methane with

488  occasional enhancements in ethanol or acetic acid that are coincident with high concentrations of

489  tracers for other sources. In contrast, there were no enhancements in methane concentrations past

490  these baseline slopes in the data. This is indicative of a singular major source of methane that is

491  clearly related to ethanol and acetic acid. This result, along with the results of Section 3.3

492  showing the agreement of dairy locations with the spatial distribution of concentrations

493  (measured via aircraft) and the statistical source footprint of both methane and ethanol, supports

494  the conclusion that dairies are the predominant source of methane in the San Joaquin Valley and

495  emissions from petroleum are minor in comparison. To calculate emission ratios, data points

496  with enhancements due to other sources (determined and shown by correlation with other tracer

497  compounds) were not considered in the emission ratio assessment. This allows ethanol and acetic

498  acid to become source-specific tracers of dairy operations. With dairy (and other cattle)

499  operations responsible for the vast majority of methane emissions observed at the Bakersfield

500  site, the emission ratios of ethanol and acetic acid to methane are effectively calculated by taking

501  the lower limit of slopes versus methane when enhancements from other sources of ethanol or

502  acetic acid are at their minimum.

503       At the Bakersfield ground site, concentrations of non-vehicular ethanol (calculated via

504  the source receptor model) were well correlated with methane except for outliers with

505  enhancements in ethanol that were coincident with large enhancements in tracers of other ethanol

506  sources (Figure 7). Other potential sources of alcohols and oxygenated gas-phase organic carbon

507  are wastewater treatment, vegetation, soil processes, motor vehicles, and landfill/composting

508  facilities. At low concentrations of these tracers, non-vehicular ethanol and methane are very

509  well correlated with a slope of 18 mmol mol$^{-1}$. Chloroform, trichloroethylene, and carbon

510  disulfide correlate with different points that deviate from the emission ratio, suggesting multiple

511  other minor sources of ethanol.

512      The results of the acetic acid versus methane assessment (Figure 8) at the Bakersfield

513  ground site produced similar results to that of non-vehicular ethanol versus methane. The

514  enhancements of acetic acid above the emission ratio slope coincided with tracers of other

515  primary and secondary sources. We calculated an emission ratio for acetic acid to methane of 1.3

516  mmol mol$^{-1}$. This value represents a lower limit of acetic acid emissions associated with dairy

517  operations. There is remaining uncertainty in this emission ratio and, based on the data shown in

518  Figure 8, the ratio of acetic acid to methane could be up to 50% greater. The diurnal profile of

519  acetic acid also suggests emissions from local/regional sources since concentrations are at their

520  maxima during the night when emissions accumulate in the nocturnal boundary layer with

521  minimal horizontal or vertical dilution. The results of our study show that there are high

522  concentrations of acetic acid that are associated with methane, formic acid, acetone, or isoprene.

523  This indicates that there are multiple major biogenic and anthropogenic sources of acetic acid in

524  the San Joaquin Valley.

525      Rice cultivation could also be an important source of light alcohols and methane (Peischl

526  et al., 2012), but there is little rice cultivation in the San Joaquin Valley. The bulk of Californian

527  rice cultivation is located in the Sacramento Valley—the northern portion of California's central

528  valley. In the San Joaquin Valley, emissions from dairy operations should far outweigh those

529  from rice cultivation. This work is focused on sources in the San Joaquin Valley, but data from

530  aircraft canister measurements suggest that dairy operations and rice cultivation have different

531  emission ratios of ethanol to methanol (Figure S6). In general, observations between the two

532  valleys are heavily influenced by the major source that dominates in each air basin (Figures 13,

533  S11).

534      Constructing an overall source profile for dairy operations is difficult since methane, light

535  alcohols, and acetic acid all have different emission rates from specific source pathways at

536  dairies. Previous studies report that methane emissions are minimal from animal waste and

537  greatest from enteric fermentation in cows. Whereas emissions of non-methane gas-phase

538  organic carbon come predominately from animal feed, followed by waste, with minor

539  contributions from the animals themselves (Chung et al., 2010; Howard et al., 2010b; Shaw et

540  al., 2007; Sun et al., 2008). Further variability is introduced by factors such as feed composition,

541 temperature, and specifics of feed and waste handling. Table 5 summarizes the average regional

542 source profile for dairy operations, determined via downwind sampling of a large collection

543 individual farms/feedlots in the San Joaquin Valley. Comparison against other studies is limited

544 by the lack of a similar set of compounds. Previous studies report high emission rates for a

545 selection of the primary compounds in Table 5, but there is no full set for comparison, and other

546 work is focused on singular emission pathways rather than the overall source profile.

547 Extrapolation to other regions must be done with caution, as the emission ratios reported here are

548 region specific. So here we compare our results to other studies to the extent that it is possible.

549       In this and other studies, emissions of ethanol are typically greater than methanol,

550 ranging 1.3-2.4 mol mol$^{-1}$. Based on the literature and our results, it is apparent that the ratios of

551 the two main alcohols to methane can vary depending on the relative amount of animals versus

552 feed and waste, and the specifics of feed/waste storage and processing. Our reported ratios

553 represent the average for the region; the ratio of ethanol to methane reported by Sun et al. (2008)

554 for lactating cows and waste (24 mmol mol$^{-1}$) is slightly higher than our value (18 mmol mol$^{-1}$).

555 Their ratio of methanol to methane (19 mmol mol$^{-1}$) was greater by 150%, but is within the range

556 observed in our analysis of aircraft data. The differences here can potentially be attributed to the

557 absence of feed, which will increase alcohol emissions. Measurements of acetic acid are less

558 common so there are few studies to compare emission ratios. Shaw et al. (2007) reported ratios

559 of acetic acid to methanol ranging from 0.05 to 0.94 mol mol$^{-1}$ for cows and their waste. In this

560 work, we observed a ratio of 0.18 mol mol$^{-1}$.

561       Emissions of other carbonyls have been reported from dairy and other livestock

562 operations in relatively minor quantities compared to the dominant compounds presented in this

563 work. There are likely small emissions of low molecular weight aldehydes (e.g. propanal,

564 butanal), ketones (eg. acetone), other alcohols (e.g. propanol, phenols), alkenes, and esters (e.g.

565 propyl acetate, propyl propionate) from dairy operations (Chung et al., 2010; Howard et al.,

566 2010b; Malkina et al., 2011). In general, a major source of many oxygenated species is

567 secondary production from the chemical oxidation of other compounds. The measurements used

568 in this study similarly suggest substantial contributions from secondary production for many of

569 the measured carbonyls and acids. At the ground site and from the aircraft, emissions of many of

570 these carbonyls from dairy operations could not be detected due to the magnitude of other

571 sources, and there were no measurements of esters or larger alcohols. In this study, dairy

572  operation emissions of these minor compounds (acetone, methyl ethyl ketone, propanal, butanal,

573  and other oxygenated VOCs measured at the Bakersfield site) make only minimal contributions

574  to total emissions of these compounds on a valley-wide basis. One potential exception is

575  acetaldehyde; previous work reported emissions equivalent to 20-110% of ethanol emissions

576  from feed and relatively minor emissions from cows and their manure (Makina et al., 2007;

577  Shaw et al., 2007). In this study, no significant correlation was observed between acetaldehyde

578  and methane in the dairy plumes measured by aircraft, and insufficient data exist from the

579  ground site to check for emissions of acetaldehyde. Also, neither methyl ethyl ketone nor acetone

580  were well correlated (r=0.55-0.65) with methane in the dairy plumes measured by the aircraft.

581  Other studies on volatile organic acids have also reported emissions of propanoic acid and

582  butanoic acid with relative emission rates ranging from an order of magnitude below acetic acid

583  to the same order of magnitude as acetic acid (Alanis et al, 2010; Shaw et al., 2007; Sun et al.,

584  2008). We did not measure propanoic or butanoic acid, but we did not observe any correlation

585  between measured concentrations of either formic or oxalic acid and the prominent compounds

586  emitted from dairies at the Bakersfield ground site. Based on our work and the literature, acetic

587  acid appears to be the most prominent acid emitted by dairy operations.

588      One of the objectives of this study was to provide a basic source profile, averaged over

589  the bulk of dairy operations in the San Joaquin Valley with the understanding that the profile can

590  potentially vary between individual operations. Methanol, ethanol, and acetic acid were the

591  predominant non-methane compounds emitted from dairy operations. Figure 9 shows

592  comparisons of the concentrations of these compounds attributed to dairy operations versus the

593  total observed concentrations for each hourly sample in Bakersfield. The percentage of each

594  compound from dairies ranged widely with some significant diurnal patterns (Figure S8). On

595  average, 27% of observed methanol was from dairies with hourly averages ranging diurnally 22-

596  37%. 28% of observed acetic acid was from dairies with a diurnal range of 11-44%. As

597  mentioned previously, the emission ratios for methanol and acetic acid are conservative

598  estimates that may tend towards lower limits. In this case, the fraction of methanol and acetic

599  acid from dairy operations will increase slightly, but since ethanol makes up a dominant fraction

600  of the non-methane source profile (Table 5) these changes will have a negligible impact on the

601  overall source profile and implications of dairy operations on air quality in the valley (Section

602  3.4). Due to the increased use of gasoline, $9.6 \pm 5.8\%$ of ethanol was emitted by gasoline-related

603  sources. Of the remainder, 48% was from dairy operations on average with a diurnal range of 30-
604  71%.

605        The diurnal average of the percent contribution from dairy sources (Fig. S8) shows

606  minima during the daytime for acetic acid and non-vehicular ethanol. These ratios vary widely

607  with time of day and meteorology. This daytime minimum can be attributed in part to biogenic

608  emissions of ethanol when emissions from natural vegetation and agriculture are likely highest.

609  For acetic acid, the minimum is likely due to secondary production from the oxidation of

610  isoprene and other reactive precursors. Methanol did not have as strong of a diurnal pattern since

611  other major day and nighttime sources have similar emission patterns (e.g. vegetation). The

612  remaining methanol observed at the Bakersfield site can be attributed to a mix of emissions from

613  anthropogenic urban sources, natural vegetation, and biogenic emissions from agriculture. A

614  recent study by Hu et al. (2011) found that 90% of methanol was biogenic during the summer in

615  the Midwestern U.S., with the remainder being anthropogenic. Heikes et al. (2002) reports a

616  similar value with primary biogenic emissions responsible for 81% of non-oceanic emissions.

617  Dairies are an important source of methanol in the San Joaquin Valley along with emissions

618  from agriculture and natural vegetation. The methods used in these studies to allocate emissions

619  will determine whether dairy (and other cattle) operations are categorized as biogenic or

620  anthropogenic sources. In this work we consider emissions from dairy operations to be

621  anthropogenic, similar to the CARB inventory.

622        Pusede et al. (2013) found that daytime average concentrations of light alcohols,

623  aldehydes, and acids at the Bakersfield site increased with daily maximum temperature. It is

624  possible that increases in ambient temperature could lead to increases in silage emissions due to

625  enhanced volatilization of some compounds (e.g. alcohols), which would change the reported

626  source profiles slightly. Yet, ethanol was the most prominent non-methane compound in our

627  source profile and results from Pusede et al. (2013) show that daytime averages of ethanol did

628  not increase between moderate and high temperatures (Table A2). So, we do not expect major

629  changes with temperature for the dairy source profile reported in this work and recommend

630  further research to identify other high-temperature sources of oxygenated compounds.

631

632  **3.3 Spatial distribution of sources**

633        Using FLEXPART-WRF meteorological data and methods for the region, distributions of

634     back-trajectories were calculated for 6 and 12 hours prior to arrival and measurement at the

635     Bakersfield site. Overall averages, as well as day and nighttime averages, are shown for the

636     entire campaign in Figure 1. The influence of local emissions near the site is important at all

637     times. Daytime measurements are largely impacted by transport from the north-northwest due to

638     consistent up-valley flows during the day. In contrast, at night the wind speeds and direction are

639     more variable and irregular with flows that arrive from all directions, but originate as up-valley

640     flows from the north-northwest. Extensive reviews of meteorology and flow patterns in the San

641     Joaquin Valley found elsewhere are consistent with the results presented in this work (Bao et al.,

642     2007; Beaver and Palazoglu, 2009). The footprint analysis used in this study provides a good

643     representation of the distribution of surface-level areas that influence parcels' contact with the

644     surface layer and associated sources, but potentially has some uncertainty given the complexities

645     of Bakersfield meteorology (Angevine et al., 2013).

646         Statistical meteorological modeling using ground site data resulted in a spatial

647     distribution of petroleum gas emissions similar to that of oil wells in the southern San Joaquin

648     Valley (Figure 10). Additionally, canister samples taken via aircraft in the region show higher

649     propane (a major component of the source profile) concentrations for some points in the southern

650     part of the valley (Figure 10C). Given the co-location of oil wells in the region and the spatial

651     distribution of elevated concentrations of petroleum gas compounds, it is probable that the

652     observed emissions occur at or near the wells during extraction, storage, and initial processing.

653         The statistical distribution of emissions of non-vehicular ethanol and methane were

654     similar for both 6 and 12 hr back-trajectories. The map of emissions is consistent with the

655     distribution of dairies in the San Joaquin Valley (Figures 11-12) and aircraft measurements of

656     ethanol and methane (Figures 13-14). While there are dairy operations within the 12 hr footprint

657     and the emitted methane and light alcohols have long atmospheric lifetimes, the dairies within

658     the 6 hr footprint are much more influential in elevated concentrations, especially at night. The

659     spatial distributions of petroleum and dairy operation emissions clearly show that they are

660     coming from different parts of the valley. The maps in this section provide strong supporting

661     evidence that the vast majority of methane is coming from dairy (and other cattle) operations.

662         The statistical emissions mapping method developed in this paper is a useful integration

663     of concentration-weighting trajectory methods with the FLEXPART-WRF modeling platform.

664     This emissions mapping tool is effective at locating point and area sources, especially for

665     prominent sources in the San Joaquin Valley. The analyses of the spatial distribution of

666     emissions from petroleum and dairy operations shown in this work are two applications of this

667     technique. For these purposes, either concentration data or modeling outputs (e.g. source receptor

668     models) can be used, both of which appear in this work.  Further development of this approach

669     will continue to improve its utility and quantitative outputs, but caution must be given to the

670     transport timescales and tracer lifetime. There is one limitation to the current version of the

671     statistical source footprint analysis. The area of analysis is limited to the distribution of sample

672     footprints across all runs, and there is likely insufficient data to assess areas outside that total

673     footprint. Nevertheless, the current method is excellent for looking at the most important sources

674     that impact an area, such as Bakersfield in this study. Coverage could be improved in other

675     studies by using data from multiple sites in a region, but care must be exercised to ensure the

676     data is properly weighted. Overall, this work demonstrates the efficacy and usefulness of this

677     tool, warrants further development, and future work should apply it on regional and continental

678     scales, as appropriate, to locate primary sources of pollution.

679

680     **3.4 Implications for Air Quality and Emissions Inventories**

681        Both petroleum and dairy (and other cattle) operations are important sources of reactive

682     organic carbon in the San Joaquin Valley. On a mass basis, observed VOC concentrations from

683     petroleum extraction/processing were on the same order as emissions from motor vehicles. Yet,

684     they represent a relatively minor contribution to potential ozone formation, as the average MIR

685     value for the source ($0.82 \ gO_3 \ g^{-1}$) is ~3-6 times less than that of motor vehicle sources. Direct

686     contributions to secondary organic aerosol (SOA) from petroleum operations source profile in

687     this study are likely to be minimal given that the yields for all of the alkanes with 8 or less

688     carbon atoms will be $0.002 \ gSOA \ g^{-1}$ at most with an organic particle loading of $10 \ \mu g \ m^{-3}$

689     (Gentner et al., 2012). The potential ozone and SOA implications of petroleum operation

690     emissions will depend greatly upon composition, which varies between regions. We did not

691     observe any aromatic content, but other studies have observed aromatic and other larger

692     compound fractions (Carter and Seinfeld, 2012; Gilman et al., 2013). Aromatics have been

693     shown to be very effective precursors to SOA and ozone (Gentner et al., 2012; Carter, 2007). So,

694     their presence in oil/gas emissions will have further implications for air quality.

695        Dairy operations in the San Joaquin Valley are largely responsible for the higher than

696 typical ethanol concentrations in the San Joaquin Valley. Based on the primary compounds

697 observed from dairy operations (ethanol, methanol, acetic acid), we infer that emissions have

698 minor impacts on SOA formation, but have a greater potential to impact ozone formation with an

699 MIR of 1.3 $gO_3$ $g^{-1}$. The inclusion of other oxygenated compounds previously observed from

700 dairy operations (e.g. Hafner et al., 2013) to the basic source profile in Table 5 may increase the

701 ozone and SOA formation potential. Yet, in this study they were minor and not significantly

702 correlated with other dairy emissions (see Section 3.2).

703 In Bakersfield during spring/summer, dairy operations were responsible for 22% of

704 anthropogenic non-methane organic carbon emissions and 13% of potential anthropogenic ozone

705 formation. Similarly, petroleum operations were responsible for 22% of anthropogenic emissions

706 and 8% of potential ozone. Motor vehicles were responsible for the remaining 56% of

707 anthropogenic emissions, 79% of anthropogenic potential ozone formation, and essentially all of

708 the potential anthropogenic SOA formation. It is important to note that emissions from petroleum

709 and dairy operations have substantial potential to impact the atmospheric chemistry leading to

710 secondary pollution, but they themselves are not a major source of SOA precursors. These results

711 apply to the emissions of VOCs from petroleum operations observed and characterized in this

712 work; other recent work on petroleum operations has reported emissions of larger hydrocarbons

713 that have higher SOA yields (Chan et al., 2013; Gilman et al., 2013). These five main sources are

714 summarized in Figure 15 and are very important sources for the San Joaquin Valley. There are

715 other anthropogenic sources that likely contribute emissions on smaller urban scales that are not

716 enumerated in this work. The contributions of biogenic sources are another major factor for air

717 quality in California's central valley.

718 In the comparison of the sources discussed in this work, the percent contribution of

719 vehicular sources is larger in Bakersfield than it would be most places in the region. In non-

720 urban areas of the San Joaquin Valley, motor vehicle emissions will still be important, but

721 emissions from petroleum and dairy operations will make up a greater fraction of non-methane

722 organic carbon in the atmosphere and will be responsible for a greater fraction of potential ozone

723 formation. These results confirm the transport and importance of emissions from dairy

724 operations throughout the San Joaquin Valley. Our results for potential ozone give a 3.5:1 ratio

725 of potential ozone from gasoline vehicles to dairy operations in Bakersfield. When considering

726 that there is a greater prevalence of motor vehicles near our measurement site and most dairy

727  emissions are outside the county (Table 7), the ratio is in agreement with the valley-wide ratio of

728  3:2 for light-duty vehicles to livestock feed modeled by Hu et al. (2012). Overall, this, and other

729  recent work (Howard et al., 2010a; Hu et al., 2012), demonstrates that motor vehicles and

730  multiple source pathways at dairy operations are major emitters of reactive ozone precursors

731  throughout the San Joaquin Valley. Elevated concentrations of non-vehicular ethanol that are

732  largely linked to dairy operations warrants further evaluation of processes involving livestock

733  silage as ethanol has been demonstrated as a primary component of those emissions (Hafner et

734  al., 2013; Howard et al., 2010a; Malkina et al., 2011).

735      Our results on the relative contributions from each source indicate a mix of influential

736  sources. Given our location in an urban area in the Southern San Joaquin Valley, where oil wells

737  are concentrated, emissions from motor vehicles and petroleum operations are likely higher than

738  other parts of the valley. The San Joaquin Valley has an abundance of agriculture and is

739  surrounded by natural vegetation that represents a large potential source of emissions following

740  transport to other parts of the valley. Comprehensive modeling assessments need to evaluate the

741  sources discussed here along with biogenic emissions of reactive organic gases from both

742  agriculture and natural vegetation.

743      Comparing different assessments for emissions from multiple sources presents challenges

744  relating to the definition of sources and spatial boundaries. Here, we provide a comparison of our

745  relative emission magnitudes at the Bakersfield site to the CARB emission inventory for the San

746  Joaquin Valley (Table 7). To promote consistency with our observed sources, we compare our

747  petroleum operations source to emissions from oil/gas production and refining, and exclude

748  petroleum marketing (and combustion from petroleum operations) since our observed source is

749  clearly related to unrefined petroleum. While there are likely some differences in emissions, it is

750  difficult to separate dairy cattle from other cattle, so we have assumed that we are observing all

751  cattle in this study and include them with dairy operations. Although in the CARB inventory,

752  dairy cattle represent almost 80% of cattle-related emissions in the San Joaquin Valley. Similarly,

753  we compare these sources to on- and off-road mobile sources as that is the best representation of

754  the observed motor vehicle sources in our source apportionment.

755      There are potential seasonal effects among the 5 sources shown in Tables 6-7 and Figure

756  15. The composition of gasoline fuel changes seasonally to reduce volatility by varying

757  formulation, which affects the composition and magnitude of emissions. In the CARB almanac,

758 VOC emissions from dairy operations and petroleum production and refining have no seasonal
759 change between summer and winter. The emissions we observe from both sources could be
760 hypothesized to volatilize more in warmer weather, but we have insufficient data to assess the
761 seasonal changes and effects other than temperature may potentially play a role.

762     The CARB emissions inventory for the San Joaquin Valley reports an average of 28 tons
763 ROG per day from petroleum operations (production and refining), which is equal to 27% of on-
764 and off-road mobile source emissions (72+32 tons per day) in the air basin (California Air
765 Resources Board, 2010). This value is consistent with daytime ratios (18-51%) observed at the
766 Bakersfield site (Figure S4) when vehicular emissions are greatest, but is smaller than nighttime
767 ratios (62-120%) and the overall ratio (39%). Nighttime ratios are significantly higher when
768 there is relatively less vehicular traffic and since Bakersfield is in much closer proximity to
769 potential petroleum operations sources compared to other parts of the air basin. A comparison on
770 a smaller scale for the portion of Kern County in the San Joaquin Valley demonstrates the local
771 importance of petroleum operations, as much of the San Joaquin Valley's petroleum operation
772 emissions are in this county. For this area, petroleum production/refining emissions in the CARB
773 inventory are equivalent to 139% of on- and off-road mobile sources (California Air Resources
774 Board, 2010). This observation is consistent with the statistical footprints shown in this work as
775 daytime footprints encompass a larger area that stretches into other counties while nighttime
776 footprints are more heavily influenced by local emissions.

777     According to the CARB emission inventory, dairy and other cattle operations in the San
778 Joaquin Valley emit 57 tons ROG per day, which is 80% of non-vegetation farming-related
779 emissions (California Air Resources Board, 2010). These emissions from dairy and cattle
780 operations are equivalent to 55% of on- and off-road motor vehicle emissions in the inventory,
781 which is higher than the average non-methane organic carbon (NMOC) mass comparison at the
782 Bakersfield measurement site (40%). The CARB inventory for the San Joaquin Valley states that
783 emissions from dairy operations are twice those from petroleum operations (dairy & other cattle
784 operations ROG emissions = 2.0 x oil/gas production and refining ROG emissions). The average
785 measured contributions from petroleum and dairy sources were equivalent at the Bakersfield site
786 (Figure 15). This is largely dependent the distribution of dairy operations relative to petroleum
787 operations, which is greatest in the southern part of the San Joaquin Valley (e.g. Bakersfield)
788 where the oil wells and related operations are concentrated. Thus, the ratio of petroleum to dairy

operation contributions goes up by several factors with decreased dilution and a greater influence of local sources (Table 6). This is likely also the reason for the greater contribution from motor vehicles relative to dairy operations at the Bakersfield site versus the inventory. The greater prevalence of motor vehicles near the site increases its impact relative to the whole valley.

A comparison of the dairy operations source profile (Table 5) with the CARB emission inventory reveals that the ratio of methane to NMOC is consistent between our results and the inventory, 93% vs. 92% methane. Additionally, the existing CARB inventory for the San Joaquin Valley reflects the difference in the magnitude of methane emissions between the two sources, with total methane emissions from dairy (and other cattle) operations being an order of magnitude greater than petroleum production operations, and responsible for at least 87% of methane emissions. Furthermore, for petroleum operations, the majority (81%) of fugitive methane (and ethane by inventory definition) emissions are from oil/gas marketing rather than production/refining (California Air Resources Board, 2010). Overall, these intercomparisons, while rough, provide validation of the CARB emission inventory for relative emission rates of dairy and petroleum operations in the San Joaquin Valley.

The San Joaquin Valley, and the central valley as a whole, contains a complex mixture of both anthropogenic and biogenic sources of reactive gas-phase organic carbon on both regional and urban scales. Our focus in this paper has been quantifying regional emissions from petroleum and dairy operations, comparing their emission rates to other anthropogenic sources, and evaluating their importance for air quality in the urban area of Bakersfield and the San Joaquin Valley. The dairy and petroleum sources are clearly relevant to air quality on both local and regional scales for ozone formation, but are likely not very important as sources of precursors to SOA. This study provides important new information expanding knowledge on the suite of compounds emitted from these sources and providing new useful information on their sources profiles.

826     **References**

827

828     Alanis, P., Ashkan, S., Krauter, C., Campbell, S., and Hasson, A. S.: Emissions of volatile
829             fatty acids from feed at dairy facilities. Atmos. Environ., 44, 5084-5092, 2010.
830     Altman, D. G., & Bland, J. M.: Standard deviations and standard errors. BMJ, 331, 903,
831             2005.
832     Angevine, W. M., Eddington, E., Durkee, K., Fairall, C., Bianco, L., Brioude, J.:
833             Meteorological Model Evaluation for CalNex 2010. Mon. Wea. Rev., 140, 3885–3906,
834             2012.
835     Angevine, W. M., Brioude, J., McKeen, S., Holloway, J. S., Lerner, B. M., Goldstein, A. H.,
836             Guha, A., Andrews, A., Nowak, J. B., Evan, S., Fischer, M. L., Gilman, J. B. and Bon,
837             D.: Pollutant transport among California regions, J. Geophys. Res. Atmos., 118, 6750–
838             6763, doi:10.1002/jgrd.50490, 2013.
839     Armendariz, A.: Emissions from Natural Gas Production in the Barnett Shale Area and
840             Opportunities for Cost-Effective Improvements, available at:
841             http://www.edf.org/sites/default/files/9235_Barnett_Shale_Report.pdf, Accessed Oct.
842             2012, 2009.
843     Atkinson, R., and Arey, J.: Atmospheric degradation of volatile organic compounds. Chem.
844             Rev., 103, 4605-4638, 2003.
845     Bao J. W., Michelson, S. A., Persson, P. O. G., Djalalova, I. V., and Wilczak, J. M.:
846             Observed and simulated low-level winds in a high ozone episode during the central
847             California ozone study, J. Appl. Meteor. Climatol., 47, 2372-2394, 2007.
848     Barletta, B., Carreras-Sospedra, M., Cohan, A., Nissenson, P., Dabdub, D., Meinard, S.,
849             Atlas, E., Lueb, R., Holloway, J. S., Ryerson, T. B., Pederson, J., VanCuren, R. A., and
850             Blake, D. R.: Emission estimates of HCFCs and HFCs in California from the 2010
851             CalNex study, J. Geophys. Res. Atmos., 118, 2019–2030, 2013.
852     Beaver S., and Palazoglu A.: Influence of synoptic and mesoscale meteorology on ozone
853             pollution potential for San Joaquin Valley of California, Atmos. Environ., 43, 1779-
854             1788, 2009.
855     Brioude, J., Angevine, W. M. M., McKeen, S. A., and Hsie, E. Y.: Numerical uncertainty at
856             mesoscale in a Lagrangian model in complex terrain. Geosci. Model Dev. Discuss., 5,
857             967-991, 2012.
858     Borbon, A., Gilman, J. B., Kuster, W. C., Grand, N., Chevaillier, S., Colomb, A.,
859             Dolgorouky, C., Gros, V., Lopez, M., Sarda-Esteve, R., Holloway, J. S., Stutz, J.,
860             McKeen, S., Petetin, H., Beekmann, M., Warneke, C., Parrish, D. D., and de Gouw, J.
861             A.: Emissions of anthropogenic VOCs in northern mid-latitude megacities: observations
862             vs. emission inventories in Los Angeles and Paris. J. Geophys. Res., 118, 2041-2057,

2013.

Buzcu B., and Fraser M. P.: Source identification and apportionment of volatile organic compounds in Houston, TX, Atmos. Envron., 40, 2385-2400, 2006.

California Air Resources Board: The California Almanac of Air Quality & Emissions – 2009 Edition, available at: http://www.arb.ca.gov/app/emsinv/fcemssumcat2009.php, accessed May 2012, 2010.

California Air Resources Board: Overview of final amendments to air regulations for the oil and natural gas industry, available at: http://www.epa.gov/airquality/oilandgas/pdfs/20120417fs.pdf, accessed March 2014, 2012.

California Air Resources Board: Final updates to requirements for storage tanks used in oil and natural gas production and transmission, available at: http://www.epa.gov/airquality/oilandgas/pdfs/20130805fs.pdf, accessed March 2014, 2013.

Carter, W. P. L.: SAPRC Atmospheric Chemical Mechanisms and VOC Reactivity Scales, available at: http://www.engr.ucr.edu/~carter/SAPRC/, accessed Aug. 2012, 2007.

Carter, W. P. L. and Seinfeld, J. H.: Winter ozone formation and VOC incremental reactivities in the Upper Green River Basin of Wyoming, Atmos. Environ., 50, 255–266, doi:10.1016/j.atmosenv.2011.12.025, 2012.

Chan, A. W. H., Isaacman, G., Wilson, K. R., Worton, D. R., Ruehl, C. R., Nah, T., Gentner, D. R., Dallmann, T. R., Kirchstetter, T. W., Harley, R. A., Gilman, J. B., Kuster, W. C., deGouw, J. A., Offenberg, J. H., Kleindienst, T. E., Lin, Y. H., Rubitschun, C. L., Surratt, J. D., Hayes, P. L., Jimenez, J. L., and Goldstein, A. H.: Detailed Chemical Characterization of Unresolved Complex Mixtures in Atmospheric Organics: Insights into Emission Sources, Atmospheric Processing and Secondary Organic Aerosol Formation. J. Geophys. Res., 118, 1-14, 2013.

Chung, M. Y., Beene, M., Ashkan, S., Krauter, C., and Hasson, A. S.: Evaluation of non-enteric sources of non-methane volatile organic compound (NMVOC) emissions from dairies. Atmos. Environ., 44, 786-794, 2010.

Crounse, J. D., McKinney, K. A., Kwan, A. J., and Wennberg, P. O.: Measurement of gas-phase hydroperoxides by chemical ionization mass spectrometry, Analytical Chemistry, 78, 6726–6732, 2006.

de Gouw, J.A., and Warneke, C.: Measurements of volatile organic compounds in the Earth's atmosphere using proton-transfer-reaction mass spectrometry, Mass Spec. Rev., 26, 223-257, 2007.

U.S. Environmental Protection Agency, EPA-CMB8.2, available at: http://www.epa.gov/scram001/receptor_cmb.htm, last accessed Jan. 2014.

Gentner, D. R., Harley, R. A., Miller, A. M., and Goldstein, A. H.: Diurnal and Seasonal Variability of Gasoline-Related Volatile Organic Compound Emissions in Riverside, California. Environ. Sci. Technol., 43, 4247-4252, 2009.

Gentner, D. R., Isaacman, G., Worton, D. R., Chan, A. W., Dallmann, T. R., Davis, L., Liu, S., Day, D. A., Russell, L. M., Wilson, K. R., Weber, R., Guha, A., Harley, R. A., and Goldstein, A. H.: Elucidating secondary organic aerosol from diesel and gasoline vehicles through detailed characterization of organic carbon emissions. P. Natl. Acad. Sci. USA, 109, 18318-23, 2012.

Gentner, D. R., Worton, D. R., Isaacman, G., Davis, L., Dallmann, T. R. Wood, E. C.

Herndon, S. C, Goldstein, A. H., and Harley, R. A.: Chemical speciation of gas-phase organic carbon emissions from motor vehicles and implications for ozone production potential. Environ. Sci. Technol., 47, 11837-11848, 2013.

Gentner, D. R., Ormeño, E., Fares, S., Ford, T. B., Weber, R., Park, J. H., Brioude, J., Angevine, W. M., Karlik, J. F., and Goldstein, A. H.: Emissions of terpenoids, benzenoids, and other biogenic gas-phase organic compounds from agricultural crops and their potential implications for air quality. Atmos. Chem. Phys. Discuss., 13, 28343-28393, 2013. (ACP in 2014)

Gilman, J. B., Lerner, B. M., Kuster, W. C., and de Gouw, J. A.: Source Signature of Volatile Organic Compounds from Oil and Natural Gas Operations in Northeastern Colorado. Environ. Sci. Technol., 47, 1297-1305, 2012.

Hafner, S. D., Howard, C., Muck, R. E., Franco, R. B., Montes, F., Green, P. G., Mitloehner, F., Trabue, S. L., and Rotz, C. A.: Emission of volatile organic compounds from silage: Compounds, sources, and implications. Atmos. Environ., 77, 827-839, 2013.

Hendler, A., Nunn, J., Lundeen, J., and McKaskle, R.: VOC Emissions from Oil and Condensate Storage Tanks, report prepared for the Houston Advanced Research Center. Available at: http://files.harc.edu/Projects/AirQuality/Projects/H051C/H051CFinalReport.pdf, Accessed Oct. 2012, 2009.

Heikes, B. G., Chang, W. N., Pilson, M. E. Q., Swift, E., Singh, H. B., Guenther, A., Jacob, D. J., Field, B. D., Fall, R., Riemer, D., and Brand, L.: Atmospheric methanol budget and ocean implication, Global Biogeochem. Cy., 16, 1133, 2002.

Howard, C. J., Kumar, A., Mitloehner, F., Stackhouse, K., Green, P. G., Flocchini, R. G., and Kleeman, M. J.: Direct Measurements of the Ozone Formation Potential from Livestock and Poultry Waste Emissions. Environ. Sci. Technol., 44, 2292-2298, 2010a.

Howard, C. J., Kumar, A., Malkina, I., Mitloehner, F., Green, P. G., Flocchini, R. G., and Kleeman, M. J.: Reactive Organic Gas Emissions from Livestock Feed Contribute Significantly to Ozone Production in Central California. Environ. Sci. Technol., 44, 2309-2314, 2010b.

Hu, J., Howard, C. J., Mitloehner, F., Green, P. G., and Kleeman, M. J.: Mobile source and livestock feed contributions to regional ozone formation in Central California. Environ. Sci. Technol., 46, 2781-2789, 2012.

Hu, L., Millet, D. B., Mohr, M. J., Wells, K. C., Griffis, T. J., and Helmig, D.: Sources and seasonality of atmospheric methanol based on tall tower measurements in the US Upper Midwest, Atmos. Chem. Phys., 11, 11145-11156, doi:10.5194/acp-11-11145-2011, 2011.

Katzenstein, A. S., Doezema, L. A., Simpson, I. J., Blake, D. R., and Rowland, F. S.: Extensive regional atmospheric hydrocarbon pollution in the southwestern United States, P. Nat. Acad. Sci. USA, 100, 11, 975–11, 979, 2003.

Kemball-Cook, S., Bar-Ilan, A., Grant, J., Parker, L., Jung, J., Santamaria, W., Mathews, J. and Yarwood, G.: Ozone impacts of natural gas development in the Haynesville Shale, Environ. Sci. Technol., 44, 9357–9363, doi:10.1021/es1021137, 2010.

Kirstine, W. V. and Galbally, I. E.: The global atmospheric budget of ethanol revisited, Atmos. Chem. Phys., 12, 545-555, doi:10.5194/acp-12-545-2012, 2012.

Klouda, G.A., Lewis, C.W., Stiles, D.C., Marolf, J.L., Ellenson, W.D., and Lonneman, W.A.: Biogenic contributions to atmospheric volatile organic compounds in Azusa, CA, J.

955          Geophys. Res., 107, doi: 10.1029/2001000758, 2002.
956    Leuchner M., Rappengluck B.: VOC source-receptor relationships in Houston during
957          TexAQS-II. Atmos. Envron., 44, 4056-4067, 2010.
958    Lillis, P. G., Warden, A., Claypool, G. E., and Magoon, L. B.: Petroleum Systems of the San
959          Joaquin Basin Province—Geochemical Characteristics of Gas Types, Ch. 10 in:
960          Petroleum Systems and Geologic Assessment of Oil and Gas in the San Joaquin Basin
961          Province, California, A.H. Scheirer (ed.), U.S. Geological Survey, available at:
962          http://pubs.usgs.gov/pp/pp1713/, 2007.
963    Liu, S., Ahlm, L., Day, D., Russell, L., Zhao, Y., Gentner, D., Weber, R., Goldstein, A.,
964          Jaoui, M., Offenberg, J., Kleindienst, T., Ru- bitschun, C., Surratt, J., Sheesley, R., and
965          Scheller, S.: Secondary organic aerosol formation from fossil fuel sources contribute
966          ma- jority of summertime organic mass at bakersfield, J. Geophys. Res., 117, D00V26,
967          doi:10.1029/2012JD018170, 2012.
968    Malkina, I., Kumar, A., Green, P. G., and Mitloehner, F.: Identification and Quantification of
969          Volatile Organic Compounds Emitted from Dairy Silages and Other Feedstuffs. J.
970          Environ. Qual., 40, 28-36, 2011.
971    Markovic, M. Z., VandenBoer, T. C., and Murphy, J. G.: Characterization and optimization
972          of an online system for the simultaneous measurement of atmospheric water-soluble
973          constituents in the gas and particle phases. J. Environ. Monitor., 14, 1872-1884, 2012.
974    Metcalf, A. R., Craven, J. S., Ensberg, J. J., Brioude, J., Angevine, W. M. M., Sorooshian,
975          A., Duong, H. T., Jonsson, H. H., Flagan, R. C., and Seinfeld, J. H.: Black carbon
976          aerosol over the Los Angeles Basin during CalNex. J. Geophys. Res., 117, D00V13,
977          2012.
978    Millet, D. B., Donahue, N. M., Pandis, S. N., Polidori, A., Stanier, C. O., Turpin, B. J., and
979          Goldstein, A. H.: Atmospheric volatile organic compound measurements during the
980          Pittsburgh Air Quality Study: Results, interpretations, and quantification of primary and
981          secondary contributions. J. Geophys. Res.,110:D07S07, 2005.
982    Pacsi, A. P., Alhajeri, N. S., Zavala-Araiza, D., Webster, M. D. and Allen, D. T.: Regional
983          air quality impacts of increased natural gas production and use in Texas., Environ. Sci.
984          Technol., 47, 3521–7, doi:10.1021/es3044714, 2013.
985    Peischl, J., Ryerson, T. B., Holloway, J. S., Trainer, M.,  Andrews, A. E., Atlas, E. L.,
986          Blake, D. R., Daube, B. C., Dlugokencky, E. J,, Fischer, M. L., Goldstein, A. H.,
987          Guha, A., Karl, T., Kofler, J., Kosciuch, E.., Misztal, P. K., Perring, A. E., Pollack, I. B.,
988          Santoni, G. W., Schwarz, J. P., Spackman, J. R., Wofsy, S. C., and Parrish' D. D.:
989          Airborne observations of methane emissions from rice cultivation in the Sacramento
990          Valley of California. J. Geophys. Res., 117, D00V25, 2012.
991    Peischl, J., Ryerson, T. B., Aikin, K. C., Andrews, A. E., Atlas, E., Blake, D., Brioude, J.,
992          Daube, B. C., de Gouw, J. A., Dlugokencky, E., Frost, G. J., Gentner, D. R., Gilman, J.
993          B., Goldstein, A. H., Harley, R. A., Holloway, J. S., Kofler, J., Kuster, W. C., Lang, P.
994          M., Novelli, P. C., Santoni, G. W., Trainer, M., Wofsy, S. C., and Parrish, D. D.:
995          Quantifying sources of methane using light alkanes in the Los Angeles basin,
996          California. J. Geophys. Res. 118, 4974-4990, 2013.
997    Petron, G., Frost, G., Miller, B. R., Hirsch, A. I., Montzka, S. A., Karion, A., Trainer, M.,
998          Sweeney, C., Andrews, A. E., Miller, L., Kofler, J., Bar-Ilan, A., Dlugokencky, E. J.,
999          Patrick, L., Moore, C. T., Jr., Ryerson, T. B., Siso, C., Kolodzey, W., Lang, P. M.,
1000         Conway, T., Novelli, P., Masarie, K., Hall, B., Guenther, D., Kitzis, D., Miller, J., Welsh,

1001          D., Wolfe, D., Neff, W., and Tans, P.: Hydrocarbon emissions characterization in the
1002          Colorado Front Range: A pilot study, J. Geophys. Res., 117, D04304,
1003          doi:10.1029/2011JD016360, 2012.
1004  Polissar, A. V., Hopke, P. K., and Harris, J. M.: Source regions for atmospheric aerosol
1005          measured at Barrow, Alaska. Environ. Sci. Technol., 35, 4214-4226, 2001.
1006  Pusede, S. E., Gentner, D. R., Wooldridge, P. J., Browne, E. C., Rollins, A. W., Min, K.-E.,
1007          Russell, A. R., Thomas, J., Zhang, L., Brune, W. H., Henry, S. B., DiGangi, J. P.,
1008          Keutsch, F. N., Harrold, S. A., Thornton, J. A., Beaver, M. R., St. Clair, J. M.,
1009          Wennberg, P. O., Sanders, J., Ren, X., VandenBoer, T. C., Markovic, M. Z., Guha, A.,
1010          Weber, R., Goldstein, A. H., and Cohen, R. C.: On the temperature dependence of
1011          organic reactivity, nitrogen oxides, ozone production, and the impact of emission
1012          controls in San Joaquin Valley California, Atmos. Chem. Phys. Discuss., 13, 28511-
1013          28560, doi:10.5194/acpd-13-28511-2013, 2013.
1014  Ryerson, T. B., Aikin, K. C., Angevine, W. M., Atlas, E. L., Blake, D. R., Brock, C. A.,
1015          Fehsenfeld, F. C., Gao, R.-S., de Gouw, J. A., Fahey, D. W., Holloway, J. S., Lack, D.
1016          A., Lueb, R. A., Meinardi, S., Middlebrook, A. M., Murphy, D. M., Neuman, J. A.,
1017          Nowak, J. B., Parrish, D. D., Peischl, J., Perring, A. E., Pollack, I. B., Ravishankara, A.
1018          R., Roberts, J. M., Schwarz, J. P., Spackman, J. R., Stark, H., Warneke, C. and Watts, L.
1019          A.: Atmospheric emissions from the Deepwater Horizon spill constrain air-water
1020          partitioning, hydrocarbon fate, and leak rate. Geophys. Res. Lett., 3, 2011.
1021  Schnell, R. C., Oltmans, S. J., Neely, R. R., Endres, M. S., Molenar, J. V. and White, A. B.:
1022          Rapid photochemical production of ozone at high concentrations in a rural site during
1023          winter, Nat. Geosci., 2, 120–122, doi:10.1038/ngeo415, 2009.
1024  Seibert, P., Kromp-Kolb, H., Baltensperger, U., Jost, D. T., Schwikowski, M., Kasper,
1025          A., and Puxbaum, H.: Trajectory analysis of aerosol measurements at high alpine sites,
1026          in: Transport and Transformation of Pollutants in the Troposphere, Borrell, P., Borrell,
1027          P. M., Cvitas, T., Kelly, K., Midgley, P., Seiler, W. (Eds), Academic Publishing, Den
1028          Haag, 689-693, 1994.
1029  Shaw, S. L., Mitloehner, F. M., Jackson, W., DePeters, E. J., Fadel, J. G., Robinson, P. H.,
1030          Holzinger, R., and Goldstein, A. H.: Volatile organic compound emissions from dairy
1031          cows and their waste as measured by proton-transfer-reaction mass spectrometry.
1032          Environ. Sci. Technol., 31, 1310-1316, 2007.
1033  Sheridan, M.: California Crude Oil Production and Imports, California Energy Commission,
1034          available at: http://www.energy.ca.gov/2006publications/CEC-600-2006-006/CEC-600-
1035          2006-006.PDF, 2006.
1036  Sun, H., Trabue, S. L., Scoggin, K., Jackson, W. A., Pan, Y., Zhao, Y., Malkina, I. L.,
1037          Koziel, J. A., and Mitloehner, F. M.: Alcohol, volatile fatty acid, phenol, and methane
1038          emissions from dairy cows and fresh manure. J. Environ. Qual., 37, 615-622, 2008.
1039  U.S. Energy Information Administration, U.S. Crude Oil Production by State, available at:
1040          http://www.eia.gov/dnav/pet/PET_CRD_CRPDN_ADC_MBBLPD_A.htm, accessed
1041          Aug. 2012, 2010.
1042  Wang, Y. Q., Zhang, X. Y., & Draxler, R. R.: TrajStat: GIS-based software that uses various
1043          trajectory statistical analysis methods to identify potential sources from long-term air
1044          pollution measurement data, Environ. Modell. Softw., 24, 938-939, 2009.
1045

1046 **Tables and Figures**

1047

1048 Table 1: Unrefined natural gas composition for thermogenic wet wells in the San Joaquin Valley
1049 from USGS samples (N=49 wells)

|  | wtC% | Std. Dev. | $k_{OH}$ | MIR |
|---|---|---|---|---|
| methane | 82.3 | 9.2 | 0.0064 | 0.014 |
| ethane | 5.33 | 3.46 | 0.248 | 0.28 |
| propane | 4.42 | 3.50 | 1.09 | 0.49 |
| isobutane | 0.920 | 0.837 | 2.12 | 1.23 |
| n-butane | 1.55 | 2.17 | 2.36 | 1.15 |
| isopentane | 0.223 | 0.401 | 3.6 | 1.45 |
| n-pentane | 0.273 | 0.405 | 3.80 | 1.31 |
| neopentane | 0.061 | 0.182 | 0.825 | 0.67 |
| n-hexane | 0.105 | 0.108 | 5.20 | 1.24 |
| n-heptane | 0.049 | 0.041 | 6.76 | 1.07 |

1050 Notes:
1051 -$k_{OH}$ is in $cm^3 s^{-1}$ molecules$^{-1}$ $\times 10^{12}$ from Atkinson and Arey, (2003)
1052 -MIR is in $gO_3 g^{-1}$ from Carter 2007
1053 -The observed source profile for petroleum gas emissions at the Bakersfield site is well
1054 represented by the composition of non-methane organic carbon shown here
1055

Table 2: Interquartile ranges and MIRs for alkanes discussed in this work

| Compound Name | # in Fig. 4 | Interquartile Range [pptv] | | WtC% of Unexplained Mass | MIR [gO$_3$ g$^{-1}$] |
|---|---|---|---|---|---|
| propane | - | 1133 - | 5602 | | 0.49 |
| n-butane | - | 230 - | 6397 | | 1.15 |
| n-pentane | - | 221 - | 2127 | | 1.31 |
| 2-2-dimethylbutane | 1 | 28.0 - | 76.6 | | 1.17 |
| 2-methylpentane & 2,3-dimethylbutane | 2 | 121.6 - | 501.0 | 9.92 | 1.2 |
| 3-methylpentane | 3 | 50.1 - | 253.9 | 7.67 | 1.80 |
| 2,4- & 2,2-dimethylpentane | 4 | 13.7 - | 54.7 | | 1.3 |
| 3,3-dimethylpentane | 5 | 4.0 - | 16.6 | | 1.20 |
| 2,3-dimethylpentane | 6 | 19.7 - | 93.0 | | 1.34 |
| 2-methylhexane | 7 | 23.2 - | 90.3 | 2.73 | 1.19 |
| 3-methylhexane | 8 | 28.0 - | 124.6 | 3.48 | 1.61 |
| 2,2-dimethylhexane | 9 | 1.0 - | 4.0 | | 1.02 |
| 2,5-dimethylhexane | 10 | 6.2 - | 35.8 | 1.44 | 1.46 |
| 2,4-dimethylhexane | 11 | 7.4 - | 32.0 | 0.84 | 1.73 |
| 2,2,3-trimethylpentane | 12 | 2.7 - | 12.1 | | 1.22 |
| isooctane | 13 | 39.1 - | 115.3 | | 1.26 |
| 2,3,4-trimethylpentane & ctc-1,2,3-trimethylcyclopentane | 14 | 31.6 - | 160.2 | 7.38 | 1.3 |
| 2,3,3-trimethylpentane & 2,3-dimethylhexane | 15 | 11.3 - | 32.8 | | 1.1 |
| 2-methylheptane | 16 | 10.2 - | 48.8 | 1.29 | 1.07 |
| 4-methylheptane | 17 | 4.3 - | 20.7 | | 1.25 |
| 3-methylheptane | 18 | 9.3 - | 43.6 | 1.79 | 1.24 |
| 2,2,5-trimethylhexane | 19 | 5.4 - | 16.3 | | 1.13 |
| 2,6-dimethylheptane | 20 | 5.4 - | 30.7 | 1.86 | 1.04 |
| 3,5-dimetylheptane | 21 | 2.2 - | 10.3 | | 1.56 |
| 2,3-dimethylheptane | 22 | 0.9 - | 4.7 | | 1.09 |
| 2- & 4-methyloctane | 23 | 2.9 - | 12.7 | | 0.9 |
| 3-methyloctane & 4-ethylheptane | 24 | 3.1 - | 12.9 | | 1.1 |
| 2,2,5-trimethylheptane | 25 | 0.7 - | 1.7 | | 1.26 |
| 2,2,4-trimethylheptane | 26 | 0.8 - | 2.6 | | 1.16 |
| C10 branched alkanes (5 unknown isomers) | 27 | 3.0 - | 11.5 | | 0.94 |
| 2,6-dimethyloctane | 28 | 0.7 - | 3.2 | | 1.08 |
| 2- & 3- & 4-methylnonane & 3- & 4-ethyloctane & 2,3-dimetyloctane | 29 | 6.9 - | 24.6 | | 0.94 |
| C11 branched alkanes (3 unknown isomers) | 30 | 0.7 - | 2.6 | | 0.73 |
| C11 branched alkanes (10 unknown isomers) | 31 | 5.4 - | 17.5 | | 0.73 |
| dimethylundecane isomer #1 | 32 | 0.8 - | 3.3 | | 0.6 |
| dimethylundecane isomer #2 | 33 | 0.8 - | 2.6 | | 0.6 |
| C13 branched alkanes (2 unknown isomers) | 34 | 2.3 - | 5.8 | | 0.6 |
| C14 branched alkanes (6 unknown isomers) | 35 | 4.4 - | 11.3 | | 0.55 |
| C16 branched alkane (unknown) | 36 | 1.3 - | 3.1 | | 0.47 |
| cyclopentane | 37 | 36.7 - | 164.5 | 4.14 | 2.39 |

| | | | | | |
|---|---|---|---|---|---|
| methylcyclopentane | 38 | 57.4 | - | 315.3 | 9.24 | 2.19 |
| cis-1,3-dimethylcyclopentane | 39 | 14.8 | - | 100.1 | 5.09 | 1.94 |
| trans-1,3-dimethylcyclopentane | 40 | 16.4 | - | 177.7 | 7.70 | 1.94 |
| ethylcyclopentane | 41 | 7.9 | - | 44.4 | 1.89 | 2.01 |
| ctc-1,2,4-trimethylcyclopentane | 42 | 5.4 | - | 52.2 | 4.09 | 1.53 |
| ctt-1,2,4-trimethylcyclopentane | 43 | 1.7 | - | 15.5 | 1.29 | 1.53 |
| Unknown methylethylcyclopentane | 44 | 0.7 | - | 4.3 | | 1.6 |
| iso-propylcyclopentane | 45 | 1.1 | - | 5.9 | 0.35 | 1.69 |
| n-propylcyclopentane | 46 | 2.1 | - | 10.0 | 0.56 | 1.69 |
| cyclohexane | 47 | 27.5 | - | 154.0 | 6.10 | 1.25 |
| methylcyclohexane | 48 | 20.4 | - | 147.0 | 7.17 | 1.70 |
| cis-1,3- & 1,1-dimethylcyclohexane | 49 | 4.6 | - | 38.4 | 2.91 | 1.4 |
| trans-1,2-dimethylcyclohexane | 50 | 4.6 | - | 42.4 | 3.27 | 1.41 |
| trans-1,3-dimethylcyclohexane | 51 | 2.9 | - | 17.8 | 0.91 | 1.52 |
| cis-1,2-dimethylcyclohexane | 52 | 1.9 | - | 9.8 | 0.51 | 1.41 |
| ethylcyclohexane | 53 | 4.8 | - | 31.9 | 2.31 | 1.47 |
| ccc-1,3,5-trimethylcyclohexane | 54 | 1.0 | - | 6.6 | | 1.15 |
| 1,1,3-trimethylcyclohexane | 55 | 2.0 | - | 20.4 | 2.26 | 1.19 |
| 1,1,4-trimethylcyclohexane | 56 | 1.1 | - | 8.8 | | 1.2 |
| ctt-1,2,4- & cct-1,3,5-trimethylcyclohexane | 57 | 0.7 | - | 3.9 | | 1.2 |
| ctc-1,2,4-trimethylcyclohexane | 58 | 1.2 | - | 9.6 | | 1.2 |
| 1,1,2-trimethylcyclohexane and isobutylcyclopentane | 59 | 0.7 | - | 2.0 | | 1.3 |
| methylethylcyclohexane isomer #1 | 60 | 0.8 | - | 4.5 | 0.32 | 1.4 |
| methylethylcyclohexane isomer #2 | 61 | 0.7 | - | 3.7 | 0.27 | 1.4 |
| iso-propylcyclohexane | 62 | 0.9 | - | 5.2 | | 1.3 |
| n-propylcyclohexane | 63 | 2.9 | - | 15.5 | | 1.29 |
| unidentified C10 cyclohexane | 64 | 2.5 | - | 7.8 | | 1.07 |
| unidentified C10 cyclohexanes | 65 | 0.7 | - | 2.7 | | 1.07 |
| unidentified C9 cycloalkane | 66 | 1.2 | - | 11.0 | 1.23 | 1.36 |

1057
1058

1059 Table 3: Observed light alkane ratios (gC gC$^{-1}$) from this and other studies

| | Unrefined SJV thermogenic wet gas [± std. err. (± std. dev.)] | Bakersfield ambient canister measurements [± std. dev. (r)] | Bakersfield ambient in situ measurements [± std. dev. (r)][a] | Colorado Front Range ambient in situ measurements [± std. err. (± std. dev.)][cd] | Colorado Front Range ambient canister measurements [range (± std. dev.) (r)][bd] | SW U.S. (fall) ambient canister measurements [± std. err.][cd] | SW U.S. (spring) ambient canister measurements [± std. err.][cd] | East Texas Condensate Tanks [± std. err. (± std. dev.)] |
|---|---|---|---|---|---|---|---|---|
| Data | This study / USGS | This study | This study | Gilman et al. (2013) | Petron et al. (2012) | Katzenstein et al. (2003) | Katzenstein et al. (2003) | Hendler et al. (2009) |
| ethane / propane | 1.2 ± 0.2 (± 1.2) | 0.6 ± 0.03 (r=0.93) | -- | 0.86 ± 0.06 (± 1.41) | -- | 1.1 ± 0.2 | 1.4 ± 0.1 | 0.64 ± 0.04 (± 0.20) |
| propane / n-butane | 2.9 ± 0.7 (± 4.6) | 1.8 ± 0.1 (r=0.98) | 1.9 ± 0.01 (r=0.98) | 1.5 ± 0.1 (± 2.6) | 1.5-1.7 (± 0.01) (r~1) | 1.7 ± 0.4 | 2.0 ± 0.3 | 1.3 ± 0.1 (± 0.4) |
| n-butane / isobutane | 1.7 ± 0.4 (± 2.8) | 1.7 ± 0.04 (r=0.99) | -- | 2.3 ± 0.2 (± 4.6) | -- | 2.2 ± 0.5 | 2.0 ± 0.3 | 1.9 ± 0.2 (± 0.8) |
| ethane / n-butane | 3.4 ± 0.8 (± 5.3) | 1.1 ± 0.1 (r=0.90) | -- | 1.3 ± 0.1 (± 2.2) | -- | 1.9 ± 0.4 | 2.9 ± 0.4 | 0.78 ± 0.07 (± 0.33) |
| Sample size (N) | 49 | 46 | 693 | 554 | 25+ [b] | 85 | 261 | 24 |

1060

1061 Notes:

1062 - Comparison done using C$_4$ alkanes and smaller as there are large contributions/interference

1063 from motor vehicle sources for C$_5$ and greater compounds at Bakersfield

1064 - Standard error (aka standard deviation of the mean) is reported as the primary uncertainty for

1065 the unrefined natural gas profile and others where appropriate, and represents the variability of

1066 the average within large highly variable datasets. Further information on statistical

1067 definitions/differences can be found in Altman & Bland (2005). Both the standard error and

1068 deviation are provided so the reader can judge the uncertainty and variability.

1069 - Results of positive matrix factorization (PMF), and similar studies are excluded from this

1070 comparison (Peischl et al., 2013; Buzcu and Fraser, 2006; Leuchner and Rappengluck, 2010)

1071 [a] Measurements of ethane and isobutane were unavailable from Bakersfield in situ data

1072 [b] Range of 5 data regressions, each with 25 or more samples and very small uncertainty. Other

1073 regressions were not reported in Petron et al. (2012).

1074 [c] Ratios calculated from mean mixing ratios and their standard deviations, with propagation of

1075 uncertainty

1076 [d] Studies focused on regions with large oil and gas operations

1077

1078   Table 4: Observed petroleum operations source profile at Bakersfield

| Compound | wtC% |
|---|---|
| ethane | 19.72 |
| propane | 34.02 |
| n-butane | 17.87 |
| n-pentane | 3.15 |
| n-hexane | 1.21 |
| n-heptane | 0.57 |
| isobutane | 10.61 |
| isopentane | 2.57 |
| neopentane | 0.70 |
| 2-methylpentane & 2,3-dimethylbutane | 0.95 |
| 3-methylpentane | 0.73 |
| 2-methylhexane | 0.26 |
| 3-methylhexane | 0.33 |
| 2,5-dimethylhexane | 0.14 |
| 2,4-dimethylhexane | 0.08 |
| 2,3,4-trimethylpentane & ctc-1,2,3-trimethylcyclopentane | 0.71 |
| 2-methylheptane | 0.12 |
| 3-methylheptane | 0.17 |
| 2,6-dimethylheptane | 0.18 |
| cyclopentane | 0.40 |
| methylcyclopentane | 0.89 |
| cis-1,3-dimethylcyclopentane | 0.49 |
| trans-1,3-dimethylcyclopentane | 0.74 |
| ethylcyclopentane | 0.18 |
| ctc-1,2,4-trimethylcyclopentane | 0.39 |
| ctt-1,2,4-trimethylcyclopentane | 0.12 |
| iso-propylcyclopentane | 0.03 |
| n-propylcyclopentane | 0.05 |
| cyclohexane | 0.58 |
| methylcyclohexane | 0.69 |
| cis-1,3- & 1,1-dimethylcyclohexane | 0.28 |
| trans-1,2-dimethylcyclohexane | 0.31 |
| trans-1,3-dimethylcyclohexane | 0.09 |
| cis-1,2-dimethylcyclohexane | 0.05 |
| ethylcyclohexane | 0.22 |
| 1,1,3-trimethylcyclohexane | 0.22 |
| methylethylcyclohexane isomer #1 | 0.03 |
| methylethylcyclohexane isomer #2 | 0.03 |
| unidentified C9 cycloalkane | 0.12 |

1079   Notes:
1080   - Source profile carbon fraction is 0.82
1081   - Uncertainties are defined as standard errors and conservatively ±20% based largely on the
1082   variability in the oil well data.
1083

1084     Table 5: Inter-quartile range ($Q_{25}$-$Q_{75}$) at Bakersfield shown with the source profile of dairy
1085     operations (determined using ground-site Bakersfield data and aircraft measurements in the San
1086     Joaquin Valley), and the ozone formation potential (MIR) of individual components

| Compound | IQR [ppbv] | wt% | MIR [$gO_3\ g^{-1}$] | % of Observed Concentrations from Dairy Operations during CalNex-Bakersfield [Avg. (Range)] |
|---|---|---|---|---|
| Methane | 1950 - 2380 | 93.3 | 0.014 | --- |
| Methanol | 9.5 - 25.5 | 1.4 | 0.67 | 27% (22-37%) |
| Ethanol | 3.9 - 14.3 | 4.9 | 1.57 | 45% (18-67%) |
| Acetic Acid | 0.79 - 2.5 | 0.45 | 0.68 | 28% (11-44%) |

1087     Note: There are potentially contributions from other organic compounds (e.g. carbonyls, larger
1088     alcohols, acids, alkenes). Based on our data, they are either minor or much more reactive than
1089     measured species as they could not be apportioned with significance in ambient measurements.
1090     Nevertheless, there are potentially other compounds emitted from dairy operations that have high
1091     ozone formation potential.
1092
1093

1094 Table 6: Quartiles [ppbC] for ambient concentrations from major petroleum-based sources
1095 measured at the Bakersfield site (does not include methane emissions) shown with maximum
1096 incremental reactivity (MIR) secondary organic aerosol (SOA) yields for each source

| | $Q_{25}$ | $Q_{50}$ | $Q_{75}$ | MIR [$gO_3$ $g^{-1}$] | SOA Yield [$gSOA$ $g^{-1}$] |
|---|---|---|---|---|---|
| Gasoline Exhaust | 12 | 21 | 35 | 4.5 | $0.023 \pm 0.007$ |
| Diesel Exhaust | 15 | 28 | 54 | 2.5 | $0.15 \pm 0.05$ |
| Non-tailpipe Gasoline | 4.1 | 8.1 | 18 | 2.0 | $0.0024 \pm 0.0001$ |
| Petroleum Gas Source | 7.6 | 20 | 89 | 0.82 | ~0 |
| Dairy Operations | 5.7 | 11 | 26 | 1.3 | ~0 |

1097 Note:
1098 -Gasoline and diesel exhaust include both emissions of unburned fuel and products of incomplete
1099 combustion. MIR and SOA yield values for motor vehicle sources shown for comparison from
1100 Gentner et al. (2013) and Gentner et al. (2012) for comparison.
1101 -Dairy operations includes other cattle farming in the San Joaquin Valley, and the MIR value is
1102 for NMOC fraction of source profile.
1103 - The average ozone formation potential (MIR) value is potentially an underestimate due to other
1104 organic compounds emitted, which may also impact the SOA formation potential (see Table 5
1105 note).
1106
1107
1108 Table 7: Comparison sources addressed in this study: Relative abundance of VOC emissions
1109 compared to CARB inventory for the San Joaquin Valley (SJV) and the portion of Kern County
1110 in the San Joaquin Valley including Bakersfield (Kern-SJV)

| | Relative Mass Abundance in Bakersfield (This Study) | Fraction of Emissions [%] in Inventory (Absolute Emission Rate [tons day$^{-1}$]) | |
|---|---|---|---|
| | | SJV Inventory | SJV-Kern Inventory |
| Petroleum Operations | 22% | 15% (28) | 53% (26) |
| Dairy Operations | 22% | 30% (57) | 9% (4.5) |
| On- & Off-Road Motor Vehicles | 56% | 55% (104) | 38% (19) |

1111 Notes:
1112 - Motor vehicle emissions are sum of on- and off-road since ambient source apportionment
1113 cannot discern between them; includes gasoline and diesel exhaust, and non-tailpipe gasoline
1114 emissions
1115 - Comparison is limited to discussed sources, biogenic emissions and other potentially important
1116 sources are excluded (for biogenic emissions from agriculture see Gentner et al., 2014)
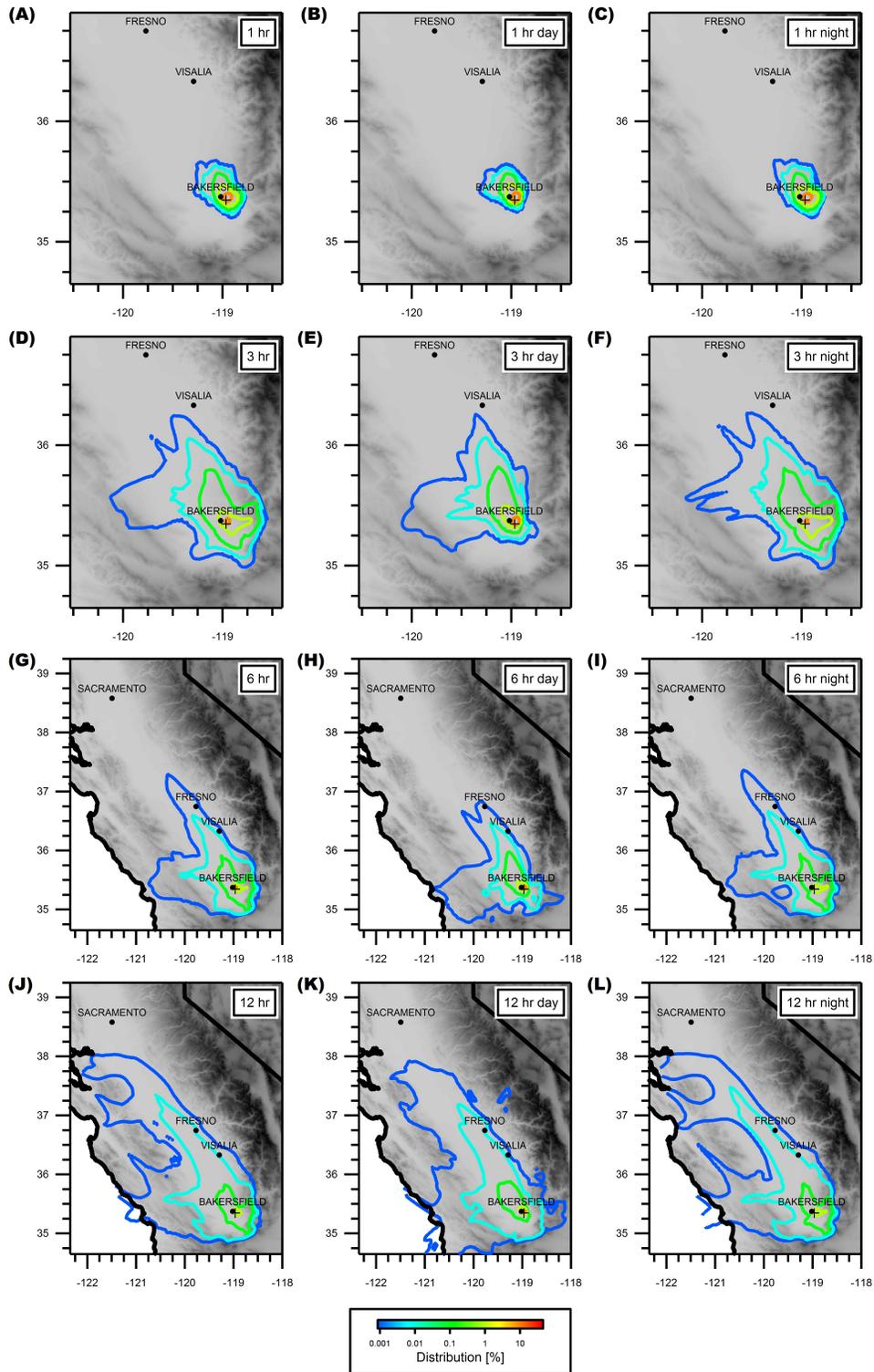
1117
Figure 1: 1, 3, 6 and 12 hour statistical footprints for the Bakersfield ground site (marked by +)
averaged across the entire CalNex campaign (y- and x-axis represent latitude and longitude). Day
**(B, E, H, K)** and nighttime **(C, F, I, L)** average are filtered for 08:00-20:00 PST and 21:00-06:00
PST, respectively, and are shown with overall averages **(A, D, G, J)**.
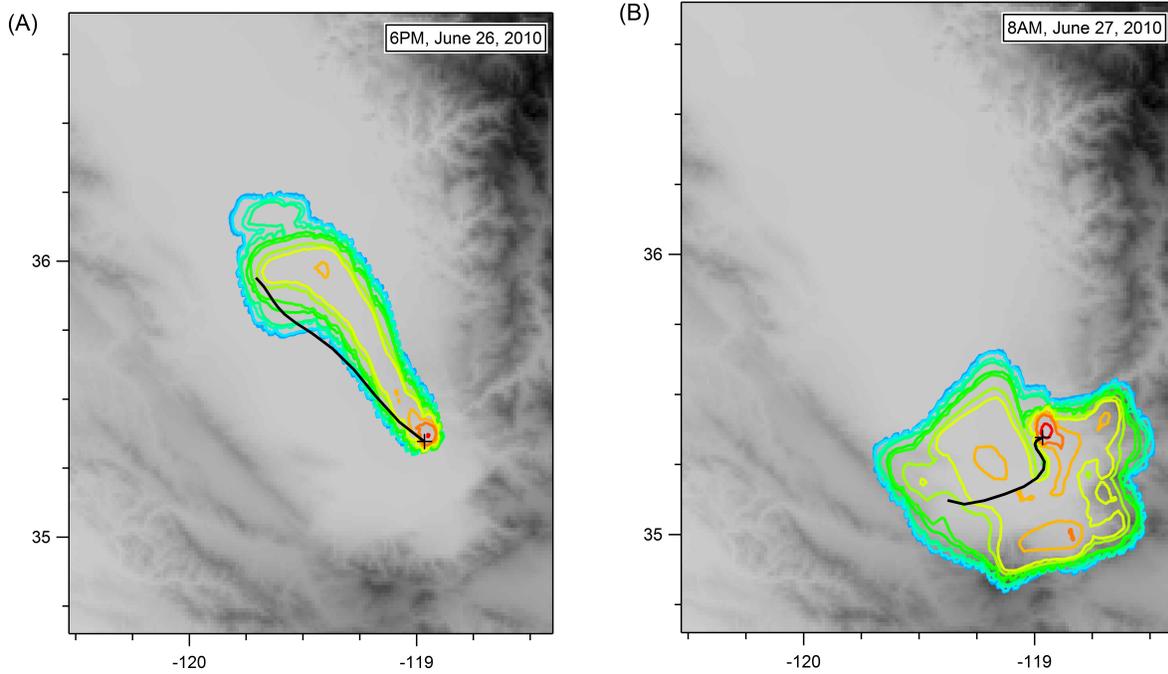
Figure 2: Examples of individual probability distribution back-trajectory footprints produced using FLEXPART-WRF (contours with log color scale – red: max, blue: min) for the Southern San Joaquin Valley with air parcels arriving at the CalNex-Bakersfield ground site. Two examples show previous 6 hours with air parcels coming **(A)** along a concentrated northwest flowpath and **(B)** a more dispersed footprint from the southern tip of the valley. Dates and arrival times are superimposed on the panels. Also shown are comparisons of single-path HYSPLIT back-trajectories (black lines) and FLEXPART-WRF footprints. Flexpart methods show some disagreement with HYSPLIT and over-simplification.
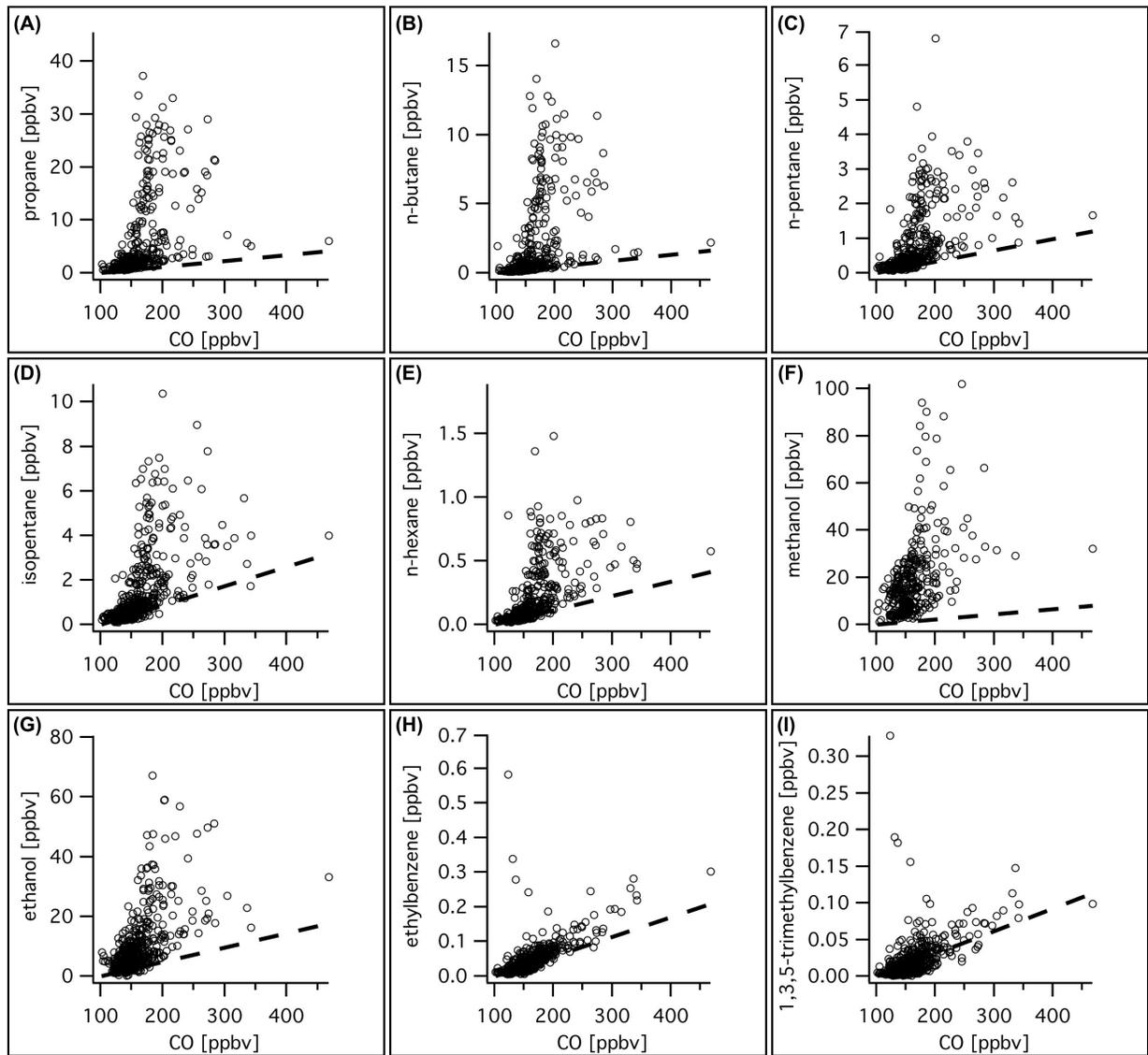
1141

1142 Figure 3: Concentrations of several compounds from Bakersfield, CA shown against carbon
1143 monoxide with the average slope of compounds vs. CO during the same time period at the
1144 CalNex-LA site in Pasadena, CA (Bourbon et al., 2012). Concentration enhancements above
1145 VOC/CO line are due to emissions from **(A-E)** petroleum operations and **(F-G)** dairy operations,
1146 neither of which emit CO. **(H-I)** are shown as examples of compounds who agree well between
1147 Bakersfield and Los Angeles.
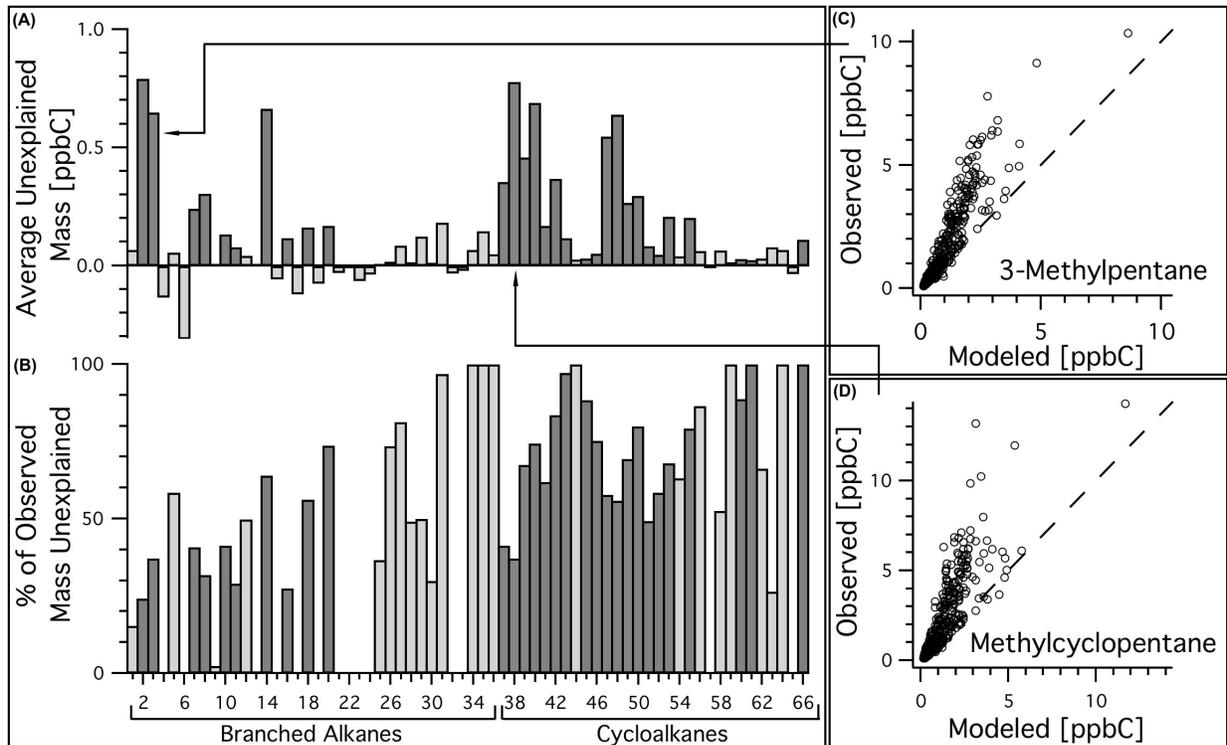
Figure 4: Many branched and cyclic alkanes exceeded predicted concentrations based on source profiles for motor vehicles. **(A-B)** The average unexplained concentration of each compound and the percentage of unexplained mass out of total observations. Compounds that are well correlated (r≥0.75) with the petroleum gas source are shown with shaded bars. A few compounds have negative residuals. **(C-D)** Examples of exceedances of observed over predicted values are shown with a 1:1 line.
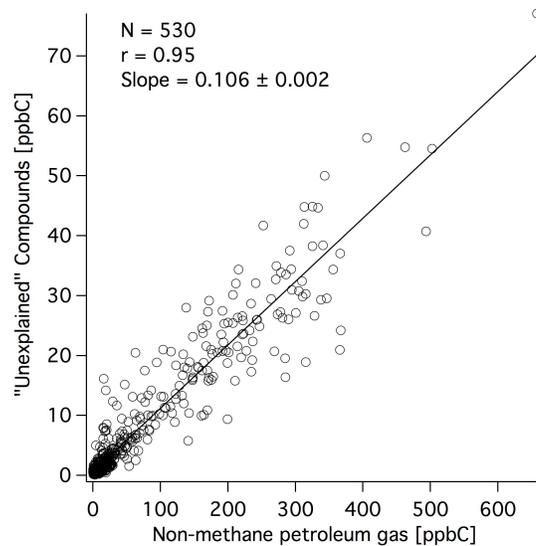


Figure 5: The sum of unexplained compounds was very well correlated with gas-phase emissions from the modeled petroleum operations source with a slope of 0.106. This increases emissions by 10.6% from the original profile

**Dairy Operation Plumes in San Joaquin Valley**

Slope = 7.4 ± 0.6 mmol mol$^{-1}$

r = 0.89

N = 47 (1 min. pts.)

_Methanol [ppbv]_ (y-axis: 5, 6, 7, 8, 9, 10, 11)

_Methane [ppbv]_ (x-axis: 1900, 2000, 2100, 2200, 2300, 2400, 2500)
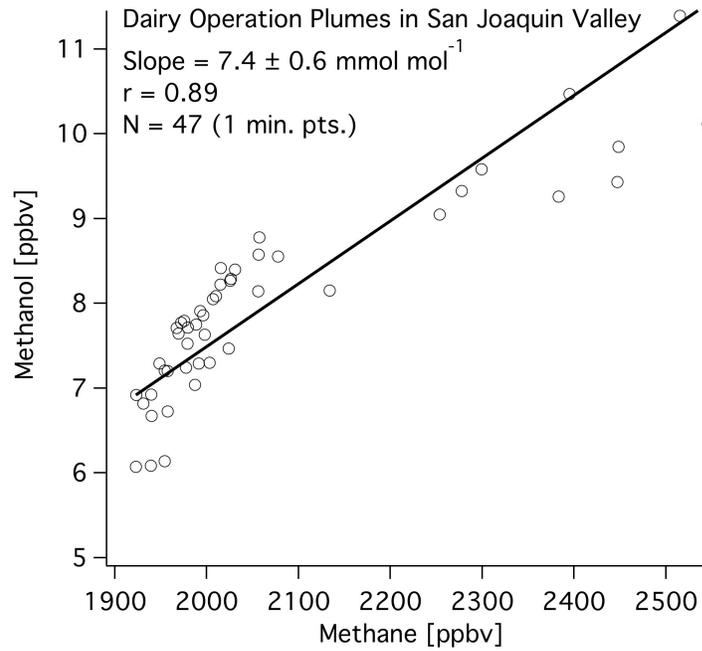
Figure 6: Methanol and methane concentrations are well-correlated in dairy operation plumes sampled via aircraft (flight dates: 5/7, 6/14, 2010). Ratios of methanol to methane average 7.4 ± 0.6 mmol mol$^{-1}$ and range up to 16 mol mol$^{-1}$ due to the heterogeneity in emission pathways at dairy operations. Note: the data shown here represent a subset of dairies in the valley measured during CalNex.

1169
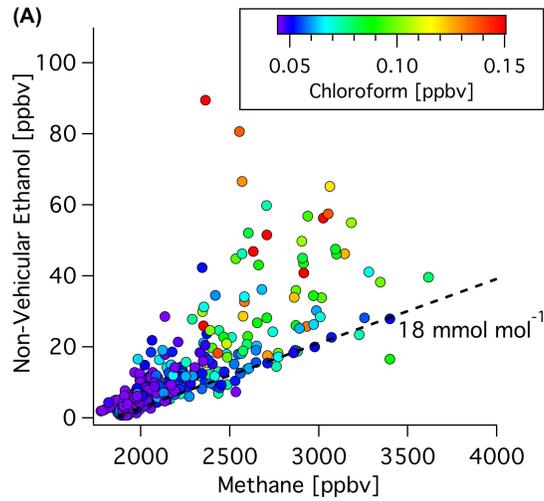
1170
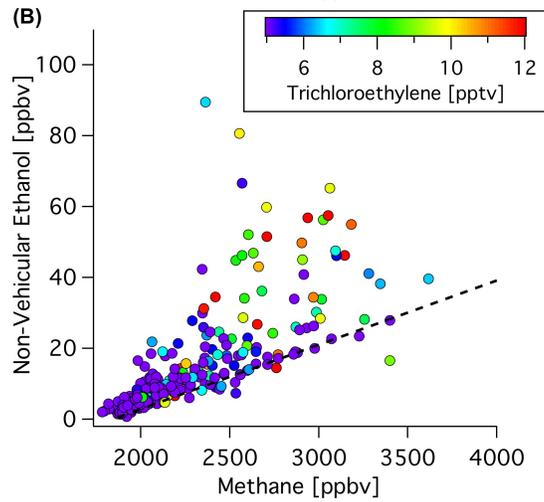
1171
Figure 7: Observations of non-vehicular ethanol vs. methane are correlated and shown with the
inferred emission ratio from dairy operations. Enhancements of ethanol from another source than
the dominant source of methane and ethanol are shown by enhancements in **(A)** chloroform, **(B)**
trichloroethylene, and **(C)** carbon disulfide. No major enhancements of methane are observed
beyond the inferred slope with non-vehicular ethanol.

1178
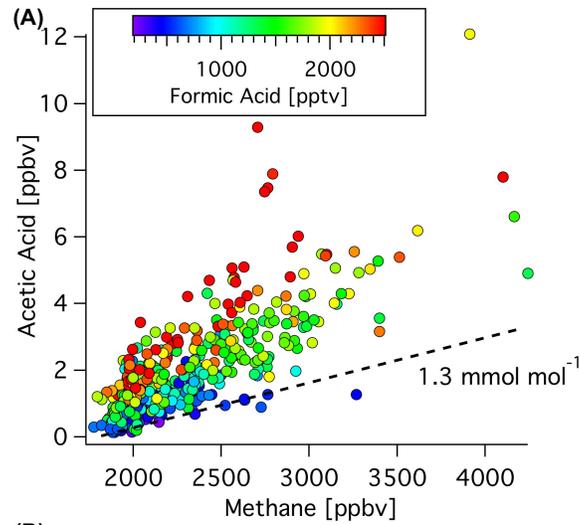


1179
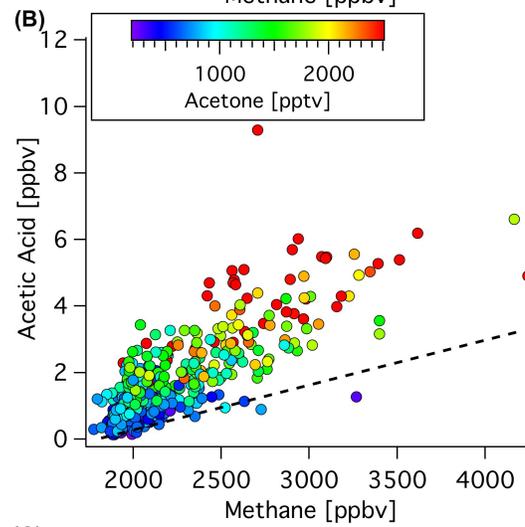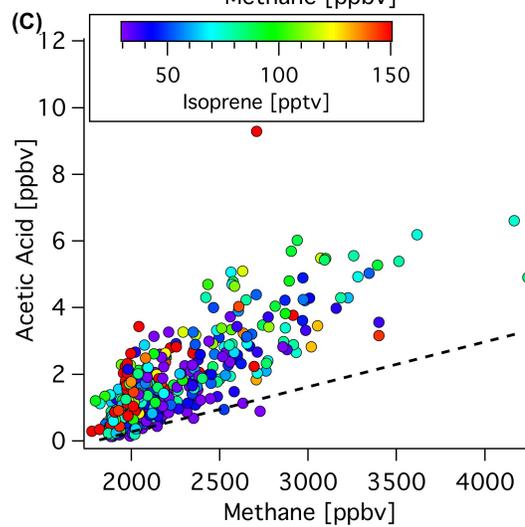


1180
1181    Figure 8: Acetic acid vs. methane shown with the inferred acetic acid:methane emission ratio
1182  from dairy operations. Acetic acid exceedances above the emission ratio are due to other sources
1183      of acetic acid coincident with emissions of **(A)** formic acid, **(B)** acetone, and **(C)** isoprene.
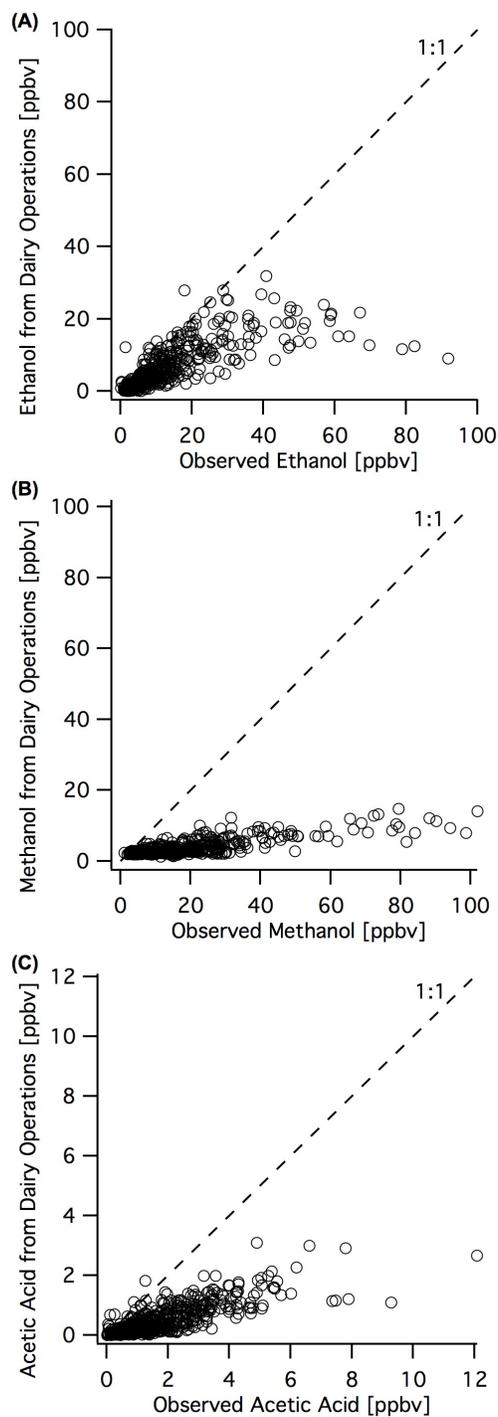
1184
1185 Figure 9: Estimated concentrations of non-methane organic compounds emitted by dairy
1186 operations shown against ambient observations at the Bakersfield ground site. Emissions are
1187 apportioned to dairy operations using emission ratios the methane determined using aircraft and
1188 ground site measurements. On average, 45% of observed **(A)** ethanol is from dairies. Whereas,
1189 smaller fractions of **(B)** methanol (27%) and **(C)** acetic acid (28%) are from dairy operations.
1190 These fractions vary with time of day and source strength. Diurnal patterns of percent
1191 contributions from dairy operations are shown in Figure S8.

1192

1193

1194
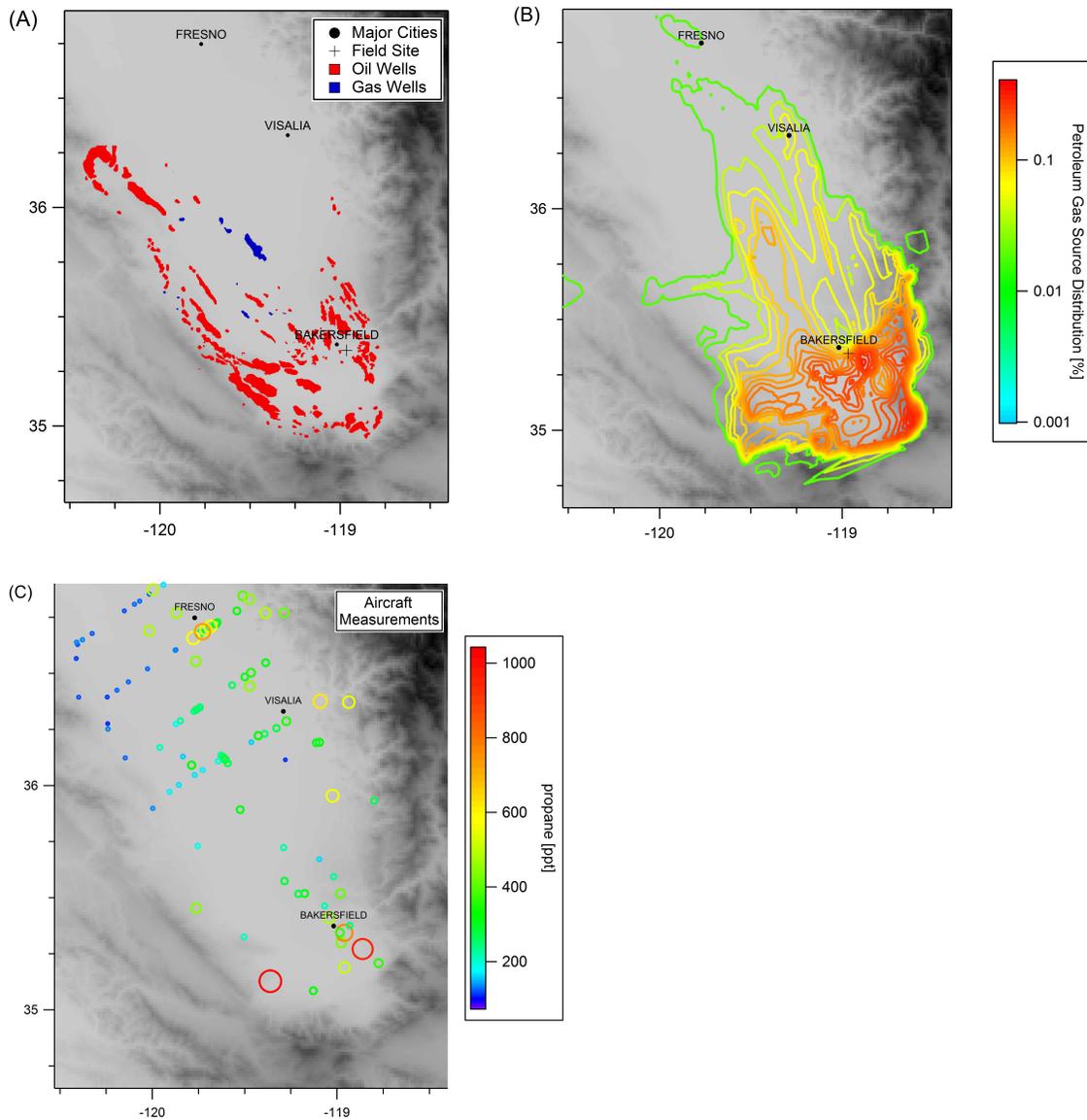


1195



1196
1197 Figure 10: Maps of southern part of the San Joaquin Valley with **(A)** the location of oil and gas
1198 wells, **(B)** the spatial distribution of petroleum gas emissions determined using statistical
1199 footprint analysis (6 hr), and **(C)** aircraft canister measurements of propane, sized and colored by
1200 concentration. Together the maps show a similar distribution of wells and emissions in the
1201 region. Note: meteorological conditions and local dilution varies between canister measurements.
1202

1203
1204
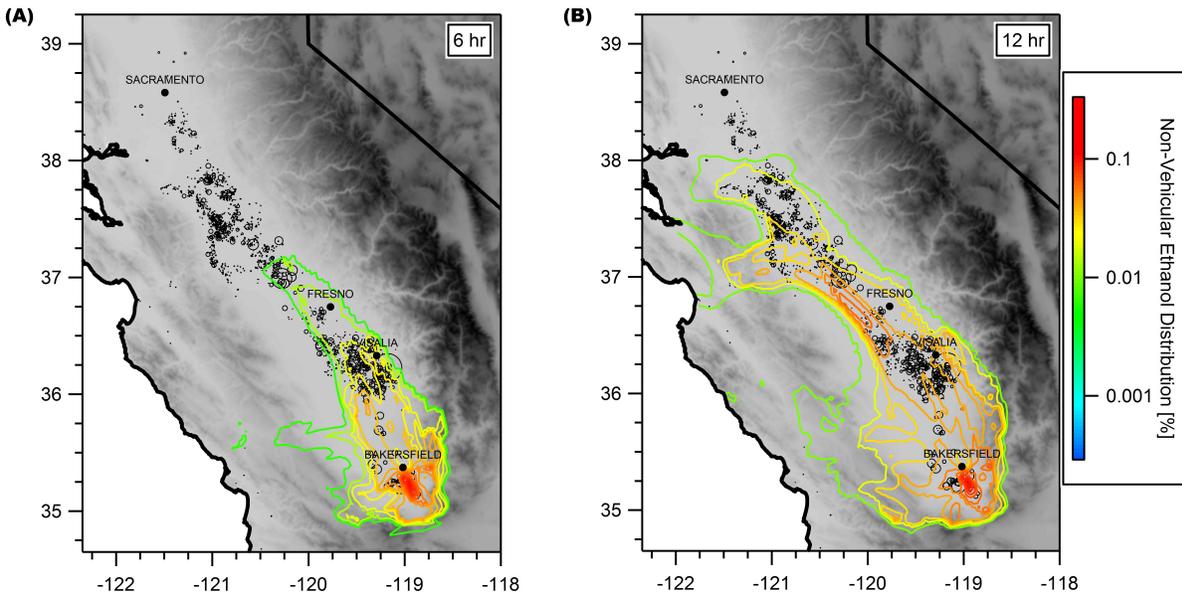


1205
1206 Figure 11: Statistical distribution of emissions of non-vehicular ethanol in the San Joaquin
1207 Valley shown as colored contours for 6 and 12-hour footprints. Modeling results shown with the
1208 location of dairies as markers (**o**) scaled by the size of each dairy.
1209
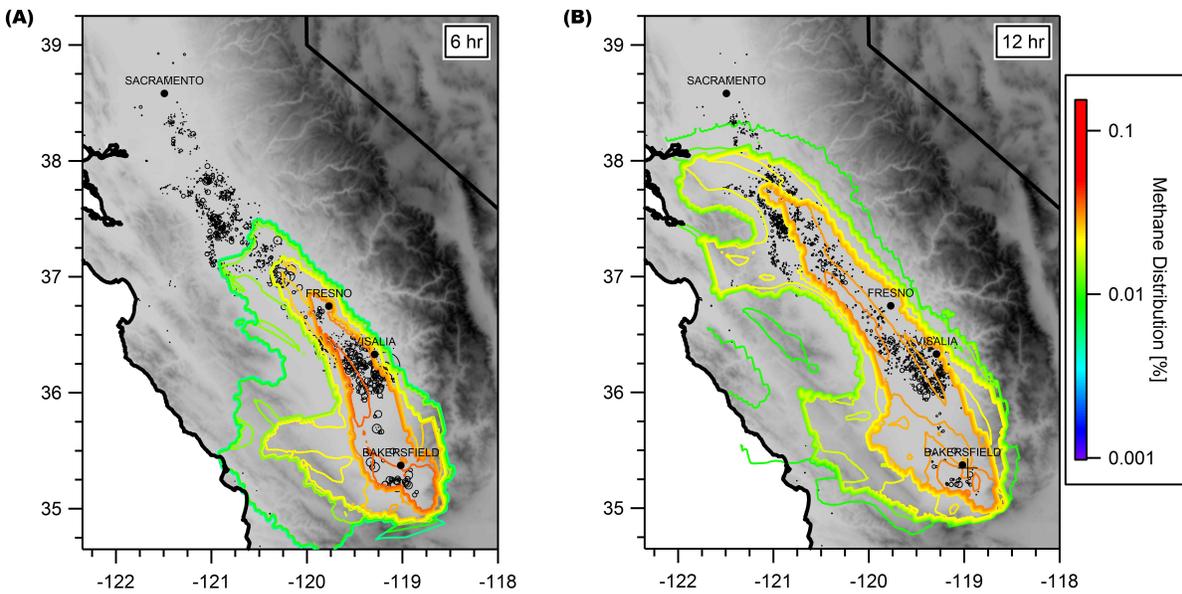


1210
1211 Figure 12: Statistical distribution of emissions of methane in the San Joaquin Valley shown as
1212 colored contours for 6 and 12-hour footprints. Modeling results shown with the location of
1213 dairies as markers (**o**) scaled by the size of each dairy.
1214
1215

1216
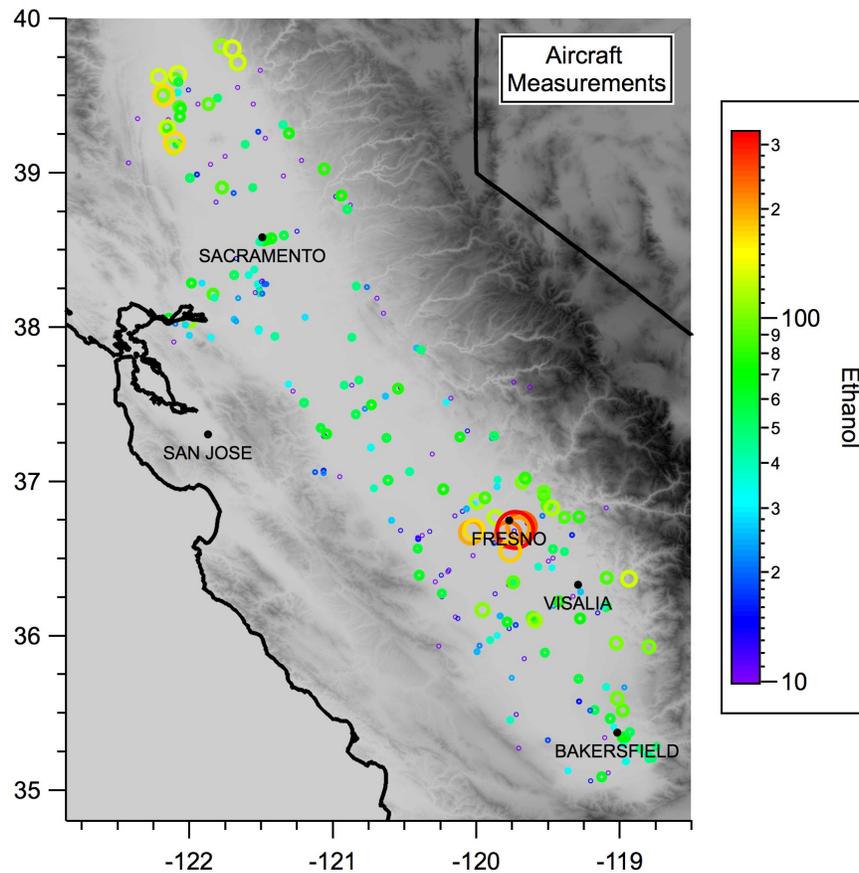


1217
1218    Figure 13: Aircraft canister measurements of ethanol in California's central valley shown as
1219    individual circles, sized and colored by ethanol concentration. Data were taken at varying
1220    altitudes above and below the boundary layer with general filter for below 1000 m. Vertical
1221    gradients are responsible for some variability, but aircraft data support conclusions of other
1222    analyses showing large ethanol sources in the central valley: dairy operations in the San Joaquin
1223    Valley and rice cultivation in the Sacramento Valley. Note: meteorological conditions and local
1224    dilution varies between canister measurements. Also, alcohol measurements made using the
1225    canisters were prone to significant losses, so their use is only relative.
1226
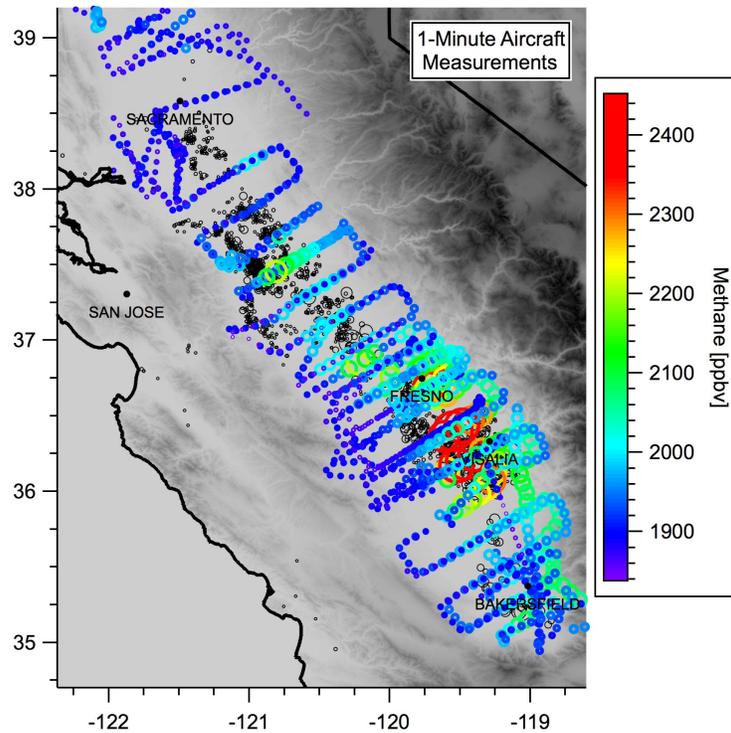
1227
1228 Figure 14: Map of observed methane concentrations over 7 flights in California's Central Valley
1229 shown as individual circles, sized and colored by methane concentration. Data were taken at
1230 varying altitudes above and below the boundary layer with general filter for below 1000 m.
1231 Vertical gradients and multiple flights are responsible for some variability, but methane
1232 enhancements in aircraft data show good correlation with the location of dairy operations (open
1233 black circles sized by bovine population). A map including the all of the Sacramento Valley can
1234 be found in the supplementary material (Figure S11).

1235

1236

## (A) Average NMOC Mass

20% Gasoline Exhaust

22% Petroleum Operations

28% Diesel Exhaust

22% Dairy Operations

8% Non-tailpipe Gasoline
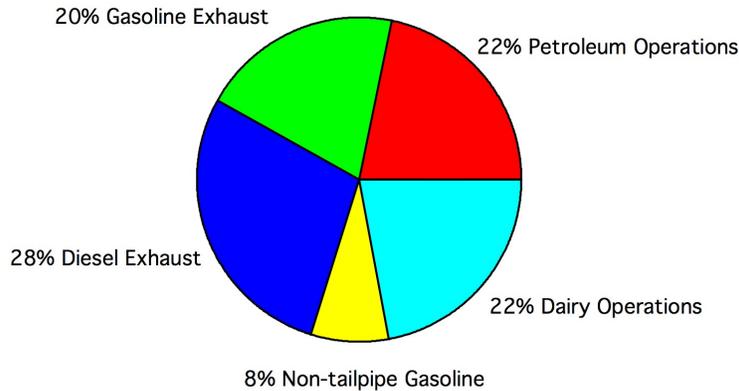
## (B) Contributions to SOA Precursors

<1% Non-tailpipe Gasoline

10% Gasoline Exhaust

90% Diesel Exhaust
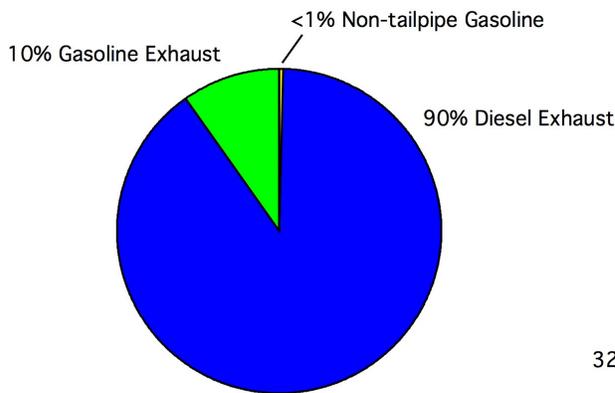
## (C) Contributions to Ozone Precursors

41% Gasoline Exhaust

8% Petroleum Operations

13% Dairy Operations

32% Diesel Exhaust

7% Non-tailpipe Gasoline
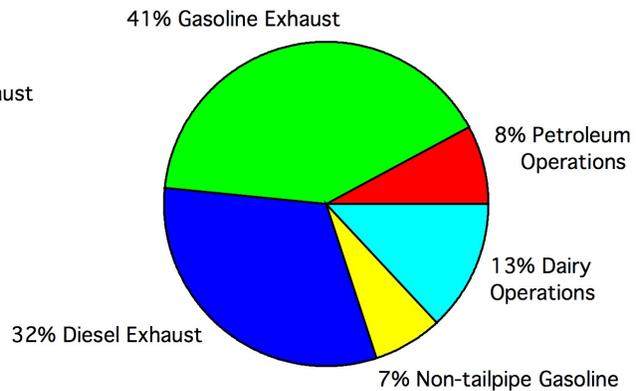
1237

1238 Figure 15: Breakdown of the contributions of prominent anthropogenic sources in Bakersfield
1239 for **(A)** total non-methane organic carbon (NMOC) mass (g), **(B)** precursors to secondary organic
1240 aerosol (SOA), and **(C)** precursors to ozone. Other sources/compounds may impact SOA
1241 formation indirectly via changes in photochemistry. The exhaust values here include unburned
1242 fuel emissions and products of incomplete combustion, and dairy operations include other cattle
1243 farming. Biogenic emissions from natural vegetation are excluded, but are likely to have
1244 important contributions to emissions and air quality in the San Joaquin Valley, but less so in the
1245 urban core of Bakersfield, CA. Note: The NMOC mass comparison mass in panel A is in terms
1246 of mass (similar to inventories), so ratios of sources will be slightly different from Table 5 where
1247 they are in mol Carbon.