



Published in final edited form as:

*Atmos Environ X*. 2024 February ; 21: . doi:10.1016/j.aeaoa.2023.100230.

## VOC source apportionment: How monitoring characteristics influence positive matrix factorization (PMF) solutions

Caroline Frischmon<sup>\*</sup>,

Michael Hannigan

Department of Mechanical Engineering, University of Colorado Boulder, Boulder, CO, 80309, USA

### Abstract

Positive matrix factorization (PMF) can be used to develop more targeted air quality mitigation strategies by identifying major sources of a pollutant in an area. This technique is dependent, however, on the ability of PMF to resolve factors that accurately represent all sources of that pollutant in an area. We investigated how the accuracy of PMF solutions might be influenced by monitoring data characteristics, such as temporal resolution, monitoring location, and species composition, to better inform the use of PMF in VOC mitigation strategies. We applied PMF to five VOC monitoring programs collected within a four-year period in Colorado and found generally consistent factors, which we identified as oil extraction, processing, and evaporation; natural gas; vehicle exhaust; and liquid gasoline/short-lived oil and gas. The main determinant influencing whether or not a dataset resolved each of these sources was whether the dataset had a comprehensive list of VOC species covering key species of each source. Pollution spikes were not well-modeled in any of the solutions. Hyperlocal and volatile chemical product factors expected to be resolved in the industrialized, urban location were also missing, highlighting three limitations of PMF analysis. Wind direction dependence and diurnal trends aided in source identification, suggesting that high-time resolution data is important for developing actionable PMF results. Based on these findings, we recommend that air monitoring for PMF-informed VOC mitigation efforts include high temporal resolution and a comprehensive array of VOC species.

### 1. Introduction

Volatile organic compounds (VOCs) are a broad category of air pollutants that originate from natural and anthropogenic sources. These pollutants affect human health in myriad

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

<sup>\*</sup> Corresponding author. caroline.frischmon@colorado.edu (C. Frischmon).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRedit authorship contribution statement

**Caroline Frischmon:** Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Michael Hannigan:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aeaoa.2023.100230>.

ways, including increased cardiovascular and respiratory conditions and cancer (Kampa and Castanas, 2008). VOCs also impact human health indirectly by contributing to secondary organic aerosol (SOA) and tropospheric ozone formation (Haagen-Smit et al., 1953; Laaksonen et al., 2008).

VOCs tend to have high spatial heterogeneity, even within the neighborhood scale (Okorn et al., 2021; Khuzestani et al., 2022). Local industrial emitters drive this spatial heterogeneity by creating highly localized VOC hotspots (Khuzestani et al., 2022; Liu et al., 2021; Healy et al., 2022). These hotspots are an environmental justice (EJ) concern, as they may lead to disproportionate VOC exposure for those living within the hotspots (Wu et al., 2012). Advancements in monitoring technology have made it possible to collect VOC data with a sufficient spatiotemporal resolution to capture this heterogeneity. As a result, VOC monitoring in EJ communities has expanded in recent years to address EJ concerns (Johnston and Cushing, 2020).

This monitoring is important for understanding VOC exposure levels, but it does not directly lead to mitigation if it is unclear where the pollution is coming from. This is especially true in EJ communities that face cumulative impacts from many industrial sources (Johnston and Cushing, 2020). Consequently, a number of studies have sought to apportion monitored VOC concentrations to specific emissions sources as a way to inform exposure reduction strategies.

Source apportionment studies have used numerous receptor models, such as chemical mass balance (CMB), principal component analysis (PCA), UNMIX, and positive matrix factorization (PMF), to attribute VOCs to sources (Miller et al., 2002; Belis et al., 2013; Hopke, 2016; Song et al., 2008; Ethirajan and Mohan, 2012; Lee et al., 2008). PMF, the most widely used model, has outperformed other source apportionment methods on simulated and actual VOC data (Miller et al., 2002; Ethirajan and Mohan, 2012; Lee et al., 2008).

PMF has successfully identified major sources of concern both regionally and locally for VOCs (Yang et al., 2022). These results have then informed recommendations for VOC control strategies (Tan et al., 2020). The source information obtained from PMF has also been used to estimate ozone formation potential (OFP) and reactivity control index (RCI), demonstrating the valuable insight PMF can provide in ozone mitigation strategies (Fu et al., 2020; Lyu et al., 2021).

Despite the potential of PMF source apportionment, there are still important limitations to consider when using the receptor model. High source collinearity increases the error in PMF predictions (Habre et al., 2011). PMF also assumes that source compositions remain constant, even though the compositions at the receptor site may change based on meteorological and photochemical conditions (Yuan et al., 2012). Recent work has shown improvements in PMF results by normalizing for dispersion and photochemical processes (Dai et al., 2020; Gu et al., 2022).

PMF results are also influenced by the temporal characteristics of the monitoring data. For example, PMF fails to capture seasonal variations in sources when the monitoring data

does not include an adequate timespan (Xie et al., 2022). When PMF is applied to coarse temporal-resolution data, sources with notable diurnal patterns, such as gasoline exhaust, or with trace species as markers have not been as well estimated (Li et al., 2020; Wang et al., 2018).

The sources identified by PMF have also depended on the input species included in the analysis. One study resolved four additional factors in the PMF analysis by including VOC and main gaseous pollutants alongside aerosol species (Chan et al., 2011). Similarly, Xie et al. found PMF results depended on if the pollutants were treated as bulk or speciated (Xie et al., 2012). They found that each species option illuminated different aspects of the sources present in the study area.

Since input data characteristics, such as temporal resolution and included species, affect the PMF source profile outputs, a single data set may not fully represent the VOC sources present at a receptor site. Furthermore, the high spatial heterogeneity of VOCs means that PMF factors from a single receptor site may not meaningfully represent an entire community. These limitations may hinder the usefulness of PMF in VOC reduction and ozone control strategies if sources are missing or poorly estimated. Therefore, it is imperative to understand the limitations of PMF results when applied to individual VOC data sets with varied monitoring characteristics.

Thus far, studies have tested the robustness of PMF by manipulating a single data set and analyzing the subsequent changes to the PMF results (Xie et al., 2012, 2022; Li et al., 2020; Wang et al., 2018; Chan et al., 2011; Tian et al., 2017). Using just one data set limits the exploration of PMF robustness, however, because the variation in input data is still limited by the initial characteristics. Typically this type of analysis only perturbs the data set across one aspect, such as temporal resolution or the number of species. No studies, to our knowledge, have taken multiple, disparate data sets from within one community to analyze the representativeness of PMF results when varying in more than one of these aspects at a time.

VOC monitoring in the Colorado Front Range, especially in Commerce City, has increased in recent years, allowing us to investigate the robustness of PMF more fully across multiple VOC datasets that feature varying temporal resolution, compounds quantified, and location within the community.

Commerce City is a disproportionately impacted community near Denver, Colorado (“Colorado enviroscreen tool). The city contains many distinct VOC sources, including a petroleum refinery, multiple highways, a railyard, and other manufacturing industries. Community concerns over industrial odors and the prevalence of health problems led to two concurrent monitoring programs near the refinery (“Suncor and third, 2022). Regulatory monitoring near a highway expansion project and in the community also occurred within the last four years in Commerce City. These four monitoring programs measured speciated VOCs either continuously or temporally aggregated using summa canisters.

PMF analysis has not yet been applied to these Commerce City-based data sets. However, prior PMF analysis of VOC monitoring throughout the Colorado Front Range has repeatedly

resolved traffic-related and oil and natural gas-related factors (Lyu et al., 2021; Hecobian and Jr, 2019; Abeleira et al., 2017; Pollack et al., 2021). Diesel combustion (Lyu et al., 2021; Hecobian and Jr, 2019), regional background (Lyu et al., 2021; Abeleira et al., 2017; Pollack et al., 2021), biogenic (Lyu et al., 2021; Hecobian and Jr, 2019; Abeleira et al., 2017), and secondary factors (Abeleira et al., 2017) have also been identified through PMF analysis.

The high density of VOC monitoring in Commerce City over the past four years provides a unique opportunity to compare the PMF results from the same geographic area but with varying monitoring characteristics. While we mainly focus on Commerce City-based datasets, this study also includes PMF results from monitoring in nearby Greeley, CO as a comparison. This investigation will provide insight into the robustness of PMF analysis. We will use these insights to develop monitoring recommendations that optimize the usefulness of PMF in VOC emission and exposure reduction efforts.

## 2. Methods

### 2.1. Data collection

**2.1.1. CW: Commerce City - West**—Commerce City - West (CW) monitoring was conducted by Boulder A. I.R. and occurred outside a home near the location shown in Fig. 1. Data included in this analysis is from 3/14/22 to 6/7/22. VOCs were separated using a gas chromatograph and then identified and quantified using a flame ionization detector (FID) and mass spectrometer (MS) detectors. The time resolution of the data is one measurement every 10 min. Wind speed and direction were recorded using a Campbell Scientific MetSENS500 and RM Young 05305-PT wind monitor.

Carbon monoxide (CO), carbon dioxide (CO<sub>2</sub>), and methane (CH<sub>4</sub>) were monitored with a Picarro G-2401 cavity ring-down spectrometer. NO<sub>x</sub> was monitored using a Teledyne Model API T200UP chem-illuminescence analyzer. These pollutants were included in the PMF analysis to aid in source identification. Additional information on this monitoring program is available from Boulder A.I.R. (“Commerce city current air conditions, 2022).

**2.1.2. CS: Commerce City - south**—For the Commerce City - South (CS) monitoring, the Denver Department of Public Health and Environment (DDPHE) and Colorado Department of Transportation (CDOT) measured C<sub>6</sub>–C<sub>12</sub> VOCs at Swansea Elementary school in the Elyria-Swansea neighborhood using a Chromatotec Airmo C<sub>6</sub>–C<sub>12</sub> analyzer. Monitoring began in 2017 and is ongoing as of January 2023. However, due to power outages and other issues, data is only used from 11/14/18 to 8/31/19. The monitoring had a time resolution of one measurement per 30 min. Wind speed and direction were recorded using an RM Young 05305 V wind monitor. Additional information on this monitoring program is available from the Denver Department of Public Health and Environment (“Swansea ambient air monitoring station).

**2.1.3. CD: Commerce City - dispersed**—Two separate datasets were examined from Montrose Air Quality Services: one from canister sampling and another from mobile monitoring. We refer to the canister sampling dataset as Commerce City - Dispersed (CD) and the mobile monitoring dataset as CD - Mobile Monitoring (CD-MM).

Canister data was collected using Entech Instruments Silonite™ CS1200E Passive Canister Samplers connected to a 6-L Summa canister. 1-hour samples were collected quarterly across approximately 10 locations shown in 1 from 7/13/21 to 9/12/22. Scheduled canister samples at three additional reference locations outside of Commerce City were also included as a comparison in this dataset. These locations were the E470-I25 Junction (JUNC), the Brighton Fire Department (BFD), and the CDPHE CAMP air monitoring station (CAMP) in downtown Denver.

Triggered 1-h canister samples were also collected at individual Commerce City sites when a continuous photo-ionic detection (PID) VOC sensor at a site measured over 1 ppm of total VOCs for over 1 min. A total of 66 scheduled canisters and 16 triggered canisters were available for PMF analysis. Of the 16 triggered canisters, two canisters had to be excluded from the analysis due to highly elevated VOC levels, which did not allow for a converged PMF solution. This is discussed further in section 4.1.

The canister sampling and analysis followed the United States Environmental Protection Agency (USEPA) Compendium Method TO-14 A “Determination of Volatile Organic Compounds (VOCs) in Ambient Air using Specially Prepared Canisters with Subsequent Analysis by Gas Chromatography” and the TO-15 entitled “Determination of Volatile Organic Compounds (VOCs) in Air Collected in Specially Prepared Canisters and Analyzed by Gas Chromatography/Mass Spectrometry (GC/MS) (EPA, 1999a; EPA, 1999b).” Analysis was completed by Enthalpy Analytical in Durham, North Carolina.

Mobile monitoring was conducted using a Mercedes 2500 Sprinter Van equipped with an Ionicon Model 6000-X2 proton transfer reaction time-of-flight mass spectrometer. The van followed a route throughout a three-mile radius of the refinery in Commerce City, driving at approximately 10 miles per hour with a sampling rate of one measurement per second. Additional information on the CD and CD-MM monitoring programs is available from Montrose Air Quality Services (“Quality Assurance Project Plan Suncor, 2021).

**2.1.4. CE and G: Commerce City - east and greeley**—A Gas Chromatography–Mass Spectrometry (GC/MS) instrument was used by the Colorado Department of Public Health and Environment (CDPHE) at Eagle Pointe Community Center in Commerce City, CO to collect speciated VOC measurements from 5/17/21–7/17/21 (CE). The instrument was also used at Bella Romero 4–8 Academy in Greeley, CO from 8/12/20 to 10/8/20 (G). The time resolution for this data is one measurement per 45 min. Wind speed and direction were measured using a Gill MaxiMet GMX501 meteorological station. Monitoring at these sites also included a cavity ring-down spectrometer for measuring CH<sub>4</sub> levels and a Teledyne T200U monitor for measuring NO<sub>x</sub> levels. These compounds were included in the PMF analysis to aid in source identification. Additional information on the CE and G monitoring programs are available from CDPHE (“Oil and gas activity monitoring).

## 2.2. PMF model

Positive matrix factorization (PMF) is an advanced multivariate factor analysis tool used widely in the source apportionment of environmental pollutants. The mathematical theory and principles of the model are described in detail elsewhere (Paatero and Tapper, 1994;

Paatero, 1997). As an overview, PMF decomposes the pollutant concentration matrix into a factor profile matrix ( $f$ ) and a factor contribution matrix ( $g$ ) according to Eq. (1).

$$X_{ij} = \sum_{k=1}^p g_{ik} f_{kj} + e_{ij} \quad (1)$$

where  $X_{ij}$  represents the pollutant concentration of the sample  $i$  of chemical species  $j$ .  $g_{ik}$  is the  $k$ th factor contribution to sample  $i$  and  $f_{kj}$  is the factor profile of species  $j$  in the  $k$ th factor.  $e_{ij}$  represents the residuals. In PMF,  $g_{ik}$  and  $f_{kj}$  are constrained to greater than or equal to 0.

PMF uses a weighted least square method to minimize the objective function  $Q$ , shown in Eq. (2), where  $s_{ij}$  is the sample uncertainty. The PMF solution is that with the lowest value of  $Q$ .

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

### 2.3. PMF implementation

We used the EPA PMF 5.0 for the source apportionment of VOCs (Norris et al., 2014). Not all of the VOC species measured in the monitoring programs were included in the PMF analysis. Species that are highly reactive (Maximum Incremental Reactivity  $> 4$  were excluded to not bias the model (Brown et al., 2007; “Tables of maximum incremental reactivity). However, following other PMF research efforts, we elected to include some reactive species in the model if they are key marker species for a factor (ethylene and propylene for vehicle exhaust and isoprene for biogenic), or if they are important for public health considerations (xylenes, ethylbenzene) (Brown et al., 2007; Zheng et al., 2018; Guo et al., 2011). Species that were often missing in the specific dataset or that had a signal-to-noise ratio of less than 0.5 are also excluded. Species with a signal-to-noise of less than 1 and greater than 0.5 were ran as “weak,” which triples the provided uncertainty (Norris et al., 2014). Finally, some species were omitted due to presumed contamination or other monitoring-related issues. A full list of species and the parameters influencing their inclusion in the PMF model is available in the supplementary information (Tables S1–S5).

Background levels of carbon dioxide, carbon monoxide, and methane, approximated as the 5th percentile concentration in each dataset, were subtracted from the samples when these species were included in the PMF analysis. Actual background levels will vary with location, wind direction, and other factors. We approximate constant background levels because each dataset has different time averaging, and there were no concurrent background measurements collected. Future studies could seek to include background measurements to more accurately subtract background pollutant levels.

For each data set, samples with missing values were removed, while values below the method detection limit (MDL) were replaced with half the value of the MDL. Samples where more than 85% of the measurements were below the MDL were removed. Uncertainty values for each measurement were set as 15% of the measurement value if the measurement was above the MDL. For values at or below the MDL, the uncertainty was set at 5/6 of the MDL. For samples that were flagged with additional uncertainty, the uncertainty value was set as 50% of the measurement.

All of the monitoring datasets include multi-point calibration and blank subtraction. One of the potential limitations of this study is that the measurements were not conducted by the authors, and therefore we are not able to fully vet the uncertainties of the measurements. This is why our study used a universal 15% uncertainty factor. Other studies have investigated the impact of uncertainty estimations on PMF solutions (Christensen and Schauer, 2008; Kim and Hopke, 2007); however, this was not the focus of our work here. Further information on data quality assurance is available directly from the monitoring programs (“Commerce city current air conditions, 2022; “Swansea ambient air monitoring station; “Quality Assurance Project Plan Suncor, 2021; “Oil and gas activity monitoring).

The number of factors for each solution was determined by investigating a combination of species residuals, factor interpretability, and error estimations. Three error estimations in the EPA PMF program were used for this process: displacement (DISP), bootstrapping (BS), and BS-DISP. Detailed descriptions of these error estimations are described elsewhere (Norris et al., 2014). PMF solutions were considered interpretable if the decrease in Q was less than 1% and no swaps occurred for the smallest dQ<sub>max</sub> in DISP and BS-DISP. Over 80% of factors also had to map in BS to be considered interpretable. FPEAK values from -1 to 1 were tested to explore the rotational stability of the solutions. None of the PMF solutions showed significant differences with FPEAK values in this range, so all solutions shown in this paper are FPEAK = 0.

### 3. Results

Two factors were identified in all converged PMF solutions: oil extraction, processing, and evaporation, and liquid gasoline/short-lived oil and gas (OG). A natural gas factor was identified in all solutions except CS, and a vehicle exhaust factor was identified in all solutions except G.

Fig. 3 shows some consistency in factor contributions across solutions, even though species vary between datasets. Overall, there are high total measured VOCs (tmVOC) contributions from natural gas and oil extraction, processing, and evaporation. The liquid gasoline/short-lived OG and the vehicle exhaust factors typically contributed the majority of observed benzene, toluene, ethylbenzene, and xylenes (BTEX) concentrations. Relatively high levels of ethane resolved in the vehicle exhaust factors, even though ethane is not associated with vehicle exhaust. This likely weights the tmVOC contributions of vehicle exhaust to be higher than the true contribution in Fig. 3.

### 3.1. Mobile monitoring

PMF could not converge on a solution for CD-MM data. The number of sources contributing to pollutant concentrations in a sample may change as the mobile monitor moves through a community. Factor profiles may also change as the mobile monitor moves closer to or further away from sources in a community, which varies the impacts of photochemical decay and dispersion. We believe the added source variability implicit in mobile monitoring did not allow PMF to converge on a solution, since the model assumes constant factor profiles and the number of factors. For PMF to succeed in future mobile monitoring campaigns, it's likely that PMF will need to account for this added variability in order to converge on a solution.

### 3.2. Oil extraction, processing, and evaporation factor

Oil extraction, processing, and evaporation factors in CW, CD, CE, and G were mainly identified by their elevated levels of ethane, propane, butanes, and pentanes (Fig. 4). C2–C5 alkanes have been attributed to oil and gas activity within Colorado, and prior PMF studies have used these key species to identify refinery emissions and oil and gas activity factors (Lyu et al., 2021; Pollack et al., 2021; Thompson et al., 2014; Buzcu and Fraser, 2006). When compared to EPA speciate profiles, the high concentrations of butanes and pentanes, along with isoprene in the CC factor, suggests this factor may also include some evaporative and biogenic emissions (Harley et al., 1992; Sadeghi et al., 2022; Sharkey et al., 2008).

CS monitoring did not quantify the key C2–C5 alkanes used to identify the other oil extraction, processing, and evaporation factors. The CS factor was instead identified by comparing its predominantly C6–C7 alkane composition with the apportionment of C6–C7 alkanes in the other PMF solutions. These species resolved a high percentage of their species sum into the oil extraction, processing, and evaporation factor. This intercomparison of solutions and related identification of the oil extraction, processing, and evaporation factor in the CS solution is only possible through the availability of multiple data sets within Commerce City. Without the key species, it would be difficult to identify this important source within the region if CS was analyzed alone. As such, it is important when conducting PMF with an individual dataset to ensure the dataset includes a full array of VOC compounds with key species of each of the likely VOC sources within an area. Otherwise, major sources might be missed or misinterpreted.

Wind direction analyses of the CE and CW solutions indicate that the factor originates in the direction of the refinery and fuel storage tanks in Commerce City (Fig. 5), indicating possible sources of oil processing and/or fuel evaporation. In the G solution, the factor shows less obvious wind dependence, which is likely because the factor originates from the many disparate extraction sites surrounding the monitoring location in Greeley, CO. The average total measured VOC concentration is much higher in the G factor compared to the Commerce City-based datasets (Fig. S1).

### 3.3. Natural gas

The natural gas factors in CW, CD, CE, and G were identified by the presence of methane, ethane, and propane (Fig. 6) (Zheng et al., 2018). CS, which did not include these three



compounds, did not resolve a natural gas factor in its solution. The CW solution separated a methane-enhanced factor from a propane and ethane-enhanced factor. A four-factor solution combined both natural gas factors into one but resulted in poorer BS and BS-DISP results, as well as significantly lower methane  $R^2$  (0.92 for 5-factor and 0.49 for 4-factor). This suggests that there is some difference in the origin of methane compared to ethane and propane at this site.

The natural gas factor could originate from leakages in the natural gas distribution system or from natural gas extraction in the surrounding region (Tollefson, 2012; Jackson et al., 2014). Wind direction analysis shows how the likely source of this factor varies by monitoring site (Fig. 7). The G natural gas factor appears to originate locally, given its high concentration at low windspeeds. This factor is likely emitted from the well pads near the monitoring site. In CW, the methane factor also is most concentrated at low windspeeds, whereas the ethane/propane factor comes from the south and the northeast at more variable windspeeds. The CE natural gas factor comes mostly from the north. These directions may point toward local leakages or regional extraction activity.

### 3.4. Vehicle exhaust

CW, CE, CD, and CS solutions resolved vehicle exhaust factors that were identified by the presence of acetylene, ethylene, propylene, nitrous oxides, carbon monoxide, and/or benzene (Fig. 8) (Abeleira et al., 2017; Harley et al., 1992). Diurnal patterns associated with rush hour were also present for all factors, further validating the identification as vehicle exhaust (Figs. S2–S3). Wind analysis indicated an association between high concentrations of this factor and wind coming from the direction of nearby major roads (Fig. 9).

27–36% of ethane resolved into the vehicle exhaust factors, even though vehicle exhaust is not expected to contain ethane. This will bias tmVOC attribution, as ethane is a major component of tmVOC concentrations in the datasets.

G is the only dataset that did not include a vehicle exhaust factor in its solution. This may be attributed in part to its monitoring location, which is much less traffic-dense than the other four monitoring locations. This dataset also did not include the key species used to identify most other vehicle exhaust factors: acetylene, ethylene, propylene, nitrous oxides, and carbon monoxide. Benzene is included in G but is not considered a key vehicle exhaust species in Greeley, CO, as Halliday et al. found that oil and gas activities, not traffic-related sources, are the primary source of benzene here (Halliday et al., 2016). Including key vehicle exhaust species may have helped PMF successfully disentangle a vehicle exhaust factor in G, as it is likely some traffic air quality impacts still exist at the monitoring location.

### 3.5. Liquid gasoline/short-lived oil and gas

A factor was resolved in each dataset that was composed of alkanes and aromatics, specifically ethylbenzene, toluene, and xylenes (Fig. 10). In the CE solution, only the higher alkanes were present, as the aromatic species mostly resolved into the vehicle exhaust factor.

Over 50% of many of the C6–C9 compounds are attributed to this factor on a species percentage basis. Prior studies have specified factors enriched in aromatics and C6–C9 alkanes as liquid or unburned gasoline fuel (Harley et al., 1992; Na and Pyo Kim, 2007). However, in areas with high oil and gas activity, this factor has also been identified as short-lived oil and gas emissions (Abeleira et al., 2017; Sadeghi et al., 2022; Bari and Kindzierski, 2018). The liquid gasoline/short-lived oil and gas (OG) factors show similar diurnal trends as the vehicle exhaust factors in some solutions (CE, CW, G) and as the oil extraction, processing, and evaporation factors in others (CS) (Fig. S2). The CW factor shows similar wind direction dependence as its oil extraction, processing, and evaporation factor, whereas the CE factor shows similar wind direction dependence as its vehicle exhaust factor (Fig. 11). Finally, correlation coefficients between all factors in a solution show inconsistent evidence of whether this factor is more correlated with oil extraction, processing, and evaporation or vehicle exhaust in each solution (Fig. S3). Given these inconclusive results and the monitoring site's proximity to both roadways and oil and gas activity, we cannot specifically identify this factor as either liquid gasoline or short-lived OG.

## 4. Analysis

Applying PMF to comparable datasets from similar regions and time periods allowed us to better evaluate the strengths and limitations of PMF. Here we outline the major findings from this analysis that can be used to inform future PMF studies seeking to inform VOC mitigation plans.

### 4.1. Assumption of constant source profile

PMF could not converge when applied to mobile monitoring data, revealing the model's limitations when applied to data with varying sources. High concentrations were also not well-modeled in any of the datasets (Fig. 12), suggesting that pollution spikes originate from infrequent sources or sources that do not hold a consistent composition profile. In CD, two triggered canister samples with very high alkane levels even had to be removed from the dataset in order for PMF to resolve a stable solution. PMF assumes constant source profiles, so any temporary changes in factor composition cannot be captured.

Poor modeling of pollution spikes in these PMF solutions has important implications for the use of PMF in VOC mitigation applications. Although PMF can resolve factors contributing to more consistently elevated VOC levels, it was not effective here at revealing the sources responsible for short, intense pollution episodes. It is worth noting that the continuous monitoring data in this analysis was collected for relatively short periods of time (less than 1 year). PMF analysis of longer-term, continuous monitoring, where several peaks occur, may be able to better capture these variable episodes. If short, intense pollution episodes are the main concern regarding VOC exposure, PMF analysis of short-term monitoring may not be the best tool to inform mitigation strategies.

### 4.2. Spatial influence

PMF does not appear to be sensitive to hyperlocal sources in this study. Commerce City, CO is an urban region containing many local VOC-emitting industrial and commercial

facilities, which can contribute to VOC hotspots (Khuzestani et al., 2022; Liu et al., 2021; Healy et al., 2022). As such, we expected the Commerce City-based datasets to show high spatial heterogeneity in the PMF factor solutions. However, each of the datasets collected throughout Commerce City showed nearly the same factors. With the exception of vehicle exhaust (see section 4.4), PMF also resolved nearly the same Commerce City-based factors in Greeley, CO; a more rural region with less diverse industrial and commercial impacts. While the expected spatial heterogeneity is not directly reflected in the PMF factors, it is likely contributing to the generally lower overall  $R^2$  values for Commerce City species as compared to Greeley species (Fig. 13).

This comparison suggests that PMF analysis in places with hyperlocal or infrequent sources may not reveal the full scope of present sources, even if PMF resolves a stable solution. PMF factors should then be interpreted more generally, rather than as a representation of hyperlocal sources at the community scale.

The CD dataset is made up of summa canister samples collected at all monitoring locations on scheduled days and at individual sites when triggered by elevated total VOC concentrations. Factor mixtures were similar across all locations each day but varied more broadly across days (Fig. 14). This suggests that the sources resolved in the PMF solutions vary more temporally than they do spatially within the area of Commerce City covered by this monitoring. The similar factor mixtures at regional sites (CAMP, JUNC, BFD) suggest that this might be true beyond Commerce City as well, but there are not enough regional canisters included in this data set to conclude beyond Commerce City.

### 4.3. Temporal resolution

All monitoring analyzed in this study occurred between 2019 and 2022, but only CD and CW overlap in time (Fig. 2). Despite the overall lack of overlap, the PMF solutions reveal similar solutions, suggesting that major sources are not changing year-to-year.

CW, CS, CE, and G contain continuous samples, while CD is made up of discrete, quarterly canister samples. There do not appear to be major differences in PMF factor profiles based on the temporal resolution of measurements here. However, there are differences in how well the PMF factors reconstruct the observed pollutant concentrations. CD shows worse fits compared to the other four datasets (Fig. 13), possibly because it could not rely on diurnal patterns to disentangle factors. This effect has been seen in other PMF studies using low-time resolution data (Li et al., 2020; Wang et al., 2018).

High-time resolution data also provided supplementary information that could be used to further inform mitigation strategies. For example, diurnal trends confirmed the identity of the vehicle exhaust factors. Diurnal trend analysis shows when factors peak throughout the day, which could also be used to target sources at critical times of the day. Wind direction analysis using high-time resolution data helped to narrow the potential sources of some factors, such as the CW oil extraction, processing, and evaporation factor. Given the profile alone, there was not enough information to distinguish whether this factor was coming from the nearby refinery and fuel storage or from surrounding oil fields. Wind direction analysis, however, clearly showed that high concentrations of the factor originated from the direction

of the refinery. Although it may not be essential solely for general source identification, the advantages of high-time resolution data are important to consider when the objective is to develop PMF-informed, source-specific mitigation strategies.

#### 4.4. Species availability

When comparing VOC PMF results across time, space, and species availability, it appears that VOC species availability is the strongest driver influencing PMF factor outcomes. Tracer species that are not VOCs, such as NO<sub>x</sub>, CO, and methane, were also important for resolving their associated factors. With the exception of the CS oil extraction and processing factor, if a dataset did not include the key species of a factor, it did not resolve that factor. CS did not resolve a natural gas source without C1–C4 alkanes. G is missing a vehicle exhaust factor without acetylene, ethylene, propylene, NO<sub>x</sub>, or CO. These missing factors may still be present at the monitoring sites but are excluded in PMF analysis. This will impact how tmVOC and BTEX concentrations are attributed at the site, leading to potentially misleading conclusions about source impacts in an area.

Oxygenated VOCs (OVOCs) and terpenes were largely missing from all monitoring datasets (CE quantified  $\alpha$ -pinene, which was excluded from PMF analysis due to missing values). OVOCs and terpenes play a key role in secondary pollutant formation globally. Additionally, these species may help to better deconvolute some sources, such as biogenic, volatile chemical product, and industrial emissions, resulting in better apportionment of all VOCs. Caution should be used with data sets that do not include comprehensive species, as it may mean the PMF solution does not represent all sources within an area. This could potentially lead to misguided mitigation strategies for both primary and secondary pollution.

#### 4.5. Volatile chemical products not resolved by PMF

Factors representing volatile chemical products (VCPs), such as cleaning agents, personal care products, adhesives, and coatings, are not present in any of the solutions, even though recent work has shown that VCPs make up approximately one-half of emitted VOCs in industrialized cities (McDonald et al., 2018). VCPs have been similarly underestimated in other source apportionment studies (Niedojadlo et al., 2007; Baudic et al., 2016). These sources may be underrepresented because VCPs have more variable source profiles, which PMF cannot capture, due to the many different products contributing to VCP emissions. VCPs may also be underrepresented because the VOC species often used in source apportionment studies are biased towards fossil fuels, rather than chemical products (McDonald et al., 2018).

### 5. Conclusion

This study provides insights into how PMF analysis can be optimized to strategically inform VOC mitigation efforts. We found that air monitoring for PMF analysis should be stationary to allow for more stable source profiles. Highly elevated pollutant concentrations, hyperlocal sources, and volatile chemical products (VCPs) were not well-modeled overall, indicating that PMF may not be a sufficient source apportionment tool when short-term pollution spikes, hyperlocal VOC hotspots, and VCPs are the main concern, at least for monitoring

data that is collected for less than 1 year. High-time resolution data was not necessary to identify the main four factors here. However, high-time resolution led to more actionable PMF outcomes by providing insight into diurnal patterns and wind direction dependence. Consequently, we recommend that high-time resolution data be used in PMF studies where the objective is to directly inform mitigation strategies, rather than just identify general source categories. Finally, we found that air monitoring for PMF analysis should include a comprehensive suite of pollutant species to ensure key species for each present source are included. The recommendations highlighted here will help to ensure that PMF solutions more accurately represent the VOC sources present near a monitoring site, leading to better-informed mitigation plans.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We would like to thank Boulder A.I.R. (Detlev Helmig), the Colorado Department of Public Health and Environment (Derek Price and Daniel Bon), the Denver Department of Public Health and Environment and the Colorado Department of Transportation (William Obermann and Tehya Stockman), and Montrose Air Quality Services (Austin Heitmann) for sharing their monitoring data for this analysis. We are also grateful to Daniel Bon and Jeff Collett, Jr. for their valuable feedback on this project. Finally, we sincerely thank the organization Cultivando for providing community insight on the Commerce City air quality concerns that motivated this work.

This research was funded by NIEHS 1R01ES033478.

Publication of this article was funded by the University of Colorado Boulder Libraries Open Access Fund.

## Data availability

The authors do not have permission to share data.

## References

- Abeleira A, Pollack IB, Sive B, Zhou Y, Fischer EV, Farmer DK, 2017. Source characterization of volatile organic compounds in the Colorado Northern Front Range Metropolitan Area during spring and summer 2015. *J. Geophys. Res. Atmos.* 122 (6), 3595–3613. [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016JD026227](https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016JD026227).
- Bari M, Kindzierski W, 2018. Ambient volatile organic compounds (VOCs) in communities of the Athabasca oil sands region: sources and screening health risk assessment. *Environmental Pollution* 235, 602–614. [PubMed: 29331893]
- Baudic A, Gros V, Sauvage S, Locoge N, Sanchez O, Sarda-Estève R, Kalogridis C, Petit J-E, Bonnaire N, Baisnée D, Favez O, Albinet A, Sciare J, Bonsang B, 2016. Seasonal variability and source apportionment of volatile organic compounds (VOCs) in the Paris megacity (France). *Atmos. Chem. Phys.* 16, 11961–11989.
- Belis CA, Karagulian F, Larsen BR, Hopke PK, 2013. Critical review and meta-analysis of ambient particulate matter source apportionment using receptor models in Europe. *Atmos. Environ.* 69, 94–108.
- Black F, High L, 1980. Passenger Car Hydrocarbon Emissions Speciation. EPA 600 2–80-085, Accessed via EPA SPECIATE.
- Brown H, 2011. Composition of natural gas for use in the oil and natural gas sector rulemaking. <https://www.regulations.gov/document?D=EPA-HQ-OAR-2010-0505-0084>. Memo to Bruce Moore, EPA/OAQPS/SPPD, EPA-HQ-OAR-2010-0505, Accessed via EPA SPECIATE.

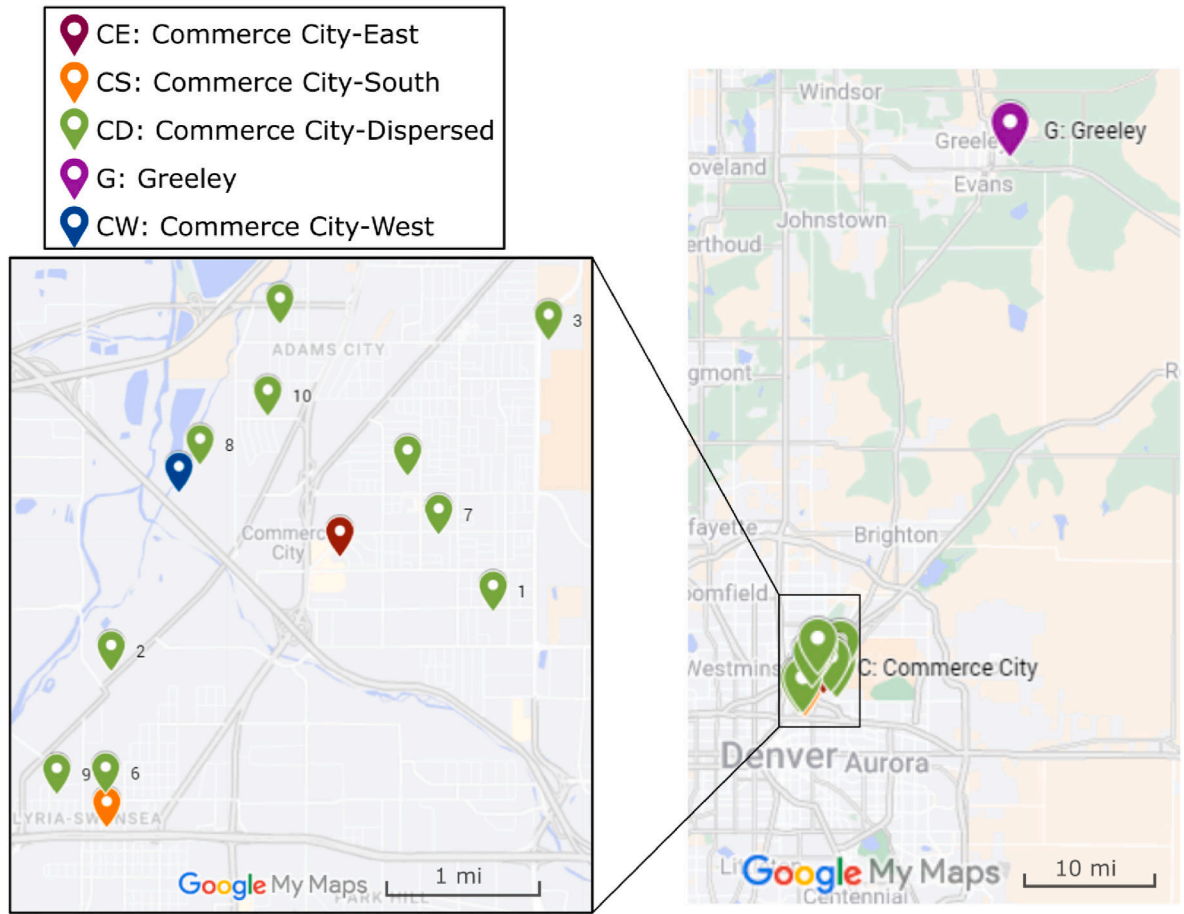
- Brown SG, Frankel A, Hafner HR, 2007. Source apportionment of VOCs in the Los Angeles area using positive matrix factorization. *Atmos. Environ.* 41, 227–237.
- Buzcu B, Fraser MP, 2006. In: *Source Identification and Apportionment of Volatile Organic Compounds*. Atmospheric Environment, Houston, TX, p. 16.
- Chan Y. c., Hawas O, Hawker D, Vowles P, Cohen DD, Stelcer E, Simpson R, Golding G, Christensen E, 2011. Using multiple type composition data and wind data in PMF analysis to apportion and locate sources of air pollutants. *Atmos. Environ.* 45, 439–449.
- Christensen WF, Schauer JJ, 2008. Impact of species uncertainty Perturbation on the solution stability of positive matrix factorization of atmospheric particulate matter data. *Environmental Science & Technology* 42, 6015–6021. [PubMed: 18767659]
- Dai Q, Liu B, Bi X, Wu J, Liang D, Zhang Y, Feng Y, Hopke PK, 2020. Dispersion normalized PMF provides insights into the significant changes in source contributions to PM<sub>2.5</sub> after the COVID-19 Outbreak. *Environmental Science & Technology* 54, 9917–9927. [PubMed: 32672453]
- EPA, 1999a. EPA air method, Toxic Organics-15 (TO-15): Compendium of methods for the Determination of Toxic organic compounds in ambient air. In: *Determination of Volatile Organic Compounds (VOCs) in Air Collected in Specially-Prepared Canisters and Analyzed by Gas Chromatography/Mass Spectrometry (GC/MS)*, second ed. EPA/625/R-96/010b.
- EPA, 1999b. Method TO-14A - Determination of Volatile Organic Compounds (VOCs) in Ambient Air Using Specially Prepared Canisters with Subsequent Analysis by Gas Chromatography. EPA 625/R-96/010b.
- Ethirajan R, Mohan S, 2012. Comparative Evaluation of VOC Source Profiles Developed by PMF and UNMIX Models. *International Journal of Environmental Science and Development*, pp. 450–455.
- Fu S, Guo M, Luo J, Han D, Chen X, Jia H, Jin X, Liao H, Wang X, Fan L, Cheng J, 2020. Improving VOCs control strategies based on source characteristics and chemical reactivity in a typical coastal city of South China through measurement and emission inventory. *Sci. Total Environ.* 744, 140825. [PubMed: 32755775]
- Gentner DR, Ford TB, Guha A, Boulanger K, Brioude J, Angevine WM, de Gouw JA, Warneke C, Gilman JB, Ryerson TB, Peischl J, Meinardi S, Blake DR, Atlas E, Lonneman WA, Kleindienst TE, Beaver MR, Clair JMS, Wennberg PO, VandenBoer TC, Markovic MZ, Murphy JG, Harley RA, Goldstein AH, 2014. Emissions of organic carbon and methane from petroleum and dairy operations in California's San Joaquin Valley. *Atmos. Chem. Phys.* 14 (10), 4955–4978. Accessed via EPA SPECIATE.
- Gu Y, Liu B, Dai Q, Zhang Y, Zhou M, Feng Y, Hopke PK, 2022. Multiply improved positive matrix factorization for source apportionment of volatile organic compounds during the COVID-19 shutdown in Tianjin, China. *Environ. Int.* 158, 106979. [PubMed: 34991244]
- Guo H, Cheng HR, Ling ZH, Louie PKK, Ayoko GA, 2011. Which emission sources are responsible for the volatile organic compounds in the atmosphere of Pearl River Delta? *J. Hazard Mater.* 188, 116–124. [PubMed: 21316844]
- Haagen-Smit AJ, Bradley CE, Fox MM, 1953. Ozone Formation in photochemical oxidation of organic Substances. *Ind. Eng. Chem.* 45 (9).
- Habre R, Coull B, Koutrakis P, 2011. Impact of source collinearity in simulated PM<sub>2.5</sub> data on the PMF receptor model solution. *Atmos. Environ.* 45, 6938–6946.
- Halliday HS, Thompson AM, Wisthaler A, Blake DR, Hornbrook RS, Mikoviny T, Müller M, Eichler P, Apel EC, Hills AJ, 2016. Atmospheric benzene observations from oil and gas production in the Denver-Julesburg Basin in July and August 2014. *J. Geophys. Res. Atmos.* 121 (18), 11,055–11,074. <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016JD025327>.
- Harley RA, Hannigan MP, Cass GR, 1992. Respeciation of organic gas emissions and the detection of excess unburned gasoline in the atmosphere. *Environmental Science & Technology* 26, 2395–2408.
- Healy RM, Sofowote UM, Wang JM, Chen Q, Todd A, 2022. Spatially resolved source apportionment of industrial VOCs using a mobile monitoring Platform. *Atmosphere* 13, 1722.
- Hecobian A, Jr JLC, 2019. *Garfield County Volatile Organic Carbon (VOC) Source Apportionment (2008–2018)*. Tech. Rep.. Fort Collins, CO.

- Hopke PK, 2016. Review of receptor modeling methods for source apportionment. *J. Air Waste Manag. Assoc.* 66, 237–259. 10.1080/10962247.2016.1140693 eprint: [PubMed: 26756961]
- Jackson RB, Down A, Phillips NG, Ackley RC, Cook CW, Plata DL, Zhao K, 2014. Natural gas Pipeline Leaks across Washington, DC. *Environmental Science & Technology* 48, 2051–2058. [PubMed: 24432903]
- Johnston J, Cushing L, 2020. Chemical exposures, health and environmental justice in communities living on the fenceline of industry. *Current environmental health reports* 7, 48–57. [PubMed: 31970715]
- Kampa M, Castanas E, 2008. Human health effects of air pollution. *Environmental Pollution* 151, 362–367. [PubMed: 17646040]
- Khuzestani RB, Liao K, Liu Y, Miao R, Zheng Y, Cheng X, Jia T, Li X, Chen S, Huang G, Chen Q, 2022. Spatial variability of air pollutants in a megacity characterized by mobile measurements. *Atmos. Chem. Phys.* 22, 7389–7404.
- Kim E, Hopke PK, 2007. Comparison between sample-species specific uncertainties and estimated uncertainties for the source apportionment of the speciation trends network data. *Atmos. Environ.* 41, 567–575.
- Kishan S, Burnette A, Fincher S, Sabisch M, Crews W, Snow R, Zmud M, Santos R, Bricka S, Fujita E, et al., 2008. Kansas City PM characterization study: Final report. Report, Assessment and Standards Division, Office of Transportation and Air Quality, US EPA, and the Eastern Research Group Inc 5608. Accessed via EPA SPECIATE.
- Laaksonen A, Kulmala M, O’Dowd CD, Joutsensaari J, Vaattovaara P, Mikkonen S, Lehtinen KEJ, Sogacheva L, Dal Maso M, Aalto P, Petäjä T, Sogachev A, Yoon YJ, Lihavainen H, Nilsson D, Facchini MC, Cavalli F, Fuzzi S, Hoffmann T, Arnold F, Hanke M, Sellegri K, Umann B, Junkermann W, Coe H, Allan JD, Alfarra MR, Worsnop DR, Riekkola M-L, Hyötyläinen T, Viisanen Y, 2008. The role of VOC oxidation products in continental new particle formation. *Atmos. Chem. Phys.* 8, 2657–2665.
- Lee S, Liu W, Wang Y, Russell AG, Edgerton ES, 2008. Source apportionment of PM<sub>2.5</sub>: comparing PMF and CMB results for four ambient monitoring sites in the southeastern United States. *Atmos. Environ.* 42, 4126–4137.
- Lewis C, 2004. In: EPA Office of Research and Development - National Exposure Research Laboratory. Personal Communication (Internal Data Collection Effort). Ying Hsu, E.H. Pechan and Associates, Inc. RTP, NC. Accessed via EPA SPECIATE.
- Li Z, Ho K-F, Yim SHL, 2020. Source apportionment of hourly-resolved ambient volatile organic compounds: influence of temporal resolution. *Sci. Total Environ.* 725, 138243. [PubMed: 32298889]
- Liu X, Yan F, Hua H, Yuan Z, 2021. Identifying hotspots based on high-resolution emission inventory of volatile organic compounds: a case study in China. *J. Environ. Manag.* 288, 112419.
- Lyu C, Capps SL, Kurashima K, Henze DK, Pierce G, Hakami A, Zhao S, Resler J, Carmichael GR, Sandu A, Russell AG, Chai T, Milford J, 2021. Evaluating oil and gas contributions to ambient nonmethane hydrocarbon mixing ratios and ozone-related metrics in the Colorado Front Range. *Atmos. Environ.* 246, 118113.
- McDonald BC, de Gouw JA, Gilman JB, Jathar SH, Akherati A, Cappa CD, Jimenez JL, Lee-Taylor J, Hayes PL, McKeen SA, Cui YY, Kim S-W, Gentner DR, Isaacman-VanWertz G, Goldstein AH, Harley RA, Frost GJ, Roberts JM, Ryerson TB, Trainer M, 2018. Volatile chemical products emerging as largest petrochemical source of urban organic emissions. *Science* 359, 760–764. [PubMed: 29449485]
- Miller SL, Anderson MJ, Daly EP, Milford JB, 2002. Source apportionment of exposures to volatile organic compounds. I. Evaluation of receptor models using simulated exposure data. *Atmos. Environ.* 36, 3629–3641.
- Na K, Pyo Kim Y, 2007. Chemical mass balance receptor model applied to ambient C<sub>2</sub>–C<sub>9</sub> VOC concentration in Seoul, Korea: effect of chemical reaction losses. *Atmos. Environ.* 41, 6715–6728.
- Niedojadlo A, Becker KH, Kurtenbach R, Wiesen P, 2007. The contribution of traffic and solvent use to the total NMVOC emission in a German city derived from measurements and CMB modelling. *Atmos. Environ.* 41, 7108–7126.

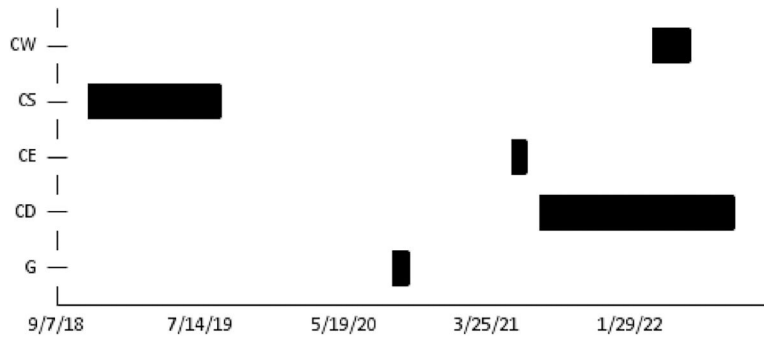
- Norris G, Duvall R, Brown S, Bai S, 2014. Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide. EPA.
- Okorn K, Jimenez A, Collier-Oxandale A, Johnston J, Hannigan M, 2021. Characterizing methane and total non-methane hydrocarbon levels in Los Angeles communities with oil and gas facilities using air quality monitors. *Sci. Total Environ.* 777, 146194. [PubMed: 34602658]
- Paatero P, 1997. Least squares formulation of robust non-negative factor analysis. *Chemometr. Intell. Lab. Syst.* 37, 23–35.
- Paatero P, Tapper U, 1994. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5 (2), 111–126. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.3170050203>.
- Pollack IB, Helmig D, O'Dell K, Fischer EV, 2021. Seasonality and source apportionment of nonmethane volatile organic compounds at Boulder Reservoir, Colorado, between 2017 and 2019. *J. Geophys. Res. Atmos.* 126 (9), e2020JD034234. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020JD034234>.
- Sadeghi B, Pouyaei A, Choi Y, Rappenglueck B, 2022. Influence of seasonal variability on source characteristics of VOCs at Houston industrial area. *Atmos. Environ.* 277, 119077.
- Sharkey TD, Wiberley AE, Donohue AR, 2008. Isoprene emission from Plants: why and how. *Ann. Bot.* 101, 5–18. [PubMed: 17921528]
- Song Y, Dai W, Shao M, Liu Y, Lu S, Kuster W, Goldan P, 2008. Comparison of receptor models for source apportionment of volatile organic compounds in Beijing, China. *Environmental Pollution* 156, 174–183. [PubMed: 18234404]
- Tan Q, Liu H, Xie S, Zhou L, Song T, Shi G, Jiang W, Yang F, Wei F, 2020. Temporal and spatial distribution characteristics and source origins of volatile organic compounds in a megacity of Sichuan Basin, China. *Environ. Res.* 185, 109478. [PubMed: 32276165]
- Thompson CR, Hueber J, Helmig D, 2014. Influence of oil and gas emissions on ambient atmospheric non-methane hydrocarbons in residential areas of Northeastern Colorado. *Elementa: Science of the Anthropocene* 3, 000035.
- Tian Y, Xiao Z, Wang H, Peng X, Guan L, Huangfu Y, Shi G, Chen K, Bi X, Feng Y, 2017. Influence of the sampling period and time resolution on the PM source apportionment: study based on the high time-resolution data and long-term daily data. *Atmos. Environ.* 165, 301–309.
- Tollefson J, 2012. Air sampling reveals high emissions from gas field. *Nature* 482, 139–140. [PubMed: 22318579]
- Wang Q, Qiao L, Zhou M, Zhu S, Griffith S, Li L, Yu JZ, 2018. Source apportionment of PM<sub>2.5</sub> using hourly measurements of Elemental Tracers and major Constituents in an urban environment: investigation of time-resolution influence. *J. Geophys. Res. Atmos.* 123 (10), 5284–5300 eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2017JD027877>.
- Wu XM, Fan ZT, Zhu X, Jung KH, Ohman-Strickland P, Weisel CP, Liyo PJ, 2012. Exposures to volatile organic compounds (VOCs) and associated health risks of socio-economically disadvantaged population in a “hot spot” in Camden, New Jersey. *Atmos. Environ.* 57, 72–79.
- Xie M, Hannigan MP, Dutton SJ, Milford JB, Hemann JG, Miller SL, Schauer JJ, Peel JL, Vedal S, 2012. Positive matrix factorization of PM<sub>2.5</sub>: comparison and implications of using different speciation data sets. *Environmental Science & Technology* 46, 11962–11970. [PubMed: 22985292]
- Xie M, Lu X, Ding F, Cui W, Zhang Y, Feng W, 2022. Evaluating the influence of constant source profile presumption on PMF analysis of PM<sub>2.5</sub> by comparing long- and short-term hourly observation-based modeling. *Environmental Pollution* 314, 120273. [PubMed: 36170893]
- Yang Y, Liu B, Hua J, Yang T, Dai Q, Wu J, Feng Y, Hopke PK, 2022. Global review of source apportionment of volatile organic compounds based on highly time-resolved data from 2015 to 2021. *Environ. Int.* 165, 107330. [PubMed: 35671590]
- Yuan B, Shao M, de Gouw J, Parrish DD, Lu S, Wang M, Zeng L, Zhang Q, Song Y, Zhang J, Hu M, 2012. Volatile organic compounds (VOCs) in urban air: how chemistry affects the interpretation of positive matrix factorization (PMF) analysis. *J. Geophys. Res. Atmos.* 117 (D24) eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2012JD018236>.



- Zheng H, Kong S, Xing X, Mao Y, Hu T, Ding Y, Li G, Liu D, Li S, Qi S, 2018. Monitoring of volatile organic compounds (VOCs) from an oil and gas station in northwest China for 1 year. *Atmos. Chem. Phys.* 18, 4567–4595.
- Colorado enviroscreen tool. [https://teeo-cdphe.shinyapps.io/COEnviroScreen\\_English](https://teeo-cdphe.shinyapps.io/COEnviroScreen_English). (Accessed 6 February 2023).
- Commerce city current air conditions site info and monitoring methods. [https://www.bouldair.com/commerce\\_city.htm](https://www.bouldair.com/commerce_city.htm), 2022-. (Accessed 31 May 2023).
- Oil and gas activity monitoring data and reports. [https://www.colorado.gov/airquality/air\\_toxics\\_repo.aspx#og](https://www.colorado.gov/airquality/air_toxics_repo.aspx#og). (Accessed 27 October 2023).
- “Quality Assurance Project Plan Suncor Energy (USA) - Commerce City Refinery Ambient Air Monitoring Program,” Tech. Rep., 2021. Montrose Air Quality Services, LLC, Commerce City, CO.
- Suncor and third-party nonprofit both monitoring air quality in Commerce City neighborhood. <https://www.9news.com/article/news/local/next/suncor-air-quality-monitoring-system/73-74c96ba0-4f09-413e-905d-3507ba293ce7>, 2022-. (Accessed 7 February 2023).
- Swansea ambient air monitoring station (SAMS) phase II air quality data report. [https://www.denvergov.org/content/dam/denvergov/Portals/771/documents/EQ/Air%20Quality/swansea-ambient-air-monitoring-site-reports/Phase\\_II\\_AirQualityDataReport.pdf](https://www.denvergov.org/content/dam/denvergov/Portals/771/documents/EQ/Air%20Quality/swansea-ambient-air-monitoring-site-reports/Phase_II_AirQualityDataReport.pdf). (Accessed 19 October 2023).
- Tables of Maximum Incremental Reactivity (MIR) Values.” EPA Title 17, Division 3, (Chapter 1), Sub(chapter 8).6, Article 1, Sections 94700–94701.



**Fig. 1.**  
Locations of all monitoring sites included in this study.



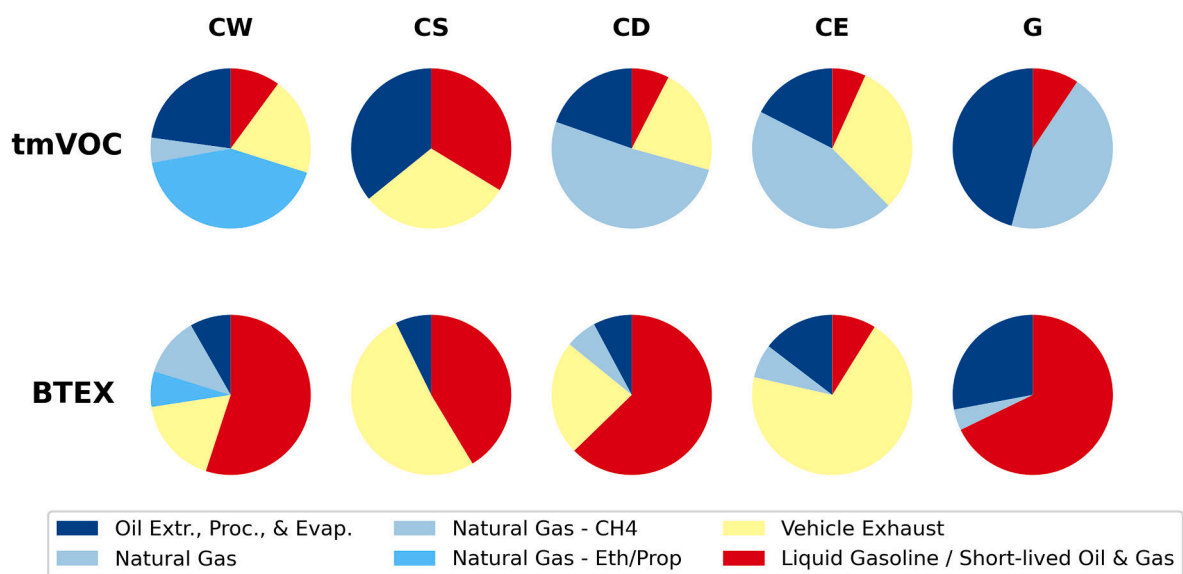
**Fig. 2.**  
Timeline of monitoring data included in this study.

Author Manuscript

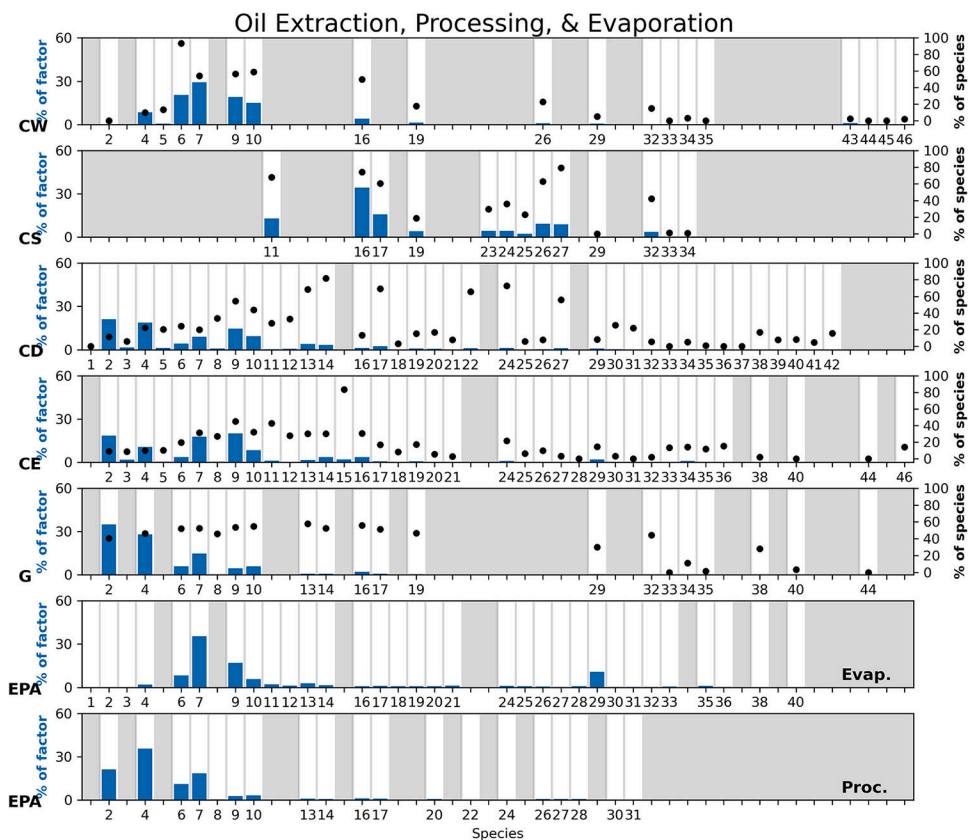
Author Manuscript

Author Manuscript

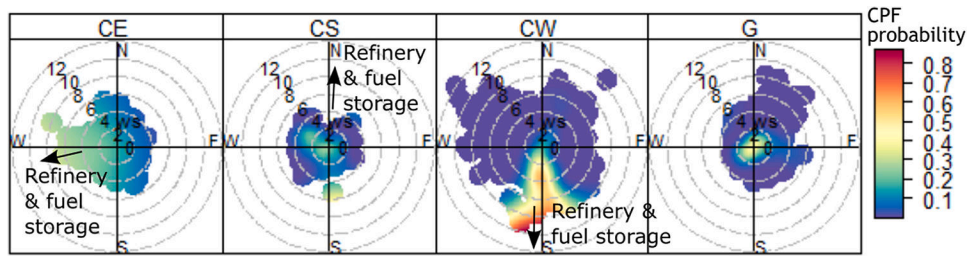
Author Manuscript



**Fig. 3.** Contributions of PMF factors in each solution to total measured volatile organic compounds (tmVOC) and benzene, toluene, ethylbenzene, and xylenes (BTEX) levels.



**Fig. 4.** Factor profiles, based on factor percent (tmVOC only) and species percent, identified as oil extraction, processing, and evaporation from CW, CS, CD, CE, and G. EPA shows an oil processing profile and a fuel evaporation profile from the EPA’s SPECIATE database (Gentner et al., 2014; Black and High, 1980). 1. acetylene, 2. ethane, 3. ethylene, 4. propane, 5. propylene, 6. isobutane, 7. n-butane, 8. cyclopentane, 9. isopentane, 10. n-pentane, 11. 2,2-dimethylbutane, 12. 2,3-dimethylbutane, 13. 2-methylpentane, 14. 3-methylpentane, 15. isoprene, 16. n-hexane, 17. methylcyclopentane, 18. 2,4-dimethylpentane, 19. benzene, 20. cyclohexane, 21. 2, 3-dimethylpentane, 22. 2-methylhexane, 23. 2,3-dimeC5 and 2-meC6, 24. 3-methylhexane, 25. 2,2,4-trimethylpentane, 26. n-heptane, 27. methylcyclohexane, 28. 2, 3,4-trimethylpentane, 29. toluene, 30. 2-methylheptane, 31. 3-methylheptane, 32. n-octane, 33. ethylbenzene, 34. m,p-xylene, 35. o-xylene, 36. n-propylbenzene, 37. isopropylbenzene, 38. n-nonane, 39. naphthalene, 40. n-decane, 41. carbon disulfide, 42. tetrachloroethene, 43. CO<sub>2</sub>\*, 44. CH<sub>4</sub>\*, 45. CO\*, 46. NO<sub>x</sub>. (\*Background subtracted).



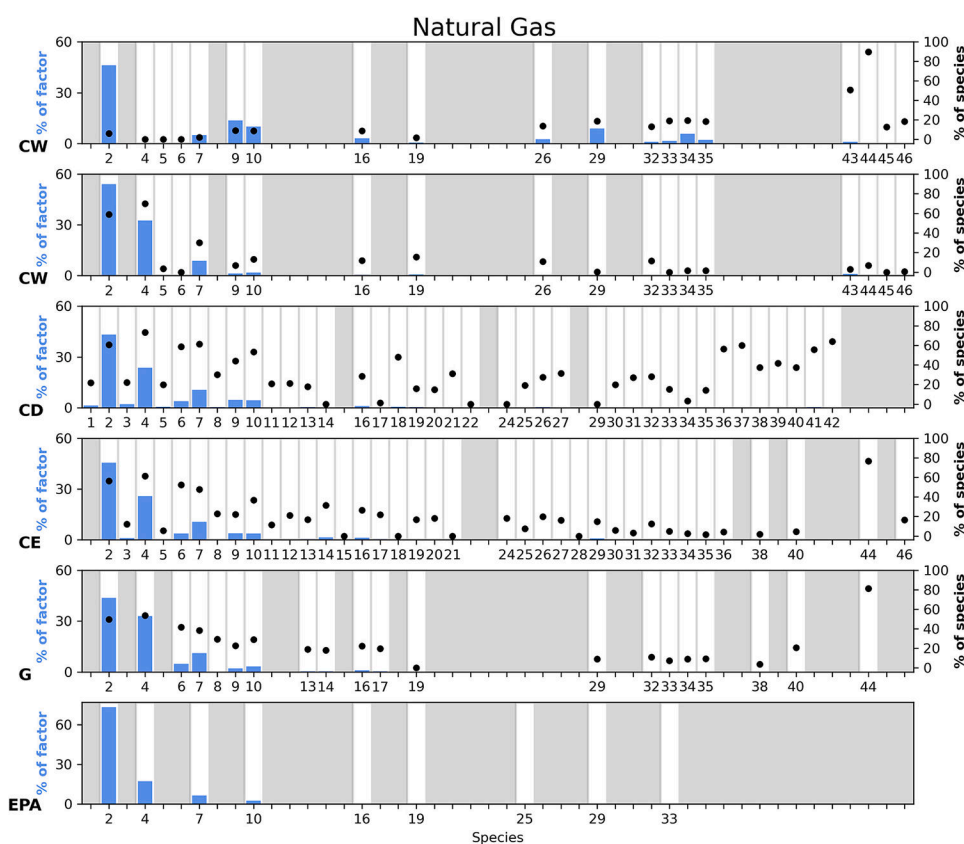
**Fig. 5.** Conditional bivariate probability function at the 75th percentile for the CW, CS, CE, and G oil extraction, processing, and evaporation factors. Arrows point towards a refinery and fuel storage tanks.

Author Manuscript

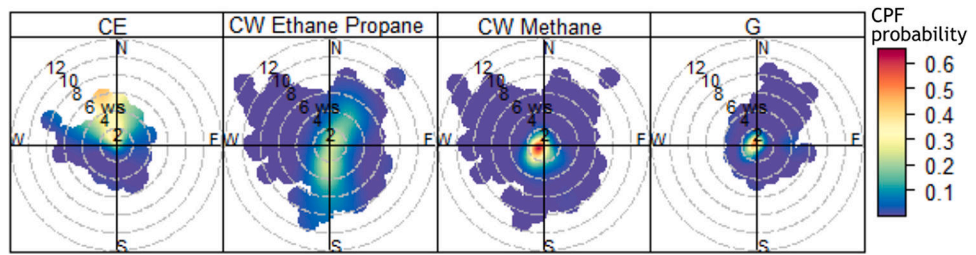
Author Manuscript

Author Manuscript

Author Manuscript

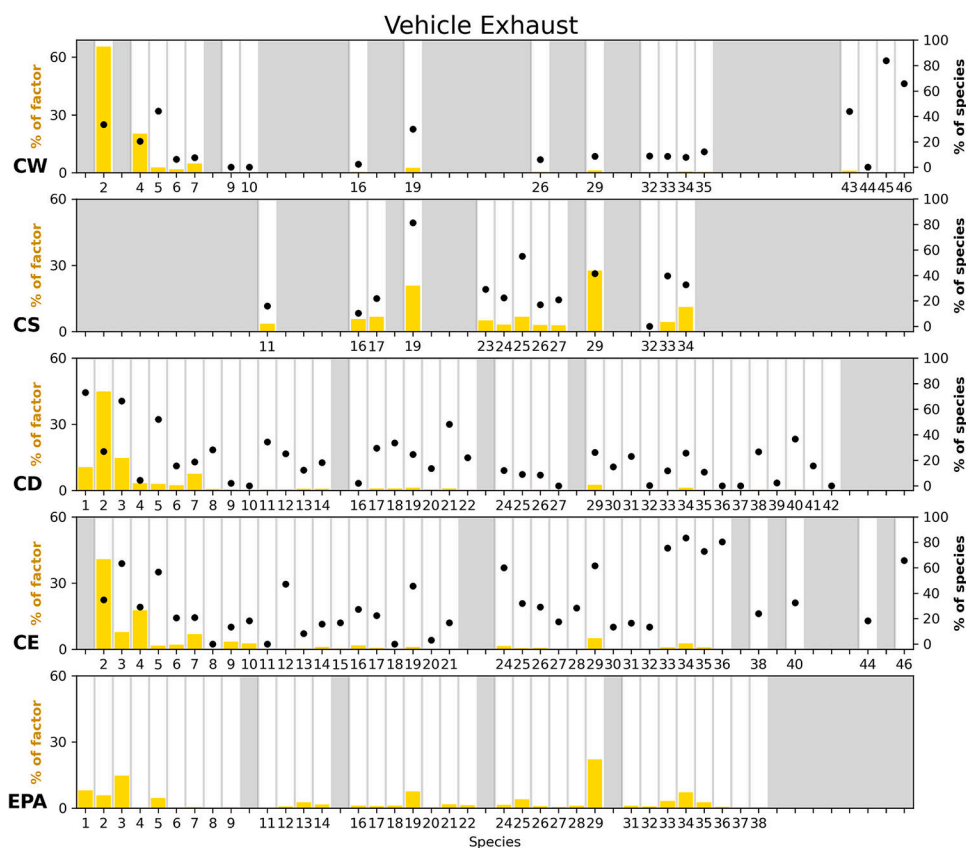


**Fig. 6.** Factor profiles, based on factor percent (tmVOC only) and species percent, identified as natural gas from CW, CD, CE, and G. CW resolved two factors related to natural gas. EPA shows a natural gas profile from the EPA's SPECIATE database (Brown, 2011). 1. acetylene, 2. ethane, 3. ethylene, 4. propane, 5. propylene, 6. isobutane, 7. n-butane, 8. cyclopentane, 9. isopentane, 10. n-pentane, 11. 2,2-dimethylbutane, 12. 2,3-dimethylbutane, 13. 2-methylpentane, 14. 3-methylpentane, 15. isoprene, 16. n-hexane, 17. methylcyclopentane, 18. 2,4-dimethylpentane, 19. benzene, 20. cyclohexane, 21. 2,3-dimethylpentane, 22. 2-methylhexane, 23. 2, 3-dimeC5 and 2-mcC6, 24. 3-methylhexane, 25. 2,2,4-trimethylpentane, 26. n-heptane, 27. methylcyclohexane, 28. 2,3,4-trimethylpentane, 29. toluene, 30. 2-methylheptane, 31. 3-methylheptane, 32. n-octane, 33. ethylbenzene, 34. m,p-xylene, 35. o-xylene, 36. n-propylbenzene, 37. isopropylbenzene, 38. n-nonane, 39. naphthalene, 40. n-decane, 41. carbon disulfide, 42. tetrachloroethene, 43. CO<sub>2</sub>\*, 44. CH<sub>4</sub>\*, 45. CO\*, 46. NO<sub>x</sub>. (\*Background subtracted). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

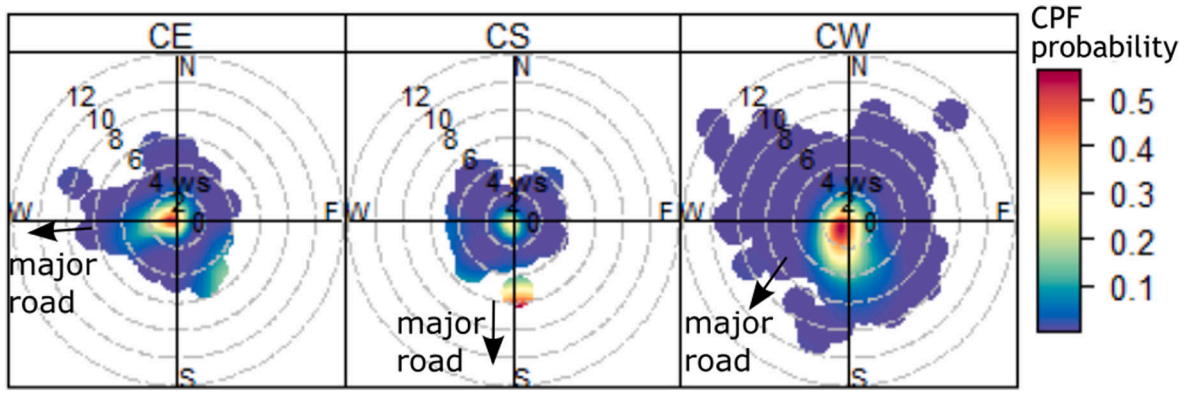


**Fig. 7.** Conditional bivariate probability function at the 75th percentile for the CW, CE, and G natural gas factors.





**Fig. 8.** Factor profiles, based on factor percent (tmVOC only) and species percent, identified as vehicle exhaust from CW, CS, CD, and CE. EPA shows a vehicle exhaust profile from the EPA's SPECIATE database (Kishan et al., 2008). 1. acetylene, 2. ethane, 3. ethylene, 4. propane, 5. propylene, 6. isobutane, 7. n-butane, 8. cyclopentane, 9. isopentane, 10. n-pentane, 11. 2,2-dimethylbutane, 12. 2,3-dimethylbutane, 13. 2-methylpentane, 14. 3-methylpentane, 15. isoprene, 16. n-hexane, 17. methylcyclopentane, 18. 2,4-dimethylpentane, 19. benzene, 20. cyclohexane, 21. 2,3-dimethylpentane, 22. 2-methylhexane, 23. 2,3-dimeC5 and 2-meC6, 24. 3-methylhexane, 25. 2,2,4-trimethylpentane, 26. n-heptane, 27. methylcyclohexane, 28. 2,3,4-trimethylpentane, 29. toluene, 30. 2-methylheptane, 31. 3-methylheptane, 32. n-octane, 33. ethylbenzene, 34. m,p-xylene, 35. o-xylene, 36. n-propylbenzene, 37. isopropylbenzene, 38. n-nonane, 39. naphthalene, 40. n-decane, 41. carbon disulfide, 42. tetrachloroethene, 43. CO<sub>2</sub>\*, 44. CH<sub>4</sub>\*, 45. CO\*, 46. NO<sub>x</sub>. (\*Background subtracted).



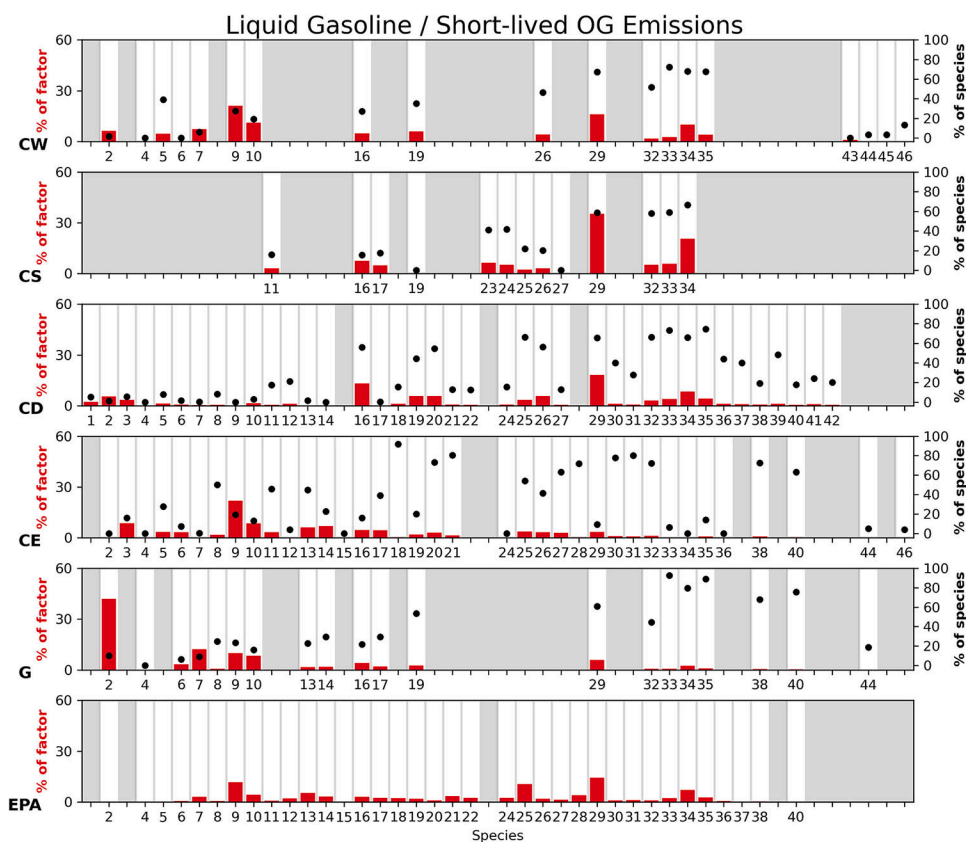
**Fig. 9.** Conditional bivariate probability function at the 75th percentile for the CW, CS, and CE vehicle exhaust factors. Arrows point towards major roads.

Author Manuscript

Author Manuscript

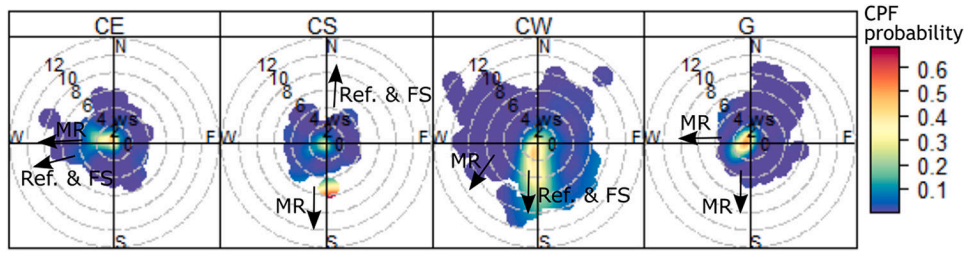
Author Manuscript

Author Manuscript



**Fig. 10.**

Factor profiles, based on factor percent (tmVOC only) and species percent, identified as liquid gasoline/short-lived OG from CW, CS, CD, CE, and G. EPA shows a liquid gasoline profile from the EPA's SPECIATE database (Lewis, 2004). 1. acetylene, 2. ethane, 3. ethylene, 4. propane, 5. propylene, 6. isobutane, 7. n-butane, 8. cyclopentane, 9. isopentane, 10. n-pentane, 11. 2,2-dimethylbutane, 12. 2,3-dimethylbutane, 13. 2-methylpentane, 14. 3-methylpentane, 15. isoprene, 16. n-hexane, 17. methylcyclopentane, 18. 2,4-dimethylpentane, 19. benzene, 20. cyclohexane, 21. 2,3-dimethylpentane, 22. 2-methylhexane, 23. 2,3-dimeC5 and 2-meC6, 24. 3-methylhexane, 25. 2,2,4-trimethylpentane, 26. n-heptane, 27. methylcyclohexane, 28. 2,3,4-trimethylpentane, 29. toluene, 30. 2-methylheptane, 31. 3-methylheptane, 32. n-octane, 33. ethylbenzene, 34. m,p-xylene, 35. o-xylene, 36. n-propylbenzene, 37. isopropylbenzene, 38. n-nonane, 39. naphthalene, 40. n-decane, 41. carbon disulfide, 42. tetrachloroethene, 43. CO<sub>2</sub>\*, 44. CH<sub>4</sub>\*, 45. CO\*, 46. NO<sub>x</sub>. (\*Background subtracted).



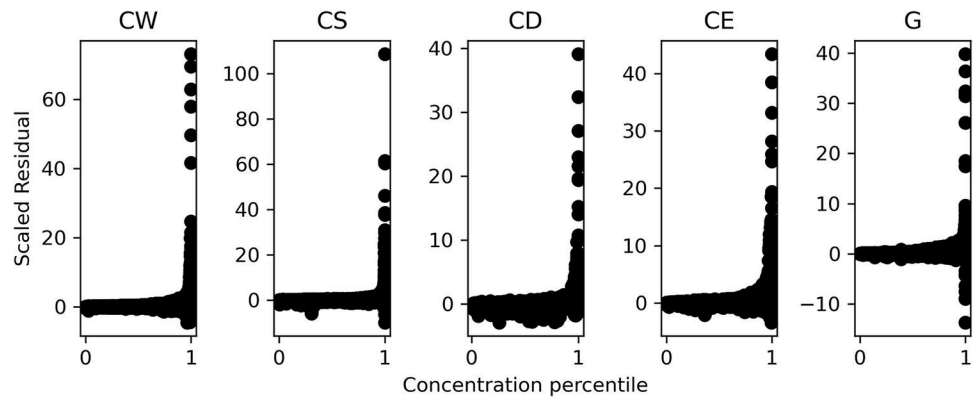
**Fig. 11.** Conditional bivariate probability function at the 75th percentile for the CW, CS, CE, and G liquid gasoline/short-lived OG factors. Arrows point towards major roads (MR) and a refinery and fuel storage (Ref. & FS).

Author Manuscript

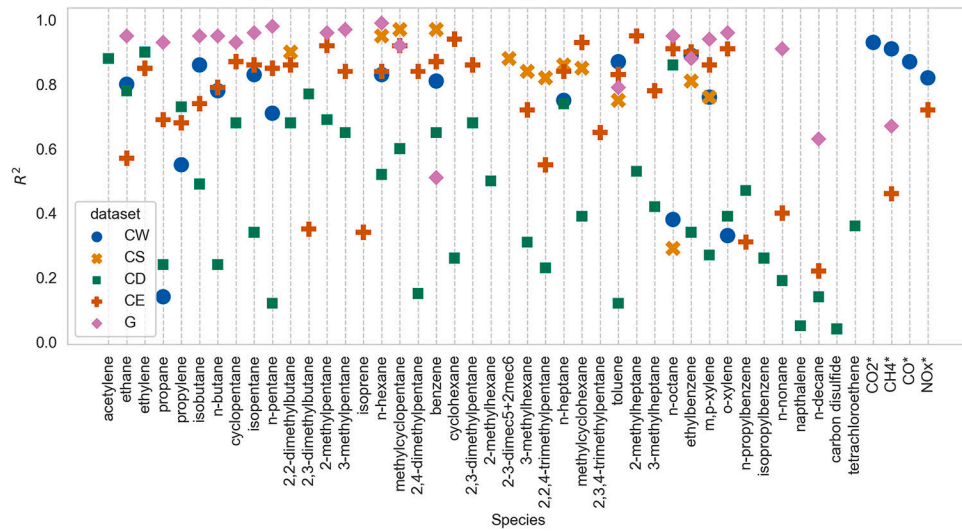
Author Manuscript

Author Manuscript

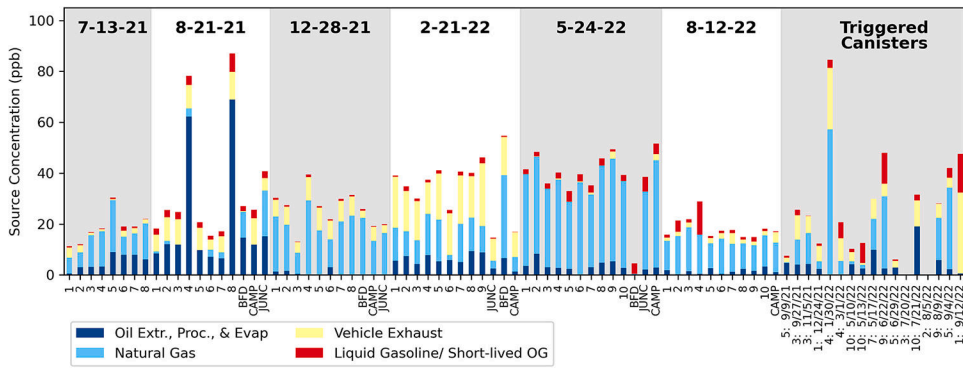
Author Manuscript



**Fig. 12.** Sample residuals based on concentration percentile for each species, scaled by respective species mean.



**Fig. 13.**  $R^2$  of PMF-modeled species concentrations versus measured species concentrations for each species across PMF datasets.



**Fig. 14.** Contributions of each CD factor sampled in each canister over the study period. Canisters are separated by scheduled days and triggered canisters.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript