Principal component analysis of summertime ground site measurements in the Athabasca oil sands with a focus on analytically unresolved intermediate volatility organic compounds

Travis W. Tokarek[1], Charles A. Odame-Ankrah[1], Jennifer A. Huo[1], Robert McLaren[2], Alex K. Y. Lee[3, 4], Max G. Adam[4], Megan D. Willis[5], Jonathan P. D. Abbatt[5], Cristian Mihele[6], Andrea Darlington[6], Richard L. Mittermeier[6], Kevin Strawbridge[6], Katherine L. Hayden[6], Jason S. Olfert[7], Elijah. G. Schnitzler[8], Duncan K. Brownsey[1], Faisal V. Assad[1], Gregory R. Wentworth[5, a], Alex G. Tevlin[5], Douglas E. J. Worthy[6], Shao-Meng Li[6], John Liggio[6], Jeffrey R. Brook[6], and Hans D. Osthoff[1*]

[1] Department of Chemistry, University of Calgary, Calgary, Alberta, T2N 1N4, Canada
[2] Centre for Atmospheric Chemistry, York University, Toronto, Ontario, M3J 1P3, Canada
[3] Department of Civil and Environmental Engineering, National University of Singapore, Singapore 117576, Singapore
[4] NUS Environmental Research Institute, National University of Singapore, Singapore
[5] Department of Chemistry, University of Toronto, Toronto, Ontario, M5S 3H6, Canada
[6] Air Quality Research Division, Environment and Climate Change Canada, Toronto, Ontario, M3H 5T4, Canada
[7] Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, T6G 1H9, Canada
[8] Department of Chemistry, University of Alberta, Edmonton, Alberta, T6G 2G2, Canada
[a] Now at: Environmental Monitoring and Science Division, Alberta Environment and Parks, Edmonton, Alberta, T5J 5C6, Canada
* Corresponding author

for Atmos. Chem. Phys.

Abstract

In this paper, measurements of air pollutants made at a ground site near Fort McKay in the Athabasca oil sands region as part of a multi-platform campaign in the summer of 2013 are presented. The observations included measurements of selected volatile organic compounds (VOCs) by a gas chromatograph – ion trap mass spectrometer (GC-ITMS). This instrument observed a large, analytically unresolved hydrocarbon peak (with retention index between 1100 and 1700) associated with intermediate volatility organic compounds (IVOCs). However, the activities or processes that contribute to the release of these IVOCs in the oil sands region remain unclear.

Principal component analysis (PCA) with Varimax rotation was applied to elucidate major source types impacting the sampling site in the summer of 2013. The analysis included 28 variables, including concentrations of total odd nitrogen (NO_y), carbon dioxide (CO_2), methane (CH_4), ammonia (NH_3), carbon monoxide (CO), sulfur dioxide (SO_2), total reduced sulfur compounds (TRS), speciated monoterpenes (including α- and β-pinene and limonene), particle volume calculated from measured size distributions of particles less than 10 μm and 1 μm in diameter (PM_{10-1} and PM_1), particle-surface bound polycyclic aromatic hydrocarbons (pPAH), and aerosol mass spectrometer composition measurements, including refractory black carbon (rBC) and organic aerosol components. The PCA was complemented by bivariate polar plots showing the joint wind speed and direction dependence of air pollutant concentrations to illustrate the spatial distribution of sources in the area. Using the 95% cumulative percentage of variance criterion, ten components were identified and categorized by source type. These included emissions by wet tailings ponds, vegetation, open pit mining operations, upgrader facilities, and surface dust. Three components correlated with IVOCs, with the largest associated with surface mining and is likely caused by the unearthing and processing of raw bitumen.

1. Introduction

The Athabasca oil sands region of Northern Alberta, Canada, has seen extraordinary expansion of its oil sands production and processing facilities (CAPP, 2016) and associated emissions of air pollutants over the last several decades (Englander et al., 2013; Bari and Kindzierski, 2015). Air emissions from these facilities have been impacting surrounding communities, including the city of Fort McMurray and the community of Fort McKay (WBEA, 2013). To assess the impact of these emissions on human health, visibility, climate, and the ecosystems downwind, it is critical to obtain an understanding of the source types from all activities associated with oil sands operations (ECCC, 2016).

Prior to 2013, there had been only a single industry-independent study of trace gas emissions from the Athabasca oil sands mining operations (Simpson et al., 2010; Howell et al., 2014). The data showed elevated concentrations in n-alkanes (30% of the total quantified hydrocarbon emissions), cycloalkanes (49%), and aromatics (15%) in plumes from an oil sands surface mining facility intercepted from a single aircraft flight. These compounds are associated with oil and gas developments including mining, upgrading, and transportation of bitumen (Siddique et al., 2006). Specifically, these activities involve the use of naphtha, a complex mixture of aliphatic and aromatic hydrocarbons in the range of C_3 to C_{14} containing n-alkanes (e.g., n-heptane, n-octane, and n-nonane) and benzene, toluene, ethylbenzene, and xylenes (BTEX).

In August 2013, a comprehensive air quality study as a part of the Joint Oil Sands Monitoring (JOSM) plan (JOSM, 2012), referred to here as the 2013 JOSM intensive study was conducted. This study was performed in northern Alberta at two ground sites in and near Fort McKay in close proximity (as close as 3.5 km) to oil sands mining operations and from a National Research Council of Canada (NRC) Convair 580 research aircraft to characterize oil sands emissions and their downwind physical and chemical transformations (Gordon et al., 2015; Liggio et al., 2016; Li et al., 2017).

One ground site, located at the Wood Buffalo Environmental Association (WBEA) air monitoring station (AMS) 13 (Fig. 1), was equipped with a comprehensive set of instrumentation to measure concentrations of a wide range of trace gases and aerosols (Table 1), yielding a unique and new data set, parts of which are presented in this paper for the first time. As part of this effort, a gas chromatograph equipped with an ion trap mass spectrometer (GC-ITMS) was deployed at AMS 13. When air masses passing over regions with industrial activities were observed (as judged from a combination of local wind direction and tracer measurements), the total ion chromatogram showed an analytically unresolved hydrocarbon signal associated with intermediate volatile organic compounds (IVOCs) with saturation concentration (C^*) in the range 10^5 µg m^{-3}< C^* < 10^7 µg m^{-3} (Liggio et al., 2016).

Emission estimates for analytically unresolved hydrocarbons range from 5×10^6 kg $year^{-1}$ to 14×10^6 kg $year^{-1}$ for the two facilities that reported such emissions (Li et al., 2017). Using aircraft measurements during the 2013 study, Liggio et al. (2016) showed that IVOCs contributed to the majority of the observed secondary organic aerosol (SOA) mass production in a similar fashion as anthropogenic VOCs contributed to SOA production during the Deepwater Horizon oil spill (de Gouw et al., 2011) and rivaling the magnitude of SOA formation observed downwind of megacities (Liggio et al., 2016), though ultimately it has remained unclear which activities are associated with IVOC emissions.

In this paper, concurrent measurements of air pollutants at the AMS 13 ground site during the 2013 JOSM intensive study are presented. The analytically unresolved hydrocarbon signal was integrated and is presented as a time series and used as an input variable in a principal component analysis (PCA) to elucidate the origin of IVOCs in the Athabasca oil sands by association. The analysis presented here is a receptor analysis focusing on the normalized variability of pollutants impacting the AMS 13 ground site and hence does not constitute a comprehensive emission profile analysis of the oil sands facilities as a whole, for which aircraft-based measurements and/or direct plume or stack measurements are more suitable. PCA was chosen over the more popular positive matrix factorization (PMF) method (Paatero

and Tapper, 1994) because it yields a unique solution and is particularly suited as an exploratory tool for identification of components without *a priori* constraints (Jolliffe and Cadima, 2016). The PCA was complemented by bivariate polar plots (Carslaw and Ropkins, 2012; Carslaw and Beevers, 2013) to show the spatial distribution of sources in the region as a function of locally measured wind direction and speed. A second PCA was performed to investigate which components correlate with (and generate) secondary pollutants, i.e., pollutants that are formed by atmospheric processes. Potential sources and processes contributing to each of the components identified by PCA are discussed.

2. Experimental

2.1 Measurement location

Measurements of air pollutants were made at AMS 13 routine air monitoring station (Fig. 1), which is operated by WBEA. The site is located at 111.6423° W longitude and 57.1492° N latitude about 3 km from the southern edge of the community of Fort McKay, 300 m west from a public road, and 1 km west of the Athabasca river. The immediate vicinity of the site consisted of mixed-leaf boreal forest with a variety of tree species, including poplar, aspen, pine and spruce trees (Smreciu et al., 2013). The site was accessible via a gravel road; traffic on this road was restricted during the study period (August - September, 2013).

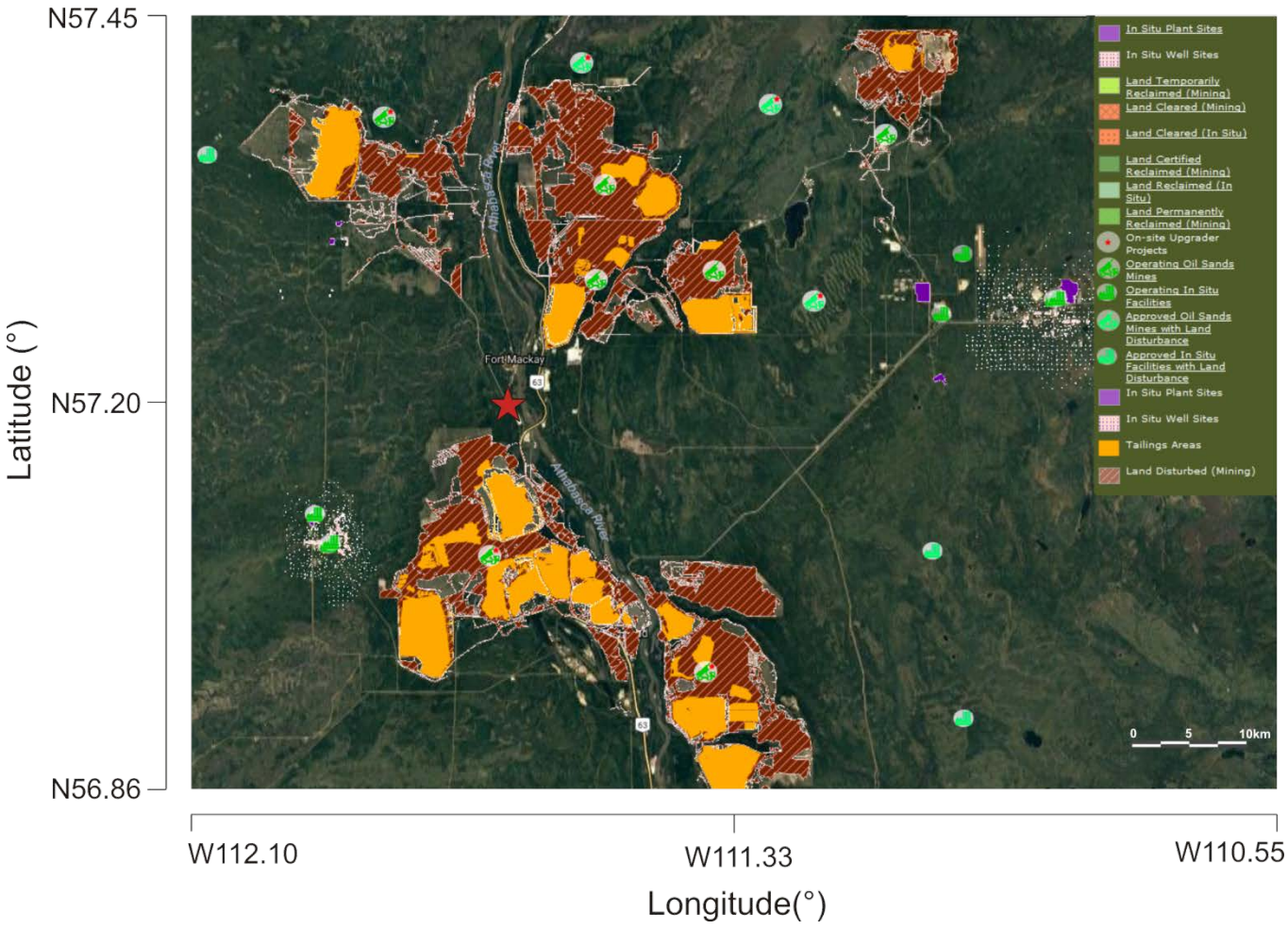
The site is impacted by emissions from nearby oil sands facilities (Table 1 and Fig. 1), including a large surface mining site operated by Syncrude Canada whose northeastern corner is located 3.5 km to the south of AMS 13 (and which is adjacent to the 5 km long Syncrude – Mildred Lake (SML) tailings pond) and from a large upgrader stack facility operated by Suncor Energy Inc. located to the Southeast. There are additional oil sands facilities operated (during the study period) by Canadian Natural Resources Limited, Imperial Oil, and Shell Canada to the North and Northeast.

Table 1. Oil sands facilities located within 30 km of AMS 13. Distances were estimated using coordinates provided in the National Pollutant Release Inventory (NPRI, 2013) and do not account for the size of each facility whose boundaries may be considerably closer to (or further away from) AMS 13. PACPRM = Petroleum and coal products refining and manufacturing; OGPS = Oil and gas pipelines and storage.

Company	Name	Type	Direction	Distance (km)
Syncrude Canada Ltd.	Mildred Lake Plant Site	PACPRM	S	12.2
Athabasca Minerals Inc.	Susan Lake Gravel Pit	Mining and Quarrying	N	15.5
Syncrude Canada Ltd.	Aurora North Mine Site	PACPRM	NE	18.7
Suncor Energy	Suncor Energy Inc. Oil Sands	PACPRM	SE	19.4
Enbridge Pipelines Inc.	MacKay River Terminal	OGPS	WSW	19.7
Suncor Energy	MacKay River, In-Situ, Oil Sands Plant	PACPRM	WSW	19.9
Enbridge Pipelines Inc.	Athabasca Terminal	OGPS	SE	21.2
Williams Energy	Fort McMurray Hydrocarbon Liquids Extraction Facility	Conventional oil and gas extraction	SE	21.6
Canadian Natural Resources Limited	Horizon Oil Sands Processing Plant and Mine	PACPRM	NNW	21.8
Shell Canada Energy	Muskeg River Mine and Jackpine Mine	PACPRM	NNE	23.7

123 **Figure 1**. Map of oil sands facilities showing locations of surface mines and tailings ponds, downloaded

124 from the Oil Sands Information Portal (Alberta, 2017). The red star indicates the location of AMS 13.



125

126 **2.2 Instrumentation**

127 A large number of instruments was deployed for this study; a partial list whose data were utilized in this

128 manuscript is given in Table 2. Detailed descriptions of these instruments and operational aspects such

129 as calibrations are given in the S.I. Sample observations of analytically unresolved hydrocarbons by GC-

130 ITMS and how these data were used in the analysis are described in section 2.2.1 below.

131 **Table 2**. Instruments used to measure ambient gas-phase and aerosol species during the 2013 JOSM

132 intensive study at AMS 13.

Instrument and Model	Species measured	Time resolution	Reference
Picarro CRDS G2401	CO, CO_2, CH_4	1 min	(Chen et al., 2013; Nara et al., 2012)
Thermo Scientific, Model 42i	NO_y	10 s	(Tokarek et al., 2014; Odame-Ankrah, 2015)
Blue diode cavity ring-down spectroscopy	NO_2	1 s	(Paul and Osthoff, 2010; Odame-Ankrah, 2015)
Thermo Scientific Model 49i	O_3	10 s	(Tokarek et al., 2014; Odame-Ankrah, 2015)
Griffin/FLIR, model 450 GC-ITMS	VOCs	1 hr	(Tokarek et al., 2017b; Liggio et al., 2016)
Thermo Scientific CON101	TS	1 min	n/a
Thermo Scientific 43iTLE	SO_2	1 min	n/a
AIM-IC	$NH_{3(g)}$, $NH_4^+{}_{(p)}$	1 hr	(Markovic et al., 2012)
Aerodyne SP-AMS	rBC, $NH_4^+{}_{(p)}$, $SO_4^{2-}{}_{(p)}$, $NO_3^-{}_{(p)}$, $Cl^-{}_{(p)}$, organics	1-5 min (variable)	(Onasch et al., 2012)
TSI APS 3321	PM_{10-1} size distribution	5-6 min (variable)	(Peters and Leith, 2003)
TSI SMPS (3081 DMA, 3776 CPC)	PM_1 size distribution	6 min	(Wang and Flagan, 1990)
EcoChem Analytics PAS 2000CE	pPAH	1 min	(Wilson et al., 1994; Burtscher et al., 1982)

133

2.2.1 Analytically unresolved hydrocarbon signature

As previously reported (Liggio et al., 2016), the total ion chromatogram of the GC-ITMS occasionally showed elevated and analytically unresolved hydrocarbons in the volatility range of C_{11} – C_{17} with saturation vapor concentration (C^*) from 10^5 µg m^{-3}< C^* < 10^7 µg m^{-3}. An example is shown in Fig. 2.

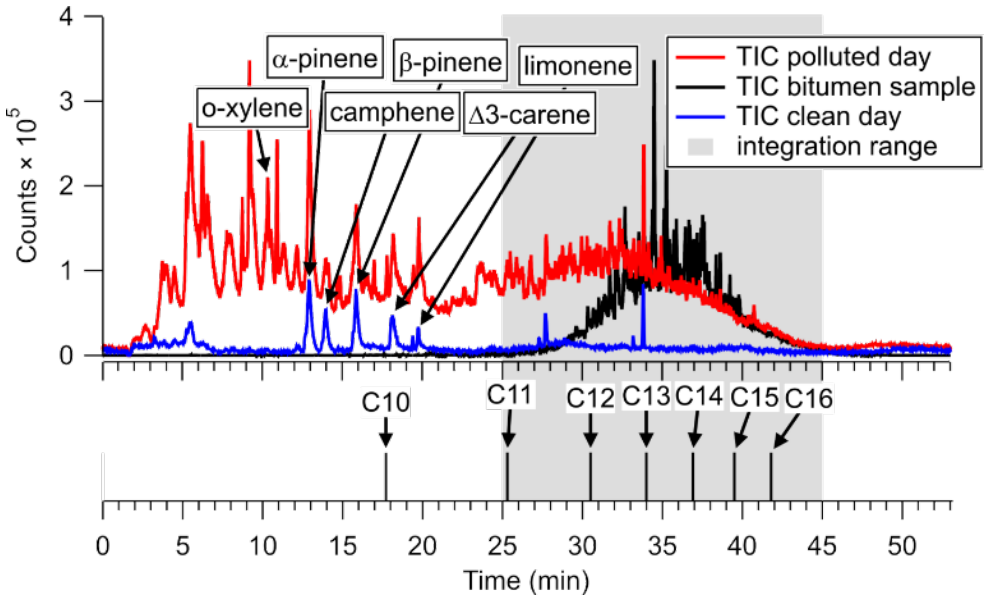


Figure 2. (**Top**) Total ion chromatograms of air samples collected on August 27, 2013 from 18:04 to 18:14 UTC (red) and on August 28, 2013 from 13:43 to 13:53 UTC (blue). The TIC of a head space sample of ground-up bitumen collected post-campaign is superimposed (black). The gray area indicates the range over which IVOC signal was integrated. (**Bottom**) Retention times of n-alkanes, determined after the measurement intensive by sampling a VOC mixture containing a C_{10} – C_{16} n-alkane ladder.

An offline analysis of the headspace above ground-up bitumen gave a similarly unresolved hydrocarbon signal (Fig. 2, black trace). In this particular case, the ambient air chromatogram also shows enhancements of lower molecular weight hydrocarbons (possibly from naphtha) that were not observed in the bitumen sample**.** The observed unresolved hydrocarbon feature is qualitatively similar to the "large chromatographic hump of unresolved complex mixtures" reported by Yang et al. (2011) during

149 their analysis of bitumen extracts.

150 The major ions contributing to the unresolved signals in Figure 2 are associated with alkanes (i.e., *m/z*
151 55, 57, 67, 69, etc. – see Fig. S-1). In contrast, counts at masses associated with aromatics (i.e., *m/z* 115,
152 $C_9H_7^+$, and *m/z* 91, $C_7H_7^+$) as reported by Cross et al. (2013) were negligible in both the bitumen head
153 space and polluted day samples. The strong resemblance of the unresolved hydrocarbon feature in
154 ambient air with the bitumen head space sample both in terms of volatility (i.e., elution time) and
155 electron impact mass fragmentation is consistent with bitumen as the source of IVOCs at this site.

156 In the interpretation of the integrated IVOC signal, it is assumed that it is of primary origin, i.e., emitted
157 directly from point sources in the vicinity of the measurement site. For the PCA, the unresolved signal
158 was integrated from a retention time of 25 min to 45 min (gray area in Fig. 2) in all ambient air
159 chromatograms.

160 The IVOCs observed in this work likely encompass a portion of the total that is emitted. For example,
161 IVOCs generated by combustion processes, such as aircraft engine exhaust, are comprised of alkanes,
162 aromatics and oxygenated compounds (Cross et al., 2013). The use of a chromatographic column in this
163 work biases the IVOC signal towards hydrocarbon-IVOCs, since oxygenated compounds (i.e., alcohols
164 and acids) will not elute from the analytical column. Furthermore, the recovery of VOCs from the pre-
165 concentration unit, while reproducible and likely complete for n-alkanes which bracket the bulk of IVOC
166 emitted and whose calibration curves were linear, is not known for late-eluting compounds, but is
167 assumed to be sufficiently reproducible to yield a semi-quantitative signal.

168

169 **2.3 Principal Component Analysis**

170 The PCA was carried out using the "Statistical Analysis System" (SAS™) Studio 3.4 software (SAS, 2015)

using a method similar to that described by Thurston et al. (2011; 1985). The source-related components and their associated profiles are derived from the correlation matrix of the input trace constituents. This approach assumes that the total concentration of each "observable" (i.e., input variable) is made up of the sum of contributions from each of a smaller number of pollution sources and that variables are conserved between the points of emission and observation.

2.3.1 Selection of variables

22 variables whose ambient concentrations are dominated by primary emissions or which are formed very shortly after emission (such as the less oxidized oxygenated organic aerosol (LO-OOA) factor observed by the SP-AMS, see below) were included in the PCA (Table 3). These variables included CO_2, CH_4, NO_y, CO, and SO_2, which are known to be emitted in the oil sands region from stacks, the mine fleet and faces, tailings ponds, and by fugitive emissions (Percy, 2013). The median NO_x (= NO + NO_2) to NO_y ratio was 0.85, consistent with the close proximity of the measurement site to emission sources and limited chemical processing. Because NO_x constituted a large fraction of NO_y, its temporal variation was captured by the latter, and it was not included as a separate variable in the PCA.

For this work, mixing ratios of all non-methane hydrocarbons (NMHCs) that were quantified (i.e., o-xylene, the n-alkanes decane and undecane, the aromatics 1, 2, 3- and 1, 2, 4-TMB, as well as limonene and α- and β-pinene) were included as variables. In addition, the aforementioned unresolved signal associated with IVOCs was included as a variable by integrating total GC-ITMS ion counts (*m/z* 50 –425) over a retention time range of 25-45 min (retention index range of 1100 to 1700).

Gas-phase ammonia was included as a variable because elevated reduced nitrogen concentrations have been observed in the region and were linked to the use of ammonia on an industrial scale, for example as a floating agent and for hydrotreating (Bytnerowicz et al., 2010). Total sulfur and total reduced sulfur

were added as tracers of upgrader stack SO_2 emissions and of "odours", believed to be emitted from oil sands tailings ponds which continue to be of concern in surrounding communities (Small et al., 2015; Percy, 2013; Holowenko et al., 2000).

Refractory black carbon was added as a variable since it is present in diesel truck exhaust and in biomass burning plumes and, hence, a combustion tracer (Wang et al., 2016; Briggs and Long). pPAHs were included because of their association with facility stack emissions and combustion particles in the area (Allen, 2008; Grimmer et al., 1987). Hydrocarbon-like organic aerosol (HOA) was included as a surrogate for fossil fuel combustion by vehicles (Jimenez et al., 2009). The LO-OOA factor was included as it is unique to the Alberta oil sands and appears to form rapidly after emission of precursors (Lee et al., 2018). Supermicron aerosol volume (PM_{10-1}, i.e., the volume of particles between PM_{10} and PM_1) was also included as a tracer of coarse particles from primary sources, which are expected to be dominated by dust emissions.

207 **Table 3.** Variables observed at the AMS 13 ground site during the 2013 JOSM campaign used for PCA.

Variable	Unit	Median[a]	Average[a,b]	Standard deviation[a,b]	LOD[e]	Min.[a]	Max.[a]	Fraction <LOD
Anthropogenic VOCs								
o-xylene	pptv[f]	5	30	69	1	< LOD	635	10%
1,2,3 - TMB	pptv	1.7	4.3	7.9	0.2	< LOD	67	27%
1,2,4 - TMB	pptv	2.1	7.7	14.7	0.2	< LOD	107	8%
decane	pptv	0.5	8.5	18.2	0.1	< LOD	125	44%
undecane	pptv	0.4	3.0	6.3	0.1	< LOD	37	39%
Biogenic VOCs								
α-pinene	pptv	477	542	401	1	19	1916	0%
ß-pinene	pptv	390	467	334	1	18	1594	0%
limonene	pptv	150	179	158	2	< LOD	711	1%
Combustion tracers								
NO_y	ppbv	1.79	4.00	5.44	0.01	0.13	41.6	0%
rBC	μg m^{-3}	0.13	0.20	0.10	0.02	< LOD	0.90	40%
CO	ppbv	117.6	120.0	18.2	5.7[h]	90.9	241.2	0%
CO_2	ppmv	420.2	433.2	39.5	0.4[h]	386.0	577.7	0%
Aerosol species								
pPAH	ng m^{-3}	1	2	2	1[c]	< LOD	14	39%
PM_{10-1}	μm^3 cm^{-3}	11.2	14.4	12.9	0.003	1.0	79.5	0%
HOA	μg m^{-3}	0.31	0.43	0.35	N/A[g]	0.04	2.32	N/A
LO-OOA	μg m^{-3}	1.19	2.00	2.26	N/A[g]	0.11	15.6	N/A
Sulfur species								
Total sulfur (TS)	ppbv	0.22	1.41	4.27	0.13	< LOD	33.3	35%
SO_2	ppbv	< LOD	1.0	4.0	0.2	< LOD	33.5	81%
Total reduced sulfur (TRS)	ppbv	0.26	0.38	1.05	0.2	< LOD	14.8	81%
Other								
IVOCs	Counts × min	1.8×10^7	3.4×10^7	4.2×10^7	N/A[g]	1.4×10^6	2.5×10^8	N/A
CH_4	ppbv	1999.2	2065.5	169.6	1.8[h]	1880	2959	0%
NH_3	μg m^{-3}	0.79	1.10	1.03	0.05	0.06	5.75	39%

[a] Values were determined only from data points included in the PCA, not from entire campaign.

[b] Average and standard deviation were calculated before zeros were replaced with 0.5×LOD.

[c] Estimated.

[e] LOD = limit of detection.

[f] parts-per-trillion by volume (10^{-12})

[g] N/A = data not available

[h] calculated using 3 × standard deviation at ambient background levels

208

To assess which components have the greatest impact on secondary product formation, a second PCA was performed which included variables mainly formed through atmospheric chemical processes and whose concentrations more strongly depend on air mass chemical age than those variables selected initially. In this PCA, odd oxygen ($O_x = O_3 + NO_2$), submicron aerosol $SO_4^{2-}{}_{(p)}$, $NO_3^-{}_{(p)}$, $NH_4^+{}_{(p)}$, a second, more-oxidized OOA factor (MO-OOA), and PM_1 volume were included, increasing the total number of variables to 28 (Table 4).

Table 4. Variables added in the second PCA. Particle-phase concentrations, i.e., $SO_4^{2-}{}_{(p)}$, $NO_3^-{}_{(p)}$, $NH_4^+{}_{(p)}$ and MO-OOA were made by aerosol mass spectrometry and account for PM_1 only.

Variable	Unit	Median	Average	Standard deviation	LOD	Min.	Max.
O_x	ppbv	7.35	11.1	10.6	1	<LOD	41.1
$SO_4^{2-}{}_{(p)}$	μg m^{-3}	0.3	0.8	1.1	0.1	<LOD	6.6
$NO_3^-{}_{(p)}$	μg m^{-3}	0.08	0.13	0.13	0.01	0.01	0.72
$NH_4^+{}_{(p)}$	μg m^{-3}	0.13	0.28	0.37	0.05	<LOD	2.21
MO-OOA	μg m^{-3}	1.65	1.83	0.960	N/A	1.41×10^{-6}	4.65
PM_1 volume	μm^3 cm^{-3}	2.48	3.77	3.72	N/A	0.35	20.9

2.3.2 Treatment of input data

Data used in the PCA were averaged to match the time resolution of the GC-ITMS VOC and IVOC measurements, i.e. over 10 minute long periods (spaced ~ 1 hr apart) set by the start and stop times of the GC-ITMS pre-concentration period. When concentrations were below their respective limit of detection (LOD; values are given in Table 3), half the reported LOD was used to minimize bias (Harrison et al., 1996; Buhamra et al., 1998). Prior to PCA, input variables were standardized to eliminate unit differences by subtracting the mean concentration $\overline{C_i}$ of pollutant *i* from the concentration of sample *k* ($C_{i,k}$) and dividing by the standard deviation (s_i) of all samples included in the PCA.

$$Z_{i,k} = \frac{C_{i,k} - \overline{C_i}}{s_i} \quad (1)$$

Here, $Z_{i,k}$ is the standardized pollutant concentration. In total, 218 data points from all identified species over the period of the campaign were used for the main PCA.

2.3.3 PCA solutions

In this work, the Varimax method (Kaiser, 1958) was used to rotate the loading matrix. This method is an orthogonal rotation (i.e., components are not expected to correlate) which minimizes the impact of high loadings, making the results easier to interpret (Kaiser, 1958). Several criteria (Table S-10) were considered for component selection: the latent root criterion, i.e., on the basis that rotated eigenvalues must be greater than unity, the (cumulative) percentage of variance criterion, where the extracted components accounts for >95% of the variance, and the Scree test (Fig. S-2) (Thurston and Spengler, 1985; Guo et al., 2004; Hair et al., 1998; Cattell, 1966). For the optimal solution presented in the main manuscript, the 95% variance criterion was chosen, providing a 10-component solution for the PCA with

only primary variables and an 11-component solution for the PCA with both primary and secondary variables. Components 1 through 4 were consistent regardless of the number of components retained. Solutions with fewer and more components are presented in the supplemental material section.

Time series of each of the components were calculated by multiplying the original standardized matrix by the rotated loading matrix and were used to generate bivariate polar plots (section 2.4).

2.4 Bivariate polar plots

The PCA was complemented by bivariate polar plots showing the wind speed and direction dependence of air pollutant concentrations. The use of these representations implies a linear relationship between local wind conditions and air mass origin, which may not be always the case (for example, during or after stagnation periods). In addition, local topography, such as the Athabasca river valley, complicates regional air flow patterns and limit the interpretability of polar plots in general and in particular to the E of AMS 13, where the river valley is located. The plots were generated with the Openair software package (Carslaw and Ropkins, 2012; Carslaw and Beevers, 2013) using the R programming language and the open-source software "RStudio: Integrated development environment for R" (RStudio Boston, 2017). The default setting (100) was used as the smoothing function.

256 **3. Results**

257 **3.1. Overview of the data set**

258 Time series of the 22 pollution tracers chosen for PCA are presented in Fig. 3, grouped approximately by
259 source type. Statistics of the data (i.e., median, average, maxima, minima, etc.) are summarized in Table
260 3.

261

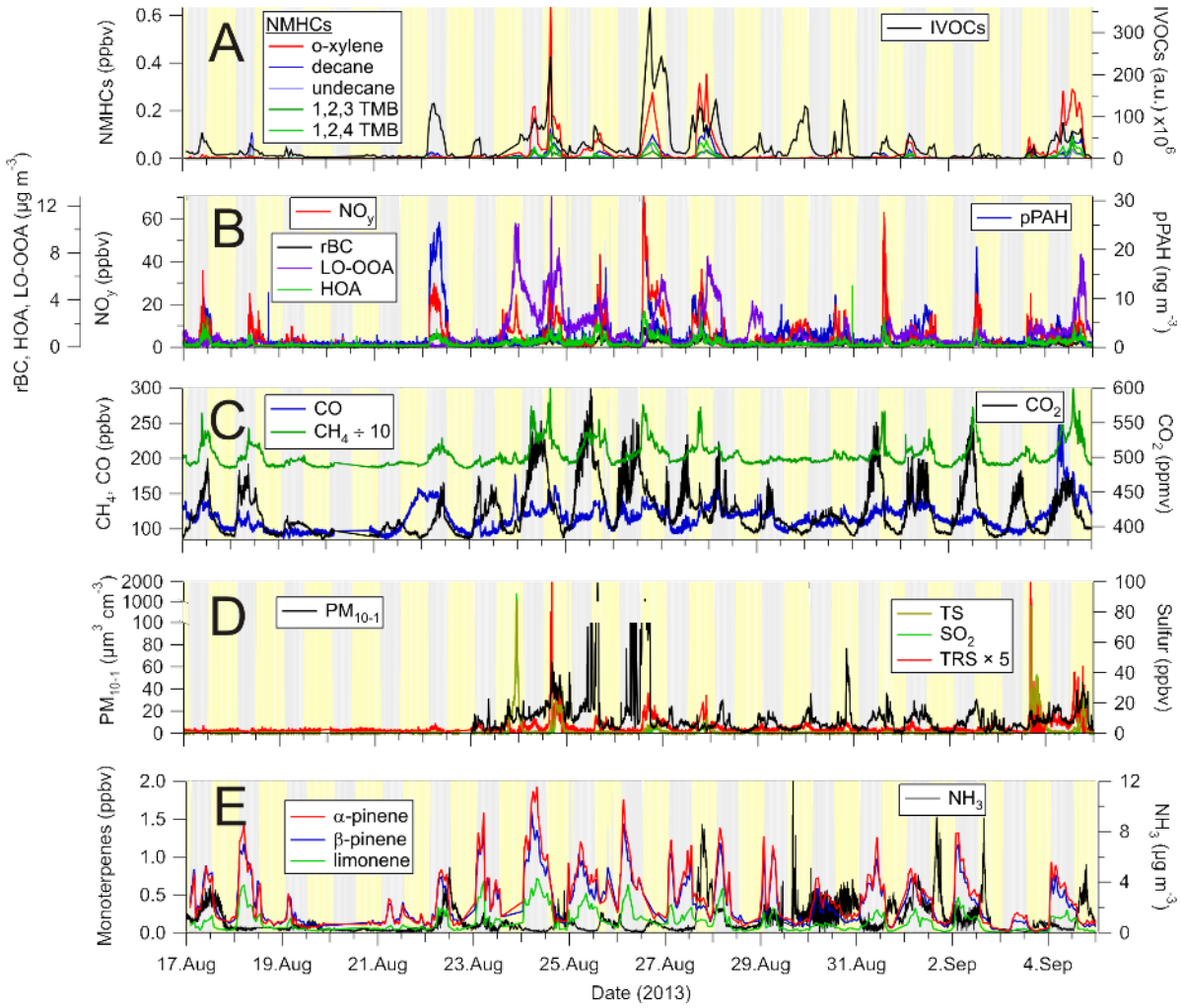


Figure 3. Time series of selected pollution tracers observed at the AMS 13 ground site in the Athabasca oil sands during the 2013 JOSM measurement intensive. The gray and yellow backgrounds represent night and day, respectively. (**A**) Selected non-methane hydrocarbons (NMHCs) and IVOCs. (**B**) Combustion product tracers: refractory black carbon (rBC), total odd nitrogen (NO_y) and particle surface bound polycyclic aromatic hydrocarbons (pPAH), and organic aerosol components: hydrocarbon-like organic aerosol (HOA) and less oxidized oxygenated organic aerosol (LO-OOA). (**C**) Methane (CH_4), carbon dioxide (CO_2) and monoxide (CO). (**D**) Total sulfur (TS), sulfur dioxide (SO_2), and total reduced sulfur (TRS) and PM_{10} particle volume. (**E**) Biogenic VOCs (α-pinene, ß-pinene and limonene) and ammonia (NH_3).

Time series of VOCs of primarily anthropogenic origin (i.e., o-xylene, 1, 2, 3- and 1, 2, 4-TMB, etc.) as well as the IVOC signature are shown in Fig. 3A. The abundances of these species, as well as the other compounds, were highly variable and varied as a function of time of day (i.e., boundary layer mixing height) and air mass origin, with higher VOC concentrations generally observed during daytime. The VOC concentrations varied between nearly pristine, remote conditions, with concentrations below detectable limits, to mixing ratios of aromatic species exceeding 100 pptv. The concentration range of o-xylene is within the extremes reported by WBEA in their 2013 annual report (WBEA, 2013), exemplifying that the data set is representative of typical pollutant levels in this region.

While there is some obvious covariance between variables (i.e., when the mixing ratios of one particular VOC increases, so do others), the ratios of hydrocarbons varied considerably. For example, on August 18, 10:50 UTC, the n-decane to o-xylene ratio was ~22:1, whereas on August 24, 07:40 UTC it was ~1:5.7. The IVOC magnitude also varied greatly and often increased and decreased in tandem with the other VOCs (e.g., on Aug 24, 16:30 UTC) but also increased independently from the other VOC abundances (e.g., on Aug 30, 01:20 UTC, and on the night of Aug 22). This behaviour suggests the presence of multiple sources with distinct signatures that are being sampled to a varying extent at different times. This, coupled with the intermittency of the highly elevated signals, presents an analysis problem frequently encountered in environmental analysis that is usually investigated through a factor or principal component analysis (Thurston et al., 2011; Guo et al., 2004).

Presented in Fig. 3B are the time series of NO_y, rBC and pPAH abundances, all of which are combustion byproducts. For example, rBC is emitted from combustion of fossil fuels, biofuels, open biomass burning, and burning of urban waste (Bond et al., 2004). Similar to the VOCs, the abundances of these species varied greatly, from very low, continental background levels (i.e., <100 pptv of NO_y, < LOD for rBC and pPAHs) to polluted concentrations (i.e., > 60 ppbv of NO_y, > 1 μg m^{-3} rBC, > 10 ng m^{-3} pPAHs) characteristic of polluted urban and industrial areas. When high concentrations of NO_y were observed,

its main component was NO_x (data not shown), which is a combustion byproduct usually associated with automobile exhaust. In the Alberta oil sands, emissions from off-road mining trucks as well as the upgrading processes are the main contributors to the NO_y burden (Percy, 2013; Watson et al., 2013).

Shown in Fig. 3C are the mixing ratios of the greenhouse gases CH_4 and CO_2 along with CO. Abundances of CO_2 were clearly attenuated by photosynthesis and respiration of the vegetation near the measurement site, as judged from the strong diurnal cycle in its concentration (not shown). Maxima typically occurred shortly after sunrise, coincident with the expected break-up of the nocturnal boundary layer. In addition to biogenic emissions from vegetation and soil, CO_2 originates from a variety of point and mobile sources in this region, including off-road mining trucks (Watson et al., 2013) and the extraction, upgrading, and refining of bitumen and on-road vehicle sources in the area (Nimana et al., 2015a, b). Concentrations of CO_2 spiked whenever these emissions were transported to the measurement site.

Concentrations of CH_4 also exhibit a diurnal cycle, with higher concentrations generally observed at night and peaking in the early morning hours. While CH_4 and CO_2 mixing ratios frequently correlated in plumes, their ratios were variable overall, suggesting they often originated from distinct sources. Potential methane point sources in the region include microbial production in tailings ponds (Siddique et al., 2012) and fugitive emissions associated with the mining and processing of bitumen (Johnson et al., 2016). Indeed, a recent analysis shows tailings ponds and open pit mining sources to be the largest sources of CH_4 in the region (Baray et al., 2018).

Similar to the anthropogenic VOCs, the abundances of CH_4 and CO_2 were highly variable and ranged from minima of 1.88 and 384 ppmv to maxima of 2.96 and 578 ppmv, corresponding to maximum enhancements of 1.63 and 1.47 relative to tropospheric global monthly means of 1.806±0.001 and 394.3±0.1 ppmv for July, 2013 (Dlugokencky, 2017b, a), respectively.

Mixing ratios of CO also varied with time but generally were not elevated greatly (median 118 ppbv) above background levels (minimum 91 ppbv), except for occasional spikes in concentration (Fig. 3C). Carbon monoxide is a tracer of biomass burning and fossil fuel combustion, in particular in automobiles with poorly performing or absent catalytic converters, but is also a byproduct of the oxidation of VOCs, in particular of methane and isoprene which are oxidized over a wide area upwind of AMS 13 (Miller et al., 2008).

Time series of sulfur species and PM_{10-1} volume are shown in Fig. 3D. The TS and SO_2 data are dominated by intermittent plumes containing SO_2 mixing ratios exceeding 5 ppbv. The highest mixing ratio observed was 92.5 ppbv (in between the preconcentration periods of the GC-ITMS). Mixing ratios of SO_2 exhibited the most variability of all pollutants, as judged from the standard deviation of each of the measurements (Table 3). TRS levels were generally small (< 1 ppbv) and variable, except for plumes; TRS abundances in plumes, however, are more uncertain since they were calculated by subtraction of two large numbers. When TS and SO_2 abundances were low (< 1 ppbv), TRS abundances were variable and occasionally exhibited spikes that did not show any obvious correlation with other variables, suggesting the presence of one or more distinct TRS sources. PM_{10} volume concentrations varied a lot as well and, just like TRS, did not show an obvious correlation with other variables. Fugitive dust emissions likely contributed to much of the PM_{10} volume in the Athabasca oil sands region (Wang et al., 2015).

Time series of monoterpene mixing ratios are shown in Fig. 3E. α-Pinene was generally the most abundant monoterpene, followed by β-pinene. Their ratio, averaged over the entire campaign was 1:0.85, though occasionally the α- to β-pinene ratio was below 1:2 (e.g., on Aug 28, 14:50 UTC and Sept 5, 12:40 UTC). Terpene mixing ratios were generally higher at night than during the day, with maxima of 1.9 and 1.6 ppbv, respectively, a diurnal pattern consistent with what has been observed at other forest locations (Fuentes et al., 1996). Monoterpenes are emitted by plants via both photosynthetic and non-photosynthetic pathways (Fares et al., 2013; Guenther et al., 2012); at night, their emissions accumulate

in a shallow nocturnal boundary layer, whereas during daytime, they are entrained aloft (above the canopy) and oxidized by the hydroxyl radical (OH) and O_3, which are more abundant during the day than at night (Fuentes et al., 1996). α- and β-pinene mixing ratios were lowest mid-day (median values at noon of 140 and 133 pptv, respectively). The largest daytime concentrations were observed on Aug 25, a cloudy day (as judged from spectral radiometer measurements of the NO_2 photolysis frequency): on this particular day, mixing ratios at noon were 687 and 850 pptv, respectively.

Also shown in Fig. 3E is the time series of ammonia. These data were dominated by spikes which were observed sporadically and did not correlate with other variables, suggesting the presence of nearby ammonia point sources. Ammonia was not as variable as some of the other pollutants (e.g., the anthropogenic VOCs, sulfur species) as judged from its standard deviation (Table 3), which suggests a geographically more disperse source or sources similar to CO or CH_4, which have a "background". This is consistent with a recent study by Whaley et al. (Whaley et al., 2018) that estimated over half (~57%) of the near-surface NH_3 during the study period originated from NH_3 bi-directional exchange (i.e. re-emission of NH_3 from plants and soils), with the remainder being from a mix of anthropogenic sources (~20%) and forest fires (~23%).

3.2. Principal component analysis

3.2.1. PCA with primary variables

The loadings of the optimum solution are presented in Table 5. The 10-component solution accounts for a cumulative variance of 95.5%. The communalities for the analysis, i.e., the fraction of total pollutant observations accounted for by the PCA are all greater than 85%, with the lowest communality obtained for the IVOCs (0.86).

In the following, an overview of the observed components is presented. Associations with r>0.7, r>0.3,

and r>0.2 are referred to as "strong", "weak", and "poor", respectively. Hypothesized identifications are given in section 4 and are summarized in Table 6 and Fig. 4.

The component accounting for most of the variance of the data, component 1, is strongly associated with the anthropogenic VOCs (r > 0.87), weakly associated with CH_4 (r = 0.59), TRS (r = 0.59), HOA (r = 0.40), LO-OOA (r = 0.45), CO (r = 0.41), and the IVOCs (r = 0.31), and poorly associated with NO_y (r = 0.27) and rBC (r = 0.30). Component 2 is strongly associated with the combustion tracers NO_y (r = 0.82), rBC (r = 0.77), HOA (r = 0.74), and pPAH (r = 0.94), weakly associated with CH_4 (r = 0.39) and IVOCs (r = 0.39), and poorly associated with ammonia (r = 0.20), and undecane and decane (r = 0.27 and 0.22, respectively). Component 3 is strongly associated (r > 0.9) with the biogenic VOCs and weakly associated with CO_2 (r = 0.48) and shows poor negative correlations with NO_y (r = -0.26) and ammonia (r = -0.24). Component 4 is strongly associated with SO_2 and TS (r = 0.97 and 0.93, respectively) and poorly with NO_y (r = 0.21) and LO-OOA (r = 0.28).

Components 1 through 4 emerged regardless of the number of components used to represent the data, whereas the structure of components 5 through 10 only fully emerged in the 10-component solution (see S.I.). Hence, components 6 through 10 are somewhat tentative as many (i.e., 7 – 9) are single variable components and have eigenvalues close to or below unity, i.e., account for less variance than any single variable. As a result, the interpretations of these components are subject to more uncertainty and are more speculative but are presented in the S.I. for the sake of completeness and transparency. For the purpose of this manuscript, this is inconsequential as components 6 – 10 are not associated with IVOCs.

Table 5. Loadings for the 10-factor, optimal solution (primary variables only). Coefficients with Pearson correlation coefficients r>0.3 are shown in bold font.

	1	2	3	4	5	6	7	8	9	10	Commu-nalities
Anthropogenic VOCs											
o-xylene	**0.88**	0.08	0.02	0.10	0.14	0.13	0.07	-0.04	0.16	**0.32**	0.95
1,2,3 - TMB	**0.93**	0.16	0.07	0.05	0.05	0.11	0.04	-0.02	0.18	-0.01	0.95
1,2,4 - TMB	**0.94**	0.14	0.01	0.10	0.11	0.08	0.07	-0.03	0.18	0.13	0.98
decane	**0.92**	0.22	-0.02	0.15	0.23	0.01	0.05	0.04	0.04	0.03	0.97
undecane	**0.87**	0.27	-0.08	0.23	0.20	-0.06	0.12	0.07	-0.04	-0.10	0.96
Biogenic VOCs											
α-pinene	-0.03	-0.08	**0.98**	-0.11	0.02	0.04	0.01	-0.08	0.02	0.01	0.98
ß-pinene	-0.02	-0.08	**0.98**	-0.12	0.02	0.03	0.02	-0.07	0.00	0.01	0.98
limonene	0.07	-0.03	**0.92**	-0.08	0.12	0.24	0.05	-0.11	0.03	-0.05	0.95
Combustion tracers											
NO_y	0.27	**0.82**	-0.26	0.21	0.22	-0.04	0.02	0.10	-0.08	0.01	0.92
rBC	**0.30**	**0.77**	0.03	0.05	**0.44**	0.10	0.09	0.13	0.12	-0.10	0.94
CO	**0.41**	0.18	0.04	0.02	0.09	0.09	0.08	0.06	**0.87**	-0.01	0.99
CO_2	0.09	0.08	**0.48**	-0.12	-0.03	**0.77**	0.25	-0.14	0.05	-0.08	0.95
Aerosol species											
pPAH	0.06	**0.94**	-0.07	-0.13	-0.11	0.07	0.01	0.13	0.10	0.04	0.95
PM_{10-1}	0.18	0.14	0.08	0.09	0.11	0.17	**0.93**	-0.03	0.07	0.08	0.98
HOA	**0.40**	**0.74**	0.02	0.12	0.25	0.15	0.23	-0.06	0.16	0.09	0.90
LO-OOA	**0.45**	0.11	0.12	0.28	**0.72**	0.05	0.25	0.00	0.10	0.04	0.91
Sulfur											
TS	0.25	0.04	-0.16	**0.93**	0.08	-0.05	0.07	-0.02	0.01	0.12	1.00
SO_2	0.12	0.03	-0.15	**0.97**	0.02	-0.04	0.03	-0.03	0.01	-0.05	0.99
TRS	**0.59**	0.04	-0.08	0.11	0.26	-0.04	0.16	0.04	-0.04	**0.71**	0.96
Other											
IVOCs	**0.31**	**0.39**	0.12	-0.08	**0.74**	-0.02	-0.02	-0.06	0.02	0.20	0.86
NH_3	0.01	0.20	-0.24	-0.05	-0.02	-0.08	-0.03	**0.94**	0.04	0.02	0.99
CH_4	**0.59**	**0.39**	0.10	-0.05	0.12	**0.59**	0.11	0.00	0.17	0.14	0.93
Eigenvalues	5.72	3.32	3.23	2.16	1.64	1.13	1.13	0.99	0.96	0.74	
% of variance	25.99	15.08	14.69	9.80	7.46	5.14	5.13	4.51	4.36	3.35	
Cumulative variance	25.99	41.07	55.76	65.56	73.02	78.16	83.30	87.81	92.17	95.52	

390 **Table 6**. Hypothesized identifications of principal components.

Component	Key observations	Possible source(s)	Relevant references
1	Enhancements of aromatics, n-alkanes, TRS, NO_y, rBC, HOA, LO-OOA, CO and CH_4	Wet tailings ponds and associated facilities	(Simpson et al., 2010; Small et al., 2015; Percy, 2013; Holowenko et al., 2000; Howell et al., 2014)
2	Enhancements of NO_y, rBC, pPAH and HOA due to engine exhaust	Mine fleet and operations	(Wang et al., 2016; Grimmer et al., 1987; Allen, 2008; Briggs and Long, 2016)
3	Enhancements of monoterpenes and CO_2, poor anticorrelation with NO_y and absence of anthropogenic VOCs	Biogenic emission and respiration	(Guenther et al., 2012; Helmig et al., 1999)
4	Enhancements of SO_2 and TS, poor correlation with NO_y and LO-OOA	Upgrader facilities	(Simpson et al., 2010; Kindzierski and Ranganathan, 2006)
5	Enhancements of IVOCs, rBC, LO-OOA, NO_y, and TRS	Surface exposed bitumen and hot-water based bitumen extraction	this work
6	Enhancements of CO_2 and CH_4, absence of combustion tracers	Mine face and soil	(Johnson et al., 2016; Rooney et al., 2012)
7	Enhancement of PM_{10-1}	Wind-blown dust	(Wang et al., 2015)
8	Enhancement of ammonia	Fugitive emissions from storage tanks and natural soil/plant emissions	(Bytnerowicz et al., 2010; Whaley et al., 2018)
9	Enhancement of CO	Incomplete hydrocarbon oxidation	(Marey et al., 2015)
10	Enhancements of TRS and o-xylene, poor association with CH_4	Composite tailings	(Small et al., 2015; Warren et al., 2016)

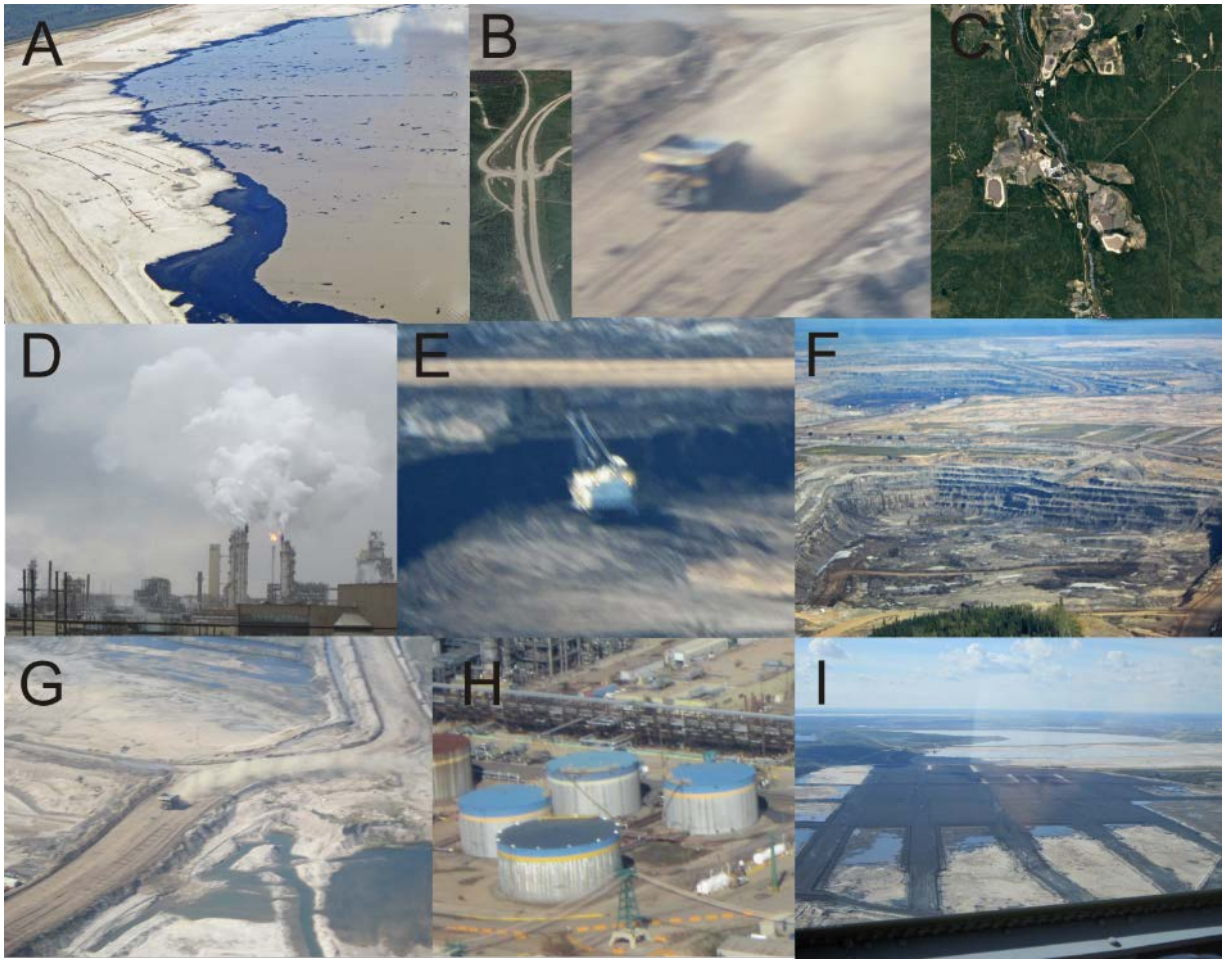


Figure 4. Images of likely sources associated with each of the principal components. From top left to bottom: **(A)** Wet tailings ponds (component 1). **(B)** Mine truck fleet and highway traffic emissions (component 2). **(C)** Biogenic emissions from vegetation (component 3). **(D)** Upgrader facilities (component 4). **(E)** Exposed bitumen on mined surfaces (component 5). **(F)** Fugitive greenhouse gas emissions from mine faces (component 6). **(G)** Wind-blown dust from exposed sand (component 7). **(H)** Fugitive emissions of ammonia from storage tanks (Component 8). **(I)** Composite (dry) tailings (component 10). No image is shown for production CO from oxidation of VOCs (component 9).

3.2.2. Extended PCA with added secondary variables

The loadings of the optimum solution that includes primary and secondary variables are shown in Table 7. In this 11-component solution, the 10 components originally identified were preserved, though their relative order was changed, with the upgrader component moving from the 4th to 2nd position. There was one new component (#6), which encompassed only secondary species, including MO-OOA (r = 0.92), O_x (r = 0.33), $NO_{3(p)}^-$ (r = 0.36), PM_1 (r = 0.31) and LO-OOA (r = 0.31).

$NH_{4(p)}^+$, $SO_{4(p)}^{2-}$, and $NO_{3(p)}^-$ are associated with the stack emissions component (#2, with r = 0.84, 0.84 and 0.44, respectively), which also weakly correlated with PM_1 (r = 0.44) and O_x (r = 0.36). The association of secondary variables with the primary components suggests rapid formation of these secondary products on a time scale that is similar to the transit time of the pollutants to the measurement site. PM_1 correlated strongly with the major IVOC component (component 5, r = 0.80), which also weakly associated with LO-OOA (r=0.66) and $NO_{3(p)}^-$ (r = 0.59), as well as $NH_{4(p)}^+$ and $SO_{4(p)}^{2-}$ (r = 0.32 and 0.33, respectively).

Table 7. Loadings for the 11-component solution with the inclusion of variables associated with secondary processes.

	1	2	3	4	5	6	7	8	9	10	11	Commu-nalities
Anthropogenic VOCs												
o-xylene	**0.89**	0.16	0.04	0.04	0.15	0.00	0.10	0.07	-0.04	0.17	0.24	0.94
1,2,3 - TMB	**0.91**	0.13	0.10	0.16	0.09	0.07	0.11	0.03	-0.03	0.16	-0.08	0.95
1,2,4 - TMB	**0.93**	0.19	0.02	0.13	0.13	0.05	0.06	0.07	-0.03	0.17	0.06	0.99
decane	**0.89**	0.25	0.00	0.22	0.26	0.05	-0.01	0.05	0.01	0.00	0.01	0.98
undecane	**0.81**	**0.35**	-0.08	0.27	0.21	0.15	-0.07	0.08	0.04	-0.12	-0.10	0.96
Biogenic VOCs												
α-pinene	0.00	-0.08	**0.98**	-0.07	0.05	0.03	0.01	0.01	-0.07	0.02	0.01	0.98
ß-pinene	0.01	-0.08	**0.98**	-0.08	0.05	0.05	0.01	0.03	-0.06	0.01	0.02	0.98
limonene	0.11	-0.02	**0.92**	-0.02	0.14	0.09	0.21	0.02	-0.10	0.02	-0.03	0.95
Combustion tracers												
NO_y	0.23	0.20	-0.27	**0.82**	0.21	-0.06	-0.07	0.03	0.10	-0.10	0.01	0.92
rBC	0.22	0.15	0.05	**0.80**	**0.43**	0.15	0.10	0.05	0.09	0.07	0.00	0.95
CO	**0.40**	0.09	0.08	0.20	0.09	0.22	0.08	0.06	0.03	**0.83**	-0.02	0.97
CO_2	0.12	-0.07	**0.50**	0.08	-0.03	0.09	**0.75**	0.28	-0.12	0.03	-0.08	0.95
Aerosol species												
pPAH	0.06	-0.10	-0.06	**0.93**	-0.07	-0.06	0.07	0.03	0.15	0.13	-0.05	0.94
PM_{10-1}	0.19	0.16	0.08	0.16	0.13	0.08	0.18	**0.91**	-0.03	0.05	0.07	0.99
PM_1	0.24	**0.44**	0.00	0.17	**0.70**	**0.31**	-0.06	0.11	-0.04	0.07	-0.14	0.90
$NH_4^+{}_{(p)}$	0.28	**0.84**	0.02	0.12	**0.32**	0.22	0.06	0.07	-0.04	0.14	-0.04	0.97
$SO_4^{2-}{}_{(p)}$	0.29	**0.84**	0.03	0.12	**0.33**	0.19	0.06	0.06	-0.05	0.12	-0.05	0.97
$NO_3^-{}_{(p)}$	**0.30**	**0.44**	0.09	0.23	**0.59**	**0.36**	0.08	0.15	-0.13	0.02	0.24	0.92
HOA	**0.37**	0.18	0.02	**0.77**	0.25	0.10	0.10	0.18	-0.08	0.13	0.14	0.93
LO-OOA	**0.37**	**0.40**	0.12	0.16	**0.66**	**0.31**	0.03	0.12	-0.06	0.00	0.27	0.97
MO-OOA	0.10	0.15	0.09	0.00	0.10	**0.92**	0.05	0.07	0.10	0.16	-0.03	0.95
Sulfur												
TS	0.27	**0.90**	-0.20	0.03	0.04	-0.04	-0.09	0.07	0.00	-0.04	0.18	0.98
SO_2	0.09	**0.96**	-0.19	0.02	-0.03	-0.01	-0.08	0.03	-0.02	-0.03	0.00	0.98
TRS	**0.65**	0.14	-0.10	0.05	0.23	-0.08	-0.07	0.17	0.06	-0.04	**0.63**	0.95
Other												
IVOCs	**0.34**	-0.01	0.12	**0.33**	**0.80**	-0.23	-0.02	0.02	0.02	0.06	0.06	0.94
NH_3	-0.03	-0.08	-0.22	0.21	-0.04	0.09	-0.07	-0.03	**0.93**	0.02	0.02	0.99
O_x	0.07	**0.36**	**-0.62**	0.01	0.27	**0.33**	**-0.41**	-0.07	-0.03	-0.14	0.12	0.91
CH_4	**0.60**	0.00	0.14	**0.42**	0.10	0.08	**0.57**	0.08	-0.04	0.13	0.16	0.94
Eigenvalues	5.85	4.30	3.71	3.51	2.78	1.58	1.24	1.09	1.01	0.94	0.75	
% of variance	20.90	15.34	13.25	12.52	9.92	5.65	4.43	3.88	3.59	3.37	2.66	
Cumulative variance	20.90	36.24	49.49	62.02	71.94	77.59	82.03	85.90	89.50	92.87	95.53	

3.3 Bivariate polar plots

Bivariate polar plots were generated for all components and their dominant, associated variables and are shown in the supplemental material section (Figs. S2-S11). Winds were predominantly from the SW but were also observed often from the S and N. Fig. 5A shows the plot for IVOCs. The highest concentrations were observed when the local wind direction was from the NE, where several facilities including the Aurora North, Musket River and Jackpine mines and large swaths of disturbed and cleared land are located in close proximity to each other (Table 1 and Fig. 1). The second highest IVOC signal intensity was observed when local wind direction was from the SSE.

The bivariate polar plots of the 3 components associated with IVOCs are shown in Fig. 5B-D. These components are associated with winds from the NE, E, SE and S at low to moderate speeds (1-3 m s^{-1}). Component 5 (Fig. 5B) was the most strongly correlated with IVOCs and shows the most spatial overlap with the distribution of the IVOC source; however, the intensities differ owing to the association of component 5 with other variables such rBC and LO-OOA.

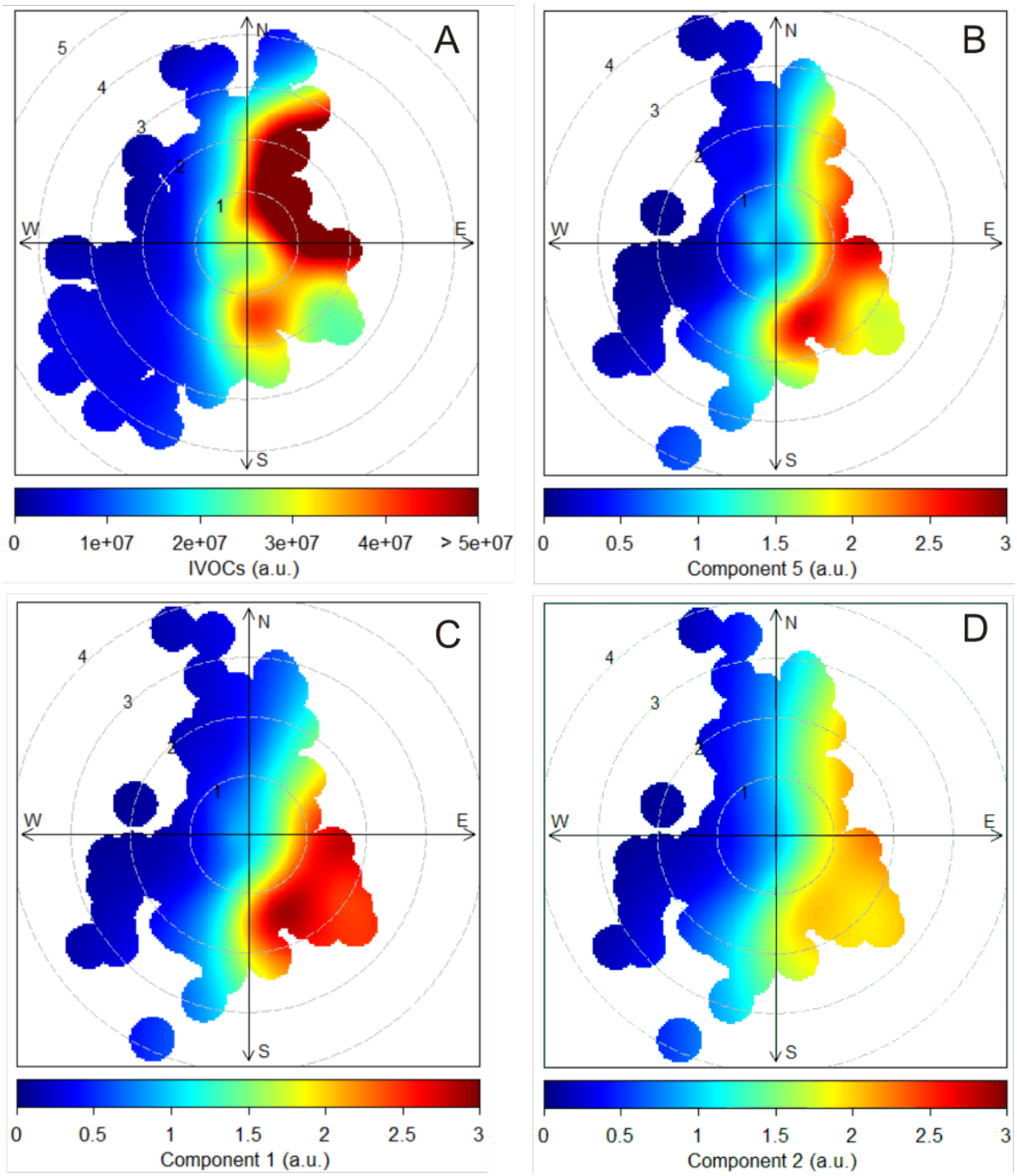


Figure 5. Bivariate polar plots related to IVOCs: **(A)** IVOCs from the complete data set. **(B)** Component 5 extracted from the main PCA (Table 5). **(C)** Component 1 extracted from the main PCA. **(D)** Component 2 extracted from the main PCA. Wind direction is binned into 10° intervals and wind direction into 30° intervals. The polar axis indicates wind speed (m s^{-1}). a.u. = arbitrary units.

4. Discussion

This work has added to the relatively few data sets of pollutants in the Athabasca oil sands region, one of the largest emitters of airborne pollutants in Canada (NPRI, 2013), that are available in the open literature. Earlier source apportionment studies in the region investigated ground level O_3 and $PM_{2.5}$ (Cho et al., 2012), examined VOCs (Bari and Kindzierski, 2018; Bari et al., 2016) and $PM_{2.5}$ (Bari and Kindzierski, 2017; Landis et al., 2017) impacting the nearby communities of Fort McKay and Fort McMurray, or investigated pollutants such as PAHs as they affect sediments (Jautzy et al., 2013) or lichens (Landis et al., 2012). The measurement suite in this work encompassed a larger variety of collocated analytical instruments closer to oil sands mining operations than these earlier studies and included a first, direct observation of airborne IVOCs, that is unique to this area and we have not observed elsewhere where we have made GC-ITMS measurements, i.e., in Calgary and on Vancouver Island (Tokarek et al., 2017a).

The main objective of this work was to elucidate the origin of the IVOC signature observed at the AMS 13 ground site downwind from the AB oil sands mining operations (Fig. 2) through a PCA. The optimum solution identified 10 components, of which three were associated with the IVOC signature: 1, 2, and 5 (Table 5). Tentative assignments of these components to source types in the oil sands are given in Table 6 and are discussed below.

Emission inventories show that the facilities that process the mined bitumen are by far the largest anthropogenic point sources in the oil sands region (NPRI, 2013), consistent with recent aircraft measurements (Baray et al., 2018; Howell et al., 2014; Li et al., 2017; Simpson et al., 2010) which have shown substantial emissions of NO_y, SO_2, CO, VOCs, CO_2, and CH_4, from these facilities and associated mining activities. No single component correlates with all of these variables, suggesting that the PCA is able to distinguish between source types within the facilities such as tailings ponds (component 1), stack

emissions (component 4), and mining (component 2).

Close-up overflights (Howell et al., 2014; Li et al., 2017; Baray et al., 2018) were able to spatially resolve various oil sands facility emission sources (i.e., tailings ponds from upgraders, fluid coking reactors, hydrocrackers and –treaters); the PCA presented in this manuscript is not expected to do this in all cases because some emissions would have frequently merged into a single plume by the time of observation at AMS 13; unless their emissions vary considerably in time, these sources could be interpreted as originating from a single source in the PCA.

The discussion below focuses on components that are associated with IVOCs (section 4.1), followed by those that are not (section 4.2). The PCAthat included 6 secondary products is discussed in section 4.3. Components which are not associated with IVOCs and have only tentatively been identified (i.e., components 6 – 10) are discussed in the S.I.

4.1 Components associated with IVOCs

4.1.1. Component 1: Tailings ponds (wet tailings)

Component 1 is strongly associated with anthropogenic VOCs ($r > 0.87$) and weakly with TRS ($r = 0.59$), and CH_4 ($r = 0.59$). These pollutants originate from tailings ponds (Small et al., 2015), though it is unclear from this analysis how large a source tailings ponds are compared to fugitive emissions of these pollutants from the nearby processing (e.g., bitumen separation and mining) facilities.

Tailings ponds cover large areas of land and are used to slowly (on a time scale of years to decades) separate solid components, or tailings, from water used in bitumen extraction. Residual bitumen often floats to the top of the settling basins. Most tailings ponds are "wet" (as they contain residual naphtha that is used as a diluent during the transfer of tailings to the ponds) and emit VOCs, CH_4, and CO_2 (Small et al., 2015). The presence of o-xylene, TMB and the n-alkanes in component 1 is consistent with the

480 fugitive release of VOCs from residual naphtha, which contains these compounds (Siddique et al., 2008;
481 Siddique et al., 2011; Small et al., 2015). Furthermore, the observation of TRS and CH_4 from this source is
482 consistent with the presence of anaerobic sulfur reducing bacteria and methanogens within the ponds,
483 which degrade not only the residual bitumen (Holowenko et al., 2000; Percy, 2013; Quagraine et al.,
484 2005) but also the various components of naphtha (Shahimin and Siddique, 2017; Small et al., 2015).
485 Overall, tailings ponds emissions explain much of the TRS and CH_4 concentration variability in this data
486 set (Table 5) and in a recent aircraft study (Baray et al., 2018).

487 While component 1 correlates with CH_4 ($r = 0.59$), it does not correlate with CO_2 ($r = 0.09$). Emissions of
488 CH_4 from tailings ponds due to methanogenic bacterial activity are well-documented (Small et al., 2015;
489 Yeh et al., 2010) and hence the correlation with CH_4 is not unexpected. On the other hand, the lack of
490 correlation with CO_2 seems inconsistent with emission inventories that generally present tailings ponds
491 as large CO_2 sources (Small et al., 2015). One plausible explanation is that tailings ponds are a relatively
492 small CO_2 source overall in the region and that other, larger CO_2 sources and sinks (such as
493 photosynthesis and respiration by the vegetation surrounding the site) dominate the variance impacting
494 the PCA results. It may also indicate that, at least on aggregate and for the particular ponds detected in
495 this work, the emissions are in a regime where the release of CH_4 dominates over CO_2, i.e., the ponds
496 have, perhaps, become more anoxic than believed to be the case in previous studies and hence emit
497 more CH_4 (Holowenko et al., 2000). For example, Small et al. (2015) showed that older tailings ponds
498 (those without the addition of fresh froth or thickening treatments) tended to emit more CH_4, while
499 newer ponds are associated with higher VOC emissions. It is likely that component 1 is dominated by the
500 nearest pond (the Mildred Lake settling basin, 6 – 11 km SSE of AMS 13) and other tailings in the SE
501 where the majority of air samples originated from. The Mildred Lake settling basin is one of the oldest in
502 the region and is still actively being used; the correlation with CH_4 and VOC emissions is hence expected.

503 Component 1 is also associated with NO_y, rBC, CO, and HOA, though these correlations are relatively

504 modest (r = 0.27, 0.30, 0.41, and 0.40, respectively). These species typically originate from combustion
505 sources, such as generators, motor vehicles, including diesel powered engines powering generators or
506 pumps; it is not obvious if and to what extent these are operated on or near tailings ponds, though.
507 Satellite observations have shown elevated concentrations of NO_2 above on-site upgrader facilities,
508 likely a result of emissions from extraction and transport sources (McLinden et al., 2012). In addition,
509 one of the major highways of the region is located adjacent to the Mildred Lake settling basin and other
510 major ponds in the region; highway traffic emissions (of CO, NO_y, rBC, and HOA) may hence also be
511 partially included in component 1.

512 The bivariate polar plot shows that component 1 was observed when local wind speeds were from the
513 SE and E of the measurement site (Fig. 5C), which is consistent with the notion that the Mildred Lake
514 settling basin and emissions along Highway 63 and, potentially, more distant facilities are sources
515 contributing to this component.

516 Component 1 is associated with the IVOC signature, though to a lesser degree than components 2 and 5.
517 The association of the IVOC signal with component 1 is slightly poorer (r = 0.31) than the association
518 with component 2 (r = 0.39), but significantly poorer than component 5 (r = 0.74). One possible
519 explanation for the association of IVOCs with tailings ponds vapor is the presence of bitumen in the
520 ponds that was not separated from the sand during the separation stage (Holowenko et al., 2000). This
521 semi-processed bitumen would be expected to emit the same IVOC vapors to those that were observed
522 in the lab (Fig. 2). Tailings ponds contain anywhere from 0.5% - 5% residual bitumen by weight
523 (Chalaturnyk et al., 2002; Holowenko et al., 2000; Penner and Foght, 2010). As illustrated in Fig. 4A,
524 some of this material floats on the ponds' surfaces, where IVOCs can partition to the air. Emission of
525 IVOCs from bitumen floating on tailings ponds would be a function of many variables (e.g., diluent
526 composition, extraction methodology, settling rate, temperature, etc.) and is thus not expected to be as
527 persistent as CH_4 partitioning from the ponds to the above air or from exposed bitumen on the mine

surface, leading to a lower overall correlation.

Component 1 is also weakly associated with the less oxidized oxygenated organic aerosol factor, LO-OOA (r = 0.45). Liggio et al. (2016) found that the observed secondary organic aerosol is dominated by an OOA factor whose mass spectrum was similar to those of aerosols formed from oxidized bitumen vapours. The organic aerosol budget in this study was also dominated by an OOA factor, the LO-OOA (Lee et al., 2018). The association of LO-OOA with component 1 is thus consistent with its association with IVOCs.

4.1.2. Component 2: Mine fleet and vehicle emissions

Component 2 strongly correlates with NO_y (r = 0.82), rBC (r = 0.77), pPAH (r = 0.94), and HOA (r = 0.74), which suggests a combustion source such as diesel engines. In the AB oil sands, there is a sizeable off-road mining truck fleet consisting of heavy aggregate haulers. In addition, there are diesel engine sources associated with generators, pumps and land moving equipment, i.e., graders, dozers, hydraulic excavators, and electric rope shovels (Watson et al., 2013; Wang et al., 2016). Most of these non-road applications have been exempt from highway fuel taxes, on-road fuel formulation requirements and after-engine exhaust treatment (Watson et al., 2013). Emissions from the hauler fleet and the stationary sources would fit the profile of component 2. Other diesel engines operated in the region include a commuter bus fleet, pickup and delivery trucks, tractor-trailers, and privately owned diesel powered automobiles used to commute from the work sites to the major residential areas around Fort McMurray, whose emissions are likely captured by component 2 as well, though the magnitude of these relative to the mining truck fleet is not known. Consistent with component 2 being associated with an anthropogenic source is its poor correlation with undecane (r =0.27), likely arising from fugitive fuel emissions.

The bivariate polar plot (Fig. 5D) for component 2 and NO_y in particular (Fig. S-4A) match the location of

Highway 63 which crosses the river to the SE of AMS 13 and bends to the E and is indicative of a line source. At the same time, some of the largest mining operations in the region, the Susan Lake Gravel Pit, Aurora North, Muskeg river, and Millennium mines are located to the NE and SE of AMS 13 as well. NO_y, rBC, and HOA (Fig. S-4A, B and D) all appear to have dominating point sources to the S and E when wind speeds are 1-2 m s^{-1}. These directions are the same as the Fort McKay industrial park to the E and the Syncrude Mildred Lake facility parking lot to the S which would have a higher concentration of vehicles emitting these pollutants in a smaller area, whose emissions would be in addition to those from industrial activities.

Component 2 is associated with the IVOCs signature and CH_4 (both r = 0.39). The mining activities bring bitumen to the surface; similar to what we had observed in lab experiments (Fig. 2, black trace), the surface exposure of bitumen during mining and on-site processing is expected to be associated with fugitive emissions of CH_4 (Johnson et al., 2016) and IVOCs.

Fine-fraction particle-surface bound PAHs (pPAH) are associated strongly with component 2, but no other components. Measurements of individual PAHs in snow and moss downwind from the oil sands facilities have identified multiple sources of PAHs in the Athabasca oil sands, which include wind-blown petroleum coke dust (also referred to as petcoke for short), a carbonaceous residual product from the upgrading of crude petroleum that is stockpiled on mine sites, and emissions from fine tailings, oil sands ore, and naturally exposed bitumen (Zhang et al., 2016; Jautzy et al., 2015; Parajulee and Wania, 2014). Given this diversity of known sources, the associations of PAHs with only a single component is surprising, though indicates that emissions from the mining fleet (which would include diesel and, perhaps, wind-blown emissions from petcoke that is being transported) gave rise to most of the variability in surface-bound PAH concentrations in this data set. The petcoke emissions identified in the studies mentioned above are likely mainly associated with larger, supermicron sized particles, whose PAH content would not be detected by the pPAH measurement in this data set.

Component 2 is not associated with LO-OOA (r = 0.11), even though IVOCs are associated with this component. This feature may indicate that the IVOCs emitted in component 2 are qualitatively different from those emitted by components 1 and 5, in that they are less likely to yield organic aerosol on the time scale of transport from emission to observation. One reason for the difference could be that the bitumen that is transported by the mining fleet is relatively freshly exposed, whereas the IVOCs released by bitumen in tailings ponds has been processed by microbes and that released by mine faces (component 5) may have been photochemically oxidized to a greater extent and hence more prone to rapid aerosol formation.

There is no association of component 2 with CO_2 (r = 0.08). This is somewhat unexpected as the trucks are expected to release CO_2 (Wang et al., 2016) but could be due to significantly larger CO_2 sources in the area dominating the observed CO_2 variability at AMS 13 (e.g., components 3 and 6). Furthermore, one would expect an association of non-road mining truck emissions with aromatics and alkanes. Component 2 exhibited only poor correlations with decane (r = 0.22) and undecane (r = 0.27) and no correlation with o-xylene (r = 0.08), suggesting that other components (i.e., component 1) explained most of the variability of their concentrations at this site.

4.1.3. Component 5: Surface-exposed bitumen and hot-water bitumen extraction

Component 5 correlates more strongly with the IVOCs (r = 0.74) than with any other component and correlates strongly with LO-OOA (r = 0.72), weakly with rBC (r = 0.44), and poorly with HOA (r = 0.25), NO_y (r = 0.22), decane (r = 0.23), undecane (r = 0.20), and TRS (r = 0.26). We interpret this profile as emissions from surface-exposed bitumen which outgases IVOCs.

One possibility is that these emissions occur on mine faces, where previously unexposed bitumen is brought to the surface as a result of mining. Only a relatively small portion of the mine faces is actively

598 mined; those parts give rise to rBC and NO_y emissions from combustion engines in heavy haulers or
599 generators powering equipment. The poor association of component 5 with TRS could be due to sulfur
600 reducing bacteria found on the surface of bitumen. However, most of the variability of TRS at AMS 13 is
601 attributed to composite or “dry” tailings ponds given their more conducive environment to microbial
602 activity.

603 Component 5 does not correlate with CO_2 (r = -0.03) or with CH_4 (r = 0.12), which is somewhat at odds
604 with the notion of mine faces as the main source of IVOCs. The mine faces give rise to substantial
605 fugitive emissions of CO_2 and CH_4 (Johnson et al., 2016) – these emissions are likely captured by
606 component 6 in this analysis (see S.I.). It is unclear to what extent these greenhouse gases are released
607 relatively quickly from "hot spots" (i.e., from a small number of locations) through surface cracks and
608 fissures or by slow release from new material that is exposed and then releases greenhouse gases
609 during material handling, transport and processing (Johnson et al., 2016). IVOCs from surface-exposed
610 bitumen are likely released by the latter mechanism and are temperature-dependent. If the mine faces
611 are indeed the main IVOC source, the analysis results presented here suggest that the IVOCs emissions
612 from surface-exposed bitumen on mine faces are decoupled from CH_4 emissions in time and appear as a
613 distinct component and hence corroborate the "hot spots" or fast release hypothesis, though clearly,
614 more work is needed to characterize greenhouse gas emissions from oil sands mine faces.

615 The association of IVOCs with component 5 may also be a result of fugitive emissions during the hot
616 water-based extraction of bitumen sand slurries during the separation phase of bitumen treatment.
617 Generally, bitumen is extracted in a weak alkaline environment by aeration of the solution to optimize
618 the separation of sand and bitumen (Masliyah et al., 2004). Unrecovered bitumen and naphtha then end
619 up in tailings. The recovered bitumen and naphtha are moved to upgrader facilities where they undergo
620 further treatment (such as coking or hydrotreatment). The magnitude of fugitive emissions during these
621 downstream extraction processes could be large, considering the bitumen is heated and actively

aerated. Future work should investigate IVOC fluxes near extraction plants and on mine faces.

Finally, it is conceivable that a "natural" background of IVOCs exists in the region (since bitumen can be found at or near the surface in many parts of the region); such a natural background would also be included in component 5. However, this "natural" bitumen would have been exposed at the surface for geological time scales and, unlike unexposed, buried bitumen, likely would have lost most of its volatile content over that period. Furthermore, the mine faces occupy large swaths of land in the region (as evident from satellite imagery). Thus, the IVOCs emissions are more likely due to anthropogenic activity than due to a natural phenomenon.

4.2. Components not associated with IVOCs

4.2.1. Component 3: Biogenic emissions and respiration

Component 3 is strongly correlated with the monoterpenes α-pinene (r = 0.98), ß-pinene (r = 0.98) and limonene (r = 0.92) and is hence identified as a biogenic emissions source. This component is also weakly associated with CO_2 (r = 0.48).

At AMS 13, CO_2 and the monoterpenes exhibit a very similar diurnal cycle: they are present in higher concentrations during the night than during the day (Fig. 3) due to a decrease in the boundary layer height (BLH) at night coupled with plant respiration of CO_2 and non-photochemical emission of monoterpenes (Fares et al., 2013; Guenther et al., 2012). During the day, mixing ratios of CO_2 are lower due to plant uptake and photosynthesis, and mixing ratios of terpenes are lower due to higher mixing heights and vertical entrainment and due to oxidation by O_3 and OH (Fuentes et al., 1996). Hence, the PCA gives a *positive* correlation of monoterpenes with CO_2 even though the physical processes, photosynthesis and respiration, work in opposite direction.

644 The bivariate polar plots (Fig. S-5A-C) show that the monoterpenes and CO_2 were observed in highest
645 concentrations when the wind speeds were low (< 1 m s^{-1}), consistent with formation of a stable
646 nocturnal boundary layer.

647 To corroborate this interpretation, the PCA was repeated with BLH estimated by a light detection and
648 ranging (LIDAR) instrument (Strawbridge et al., in prep.) added as a variable (Table S-9 in the S.I.). Since
649 BLH is not "emitted" by any source, it appears as a single variable component (r = 0.90). The only other
650 component that BLH (anti)correlates with is the biogenic component 3 (r = -0.35).

651 The dominant monoterpene species observed was α-pinene, followed by ß-pinene and limonene,
652 though occasionally there was twice as much ß-pinene than α-pinene in the sampled air. Some
653 variability of this ratio is expected since emission factors vary considerably between tree species (Geron
654 et al., 2000) which are not homogeneously distributed throughout the region (e.g., Fig. S1 of Rooney et
655 al. (2012)).

656 Simpson et al. (2010) observed enhancements of α-pinene and, to a greater extent, β-pinene over the
657 oil sands (up to 217 pptv and 610 pptv) compared to background levels of 20±7 and 84±24 pptv,
658 respectively, during mid-day overflights (which occurred between 11:00 and 13:00 local time). Similar
659 enhancements were also reported by Li et al. (2017) who observed emissions of biogenic hydrocarbons
660 in the four facilities sampled, three of which showed a higher β- than α-pinene concentration. The PCA
661 (Table 5) showed no significant correlation of α- and β-pinene with any of the anthropogenic
662 components, which implies that the biogenic source strength is simply too large for any anthropogenic
663 emissions of terpenes to be picked up in the analysis, especially considering that terpenes are relatively
664 short-lived.

665 The biogenic source shows poor anticorrelations with NO_y (r = -0.26) and NH_3 (r = -0.24). Many NO_y
666 species (i.e., NO_2, HONO, peroxycarboxylic nitric anhydrides or PAN, and HNO_3) deposit to the forest

canopy (Hsu et al., 2016; Min et al., 2014; Fenn et al., 2015); at night, when mixing heights are lower, their concentrations are expected to decrease faster than during the day and are thus out of phase with the CO_2 and terpene concentrations. The poor anticorrelation with NH_3 likely arises because the NH_3 emissions from plants are mainly stomatal and scale with temperature and are hence larger during the day than at night, anticorrelated with the terpene source (Whaley et al., 2018).

4.2.2 Component 4: Upgrader emissions

Component 4 is strongly correlated with SO_2 (r = 0.97) and total sulfur (r = 0.93). By far the largest source of SO_2 in the region are upgrader facilities, which emit as much as 6×10^7 kg annually according to emission inventories (ECCC, 2013). Significant SO_2 emissions from upgrader facilities have recently been confirmed by aircraft studies (Simpson et al., 2010; Howell et al., 2014; Liggio et al., 2016). Component 4 is also poorly correlated with NO_y (r = 0.21) but not with rBC (r = 0.05), consistent with a non-sooty (i.e., lean) combustion source such as upgrader stacks. Strong enhancements in SO_2 were only observed intermittently as "spikes", which is expected when sampling emissions from relatively few and discrete point sources.

Component 4 isnot associated with CO_2 (r = -0.12), even though inventories indicate that the upgrading facilities are the largest CO_2 source in the region (Furimsky, 2003; Englander et al., 2013; Yeh et al., 2010). In this data set, the lack of correlation of component 4 with CO_2 (and to some extent with PM_{10-1} as well) likely arises mainly from a sampling bias as stack emissions were only observed during daytime, likely due to diurnal variability of the atmospheric boundary layer structure as explained below.

Most of the variability in CO_2 concentration at AMS 13 is due to surface-based sources that originate from large areas, especially biogenic processes (photosynthesis during the day and respiration at night, component 3) and anthropogenic surface sources such as those captured by component 6 (section 4.2.3). Other anthropogenic pollutants, such as SO_2, NO_y, and CH_4, are not subject to large biogenically

690 driven processes and are less affected than CO_2.

691 In contrast to surface sources, emissions from the > 100 m tall stacks are comparatively undersampled
692 and observed mainly during daytime, when vertical mixing brings elevated plumes to the surface, yet
693 CO_2 concentrations are generally much lower than during the night due to uptake by vegetation. At
694 night, pollutants emitted from stacks are injected above the likely very shallow nocturnal surface layer
695 and were hence not observed at the surface. Vertical profile measurements of SO_2 stack plumes by a
696 Pandora spectral sun photometer at Fort McKay during daytime have shown considerable vertical
697 gradients and only occasional transport of SO_2 all the way to the surface (Fioletov et al., 2016).

698 The association of component 4 with CO_2 is negative because the stack emission source is observed only
699 during the day when the large biogenic sink dominates and effectively masks the relatively small
700 increase due to anthropogenic CO_2. In contrast, background concentrations of SO_2 are comparatively
701 low, and the increase in SO_2 concentrations is readily picked up the PCA.

702 It would be interesting to conduct a future study in winter when biogenic activities decrease; a
703 wintertime PCA of surface measurements might be able to associate CO_2 enhancements with upgraders,
704 though boundary layer mixing heights would decrease as well, which would make a PCA using surface
705 data even more challenging.

706 Component 4 does not correlate with PM_{10-1} volume ($r = 0.09$). It is clear that the emitted SO_2 will
707 contribute to secondary aerosol formation downwind, such that a correlation of stack emissions with
708 PM_{10-1} volume might be expected. However, these secondary contributions will likely mostly be in the
709 submicron aerosol fraction, which adds relatively little to PM_{10-1} volume. Further, PM_{10-1} volume is
710 dominated by coarse particles from other primary sources, mostly wind-blown emission of sand from
711 the mine surfaces, roadways and, perhaps, bioaerosol (component 7, see S.I.). These effects make PM_{10-1}
712 volume from stacks appear comparatively small, such that the variability of the larger, surface-based

sources likely masks the contribution of stacks emissions to PM_{10-1} variability.

The bivariate polar plot of component 4 (Fig. S-6D) shows that the largest magnitudes were observed when local winds were from the SE. The corresponding plot of SO_2 (Fig. S-6A) reveals two more distinct sources: a larger one from the E and a smaller one from the SSE. However, only two facilities (Sunrise and Firebag) are located to the E at relatively large distances of 37 km and 47 km respectively. The largest known upgraders and SO_2 sources in the area (i.e., upgraders located at the Mildred Lake and Suncor base plants) are located to the S and SE of AMS 13. Considering that the stack emissions are only observed intermittently, we speculate that there exists a mesoscale transport pattern in the Athabasca river valley which channel emissions, such that the local wind direction and speed may be misleading as to the true location of these sources. For more extensive data sets, such phenomena may very well average out but perhaps did not in this case.

4.3. Extended PCA with added secondary variables

The extended analysis (Table 7) qualitatively preserves the structure (with the exception of an added "Aged" component, # 6) of the original 10-component solution but allows an assessment of which components most result in formation of secondary products such as SOA, which has implications for health (Bernstein, 2004) and climate (Charlson et al., 1992). Secondary products vary considerably as a function of air mass chemical age (which depends, amongst other components, on time of day and synoptic conditions, including wind speed) and are hence expected to add considerable noise and scatter to the results leading to lower correlations. On the other hand, the distance between the measurement site and sources is fixed, such that this variability should average out over time. This indeed appears to have happened in this data set in spite of the relatively low sample size.

The analysis indicates that the component with the strongest IVOC source (Component 5) also has the highest association with PM_1 ($r = 0.70$; Table 7). Aircraft measurements combined with a modelling

736 study have required a group of IVOC hydrocarbons to explain the significant SOA formation and growth
737 downwind of the oil sands region (Liggio et al., 2016). The association of IVOCs with PM_1 volume is
738 consistent with the hypothesis that oxidation of IVOCs observed at AMS 13 leads to SOA generation and
739 appears to have a significant impact on the variation in PM_1 mass.

740 The second component influencing PM_1 is that from stack emissions (Component 4 in the primary PCA;
741 Component 2 in the secondary PCA) (Tables 5 and 7). It is well established that the oxidation of SO_2 to
742 sulfate will lead to formation of fine particulate matter. This apparently occurs, at least partially, on the
743 time scale between the point of emission and the AMS 13 site (assuming a wind speed of 3 m/s and a
744 distance of 11 km, the transit time is 1 hour), though some fraction of SO_4^{2-}(p) is likely directly emitted.

745

746 **5. Summary and conclusions**

747 A PCA was applied to continuous measurements of 22 primary pollutant tracers at the AMS 13 ground
748 site in the Athabasca oil sands during the 2013 JOSM intensive study to elucidate the origins of airborne
749 analytically unresolved hydrocarbons that were observed by GC-ITMS. The analysis identified 10
750 components. Three components correlated with the IVOC signature and were tentatively assigned to
751 mine faces and, potentially, hot-water bitumen extraction facilities, the mine hauler fleet, and wet
752 tailings ponds emissions. All three are anthropogenic activities that involve the handling of raw bitumen,
753 i.e., the unearthing, mining and transport of crude bitumen, and the disposal of processed material that
754 contains residual bitumen in wet tailings ponds. The PCA results are consistent with our previous
755 interpretation that the unresolved hydrocarbons originate from bitumen based on the similarity of the
756 chromatograms with those obtained in a head space vapor analysis of ground-up bitumen in the
757 laboratory.

758 Liggio et al. (2016) showed that these hydrocarbons constitute a group of IVOCs in the saturation vapor

concentration (C^*) range 10^5 µg m^{-3}< C^* < 10^7 µg m^{-3} that contribute significantly to secondary organic aerosol formation and growth downwind of the oil sands facilities. The correlation of LO-OOA with two of the three IVOC components in the main PCA and with PM_1 in the extended analysis is consistent with the high SOA formation potential of IVOCs and suggests that further differentiation may be needed and stresses the need for IVOCs to be routinely monitored. In particular, direct measurements of emissions throughout the processing of raw bitumen are needed to pinpoint source contributions more accurately and aid in the development of potential mitigation strategies.

The PCA in this study suffered from several limitations. For instance, PCA does not provide insight into emission factors of individual facilities, though it does capture what conditions change ambient concentrations the most. Further, the receptor nature of PCA did not always discern between large source areas that may have many individual point sources coming together at the point of observation. For example, component 1 contains an obvious tailings pond signature because of its high correlation with anthropogenic VOCs, methane and TRS, but also includes several combustion sources, making interpretation of this IVOC source location more challenging. A longer continuous data set with a greater number of variables would have perhaps been able to resolve these different sources, including the various tailings ponds, of which there are 19 in the region, all with slightly different emission profiles (Small et al., 2015) .

Another limitation is the bias of this (and most) ground site data set towards surface-based emissions and the undersampling of stack emissions. Facility stacks were only observed in the daytime because at night the mixing height is so low that the stacks are emitting directly into the residual layer. These emissions could be quantified using aircraft based platforms (Howell et al., 2014; Li et al., 2017; Baray et al., 2018). The PCA struggled most with the allocation of greenhouse gases. Mixing ratios of CO_2, in particular, were difficult to reconcile in this analysis due to a high background and large attenuation by biogenic activity and boundary layer meteorology. Forests greatly affected CO_2 levels in the region

783 because it is taken up during the day when plants are photosynthetically active and emitted at night
784 when plants undergo cellular respiration. This CO_2 source and sink appears to dominate the PCA,
785 effectively masking relatively small emissions from tailings ponds, facilities, and tail pipes in particular
786 from the mine hauling fleet.

787 Finally, there is a need for improved monitoring methods for IVOCs. For instance, future studies should
788 focus on characterizing the VOCs in the above mentioned volatility range using a greater mass and time
789 resolution instrument, such as a time-of-flight mass spectrometer (TOF-MS) or higher resolution
790 separation methods (e.g., multi-dimensional gas chromatography), and also include measurement of
791 speciated aerosol organic composition by, for example, thermal desorption aerosol GC (TAG) analysis
792 (Williams et al., 2006). Future studies should also investigate how IVOC volatility distributions vary with
793 source type and chemical age.

Acknowledgments

Funding for this study was provided by Environment and Climate Change Canada and the Canada-Alberta Oil Sands Monitoring program. The GC-ITMS used in this work was purchased using funds provided by the Canada Foundation for Innovation and matching funds by the Alberta government. TWT, JAH, DKB, FVA and GRW acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Training Experience Program (CREATE) program Integrating Atmospheric Chemistry and Physics from Earth to Space (IACPES).

6. References

Government of Alberta, oil sands information portal: http://osip.alberta.ca, access: 23-FEB-2017, 2017.

Allen, E. W.: Process water treatment in Canada's oil sands industry: II. A review of emerging technologies, J. Environ. Eng. Sci., 7, 499-524, 10.1139/s08-020, 2008.

Baray, S., Darlington, A., Gordon, M., Hayden, K. L., Leithead, A., Li, S. M., Liu, P. S. K., Mittermeier, R. L., Moussa, S. G., O'Brien, J., Staebler, R., Wolde, M., Worthy, D., and McLaren, R.: Quantification of methane sources in the Athabasca Oil Sands Region of Alberta by aircraft mass balance, Atmos. Chem. Phys., 18, 7361-7378, 10.5194/acp-18-7361-2018, 2018.

Bari, M., and Kindzierski, W. B.: Fifteen-year trends in criteria air pollutants in oil sands communities of Alberta, Canada, Environ. Int., 74, 200-208, 10.1016/j.envint.2014.10.009, 2015.

Bari, M. A., Kindzierski, W. B., and Spink, D.: Twelve-year trends in ambient concentrations of volatile organic compounds in a community of the Alberta Oil Sands Region, Canada, Environ. Int., 91, 40-50, 10.1016/j.envint.2016.02.015, 2016.

Bari, M. A., and Kindzierski, W. B.: Ambient fine particulate matter (PM2.5) in Canadian oil sands communities: Levels, sources and potential human health risk, Sci. Tot. Environm., 595, 828-838, 10.1016/j.scitotenv.2017.04.023, 2017.

Bari, M. A., and Kindzierski, W. B.: Ambient volatile organic compounds (VOCs) in communities of the Athabasca oil sands region: Sources and screening health risk assessment, Environ. Pollut., 235, 602-614, 10.1016/j.envpol.2017.12.065, 2018.

Bernstein, D. M.: Increased mortality in COPD among construction workers exposed to inorganic dust, Eur. Resp. J., 24, 512-512, 10.1183/09031936.04.00044504, 2004.

Bond, T. C., Streets, D. G., Yarber, K. F., Nelson, S. M., Woo, J. H., and Klimont, Z.: A technology-based global inventory of black and organic carbon emissions from combustion, J. Geophys. Res., 109, D14203, 10.1029/2003JD003697, 2004.

825 Briggs, N. L., and Long, C. M.: Critical review of black carbon and elemental carbon source
826 apportionment in Europe and the United States, Atmos. Environ., 144, 409-427,
827 10.1016/j.atmosenv.2016.09.002, 2016.
828 Buhamra, S. S., Bouhamra, W. S., and Elkilani, A. S.: Assessment of air quality in ninety-nine residences of
829 Kuwait, Environ. Technol., 19, 357-367, 10.1080/09593331908616691, 1998.
830 Burtscher, H., Scherrer, L., Siegmann, H. C., Schmidtott, A., and Federer, B.: Probing aerosols by
831 photoelectric charging, J. Appl. Phys., 53, 3787-3791, 10.1063/1.331120, 1982.
832 Bytnerowicz, A., Fraczek, W., Schilling, S., and Alexander, D.: Spatial and temporal distribution of
833 ambient nitric acid and ammonia in the Athabasca Oil Sands Region, Alberta, J. Limnol., 69, 11-21,
834 10.3274/jl10-69-s1-03, 2010.
835 The facts on Canada's oil sands: http://www.capp.ca/publications-and-statistics/publications/296225,
836 access: April 20, 2017, 2016.
837 Carslaw, D. C., and Ropkins, K.: openair - An R package for air quality data analysis, Environ. Modell.
838 Softw., 27-28, 52-61, 10.1016/j.envsoft.2011.09.008, 2012.
839 Carslaw, D. C., and Beevers, S. D.: Characterising and understanding emission sources using bivariate
840 polar plots and k-means clustering, Environ. Modell. Softw., 40, 325-329,
841 10.1016/j.envsoft.2012.09.005, 2013.
842 Cattell, R. B.: The Scree Test For The Number Of Factors, Multivariate Behavioral Research, 1, 245-276,
843 10.1207/s15327906mbr0102_10, 1966.
844 Chalaturnyk, R. J., Don Scott, J., and Ozum, B.: Management of oil sands tailings, Petroleum Science and
845 Technology, 20, 1025-1046, 10.1081/lft-120003695, 2002.
846 Charlson, R. J., Schwartz, S. E., Hales, J. M., Cess, R. D., Coakley, J. A., Hansen, J. E., and Hofmann, D. J.:
847 Climate forcing by anthropogenic aerosols, Science, 255, 423-430, 10.1126/science.255.5043.423
848 1992.

849 Chen, H., Karion, A., Rella, C. W., Winderlich, J., Gerbig, C., Filges, A., Newberger, T., Sweeney, C., and
850 Tans, P. P.: Accurate measurements of carbon monoxide in humid air using the cavity ring-down
851 spectroscopy (CRDS) technique, Atmos. Meas. Tech., 6, 1031-1040, 10.5194/amt-6-1031-2013, 2013.
852 Cho, S., Morris, R., McEachern, P., Shah, T., Johnson, J., and Nopmongcol, U.: Emission sources
853 sensitivity study for ground-level ozone and PM2.5 due to oil sands development using air quality
854 modelling system: Part II – Source apportionment modelling, Atmos. Environm., 55, 542-556,
855 10.1016/j.atmosenv.2012.02.025, 2012.
856 Cross, E. S., Hunter, J. F., Carrasquillo, A. J., Franklin, J. P., Herndon, S. C., Jayne, J. T., Worsnop, D. R.,
857 Miake-Lye, R. C., and Kroll, J. H.: Online measurements of the emissions of intermediate-volatility and
858 semi-volatile organic compounds from aircraft, Atmos. Chem. Phys., 13, 7845-7858, 10.5194/acp-13-
859 7845-2013, 2013.
860 de Gouw, J. A., Middlebrook, A. M., Warneke, C., Ahmadov, R., Atlas, E. L., Bahreini, R., Blake, D. R.,
861 Brock, C. A., Brioude, J., Fahey, D. W., Fehsenfeld, F. C., Holloway, J. S., Le Henaff, M., Lueb, R. A.,
862 McKeen, S. A., Meagher, J. F., Murphy, D. M., Paris, C., Parrish, D. D., Perring, A. E., Pollack, I. B.,
863 Ravishankara, A. R., Robinson, A. L., Ryerson, T. B., Schwarz, J. P., Spackman, J. R., Srinivasan, A., and
864 Watts, L. A.: Organic Aerosol Formation Downwind from the Deepwater Horizon Oil Spill, Science,
865 331, 1295-1299, 10.1126/science.1200320, 2011.
866 Trends in atmospheric methane: www.esrl.noaa.gov/gmd/ccgg/trends_ch4/, access: April 11, 2017,
867 2017a.
868 Trends in atmospheric carbon dioxide: www.esrl.noaa.gov/gmd/ccgg/trends/, access: April 11, 2017,
869 2017b.
870 National pollutant release inventory (NPRI): http://open.canada.ca/data/en/dataset/e40099ae-b116-
871 4c48-9475-f3806fe5a6a6, access: October 5, 2016, 2013.

ECCC: Joint oil sands monitoring program emissions inventory compilation report, Environment and Climate Change Canada, Downsview, 2016.

Englander, J. G., Bharadwaj, S., and Brandt, A. R.: Historical trends in greenhouse gas emissions of the Alberta oil sands (1970-2010), Environm. Res. Lett., 8, 044036, 10.1088/1748-9326/8/4/044036, 2013.

Fares, S., Schnitzhofer, R., Jiang, X., Guenther, A., Hansel, A., and Loreto, F.: Observations of Diurnal to Weekly Variations of Monoterpene-Dominated Fluxes of Volatile Organic Compounds from Mediterranean Forests: Implications for Regional Modeling, Environm. Sci. Technol., 47, 11073-11082, 10.1021/es4022156, 2013.

Fenn, M. E., Bytnerowicz, A., Schilling, S. L., and Ross, C. S.: Atmospheric deposition of nitrogen, sulfur and base cations in jack pine stands in the Athabasca Oil Sands Region, Alberta, Canada, Environ. Pollut., 196, 497-510, 10.1016/j.envpol.2014.08.023, 2015.

Fioletov, V. E., McLinden, C. A., Cede, A., Davies, J., Mihele, C., Netcheva, S., Li, S. M., and O'Brien, J.: Sulfur dioxide (SO_2) vertical column density measurements by Pandora spectrometer over the Canadian oil sands, Atmospheric Measurement Techniques, 9, 2961-2976, 10.5194/amt-9-2961-2016, 2016.

Fuentes, J. D., Wang, D., Neumann, H. H., Gillespie, T. J., DenHartog, G., and Dann, T. F.: Ambient biogenic hydrocarbons and isoprene emissions from a mixed deciduous forest, J. Atmos. Chem., 25, 67-95, 10.1007/BF00053286, 1996.

Furimsky, E.: Emissions of carbon dioxide from tar sands plants in Canada, Energy Fuels, 17, 1541-1548, 10.1021/ef0301102, 2003.

Geron, C., Rasmussen, R., Arnts, R. R., and Guenther, A.: A review and synthesis of monoterpene speciation from forests in the United States, Atmos. Environm., 34, 1761-1781, 10.1016/S1352-2310(99)00364-7, 2000.

Gordon, M., Li, S. M., Staebler, R., Darlington, A., Hayden, K., O'Brien, J., and Wolde, M.: Determining air pollutant emission rates based on mass balance using airborne measurement data over the Alberta oil sands operations, Atmos. Meas. Tech., 8, 3745-3765, 10.5194/amt-8-3745-2015, 2015.

Grimmer, G., Brune, H., Deutschwenzel, R., Dettbarn, G., Jacob, J., Naujack, K. W., Mohr, U., and Ernst, H.: Contribution of polycyclic aromatic-hydrocarbons and nitro-derivatives to the carcinogenic impact of diesel-engine exhaust condensate evaluated by implantation into the lungs of rats, Cancer Lett., 37, 173-180, 10.1016/0304-3835(87)90160-1, 1987.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5, 1471-1492, 10.5194/gmd-5-1471-2012, 2012.

Guo, H., Wang, T., and Louie, P. K. K.: Source apportionment of ambient non-methane hydrocarbons in Hong Kong: Application of a principal component analysis/absolute principal component scores (PCA/APCS) receptor model, Environ. Pollut., 129, 489-498, 10.1016/j.envpol.2003.11.006, 2004.

Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C.: Multivariate data analysis, in, 7th edition ed., Prentice-Hall, Upper Saddle River, NJ, pp. 108 -110, 1998.

Harrison, R. M., Smith, D. J. T., and Luhana, L.: Source apportionment of atmospheric polycyclic aromatic hydrocarbons collected from an urban location in Birmingham, UK, Environ. Sci. Technol., 30, 825-832, 10.1021/es950252d, 1996.

Helmig, D., Klinger, L. F., Guenther, A., Vierling, L., Geron, C., and Zimmerman, P.: Biogenic volatile organic compound emissions (BVOCs) I. Identifications from three continental sites in the U.S, Chemosphere, 38, 2163-2187, 10.1016/S0045-6535(98)00425-1, 1999.

Holowenko, F. M., MacKinnon, M. D., and Fedorak, P. M.: Methanogens and sulfate-reducing bacteria in oil sands fine tailings waste, Canadian Journal of Microbiology, 46, 927-937, 10.1139/cjm-46-10-927, 2000.

Howell, S. G., Clarke, A. D., Freitag, S., McNaughton, C. S., Kapustin, V., Brekovskikh, V., Jimenez, J. L., and Cubison, M. J.: An airborne assessment of atmospheric particulate emissions from the processing of Athabasca oil sands, Atmos. Chem. Phys., 14, 5073-5087, 10.5194/acp-14-5073-2014, 2014.

Hsu, Y. M., Bytnerowicz, A., Fenn, M. E., and Percy, K. E.: Atmospheric dry deposition of sulfur and nitrogen in the Athabasca Oil Sands Region, Alberta, Canada, Sci. Tot. Environm., 568, 285-295, 10.1016/j.scitotenv.2016.05.205, 2016.

Jautzy, J., Ahad, J. M. E., Gobeil, C., and Savard, M. M.: Century-Long Source Apportionment of PAHs in Athabasca Oil Sands Region Lakes Using Diagnostic Ratios and Compound-Specific Carbon Isotope Signatures, Environm. Sci. Technol., 47, 6155-6163, 10.1021/es400642e, 2013.

Jautzy, J. J., Ahad, J. M. E., Gobeil, C., Smirnoff, A., Barst, B. D., and Savard, M. M.: Isotopic Evidence for Oil Sands Petroleum Coke in the Peace-Athabasca Delta, Environm. Sci. Technol., 49, 12062-12070, 10.1021/acs.est.5b03232, 2015.

Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, A. S. H., Zhang, Q., Kroll, J. H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, A. C., Docherty, K. S., Ulbrich, I. M., Grieshop, A. P., Robinson, A. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V. A., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., E., Dunlea, J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, A. M., Williams, L. R., Wood, E. C.,

941 Middlebrook, A. M., Kolb, C. E., Baltensperger, U., and Worsnop, D. R.: Evolution of Organic Aerosols
942 in the Atmosphere, Science, 326, 1525-1529, 10.1126/science.1180353, 2009.
943 Johnson, M. R., Crosland, B. M., McEwen, J. D., Hager, D. B., Armitage, J. R., Karimi-Golpayegani, M., and
944 Picard, D. J.: Estimating fugitive methane emissions from oil sands mining using extractive core
945 samples, Atmos. Environm., 144, 111-123, 10.1016/j.atmosenv.2016.08.073, 2016.
946 Jolliffe, I. T., and Cadima, J.: Principal component analysis: a review and recent developments,
947 Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,
948 374, 10.1098/rsta.2015.0202, 2016.
949 Kaiser, H. F.: The varimax criterion for analytic rotation in factor-analysis, Psychometrika, 23, 187-200,
950 10.1007/bf02289233, 1958.
951 Kindzierski, W. B., and Ranganathan, H. K. S.: Indoor and outdoor SO2 in a community near oil sand
952 extraction and production facilities in northern Alberta, J. Environ. Eng. Sci., 5, S121-S129,
953 10.1139/s06-022, 2006.
954 Landis, M. S., Pancras, J. P., Graney, J. R., Stevens, R. K., Percy, K. E., and Krupa, S.: Chapter 18 - Receptor
955 Modeling of Epiphytic Lichens to Elucidate the Sources and Spatial Distribution of Inorganic Air
956 Pollution in the Athabasca Oil Sands Region, in: Developments in Environmental Science, edited by:
957 Percy, K. E., Elsevier, 427-467, 2012.
958 Landis, M. S., Pancras, J. P., Graney, J. R., White, E. M., Edgerton, E. S., Legge, A., and Percy, K. E.: Source
959 apportionment of ambient fine and coarse particulate matter at the Fort McKay community site, in
960 the Athabasca Oil Sands Region, Alberta, Canada, Sci. Tot. Environm., 584, 105-117,
961 10.1016/j.scitotenv.2017.01.110, 2017.
962 Lee, A. K. Y., Adam, M. G., Liggio, J., Li, S.-M., Li, K., Willis, M. D., Abbatt, J. P. D., Tokarek, T. W., Odame-
963 Ankrah, C. A., Huo, J. A., Osthoff, H. D., Strawbridge, K. B., and Brook, J. R.: A large contribution of
964 anthropogenic organonitrate to secondary organic aerosol in Alberta oil sands, in prep., 2018.

965 Li, S.-M., Leithead, A., Moussa, S. G., Liggio, J., Moran, M. D., Wang, D., Hayden, K., Darlington, A.,
966 Gordon, M., Staebler, R., Makar, P. A., Stroud, C. A., McLaren, R., Liu, P. S. K., Oâ€™Brien, J.,
967 Mittermeier, R. L., Zhang, J., Marson, G., Cober, S. G., Wolde, M., and Wentzell, J. J. B.: Differences
968 between measured and reported volatile organic compound emissions from oil sands facilities in
969 Alberta, Canada, Proceedings of the National Academy of Sciences, 114, E3756-E3765,
970 10.1073/pnas.1617862114, 2017.
971 Liggio, J., Li, S.-M., Hayden, K., Taha, Y. M., Stroud, C., Darlington, A., Drollette, B. D., Gordon, M., Lee, P.,
972 Liu, P., Leithead, A., Moussa, S. G., Wang, D., O'Brien, J., Mittermeier, R. L., Brook, J., Lu, G., Staebler,
973 R., Han, Y., Tokarek, T. W., Osthoff, H. D., Makar, P. A., Zhang, J., Plata, D., and Gentner, D. R.: Oil
974 Sands Operations as a Large Source of Secondary Organic Aerosols, Nature, 534, 91-94,
975 10.1038/nature17646, 2016.
976 Marey, H. S., Hashisho, Z., Fu, L., and Gille, J.: Spatial and temporal variation in CO over Alberta using
977 measurements from satellites, aircraft, and ground stations, Atmos. Chem. Phys., 15, 3893-3908,
978 10.5194/acp-15-3893-2015, 2015.
979 Markovic, M. Z., VandenBoer, T. C., and Murphy, J. G.: Characterization and optimization of an online
980 system for the simultaneous measurement of atmospheric water-soluble constituents in the gas and
981 particle phases, J. Environ. Monit., 14, 1872-1884, 10.1039/C2EM00004K, 2012.
982 Masliyah, J., Zhou, Z. J., Xu, Z. H., Czarnecki, J., and Hamza, H.: Understanding water-based bitumen
983 extraction from athabasca oil sands, Can. J. Chem. Eng., 82, 628-654, 10.1002/cjce.5450820403,
984 2004.
985 McLinden, C. A., Fioletov, V., Boersma, K. F., Krotkov, N., Sioris, C. E., Veefkind, J. P., and Yang, K.: Air
986 quality over the Canadian oil sands: A first assessment using satellite observations, Geophys. Res.
987 Lett., 39, 8, 10.1029/2011gl050273, 2012.

Miller, S. M., Matross, D. M., Andrews, A. E., Millet, D. B., Longo, M., Gottlieb, E. W., Hirsch, A. I., Gerbig, C., Lin, J. C., Daube, B. C., Hudman, R. C., Dias, P. L. S., Chow, V. Y., and Wofsy, S. C.: Sources of carbon monoxide and formaldehyde in North America determined from high-resolution atmospheric data, Atmos. Chem. Phys., 8, 7673-7696, 10.5194/acp-8-7673-2008, 2008.

Min, K. E., Pusede, S. E., Browne, E. C., LaFranchi, B. W., and Cohen, R. C.: Eddy covariance fluxes and vertical concentration gradient measurements of NO and NO2 over a ponderosa pine ecosystem: observational evidence for within-canopy chemical removal of NOx, Atmos. Chem. Phys., 14, 5495-5512, 10.5194/acp-14-5495-2014, 2014.

Nara, H., Tanimoto, H., Tohjima, Y., Mukai, H., Nojiri, Y., Katsumata, K., and Rella, C. W.: Effect of air composition (N_2, O_2, Ar, and H_2O) on CO_2 and CH_4 measurement by wavelength-scanned cavity ring-down spectroscopy: calibration and measurement strategy, Atmos. Meas. Tech., 5, 2689-2701, 10.5194/amt-5-2689-2012, 2012.

Nimana, B., Canter, C., and Kumar, A.: Energy consumption and greenhouse gas emissions in the recovery and extraction of crude bitumen from Canada's oil sands, Appl. Energy, 143, 189-199, 10.1016/j.apenergy.2015.01.024, 2015a.

Nimana, B., Canter, C., and Kumar, A.: Energy consumption and greenhouse gas emissions in upgrading and refining of Canada's oil sands products, Energy, 83, 65-79, 10.1016/j.energy.2015.01.085, 2015b.

Detailed facility information: http://www.ec.gc.ca/inrp-npri/donnees-data/index.cfm?do=facility_information&lang=En&opt_npri_id=0000002274&opt_report_year=2013, access: April 13, 2017, 2013.

Odame-Ankrah, C. A.: Improved detection instrument for nitrogen oxide species, Ph.D., Chemistry, University of Calgary, http://hdl.handle.net/11023/2006, 10.5072/PRISM/26475, Calgary, 2015.

1010 Onasch, T. B., Trimborn, A., Fortner, E. C., Jayne, J. T., Kok, G. L., Williams, L. R., Davidovits, P., and
1011 Worsnop, D. R.: Soot Particle Aerosol Mass Spectrometer: Development, Validation, and Initial
1012 Application, Aerosol Sci. Technol., 46, 804-817, 10.1080/02786826.2012.663948, 2012.
1013 Paatero, P., and Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal
1014 utilization of error estimates of data values, Environmetrics, 5, 111-126,
1015 doi:10.1002/env.3170050203, 1994.
1016 Parajulee, A., and Wania, F.: Evaluating officially reported polycyclic aromatic hydrocarbon emissions in
1017 the Athabasca oil sands region with a multimedia fate model, Proceedings of the National Academy
1018 of Sciences, 111, 3344-3349, 10.1073/pnas.1319780111, 2014.
1019 Paul, D., and Osthoff, H. D.: Absolute Measurements of Total Peroxy Nitrate Mixing Ratios by Thermal
1020 Dissociation Blue Diode Laser Cavity Ring-Down Spectroscopy, Anal. Chem., 82, 6695-6703,
1021 10.1021/ac101441z, 2010.
1022 Penner, T. J., and Foght, J. M.: Mature fine tailings from oil sands processing harbour diverse
1023 methanogenic communities, Canadian Journal of Microbiology, 56, 459-470, 10.1139/w10-029, 2010.
1024 Percy, K. E.: Ambient Air Quality and Linkage to Ecosystems in the Athabasca Oil Sands, Alberta, Geosci.
1025 Can., 40, 182-201, 10.12789/geocanj.2013.40.014, 2013.
1026 Peters, T. M., and Leith, D.: Concentration measurement and counting efficiency of the aerodynamic
1027 particle sizer 3321, J. Aerosol Sci., 34, 627-634, 10.1016/s0021-8502(03)00030-2, 2003.
1028 Quagraine, E. K., Headley, J. V., and Peterson, H. G.: Is biodegradation of bitumen a source of recalcitrant
1029 naphthenic acid mixtures in oil sands tailing pond waters?, J. Environ. Sci. Health Part A-Toxic/Hazard.
1030 Subst. Environ. Eng., 40, 671-684, 10.1081/ese-200046637, 2005.
1031 Rooney, R. C., Bayley, S. E., and Schindler, D. W.: Oil sands mining and reclamation cause massive loss of
1032 peatland and stored carbon, Proc. Natl. Acad. Sci. U.S.A., 109, 4933-4937, 10.1073/pnas.1117693108,
1033 2012.

RStudio Boston, M.: Integrated development environment for R, 2017.

Shahimin, M. F. M., and Siddique, T.: Sequential biodegradation of complex naphtha hydrocarbons under methanogenic conditions in two different oil sands tailings, Environ. Pollut., 221, 398-406, 10.1016/j.envpol.2016.12.002, 2017.

Siddique, T., Fedorak, P. M., and Foght, J. M.: Biodegradation of short-chain n-alkanes in oil sands tailings under methanogenic conditions, Environ. Sci. Technol., 40, 5459-5464, 10.1021/es060993m, 2006.

Siddique, T., Gupta, R., Fedorak, P. M., MacKinnon, M. D., and Foght, J. M.: A first approximation kinetic model to predict methane generation from an oil sands tailings settling basin, Chemosphere, 72, 1573-1580, 10.1016/j.chemosphere.2008.04.036, 2008.

Siddique, T., Penner, T., Semple, K., and Foght, J. M.: Anaerobic Biodegradation of Longer-Chain n-Alkanes Coupled to Methane Production in Oil Sands Tailings, Environm. Sci. Technol., 45, 5892-5899, 10.1021/es200649t, 2011.

Siddique, T., Penner, T., Klassen, J., Nesbo, C., and Foght, J. M.: Microbial Communities Involved in Methane Production from Hydrocarbons in Oil Sands Tailings, Environ. Sci. Technol., 46, 9802-9810, 10.1021/e53022024, 2012.

Simpson, I. J., Blake, N. J., Barletta, B., Diskin, G. S., Fuelberg, H. E., Gorham, K., Huey, L. G., Meinardi, S., Rowland, F. S., Vay, S. A., Weinheimer, A. J., Yang, M., and Blake, D. R.: Characterization of trace gases measured over Alberta oil sands mining operations: 76 speciated C2–C10 volatile organic compounds (VOCs), CO_2, CH_4, CO, NO, NO_2, NO_y, O_3 and SO_2, Atmos. Chem. Phys., 10, 11931-11954, 10.5194/acp-10-11931-2010, 2010.

Small, C. C., Cho, S., Hashisho, Z., and Ulrich, A. C.: Emissions from oil sands tailings ponds: Review of tailings pond parameters and emission estimates, Journal of Petroleum Science and Engineering, 127, 490-501, 10.1016/j.petrol.2014.11.020, 2015.

1058 Thurston, G. D., and Spengler, J. D.: A quantitative assessment of source contributions to inhalable
1059 particulate matter pollution in metropolitan Boston, Atmos. Environ., 19, 9-25, 10.1016/0004-
1060 6981(85)90132-5, 1985.
1061 Thurston, G. D., Ito, K., and Lall, R.: A source apportionment of U.S. fine particulate matter air pollution,
1062 Atmos. Environ., 45, 3924-3936, 10.1016/j.atmosenv.2011.04.070, 2011.
1063 Tokarek, T. W., Huo, J. A., Odame-Ankrah, C. A., Hammoud, D., Taha, Y. M., and Osthoff, H. D.: A gas
1064 chromatograph for quantification of peroxycarboxylic nitric anhydrides calibrated by thermal
1065 dissociation cavity ring-down spectroscopy, Atmos. Meas. Tech., 7, 3263-3283, 10.5194/amt-7-3263-
1066 2014, 2014.
1067 Tokarek, T. W., Brownsey, D. K., Jordan, N., Garner, N. M., Ye, C. Z., Assad, F. V., Peace, A., Schiller, C. L.,
1068 Mason, R. H., Vingarzan, R., and Osthoff, H. D.: Biogenic emissions and nocturnal ozone depletion
1069 events at the Amphitrite Point Observatory on Vancouver Island, Atmosphere-Ocean, 55, 121-132,
1070 10.1080/07055900.2017.1306687, 2017a.
1071 Tokarek, T. W., Brownsey, D. K., Jordan, N., Garner, N. M., Ye, C. Z., Assad, F. V., Peace, A., Schiller, C. L.,
1072 Mason, R. H., Vingarzan, R., and Osthoff, H. D.: Biogenic Emissions and Nocturnal Ozone Depletion
1073 Events at the Amphitrite Point Observatory on Vancouver Island, Atmosphere-Ocean, 1-12,
1074 10.1080/07055900.2017.1306687, 2017b.
1075 Wang, S. C., and Flagan, R. C.: Scanning electrical mobility spectrometer, Aerosol Sci. Technol., 13, 230-
1076 240, 10.1080/02786829008959441, 1990.
1077 Wang, X. L., Chow, J. C., Kohl, S. D., Percy, K. E., Legge, A. H., and Watson, J. G.: Characterization of
1078 PM2.5 and PM10 fugitive dust source profiles in the Athabasca Oil Sands Region, J. Air Waste Manag.
1079 Assoc., 65, 1421-1433, 10.1080/10962247.2015.1100693, 2015.

1080 Wang, X. L., Chow, J. C., Kohl, S. D., Percy, K. E., Legge, A. H., and Watson, J. G.: Real-world emission
1081 factors for Caterpillar 797B heavy haulers during mining operations, Particuology, 28, 22-30,
1082 10.1016/j.partic.2015.07.001, 2016.
1083 Warren, L. A., Kendra, K. E., Brady, A. L., and Slater, G. F.: Sulfur Biogeochemistry of an Oil Sands
1084 Composite Tailings Deposit, Front. Microbiol., 6, 14, 10.3389/fmicb.2015.01533, 2016.
1085 Watson, J., Chow, J., Wang, X., Zielinska, B., Kohl, S., and Gronstal, S.: Characterization of real-world
1086 emissions from nonroad mining trucks in the Athabasca Oil Sands Region during September, 2009,
1087 2013.
1088 WBEA: WBEA annual report 2013, Wood Buffalo Environmental Association, 2013.
1089 Whaley, C. H., Makar, P. A., Shephard, M. W., Zhang, L., Zhang, J., Zheng, Q., Akingunola, A., Wentworth,
1090 G. R., Murphy, J. G., Kharol, S. K., and Cady-Pereira, K. E.: Contributions of natural and anthropogenic
1091 sources to ambient ammonia in the Athabasca Oil Sands and north-western Canada, Atmos. Chem.
1092 Phys., 18, 2011-2034, 10.5194/acp-18-2011-2018, 2018.
1093 Williams, B. J., Goldstein, A. H., Kreisberg, N. M., and Hering, S. V.: An in-situ instrument for speciated
1094 organic composition of atmospheric aerosols: Thermal Desorption Aerosol GC/MS-FID (TAG), Aerosol
1095 Sci. Technol., 40, 627-638, 10.1080/02786820600754631, 2006.
1096 Wilson, N. K., Barbour, R. K., Chuang, J. C., and Mukund, R.: Evaluation of a real-time monitor for fine
1097 particle-bound PAH in air, Polycycl. Aromat. Compd., 5, 167-174, 10.1080/10406639408015168,
1098 1994.
1099 Yang, C., Wang, Z., Yang, Z., Hollebone, B., Brown, C. E., Landriault, M., and Fieldhouse, B.: Chemical
1100 Fingerprints of Alberta Oil Sands and Related Petroleum Products, Environmental Forensics, 12, 173-
1101 188, 10.1080/15275922.2011.574312, 2011.

Yeh, S., Jordaan, S. M., Brandt, A. R., Turetsky, M. R., Spatari, S., and Keith, D. W.: Land Use Greenhouse Gas Emissions from Conventional Oil Production and Oil Sands, Environm. Sci. Technol., 44, 8766-8772, 10.1021/es1013278, 2010.

Zhang, Y., Wang, Y., Chen, G., Smeltzer, C., Crawford, J., Olson, J., Szykman, J., Weinheimer, A. J., Knapp, D. J., Montzka, D. D., Wisthaler, A., Mikoviny, T., Fried, A., and Diskin, G.: Large vertical gradient of reactive nitrogen oxides in the boundary layer: Modeling analysis of DISCOVER-AQ 2011 observations, J. Geophys. Res.-Atmos., 121, 1922-1934, 10.1002/2015jd024203, 2016.