

General Comments: This manuscript describes an effort to apportion PM_{2.5} elements to different sources in the Athabasca Oil Sands Region. Two types of data were used: 24-hr filter samples from three near-mine monitoring stations over a two-year period from Dec. 2010 to Nov. 2012, and hourly elemental concentration from a portable elemental analyzer and 24-hr filters during an intensive study (Aug. 2013). Source apportionment was accomplished by Positive Matrix Factorization (PMF). Correlations with gases, time series, conditional probability function, and trajectory analyses were used to support the rationality the PMF factors. Elevated concentrations of elements have been observed in snow, river water, and lichen samples in the oil sands region (Bari et al., 2014; Graney et al., 2012; Huang et al., 2016; Kelly et al., 2010; Landis et al., 2012). Understanding the concentrations and sources of elements in PM_{2.5} in the region is important for assessing their environmental impacts and implementing control strategies. Therefore, the topic described in this manuscript is of great interest to the environmental science community, particularly to stakeholders in the oil sands region.

I have three major concerns on the data and results:

- 1. The source apportionment only used elemental composition. As can be seen from Figs. 2 and 3, many elements are present in multiple PMF factors, and most PMF factors are lacking specific elemental source markers. This collinearity make source identification and contribution apportionment less specific and more uncertain. The similarity in the temporal trend (e.g., Fig. S10) of different PMF factors is an indication of the difficulty in resolving these similar factors. The authors indicate that follow up analyses will include other chemical components (Page 20). I would prefer a source apportionment paper using all available chemical species (e.g., ion, carbon, isotopes, organic speciation, etc.) to create more specific and confident source apportionment results.*

We thank the reviewer for their helpful comments. The difficulty in adding other species lies in the variance of the sampling strategies and techniques between and within the two campaigns. In order to include as much data as possible, which increases the accuracy of the PMF analysis as it provides the program with more patterns to analyse, the filter data from all three sites and both campaign periods were included in the PMF run. However, only elemental composition data were available for these filters thus only trace element data was used for the high time-resolution study as well. As a main goal of the paper was to compare the source apportionment identified through a long-term, low time-resolution study and a short-term, high time-resolution study, the species analysed were kept as similar as possible. Follow-up research examining far more species would likely have to focus on the intensive period, when far more species were studied.

- 2. The Xact malfunctioned during the intensive study. While the authors tried their best to retain the Xact data as much as possible, the data quality is still in question for the following reasons:*
 - a) There was a drop in the internal measurement values for Pd, Pb, Cr, and Cd between 8/25 and 9/2/2013. It was assumed that the regression Eqns in Fig. S1 were applicable to all metals within its energy level. However, Fig. S2 shows that such correction caused the slope of Xact sulfur vs. AMS sulfur to change from 2.75 to 3.57, a significant 30% difference. The authors speculate that “could have been due to an increase in sulfate size distribution” but did not provide such evidence. Did the dichotomous sampler coarse channel show higher sulfate after 8/25?*

As stated in the Supplementary, estimated AMS PM1 sulfur and AIM-IC PM2.5 sulfur data were used to evaluate the instrument performance and the correction of Xact internal signals. We found more than 60% differences between AMS and AIM-IC sulfur data. Different cut-diameters and collection efficiencies of these two instruments could be a main reason for the difference. Thus we speculated that the change of the slope between AMS sulfur and Xact sulfur also might be related to a poor collection efficiency of the AMS as sulfate size distribution increases between near PM1 and PM2.5. The dichot sampler provides fine (PM2.5) and coarse mode (PM2.5-10) speciation data. However, the coarse mode sulfate data did not provide any stronger evidence of a shift in the sulfate size distribution between near PM1 and PM2.5.

Another possible explanation of the slope change before and after the Xact correction was identified and explored: the range of sulfur concentrations was quite different between the two periods. Before August 25, AMS sulfur increased to $\sim 3500 \text{ ng/m}^3$, which was almost one order of magnitude higher than the sulfate range after August 25. In order to examine the effect of sulfur concentrations, AMS sulfur data before the 25th were regrouped into two ranges, $< 300 \text{ ng/m}^3$ and $> 300 \text{ ng/m}^3$. As shown in Figure R1a, the slope for the lower concentrations range was closer to the slope after Aug. 25 (3.26 vs. 3.57). At the high range the agreement between Xact and AMS sulfur was better (Figure R1b), which was close to the slope for the entire concentrations ranges (2.71 vs. 2.75). The reason of the concentration dependency is unclear, it could be due to the change of the Sulfate/Sulfur ratio varying by sources. However, the agreement of the Xact with the AMS is not the key point here as the agreement between these two instruments is still not well established. Rather, it is the similarity in the agreement of the Xact before and after the correction that is the point. Here the similar slope ($\sim 10\%$ difference) for S before and after the Xact correction clearly supports the Xact correction is valid. Furthermore, Xact daily sulfur averages were compared to the filter XRF data before and after the correction. As shown in Figure R2, the corrected Xact S data lie in the range of the 95% prediction interval, indicating they are in the uncertainty levels of the analytical methods. In the revised supplement, Figure S2 was changed to display the slope at the lower range and we revised the statement on the comparison of the Xact data with SP-AMS.

“Overall, there was a large change in the slope of the line from 2.75 before August 25 to 3.57. However, the corrected slope was comparable to the slope in a similar concentration range (Figure S2b) before August 25 (3.26 vs. 3.57), suggesting that the Xact data correction was reasonable.”

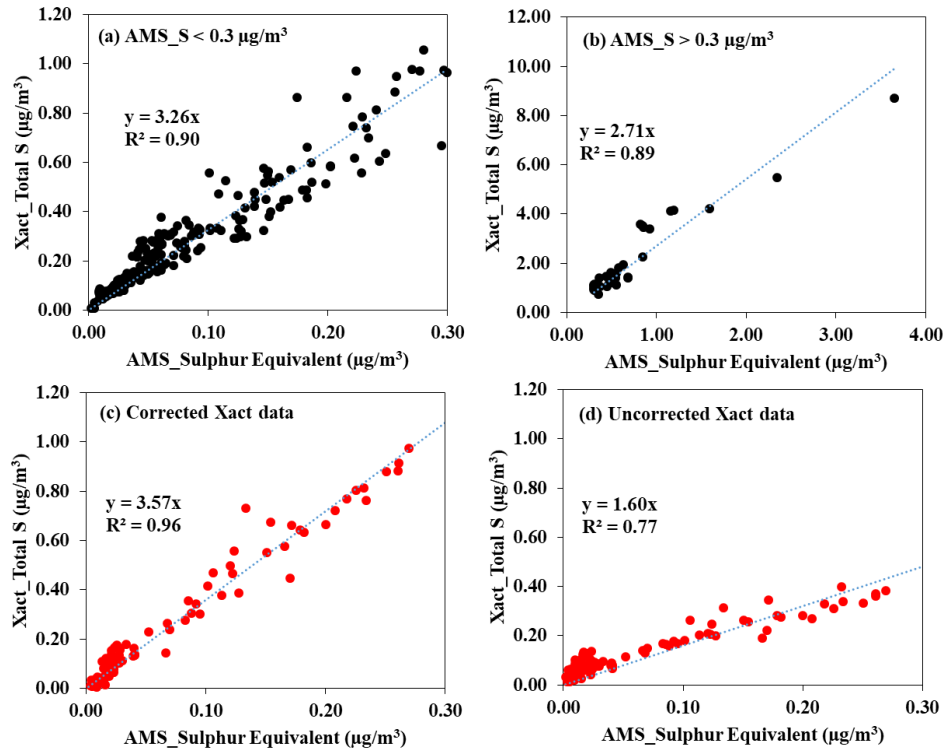


Figure R1. Comparison of SP-AMS sulphur equivalent to Xact S before Aug. 25, 2013 (a, b) and after Aug. 25, 2013 (c, d). All SP-AMS sulphate concentrations were divided by three to determine equivalent sulphur mass concentrations.

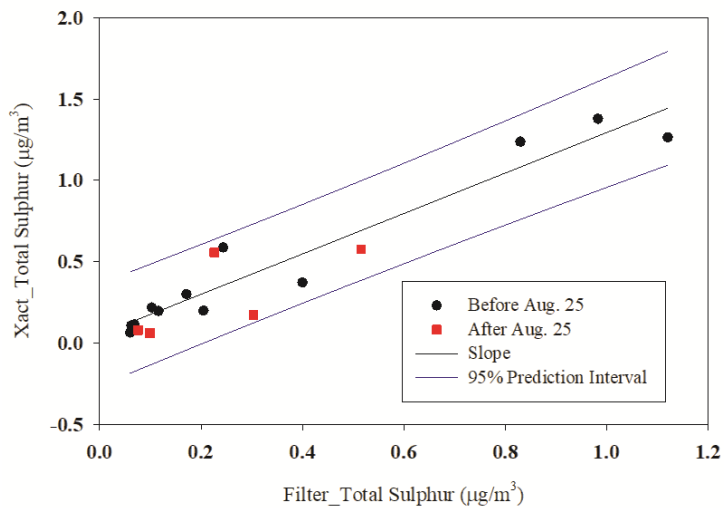


Figure R2. Comparison of daily integrated filter sulphur to Xact sulphur before and after August 25, 2013.

b) Figs. S3d and S3e show that sulfur from filter and AIM-IC measurements were comparable (indicating non-sulfate sulfur may not be significant), but Xact was ~40% higher. Such discrepancies also exist for several other high concentration elements (Fig. S3d), causing concerns about the Xact data quality. c) The Xact and filter data had various regression slopes according to concentration ranges (0.77 to 1.43; Figs. S3b-3d). From the assumption of using Fig. S1 to correct elements in the same energy level, I would expect the slopes to be similar for those within the same energy level, but not for those in the same concentration range. Both K and S belong to Energy Level 1 (Table S1), but the slopes are 0.77 and 1.43, respectively (Fig. S3c and 3d). This data is puzzling. Any explanations?

We agree that some hourly element concentrations measured by Xact were higher than the 23-hr integrated filter measurements. This difference is puzzling and thus we probed it in as many ways as we could and documented this in the supplementary. There could be many reasons for the discrepancy due to possible measurement biases of both methods (i.e., sampling time, inlet location, internal calibration factor). A further study for the comprehensive comparison of Xact and filter measurements is required to resolve confounding factors. In this study, the correlations between Xact and Filter for high concentration elements were very good ($r^2 > 0.85$), which was higher than the correlation between AIM-IC and Filter sulfur ($r^2 = 0.68$). We believe that this possible inaccuracy in the Xact data did not detract from the PMF analysis and the majority of elements included agreed well based on the filter comparison result.

The upscaling correction using Figure S1 equally affected all elements in the same energy level as the reviewer pointed out. However, each element in the Xact data has a specific internal calibration factor determined by standard element checks. Thus, the slope could be changed by elements and instruments.

- 3. While trying to explain the results, there are quite a few hand waving speculations, which shows insufficient understanding of the potential pollution sources in the oil sands region. For example, the high sulfur in the stack flue gas was assumed to be caused by #4 boiler fuel instead of the sulfur origin from upgrading process. Several other speculations are commented later. More specific and detailed comments are given below.*

We have tightened up the language. In particular, we agree that Boiler Fuel #4 was added as an example of a potential source. This has been edited for clarity and accuracy as follows:

“The profile is suggestive of a mixed-combustion source (Lee et al., 2000; Van et al., 2008), such as coke, or the process gasses, which are comprised of effluent from the sulphur recovery units (Wang et al., 2012).”

Specific Comments:

- 1. Title: “Source of Particulate Matter. . .” This is paper is about PM2.5 elements. Modify the title to be accurate.*

Since we didn't use major PM2.5 chemical speciation in PMF and multi linear regression of the source contributions against the total PM2.5 mass, the source apportionment and contributions are just related to the total metal mass, which are trace amounts within PM2.5. As such, the use of “Particulate Matter” encapsulates the purpose of the study for the title, as metals comprise such a small amount.

- 2. The authors generally referred the elements as metal (e.g., Page 3 last paragraph), which is not accurate because there are non-metal elements (e.g., S and Si). I would suggest to call them element to be accurate.***

Agreed, and done

- 3. There is a gap between the long-term filters and intensive study (Dec. 2012-July 2013). It would be logic to have the long-term sampling overlap the intensive study.***

It is agreed that it would be have been ideal to have a complete time series between the intensive and long-term campaign. However, at time of research and writing, these data were not available.

- 4. Page 3, 2nd paragraph. The review of past PM_{2.5} research in the oil sands is missing a large body of studies organized by the Wood Buffalo Environmental Association, such as air quality trend, emission source characterization, and elemental composition in lichen (Landis et al., 2012; Landis et al., 2017; Percy et al., 2012; Wang et al., 2012; 2015a; 2015b; 2016; Watson et al., 2012). These are very relevant to this manuscript.***

It is agreed that the second paragraph of page three does not include a full account of all studies done in the area, as it is simply an introduction to some of the work done in the area, not a full synopsis. Some of these papers had been added to other parts of the paper, some have been added as per further comments. In order to provide a more full reference to the amount of research in the area, the section have been altered as follows:

“Past research on PM_{2.5} within the Athabasca Region has included overall and comparative emission and air quality analyses (Kindzierski and Bari, 2011; Kindzierski and Bari, 2012; Wang et al., 2012; Percy et al., 2012; Howell et al., 2014; Wang et al., 2015; Landis et al., 2017). Further studies have developed into modelling the emission sources through both computer-based (Cho et al., 2012) and measurement-based methods (Landis et al., 2012)”

- 5. Page 4 Line 3. “. . . V and Ni are often indicative of oil combustion. . .” While this statement might be true in general, it may not be accurate in the oil sands region because bitumen is enriched with V and Ni (Shotyk et al., 2016). Attributing V and Ni to oil combustion will not reflect the bitumen-rich environment of the oil sands region. Citation of marker species should consider the local sources and chemical nature in the study area. This also raises question about the reason for attributing the factor “Upgrader Emission to oil combustion due the higher percentage of V and Ni. This may not be due to oil combustion, but due to bitumen processing. Therefore, it is important to find out if oil is largely used in combustion in the mining/upgrading facilities.***

This paragraph is intended as an example of how metals can and have been grouped together for source apportionment in the past, not as evidence of results. As bitumen is an oil derivative, this similarity would not be unexpected. Paragraph reworded to elucidate:

“However, compositional analysis of PM_{2.5} has helped elucidate sources and processes that contribute to PM_{2.5} mass concentrations in other regions. The element species in PM_{2.5} are of particular importance because they can be source-specific and are typically preserved in the aerosol phase during transport. For

example, V and Ni are often indicative of oil combustion (Becagli et al., 2012), as well as oil derivatives (Shotyk et al., 2016),”

6. ***Page 4 Line 3-4. “Al, K, Mg, and Cr are indicative of road dust. . .” This sentence was cited from a Mediterranean study, and the statement contradicts the results. For example, Al is abundant in almost all dust sources in the oil sands region (Wang et al., 2015a), K is a marker for biomass burning (Fig. 3) and Cr is not abundant in the haul road dust factor (Fig. 2).***

It is agreed that this is a source from a Mediterranean study. The purpose of this paragraph is to introduce how metals have been grouped together in the past for source apportionment, this has been edited for clarity as follows:

“while Al, Mg, and Cr, when grouped together have been indicative of dust, in the past specifically that associated with transportation (Amato et al., 2014)”

7. ***In the quality assurance and quality control sections (maybe in corresponding supplemental materials), I suggest adding: a) A comparison of overlapping elements that were measured by both ICP-MS and XRF. b) Describe in more details about how PMF factors were optimized and their uncertainties were estimated (e.g., Landis et al., 2017; Reff et al., 2007). Each PFM factor should show uncertainty based on bootstrapping. The similarity of profiles may be examined with Chi-square in addition to correlation coefficients.***

Scatter plots for Al, Ti, V, Mn, Fe, and Zn measured by ICP-MS and ED-XRF have been added in the supplementary. The correlation coefficients (r^2) ranged from 0.81 to 0.96 with good agreements.

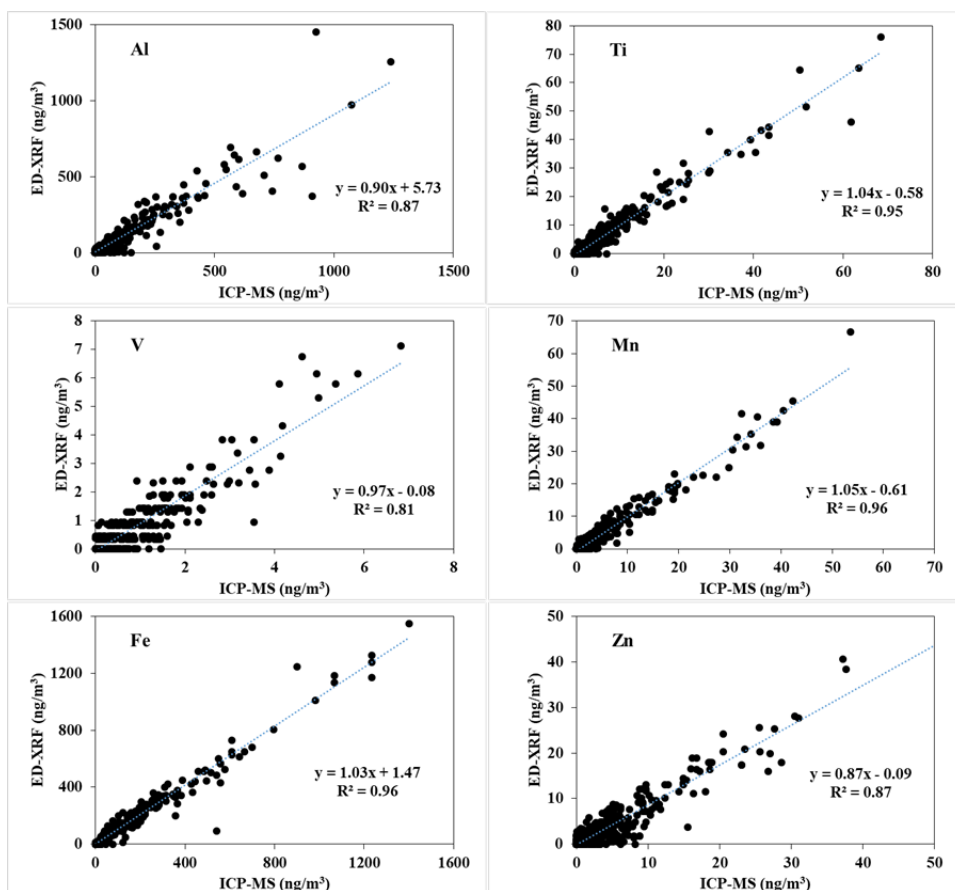


Figure R3. Comparison of ED-XRF and ICP-MS measurements at AMS 5, 11, and 13 from Dec. 16, 2010 to Nov. 29, 2012. Nine elements with more than 80% of data below the minimum detection limit were excluded from the comparison.

In the revised manuscript, we re-examined the PMF solution using EPAPMF 5 and calculated comprehensive error estimates of the PMF solution. Possible solutions were compared to determine the best solution based on their stabilities and uncertainties. Detailed information about the solution evaluation has been added in supplementary as follows:

“In order to estimate uncertainties and evaluate the robustness and rotational ambiguity of PMF modeling results, the solutions were evaluated using the error estimation methods of EPA PMF 5; bootstrap analysis (BS), displacement analysis (DISP), and bootstrap enhanced by displacement (BS-DISP). Bootstrap analysis (BS) was performed to quantify the uncertainty of a PMF-resolved solution. In addition, 100 bootstrap iterations were conducted to obtain the percentage of factors assigned to each base case factor (i.e. bootstrap mapping) and determine unstable factors in the PMF solutions. With the displacement analysis (DISP), each element in source profile is displaced from its fitted value in a PMF solution to estimate the uncertainties for each element in each factor profile. Based on the result of the displacement analysis of a PMF solution, the rotational ambiguity of PMF solutions was assessed (i.e. number of swaps at the lowest predetermined Q levels). BS-DISP, a combination of BS and DISP, estimates the error associated with both random and rotational ambiguity (Paatero et al., 2014; Brown et

al., 2015). A discussion of diagnostic results of the error estimation methods for possible PMF solutions is provided in Supplement S.3.”

“The bootstrap (BS) analysis was conducted to evaluate the uncertainties (i.e. random error in data values) of the source profiles and the reproducibility of factors in every bootstrap (Paatero et al., 2014; Brown et al., 2015). In the BS analysis, the BS factors are compared with the base run factors and then mapped to the base factor if the correlation is higher than a threshold ($r^2=0.8$ in this study). Tables S4 and S5 summarize the diagnostics of the error estimation for three PMF solutions (i.e. 4-, 5-, 6-factor solution) for the intensive Xact data and the long-term filter data, respectively. In the 5-factor solution of both Xact and filter data, we found most bootstrap factors were well assigned to base factors in >96% of every bootstrap. Overall reproducibility (i.e. average BS mapping percentages) for each factor in the 5-factor solution was higher than other solutions, suggesting the 5-factor solution was very reproducible and the optimal solution. The displacement (DISP) analysis was conducted to evaluate rotational ambiguity in the PMF solution as well. Multiple solutions may be generated with the same value of the object function Q due to rotational ambiguity. In DISP, each fitted element (only good species) in a source profile is displaced in turn from its fitted value until Q increases by a predetermined maximum change in Q. An uncertainty estimate for each element in each factor profile is thereby yielded and factor swaps may occur if factors change too much. A comprehensive error estimate method, bootstrap enhanced by displacement (BS-DISP) combine the strengths of BS and DISP, which evaluate both the robustness to data errors and rotational uncertainty. Overall, no change in DISP Q (%dQ) was found for the 5-factor solutions. Furthermore, no swapped factor was found in DISP BS-DISP runs, indicating the 5-factor solution was a global minimum and well-defined PMF solution.”

“The source profiles of the 4- and 6-factor solution for the intensive Xact PMF analysis are shown in Figures S5 and S6. In the 4-factor solution, the Soil and Haul Road Dust factors can be combined, but the reproducibility of the solution was poor and there were factor swaps in the BS-DISP runs. In the 6-factor solution, the Soil factor from the 5-factor solution split into two similar soil factors which have poor BS mapping reproducibility and very high factor swaps in the BS-DISP analysis. Another solution could be possible in the 6-factor solution which was characterized by additional resolution of the Mixed sources in the 5-factor solution. As shown in Figure S6b, two Mixed sources were characterized by high loadings of Cu (Mixed Sources I) and Br and Se (Mixed Sources II). Thus, due to the robustness of the solution and the physically meaningfulness of the factor profiles, the 5-factor solution was clearly acceptable for the intensive campaign.”

“In the long-term campaign the Haul Road Dust and Soil factors in the 5-factor solution can be combined into one factor in the 4-factor solution (Figure S7a). However, there was an alternative solution including a combined factor of Mixed and Upgrader Emissions (Figure S7b). Due to the instability, there were factor swaps in BS-DISP and the reproducibility of the 4-factor solution was poor. With 6 factors, an additional Soil factor characterized by high loadings of rare earth elements and vanadium was found. However, this second Soil factor is only found 56% of the BS resamples and 85% of the BS-DISP runs were accepted with high factor swaps. In the 7-factor solution, additional factor containing high Pb and Br can be isolated from the Upgrader Emissions factor, but it’s stability was very poor and there was no reasonable source for Pb and Br only. The BS resamples and BS-DISP runs of the 5-factor solution was better than the 4- and 6-factor solutions for the long-term filter data (Table S5). These results indicate that the 4- and 6-factor solution are much less certain than the 5-factor solution. As a result, the 5-factor solution was chosen as the most reasonable and stable solution for the filter data.”

“In order to determine the relative weights of the different factors, a multiple linear regression of the time series of each factor for both campaigns was run against both the summed metal concentrations as well as

the PM_{2.5} concentrations (obtained from WBEA). As trace metals only account for a small percentage of the overall PM_{2.5} mass, the results of the PM_{2.5} regression proved to be a poor fit both statistically ($r^2 < 0.8$) and physically, as it resulted in negative relative weights, to the metal speciation factor solutions. Because of this, the total metals concentration was used to determine the relative weights of the different factors, which resulted in a much better fit ($r^2 > 0.99$).”

The chi-square value is the weighted sum of squares of the differences between the measured (i.e., Landis et al. 2012) and PMF-modelled element concentrations. However, no Chi-square values are available in this study since the Landis’s source profiles were reported as relative contributions (i.e., $\mu\text{g/g PM}$). In the revised manuscript, uncentered correlation coefficient has been used as an unbiased metric to evaluate the level of similarity between the profiles of sources. This metric is particularly useful as it takes into account the similarity of minor peaks.

Table S4. Summary of error estimation diagnostics for intensive Xact data.

	4-Factor Solution	5-Factor Solution	6-Factor Solution
Robust Mode	Yes		
Seed Value	Random		
# of Bootstraps in BS	100		
R ² in BS	0.8		
DISP active species	Si, S, K, Ca, Ti, V, Mn, Fe		
BS-DISP active species	S, K, Ti, V, Fe		
Factors with BS mapping < 100%	Upgrader Emissions II(51%), Mixed (99%)	Upgrader Emissions II (98%)	Soil (53%), Soil II (93%)
DISP %dQ	0	0	0
DISP # of swaps	0	0	0
BS-DISP % of Cases Accepted	95	97	27
BS-DISP # of swaps	4	0	198

Table S5. Summary of error estimation diagnostics for long-term combined filter data.

	4-Factor Solution	5-Factor Solution	6-Factor Solution
Robust Mode	Yes		
Seed Value	Random		
# of Bootstraps in BS	100		
R ² in BS	0.8		
DISP active species	Si, S, K, Ca, Ti, Fe, Cu, Sr, Al, Cd, Ce, La, Pr, Nd, Sm, Gd, Pb, U		
BS-DISP active species	Si, S, K, Ca, Fe, Cu, La		
Factors with BS mapping < 100%	Mixed (29%), Biomass Burning (98%)	Mixed (96%), Soil (99%)	Soil (56%), Haul Road Dust (97%), Mixed (99%)
DISP %dQ	9.3E-5	0	1.8E-5
DISP # of swaps	0	0	0
BS-DISP % of Cases Accepted	87	97	85
BS-DISP # of swaps	10	0	12

8. *The calibration with metals standards showed good accuracy (Table S1) after the Xact returned to the laboratory after field campaign. Was there anything done to fix the Pd signal drop during the field campaign? If so, what was cause of the problem and what was the fix? If not, did the instrument somehow fix by itself sometime between 8/25 and laboratory tests?*

Many parts including high voltage power supply (HVPS), x-ray tube, and filament cables had to be checked. We finally found a loose wire connection from the National Instruments AD converter and suspected that the position of the Pd rod changed slightly because of vibration during shipping. After the field campaign the HVPS had to be replaced. It appears that shipping can be tough on this instrument.

9. *Page 11 Line 17-19. Please verify if No. 4 boiler fuels were used in upgrading facilities. From an earlier publication on oil sands stack emission (Wang et al., 2012), the boilers were fired with natural gas, process gas, and/or coke. Process gases (e.g., effluent from the sulfur recovery units) and coke, instead of No. 4 fuels, are likely the main sources of sulfur in that area. Careful survey and understanding of the industrial operations are required to reduce speculations.*

We agree and thank the reviewer. The comparison was simply to show an example of a mixed source combustion fuel, this section has been edited for accuracy and clarity as follows:

“The profile is suggestive of a mixed-combustion source (Lee et al., 2000; Van et al., 2008), such as coke, or the process gasses, which are comprised of effluent from the sulphur recovery units (Wang et al., 2012).”

10. *Page 12. “It is speculated that this factor may have arisen from short term changes in fuel, such as a switch to oil combustion for heat/energy or the burning of coke. . .” This speculation of the source of “Upgrader Emission II” is not supported by hard evidence.*

We agree and had worked this text accordingly so as to emphasise that this was speculation and that this was raising possibilities rather than making any definitive claim

For clarity, this has been changed to: “It is speculated that this factor may have been due to two different stacks from within the upgrading process/facility. A less likely, but more fortuitous, possibility is that the Upgrader Emissions II factor was due to a short-term change in upgrader fuel that occurred only during the intensive measurement campaign and not during the long-term campaign”

11. *Sections 3.2.3 and 3.2.4. I am not convinced that the soil factor and haul road factor can be reliably separated based on elemental composition. The profiles Figs. 2-3 are very similar. The overburden itself is stable and is not a large source of fugitive dust (Wang et al., 2015b). Dust is emitted from dikes built by overburden when heavy haulers are travelling on them. Then the emissions are similar to the haul road. Emissions of haul road dust occur when heavy haulers are moving on the unpaved haul road and when wind speed is high. Since diesel-fueled heavy haulers are the largest emitters of NO_x, I would expect the haul road dust to better correlate with NO_x than the “Soil” factor. Large dust plumes are often observed from tailings beaches under high wind conditions. Also there are several unpaved roads around the sampling sites, which are large dust sources. As shown in Figure 6 and its related discussion, both Soil and*

Haul Road factors increased during the day with very similar temporal patterns and are not driven by wind speed. These indicate that these two factors are likely driven by road dust. A single “dust” factor instead two may be more appropriate.

While there are overlaps in the elements within the haul road dust and soil factors, there are enough differences to separate them as two separate sources. In fact, the source profiles for the Soil and Haul Road Dust were only weakly correlated (uncentered $r=0.129$), as is stated in the paper. Figure 6 simply shows the averaged-daily trends, not the overall hour-by-hour trends that PMF used, which are seen in Figure S9. These overall temporal trends do show clear differences in the time series of the two factors. Further, the speculated sources of the Haul Road Dust and Soil factors are due to on and off-road transportation. As the overburden dump is comprised of the overburden (top soil) in the area, it would be expected that any off-road transportation, whether on the dikes made of the overburden dump, or undisturbed overburden (top soil), would kick up dust comprised of the same material as the overburden dump. However, dust kicked up by vehicles on a road would be expected to have a slightly different elemental composition than that of the dust kicked up by the off-road vehicles, thus the two factors. Further the two factors, particularly in the intensive campaign, were very stable. These two factors can be combined in the 4-factor solution, but the reproducibility of the solution was poor and there were more factor swaps (please refer to the uncertainty analysis as shown earlier). This stability, along with lack of correlation, justifies the separation into two factors.

12. Table 2. Have the correlations between PMF factors and CO been looked at? CO is an indicator of biomass burning and/or vehicle emissions and may offer added evidence to the PMF factors.

While this would be a useful comparison, of the species available to us at the sites, as well as during the two campaigns, CO was not a gas that available to us.

13. Page 20 Line 16-17. It is not accurate to categorize the Soil and Haul Rod Dust factors as “two with the transportation of the bitumen-rich oil”. This excludes on-road vehicle and other dust sources.

We agree that “the transportation of bitumen-rich oil” does not entirely encapsulate all of the transportation. To account for this, the sentence has been adjusted to read as follows:

“two with on and off-road transportation”

14. Figure 6 and S10. It might be useful to examine the diurnal variation each factor.

While the Soil and Haul Road Dust Factors showed what appeared to be regular, diurnal trends (See Figure S9), the other three factors did not exhibit any patterns that resemble a daily trend. Taking daily averages of such episodic occurrences would not aid in the identification of the factor.

Technical Details

1. Figs 1, 7, S14 etc. All maps should have a scale bar to infer distances.

Agreed, and they have been added to the Figures.

- 2. Figs. 2, 3, S5, and S6. Add a note to indicate if the bar and symbol refers to the left or right hand side of the Y-axis.**

Agreed. For clarity, the following has been added to the figure descriptions:

“Factor concentrations depicted as bars, percentages depicted as circles.”

- 3. Page 6 Line 4. I suggest changing “dirt” to “oil sands”.**

While this wording was constructed in a way to describe what “oil sand” is, the paper has been adjusted as follows:

“Bitumen-rich dirt” has been converted to “Oil Sands”

- 4. Page 6 Lines 7-10. Each oil sands mining facility has multiple stacks that are connected to different upgrading processes. For example, Syncrude has a main stack, a FGD stack, and several smaller stacks. Furthermore, most stacks are equipped with pollution control devices. Some particles, e.g., ammonium sulfate, are formed in the FGD process designed to remove SO₂ (Wang et al., 2012). Rewrite this sentence to accurately reflect this information.**

It is agreed that while there are numerous stacks, our comparison was to the measurements of the main stack taken by (Landis et al., 2012). However, in order to better reflect the totality of what happens, the wording has been changed to the following:

“Once at the upgrading facility the bitumen is separated from the slurry in large settling vessels, after which it is upgraded into different hydrocarbon streams using steam, vacuum distillation, fluid cokers, and hydrocrackers, these processes produce aerosol particles which are emitted to the air through both the main stack (Landis et al., 2012), as well as numerous other secondary stacks. Within these stacks, some particles are directly emitted from the upgrading processes, while others are created by the pollution control devices installed within the stacks (Wang et al., 2012).”

- 5. Page 6 Lines 10-13. Evaporative emissions from tailings ponds may be an importance source of VOCs and secondary particles, but may not be an important source for primary particles. Instead, windblown dust from tailings beaches and dikes is a significant particle source (Wang et al., 2015b).**

It is agreed that evaporative emissions is likely not a significant source of primary particles. As such, the sentence has been altered to read:

“Other known sources of particles are: the large fleets of on and off-road vehicles, dust re-suspended by mining activities, windblown dust from the tailings ponds and dikes (Wang et al., 2015), and dust re-suspended from open petroleum coke piles.”

6. Fig. S3b-S3d showed 9 metals. Where are other metals?

In order to do a comparison, hourly measurements were turned into daily averages. If in the course of the day over 50% of the hourly measurements were below detection limit or if the average itself was below detection limit, this daily average data point was removed due to increase unreliability associated with below detection limit measurements. In some cases this removed all daily average measurements for a metal.

7. Write “Eqn 1” in supplemental materials in an equation form.

Equations have been denoted as Arabic numerals in parentheses as suggested by the journal guideline.

8. There are two Table S4

Agreed, in sections S.5 and S.6, the table numbers have been altered to be “Table S6 and Table S7”

9. Page 10 Section 3.1 and Section 3.5. I don’t think it is fair to compare the 90th percentile in the oil sands to the average city values. If such comparison should be done, the 90th percentile of city values should be used. Also the three monitoring stations are considered as near-source monitoring due to their close distance to mining facilities. It is not surprising that their elemental concentrations are higher than many cities, considering the abundance of dust in the region (Shotyk et al., 2016).

We agree that the 90th percentile does provide a different comparison than a straight mean comparison table. The main reason for including this table was to provide some context for the levels of PM2.5 metals relative to other sites in Canada where this type of data is available, which all happen to be cities. The second reason is to identify which metals are likely related to activities in the region. To do this we looked at the 90th percentile values which are more likely representative of intermittent plumes, than day to day averages.

To increase the clarity of these points, we made the following changes:

Within Section 3.1:

In order to clarify that the only reason that the oil sands concentrations were compared to the cities alone was due to lack of Canadian trace element data for remote sites, “cities” has been changed to “sites” in many places in order to clarify the intent of the comparison. For example, the first sentence, which reads “Average metal concentrations from the filter data were compared to measurements taken by the NAPS program at seven different Canadian sites (Environment and Climate Change Canada, 2015) (Table S6), which, due to a lack of comparable rural or remote baseline data, corresponded to different Canadian cities.”

In order to clarify the use of the 90th percentile as a means to predict which metals come from anthropogenic sources, the end of section 3.1 has been altered to read:

“In the oil sands, large swaths of forest are broken up by the occasional mine or upgrader. When the wind comes from one of these directions, particularly the upgraders, there is a noticeable difference in the air quality. To illustrate this large variability, the 90th percentile of the various elements was calculated and

analysed, as species that show a high degree of variability are more likely from these intermittent pollution sources. The results of this showed that the previously discussed elevated elements showed peaks indicating large variability. Additionally, at the highest peaks the concentrations of S, Ba, Br, and Mn also showed large increases, which indicates that in the Oil Sands, they are likely caused by anthropogenic sources.”

Within Section 3.5:

The second and third sentences have been altered to read “In particular, the concentrations of Si, Ti, K, Fe, Ca, and Al were, on average, higher than those measured in major Canadian cities. At their highest peaks the concentrations of S, Ba, Br, and Mn also exhibited large increases in concentrations.”

10. Page 13 Line 17. “lower concentrations of Mn and Fe” should be “higher. . .”

We agree that the “lower” should be changed to “higher”. This has been changed.

The sentence in the “Haul Road Dust” Section now reads as “What differentiated this factor from the Soil factor were the higher concentrations of Mn and Fe, and Ca. (Fig. 2).”

11. Page 20 Line 4. Incomplete sentence: including. . .

We agree that this is an incomplete sentence. And the “including” has been dropped from that sentence.

The sentence in “Implications of the Source Identification for Element Species” section now reads “As these factors represented a combination of multiple sources, the individual sources causing elevation of these elements is still not known; this limitation may help direct further studies.”

References

- Bari, M., Kindzierski, W.B., Wallace, L.A., Wheeler, A.J., MacNeill, M., Héroux, M.-È.; Indoor and outdoor levels and sources of submicron particles (PM1) at homes in Edmonton, Canada, *Environ. Sci. Technol.* (49), 6419-6429, 2015.
- Bari, M., and Kindzierski, W.: Ambient fine particulate matter (PM2.5) in Canadian oil sands communities: Levels, sources and potential human health risk, *Sci. Total Environ.*, 828-838, 2017.
- 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, doi:10.5194/acp-14-5073-2014, 2014.
- Kindzierski, W. and Bari, M.: Long-term temporal trends and influence of criteria pollutants on regional air quality in Fort McKay, Alberta, in: *Proceedings of the IASTED International Conference on Unconventional Oils and the Environment*, Calgary, Alberta, July 4-6, 2011, 162-168, 2011.
- Kindzierski, W. and Bari, M.: Air quality in Alberta oil sands community of Fort McKay compared to urban residential locations in Canada in 2009, in: *Proceedings of the Canadian Society for Civil Engineers 2012, Leadership in Sustainable Infrastructure*, Edmonton, AB, 974-983, 2012.

Landis, M., Patrick Pancras, J., Graney, J., White, E., Edgerton, E., Legge, A., & Percy, K. Source apportionment of ambient fine and coarse particulate matter at the Fort McKay community site, in the Athabasca Oil Sands Region, Alberta, Canada. *Sci. Total Environ.* , 584-585:105-117, 2017.

Percy, K., Hansen, M., and Dann, T.: Air quality in the Athabasca Oil Sands Region 2011. In *Alberta Oil Sands: Energy, Industry, and the Environment*, Elsevier Press, Amsterdam, The Netherlands, 47-91, 2012.

Shotyk, W., Bicalho, B., Cuss, C., Duke, M., Noernberg, T., Pelletier, R., Pelletier, R., Zaccone, C.: Dust is the dominant source of “heavy metals” to peat moss (*Sphagnum fuscum*) in the bogs of the Athabasca Bituminous Sands region of northern Alberta, *Environ. Int.*, 92-92, 494-506, 2016.

Wang, X. L., Watson, J. G., Chow, J. C., Kohl, S. D., Chen, L.-W., Sodeman, D. A., Legge, A. H., Percy, K. E.: Measurement of real-world stack emissions with a dilution sampling system. In *Alberta Oil Sands: Energy, Industry, and the Environment*, Elsevier Press, Amsterdam, The Netherlands, 171-912, 2012.

Wang, X. L., Chow, J., Kohl, S., Yatavelli, L., Percy, K., Legge, A., and Watson, J.: Wind erosion potential for fugitive dust sources in the Athabasca Oil Sands Region, *Aeolian Res.*, 18, 121-134, 2015.