**Response to Reviewers**
**A chemical transport model study of plume rise and particle size distribution for the Athabasca oil sands**

Original Reviewer comments are in normal font, responses in *italics.*

**Anonymous Referee # 1**
The manuscript by Akingunola et al. presents an interesting set of sensitivity simulations to plume rise modelling from stacks in the framework of the Canadian mesoscale chemistry-transport eulerian model GEM-MACH. The study is timely since simulation of the subgrid plume dynamic processes are still affected by significant uncertainties, as also confirmed by this work, and it is relevant for air quality applications considering the large role that emissions from elevated stacks plays nowadays and will play also in the near future. I found the manuscript generally well written and clear and I recommend publication to ACP after some minor corrections and clarifications as detailed below.

*We thank the reviewer for the helpful comments; our responses to the minor corrections and clarifications follows.*

1. P. 1, L. 19: ". . .reducing the magnitude of the original surface PM2.5 negative biases by 32%". Would be more clear to specify the range of change: from bias x to bias y.

*The text has been modified to read "...reducing the magnitude of the original surface PM2.5 negative biases 32%, from -2.62 to -1.72 $\mu g\ m^{-3}$."*

2. P. 1, L. 24: ". . .with 39 to 60% of predicted plume heights . . .". I suggest to specify what the given range is referring to, e.g. rephrasing the sentence adding a compact summary of what are the best and worst performing cases.

*Our submitted draft referred to the companion paper's results in the abstract – rather than take examples from that work, we've carried out additional analysis from our scatterplot figure (Figure 8 in the revised manuscript), and based on that analysis, we have modified the sentence in the abstract to read:*
*"As in our companion paper (Gordon et al., 2018), we found that Briggs algorithms based on estimates of atmospheric stability at the stack height resulted in under-predictions of plume rise, with 116 out of 176 test cases falling below the model:observation 1:2 line, 59 cases falling within a factor of 2 of the observed plume heights, and an average model plume height of 289 m compared to an average observed plume height of 822 m. We used a high resolution meteorological model to confirm the presence of significant horizontal heterogeneity in the local meteorological conditions driving plume rise. Using these simulated meteorological conditions at the stack locations, we found that a layered buoyancy approach for estimating plume rise in stable to neutral atmospheres, coupled with the assumption of free rise in convectively unstable atmospheres, resulted in much better model performance relative to observations (124 out of 176 cases falling within a factor of two of the observed plume height, with 69 of these cases above and 55 of these cases below the 1:1 line and within a factor of two of observed values). This is in contrast to our companion paper, wherein this layered approach (driven by meteorological*

*observations not co-located with the stacks) showed a relatively modest impact on predicted plume heights. "*
*We have also included a new table (Table 7 in the revised manuscript) which includes the distribution of values about the 1:2, 1:1 and 2:1 lines of the scatterplots, as well as the predicted average plume heights, in order for this comparison to be more quantitative.*

3. P. 1, L. 28: ". . .between the surface and 1km elevation". I suggest to clarify the concept. From my understanding this refers to the bias in the simulated lapse rate dT/dz as compared to observations.

*We have changed the sentence to read, "Persistent issues with over-fumigation of plumes in the model were linked to a more rapid decrease in simulated temperature with increasing height than was observed. This in turn may have led to overestimates of near-surface diffusivity, resulting in excessive fumigation."*

4. P. 2, L. 6: ". . .(it is not created by chemistry)". Suggest to change "chemistry" in "photochemical reactions in the atmosphere".

*We've followed your suggestion; the sentence now reads: "$SO_2$ is a primary emitted pollutant (it is not created by photochemical reactions in the atmosphere), with the majority of anthropogenic $SO_2$ emissions in the study region coming from large smokestacks (Zhang et al., 2018)."*

5. P. 2, L. 10-11: "Anthropogenic SO2 emissions are the main source of most atmospheric sulphur deposition". Suggest to add a reference for this statement.

*We have added a reference to Mylona, 1996 in the updated manuscript to support the above statement.*

6. P. 3, L. 17-18: Please specify to what conditions/cases the given ranges (34 to 52% and 0 to 11%) are referred to.

*The sentence refers to the companion paper, which did not segregate the extent of over/underprediction by stability class; we have added to the original statement, viz: "There we found that the Briggs (1984) plume rise parameterization significantly underpredicted plume heights in the vicinity of the multiple large $SO_2$ emissions sources in the Canadian Athabasca oil sands, with 34 to 52% of the parameterized heights falling below half of the observed height, compared to 0 to 11% of predicted plume heights being above twice the observed height, over conditions ranging from neutral, through stable to unstable."*

7. P. 3, L. 22: typo "Sulpher" should be "Sulphur"
*Corrected.*

8. P. 4, L. 15: typo "as-phase" should be "gas-phase"
*Corrected.*

9. P. 5, L. 2: Would be more informative to add the height of the levels in the bottom 1

km of the model.

*The model uses a hybrid coordinate system so the heights above ground change with location. However, the revised manuscript includes a new Figure (Figure 2 in the revised manuscript) which shows the model heights at a number of locations in the portion of the 2.5km domain studied here.*

10. P. 5: Moreover, given the relevance for the results, I recommend to add a description of the parameterizations adopted in the model for the PBL and the surface layer turbulence.

*The revised manuscript contains a new table (Table 1) which includes the main parameterizations in the meteorological model, including the Moist TKE scheme used for turbulence.*

11. P. 8, eq. 6: Please check the second condition "0.5 < H < 1.5" since the range seems to refer to a unitless quantity, but here only H is given.

*This was a typo on our part, and has been corrected in the revised manuscript.*

12. P. 9, L. 6: "top of the atmosphere" is confusing: is it perhaps the top of the PBL?

*An error on our part – the portion of the sentence should have read (and has been corrected to) "between $h_t$ and $h_b$".*

13. P. 9, L. 8: "value of hs was assumed", perhaps is "value AT hs was assumed". Moreover, the "s" of "hs" should be a subscript.

*The phrase was modified to read "centered on $h_s$"*

14. P. 10, L. 5: typo "and = hs", please check the left-hand side.

*The sentence has been modified to "At the stack height, $F_{j=0} = F_b$, and $z_j = 0$ (that is, the vertical distances are relative to the top of the stack)."*

15. P. 11, L. 21: "modstat" should be "modStat"

*Done.*

16. P. 12, L. 27: ". . .negative bias has decreased by 34%" it is not perfectly clear here and in the following if these bias changes are actual relative changes or absolute changes of the normalized mean bias. Please clarify.

*The sentence has been changed to ". For example, the magnitude of the mean bias has decreased from -2.623 to -1.725 $\mu g\ m^{-3}$, a reduction of 34%, indicating that a sizeable fraction of particulate under-predictions in 2-bin simulations may be due to poor representation of particle*

*microphysics through the use of the 2-bin distribution, despite sub-binning being used in some microphysics processes. "*

17. P. 12, L. 31-32: "Figure 2 shows that . . . less than 5 um diameter . . .". Please double check this statement. The figure shows the PM2.5 concentrations binned as a function of CONCENTRATION not SIZE.

*Thanks for catching this – the reviewer is quite right; this has been corrected.*

18. P. 15, Table 2: Please check the values that should be given in Italics, since not all the rows seem to contain it.

*Corrected.*

19. P. 16: referred to the discussion of SO2 overestimation and SO4 underestimation: can the two things be linked? E.g. by slow SO2 to SO4 conversion in the model, perhaps by slow aqueous chemistry?

*The aircraft observations were conducted under clear-sky conditions, so the potential for aqueous chemistry being the main issue is unlikely. Rather, noting that both predicted SO2 and SO4 aloft had negative biases, while predicted SO2 at the surface was biased high, it seems more likely that the main cause of the SO2 and SO4 negative biases aloft was a tendency for the model to overpredict fumigation to the surface, as noted in the original and revised manuscript.*

20. P. 17, L. 4-7: the paragraph seems to imply the presence of at least a (b) point, but only (a) is given. Please check or rephrase.

*The (a) has been removed (holdover from an earlier version of the manuscript, missed on checking prior to submission).*

21. P. 19, L. 11: ". . .took place between 16:30 and 20:30 on Aug 24th. . .". Although I am assuming the intervals are given in local time and not in UTC, it would be useful to have a confirmation in the paper. Here and also at least in the caption of the first figure showing time series (Figure 5).

*Corrected. The first pair of times are actually UTC (local times have been added in the revised manuscript), and Figure 5 (now Figure 6 in the revised manuscript) were also in UTC; this has been mentioned explicitly in the revised figure caption.*

**Anonymous Referee # 2:**

This paper introduces a very interesting and potentially highly useful field campaign. It also provides some important insights into the performance of the operational Canadian model. However, the paper has some weaknesses listed below.

*We thank the reviewer for the detailed comments on the paper – in addressing these, we feel that these comments have resulted in a significantly improved manuscript. The reviewer's comments and our responses (in italics) follow:*

1. In the same special issue for which this paper is submitted, there is another paper by the same group of authors, Gordon et al. 2018, which is also devoted to the plume rise topic, and it is said to have found opposite results. Neither is the reason for that clearly resolved, nor does it become clear why the plume rise topic is split between two papers.

*In our observation companion paper (Gordon et al, 2018) we saw that the Briggs algorithms, including the layered approach, tended to significantly underestimate plume rise. However, in that work we also noted that the observations themselves showed significant horizontal heterogeneity in the meteorological data used to drive the plume rise equations, with the corollary that the conditions at the actual location of the stacks may be sufficiently different from the surrounding meteorological towers to influence the predicted heights. No observations are available at the stack locations themselves – we therefore investigate this potential local influence further, using the high resolution on-line chemical transport model, GEM-MACH. As a demonstration of the model's ability to capture the local heterogeneity, we show in Figure 2(a) of the revised manuscript a snapshot of the PBL height at Saturday August 24, 2013 at 1 pm local time, and Figure 2(b) the corresponding model generated temperature profiles at the local meteorological towers AMS03, AMS05, the windRASS instrument and at three of the stacks examined in the companion paper. The model shows a significant variation in the temperature profiles between the tower and windRASS locations where the observations are available, and the stack locations. The temperature profiles suggest strong differences in both the strength of the inversion and its vertical location. This confirms the potential for spatial variability to have a significant influence on predicted plume heights relative to the meteorological observation locations in our companion paper. We therefore investigated the plume rise algorithms again within the current work, in order to determine the extent to which this local variability may influence predicted plume heights. In contrast to the companion paper, we found that the "layered" approach of calculating local stability residuals through successive model layers resulted in significantly improved plume heights relative to the more standard Briggs approaches which employ stability estimates at the top of the stacks.*

*We have included the new Figure 2 in the revised manuscript, as well as some discussion in the Introduction section of the manuscript:*

*"…Our companion paper made use of different sources of meteorological observations to drive the Briggs (1984) plume rise algorithms, as well as CEMS data and aircraft observations of $SO_2$ plumes from multiple sources over a 29-day period. There we found that the Briggs (1984) plume rise parameterization significantly underpredicted plume heights in the vicinity of the multiple large $SO_2$ emissions sources in the Canadian Athabasca oil sands, with 34 to 52% of the parameterized heights falling below half of the observed height, compared to 0 to 11% of predicted plume heights being above twice the observed height, over conditions ranging from neutral, through stable to unstable.*

*However, in our companion paper we also noted the presence of considerable spatial heterogeneity in the meteorological observations used for the algorithm tests. Temperature profiles and other data used to define the input parameters for the Briggs algorithms were taken from two tall meteorological towers, a windRASS, and a research aircraft, and showed a substantial variation in the resulting plume height predictions, despite relatively close physical proximity of these sources of meteorological data (e.g. 8 km distance between the two meteorological towers). The region under study is subject to complex meteorological conditions due to the nature of the terrain (a river valley with up to 800 m of vertical relief, and open pit mines and settling ponds which may each be tens of $km^2$ in spatial extent). This heterogeneity cast some uncertainty on the results of the companion paper, in that the best application of the plume rise algorithms would be driven by the meteorology at the location of the stacks, rather than the location of the available meteorological instruments, and the latter suggested substantial local changes in meteorological conditions. As we show in the sections which follow, the spatial heterogeneity of meteorological conditions has a controlling factor on the predicted plume rise, and, in contrast to our companion paper, an approach making use of local temperature gradients between individual model layers has greatly improved accuracy in comparison to those inferring atmospheric stability conditions from the conditions at the top of the emitting stacks."*

*The new Figure 2 is described via the following discussion in the revised text:*

*"We noted in our companion paper (Gordon et al., 2018) that meteorological observations varied substantially in the study region depending on location, citing this as a possible confounding factor on the results of tests of the plume rise algorithms. This spatial heterogeneity was well captured by the high resolution GEM-MACH simulations, as is demonstrated by the example depicted in Figure 2, which shows the typical local variation in planetary boundary layer height (Figure 2(a)), ranging from about 1200m to 400m, the lower values corresponding to the main cleared areas (open pit mines, settling ponds) of the industrial facilities. The corresponding temperature profiles in several locations marked in Figure 2(a) are given in Figure 2(b): These show a substantial difference in model predicted stability at the three meteorological observation locations of Gordon et al. (2018) (windRASS, AMS03, and AMS05), and substantial differences between these and the locations of the main stacks of some of the facilities (Syncrude 1, CNRL, and Suncor). The temperature profiles show that the height and strength of the inversion may vary by over 100m in the vertical, and that the profiles do not merge with the larger scale flow until an elevation of 750m asl (450m agl) is reached. Given this level of variation, we might expect potential errors in calculated plume heights when applying the meteorological observations to plume rise at the stack locations, in turn suggesting that a re-examination of plume rise using the model results is worthwhile."*

2. The model performance is not only influenced by the aspects forming the focus of this paper, but also by the accuracy of the meteorological part of the model, and by the numerics of

transport, notably the vertical diffusion and the handling of the point sources in the Eulerian framework. Their role is discussed only at the very end and, in my opinion, not sufficently in depth. In order to evaluate specific model aspects, one first needs to understand the performance of the model in general, with its strengths and weaknesses.

*The focus of this work is not to evaluate the overall model performance, but to evaluate how specific updates to the representation of the aerosol size distribution and the plumerise algorithm contribute to better model performance when compared to available observations. An evaluation of each aspect of a complex reaction-transport model is beyond the scope of a single paper. Nevertheless, as we already stated in the discussion section of the manuscript, we have carried out a sensitivity run which showed that variations in the magnitude of model diffusivity had a minimal impact the predicted plume behaviour and on the vertical distribution of $SO_2$ plumes at the point of release from the stacks, though we acknowledge that the model's tendency to overpredict the rate of decrease of air temperature with height may influence the shape of the diffusivity profile. We have also added additional references on the description and evaluation of the vertical diffusion scheme used in the meteorological portion of the model (Mailhot and Benoit 1982), as well as a more recent publication on the overall description and performance of the meteorological model as a whole (Girard et al., 2014), in Table 1 of the revised manuscript.*

3. The statistical approach chosen for the evaluation of the model options relies on metrics which exclusively are based on "match in time and space" data pairs. It is well known in air-pollution modelling that for near-source conditions (which is what we find here), there is often too much "noise" in the data (be it due to the stochastic nature of the plume, be it due to unresolved meteorological variability) to give meaningful results. Correspondingly, some of the statistical parameters are not very good. Therefore, global comparisons (such as deviations from the cumulative frequency distribution, statistics of cross-wind integrated values, or average dependency on key parameters such as stability and wind speed) are often used to assess models in a more robust way.

*We have added a simple table (Table 7, revised manuscript) showing the frequency distribution of the predicted versus observed plume rise from the three different variations on plume rise examined here. This new table is in agreement with the measurement statistics in that both show that the layered approach provides a better fit to observations, with a distribution more centered around the model:observation 1:1 line. While we agree with the reviewer regarding the difficulties associated with use of matching pairs for near-source conditions, we nevertheless respectfully hold that improvements in these statistics represent real improvements in model performance. For example, while a mean bias score is the average deviation between model and observed pairs, this average is over a large set of conditions, hence should be subject to less issues associated with the stochastic nature of the plume on any given hour. While near-source comparisons are often difficult due to the nature of the near-source region (as the reviewer suggests), improvements in these statistics nevertheless imply real improvements in model performance.*

4. The paper is written well on the "small scale" (apart from numerous technical deficiencies as listed below), but the broad topics could be worked out more clearly. In the end, the findings are: twelve aerosol size bins are better than two (not surprising, but good to see it quantified), there is

an improvement by using the model's vertical profile information for plume rise calculation but given the model's deficiencies the overall conclusion seems to be not so clear, and no improvement was found for using hourly stack data, but it remains unresolved why. We may wonder whether the work is mature enough for publication if we consider this state of the quintessential findings.

*We note that the results presented in the work show that the use of stack-location-specific meteorological information combined with the residual buoyancy calculations provides a considerably more accurate estimate of plume rise than the top-of-stack stability parameterizations often used in air-quality models. We have provided an additional table which shows that the distribution of plumes is better represented with the residual buoyancy calculation than with the top-of-stack stability plume rise calculation. The average plume rise calculated using the CEM-based data is closer to the observations than the annual totals, from both the original analysis and the additional table. While both the CEMS (using the hourly stack data) and the non-CEMS model scenarios are very close, the key point is that the revised, residual buoyancy plume rise algorithm has much better performance than the original algorithm. We have noted in the revised manuscript that "the relatively small differences between Figure 8(b) and 8(c), and between the last two columns of Table 7, imply that the residual buoyancy approach of equations 9 was relatively insensitive to the range of the initial buoyancy flux resulting from the two sets of emissions data used here, compared to the temperature gradients in equation (5)." We have also mentioned the insensitivity of the residual buoyancy calculation to the range of initial buoyancy flux in the revised conclusions; "However, the latter approach was also shown to be relatively insensitive to the range of initial buoyancy fluxes resulting from the two different emissions estimates, with the use of hourly observed (and presumed more accurate) stack parameters resulted in a slight degradation of performance relative to the use of annual reported values for these parameters."*

*Both sources of emissions data are limited by the model resolution and the independent verification of the accuracies of either is not available. We also point out in the revised manuscript that both sources of data have inherent errors. For example, as mentioned in the emissions paper referenced by this work (Zhang et al., 2018), and clarified/noted in the revised manuscript, the data referred to as CEMS here also contains engineering estimates of "upset conditions" wherein facility emissions are redirected to a flare stack for which direct emissions observations are not possible. That is, what we have referred to as "CEMS" data incorporates considerable associated uncertainties – this should have been included in the original manuscript and not left as a reference to the emissions paper alone. This has been mentioned in the description of the CEMS data as follows, "…and second with emissions information derived from a combination of CEMS hourly stack parameters as well as engineering estimates of emissions during "upset conditions" in which the effluent is redirected to flare stacks (the latter estimates are considerably more uncertain than the CEMS information, but are nevertheless included here since they result in substantial changes in pollutant emissions and plume characteristics, see Zhang et al, 2018)."*

**RC2: Specific Comments**

1. Page 2, l. 18: Why are you thinking that reasons for weak performance include only those meteorological variables that are used for the plume rise calculation, but not, for example, wind direction?

*We acknowledge that the model's performance is of course the result of many factors. The intent of our work is rather to evaluate the relative impact of the plume rise calculations on the results. The given sentence has been modified to "(iii) errors in meteorological forecast variables (including wind speed and direction, etc., as well as those used in calculating plume rise)".*

2. The model overview section lacks information on the numerical scheme used for vertical diffusion even though this is crucial in the context of study (cf. discussion on p. 24). The main reference for the MACH model seems to be Makar et al. (2010) – an extended abstract that would not be available for most people who haven't attended the conference as it is not freely accessible. Is there no more detailed and open description of this model? Note that also the Coté et al. citation is one of those for which the reference is missing. In addition, the handling of the point sources is not described (usually, Eulerian models use some sub-model to track plumes until they match the size of the grid cells).

*The revised manuscript includes a new Table (Table 1) which gives the main references for the meteorological (GEM) components for the model, including the reference for the Moist TKE approach used for calculating vertical diffusivity. Regarding the online air quality GEM-MACH model, the first overall description reference is Moran et al. (2010) (and not Makar et al. (2010) as stated by the reviewer), and hence we feel obliged to include it in published descriptions of GEM-MACH. However, this is not the only description of the model or its components, and others from the journal literature appeared in the original manuscript; we cited Gong et al (2003) for the aerosol microphysics, Gong et al (2006) for the aqueous-phase chemistry, Makar et al, (2003) for the inorganic heterogeneous chemistry, Lurmann et al. (1986) and Stroud et al. (2008) for the gas-phase mechanism , Zhang et al. (2001, 2002, 2003) and Makar et al (2018) for the gas and particle deposition. We also cited the Air Quality Model Evaluation International Initiative papers Im et al (2015a,b) and Makar et al (2015(a,b)) papers, which contain detailed descriptions of the model, its chemical and physical parameterizations, and its performance relative to other models of its type. The revised reference list has been double-checked to include all references (including the papers by Côté et al).*

*While some air-quality models include a form of "plume in grid" parameterization which track emitted puffs in a Lagrangian sense or employ a Gaussian dispersion model at the sub-grid scale, these approaches have not become predominant for three main reasons: (1) they ultimately rely on the driving large-scale meteorology (which may be inaccurate, as already pointed out in our work and by the reviewer, reducing their potential advantages); (2) they may add considerably to the processing time (particularly if a large number of point sources, chemical reactions and multiple species are considered), and (3) most models employ a self-nesting capability which allow the models to locally go to higher resolution, negating some of the advantages to a plume-in-grid approach. Consequently, most air-quality models have continued to rely on the handling of point sources using a combination of plume rise algorithms, and nesting to higher resolutions, as has been done in our work.*

3. The model set-up description in section 2.2 is not easy to follow. It might be helpful to move some of the information into a table and to shorten the text.

*A table summarizing the model description has been added as suggested.*

4. Page 8, line 1: The plume's buoyancy flux is **not** dependent on the stack height (at least not directly).

*This typographical error has been corrected.*

5. From the sentence beginning on p. 11, l. 7, on, the text does not really belong to the section 2.2.3. It should become a section of its own, as it introduces the simulations forming the base of the rest of the paper (maybe merge with some parts of the 2.2 chapeau).

*Section 2.2.3 has been renamed "Sources of Emissions Data", and the remainder of the previous section 2.2.3 past the point noted by the reviewer has been split off and renamed "2.2.4 Simulation Scenarios"*

6. Page 11, Section 3.1: What is the justification for removing measurements with values exceeding some threshold? Without proper justification this would not be acceptable.

*The key phrase in the original manuscript was that "extreme single-hour measurements" have been removed. That is, if the time series jumps from a background value to something greater than 150ppbv ($SO_2$, $NO_2$, and $O_3$) or 150 $\mu g\ m^{-3}$ (PM2.5), then immediately back again in the next hour, that jump is assumed to be due to instrumentation error and/or calibration times in the measurement record. In contrast, a rise above these levels for more than a single hour is retained. This has been clarified in the revised manuscript, viz: "The observation data have been filtered to remove extreme single-hour measurements that are greater than 150ppbv for $SO_2$, $NO_2$, and $O_3$, and 150 $\mu g\ m^{-3}$ for $PM_{2.5}$ (single-hour spikes of this nature in hourly records are assumed to correspond to instrumentation errors or calibration times for the instruments)."*

7. Page 12, Section 3.2: The phrase 'spatial linearly interpolated model values at the models chemistry time resolution of 2 minutes' is awkward. If you have 10 s data as said before, why do you need to interpolate for obtaining 2 min data? Also, it would be good to know which distance corresponds to both 10 s and 2min flight data, and how this compares to the model's grid size.

*The portion of the sentence has been corrected to "linearly interpolated values in time and space from the model's 2 minute time step and 2.5km resolution". We have also added the sentence: "The nominal cruise speed of the National Research Council Convair 580 used in the experiment is 550 km/hour; a 10 second time interval thus represents an observation integration distance of 1.528 km, and a two minute time interval an observation integration distance of 18.3 km."*

8. Section 4 (Results and Discussion) needs to be structured into subsections.

*This has been done, with subsections 4.1 Spatial Heterogeneity of Meteorological Conditions, 4.2 Two-bin versus Twelve-bin Evaluation, and 4.3 Plume Rise Algorithm Evaluation.*

9. Table 1: Apart from widely used or self-explanatory metrics such as FAC2, RMSE or r , the metrics parameters need to be defined.

*We have included a new Table; Table 2, which is now referenced at the start of section 3, and includes the mathematical definitions of all of the metrics.*

10. Page 12, l. 31: "Figure 2 shows that the model simulations are biased high for particles less than 5 μm diameter, and biased low for the larger particle sizes." As this figure only shows results for PM2.5, a statement on larger particles can't be based on it

*This was a typo; the text should have referred to 'concentration' and not 'particle sizes'; this has been corrected in the updated manuscript.*

.

11. Page 13, l. 14: Information on the bin sizes belongs to the model description section, not the result section.

*We have added a description of the cut sizes for both 2 and 12 bin simulations to the model description section. However, the mention of the 2 bin cut sizes is necessary here to explain why a comparison between the 2-bin model results with the aircraft observations is not appropriate (the 2-bin model lacks the size cut resolution to be able to simulate the PM1 observed by the aerosol mass spectrometer aboard the aircraft). The sentence has been changed to "The aircraft's AMS instrument measures speciated atmospheric particle concentrations for particles less than 1μm size, and therefore cannot be compared with the 2-bin model results because the smaller size bin (with upper diameter size cut 2.56 μm) will be biased high relative to the 1 μm size cut of the AMS."*

12. Page 13: The second paragraph on this page contains a number of statements about results without pointing to the figures or tables which show them.

*The revised manuscript references Table 4 for this paragraph.*

13. Concerning the model performance for PM, it should be discussed that even though the twelve-bin version leads to significant improvements, major discrepancies to observations remain.

*The last sentence of the new section 4.2 has been modified to read: "The use of the 12-bin size distribution (purple histogram bars, Figure 3) improves the fit to the observations (blue histogram bars), in comparison to the 2-bin distribution results (red histogram bars), though significant over-predictions of the frequency of low concentration events and under-prediction of high concentration events, remain.".*

14. A number of tables are presented where several metrics are used to compare various model versions, with the best one being emphasised by bold print. Sometimes, differences are tiny and probably insignificant. Only those values that are significantly better should be highlighted to

avoid a wrong impression of the results (for example, in Table 3 the model version seems to have no impact for O3 but we get the impression that the simpler model is better.)

*The problem with this suggestion is that different readers may have different ideas regarding what is considered a 'significant' change, what is considered to be a 'tiny' difference, etc. We feel that the readers will look at both the numbers themselves as well as the highlighting, as the reviewer did, to note the relative level of differences. We mentioned some of these relative differences in the original text as well, as a caution to the reader not to base their judgement on the highlighting alone, e.g., with reference to the ozone evaluation in Table 3: "Ozone, in contrast, is created or destroyed through secondary chemistry over relatively longer time-spans than the transport time from the sources in this comparison (spatial scales on the order of 10's of km). Accordingly, the impact of the plume rise of $NO_x$ on ozone formation is relatively minor, usually in the third decimal place (though first decimal place improvements occur for the mean bias with the use of the new plume rise algorithm)."*

15. Why is the use of hourly emission data beneficial for NO2 but detrimental for SO2?

*One significant difference between $SO_2$ and $NO_2$ for the study region is that the latter originates almost completely in major point sources, while only 40% of the latter originates in major point sources, the rest in area sources (heavy hauler fleets used by the open pit mine operators). The $NO_2$ values will thus be due to a combination of sources, with the possibility of compensating errors at the emissions level influencing the net model $NO_2$ concentration .*

16. The discussion paper does not comply with the ACP Data Policy; it does not have a "Data availability" section and says nothing about data availability

*This has been added to the revised manuscript (in our other papers with ACP, this has come as a request from the Editor following the completion of the review process, sorry for not having added it to the submission): all of the data used here are publicly available on the oil sands data archive or the website of the Wood Buffalo Environmental Association. We have also added the standard Author Contributions, Competing Interests, Special Issue Statement, and Acknowledgements, to the revised manuscript.*

# A chemical transport model study of plume rise and particle size distribution for the Athabasca oil sands

Ayodeji Akingunola[1], Paul. A. Makar[1], Junhua Zhang[1], Andrea Darlington[2], Shao-Meng Li[2], Mark Gordon[3], Michael D. Moran[1], Qiong Zheng[1]

[1]Modelling and Integration Section, Air Quality Research Division, Environment and Climate Change Canada
[2] Processes Research Section, Air Quality Research Division, Environment and Climate Change Canada
[3]Centre for Research In Earth And Space Engineering, York University, Toronto Canada

*Correspondence to*: Ayodeji Akingunola (Ayodeji.akingunola@canada.ca), Paul Makar (paul.makar@canada.ca)

**Abstract.**

We evaluate four high-resolution model simulations of pollutant emissions, chemical transformation and downwind transport for the Athabasca oil sands using the Global Environmental Multiscale – Modelling Air-quality and Chemistry (GEM-MACH) model, and compare model results with ~~using~~ surface monitoring network and aircraft observations of multiple pollutants, for simulations spanning a time period corresponding to an aircraft measurement campaign ~~in the region~~ in the summer of 2013. We have focussed here on the impact of different representations of the model's aerosol size distribution and plume-rise parameterization on model results.

The use of a more finely resolved representation of the aerosol size distribution was found to have a significant impact on model performance, reducing the magnitude of the original surface $PM_{2.5}$ negative biases ~~by 32%.~~32%, from -2.62 to -1.72 μg $m^{-3}$.

We compared model predictions of $SO_2$, $NO_2$, and speciated particulate matter concentrations from simulations employing the commonly-used Briggs (1984) plume-rise algorithms to redistribute emissions from large stacks, with stack plume observations. As in our companion paper (Gordon *et al*., 2018), we found that Briggs algorithms based on estimates of atmospheric stability at the stack height resulted in under-predictions of plume rise, with 116 out of 176 test cases falling below the model:observation 1:2 line, 59 cases falling within a factor of 2 of the observed plume heights, and an average model plume height of 289 m compared to an average observed plume height of 822 m. We used a high resolution meteorological model to confirm the presence of significant horizontal heterogeneity in the local meteorological conditions driving plume rise. Using these simulated meteorological conditions at the stack locations, we found that a layered buoyancy approach for estimating plume rise in stable to neutral atmospheres, coupled with the assumption of free rise in convectively unstable atmospheres, resulted in much better model performance relative to observations (124 out of 176 cases falling within a factor of two of the observed plume height, with 69 of these cases above and 55 of these cases below the 1:1 line and within a factor of two of observed values). This is in contrast to our companion paper, wherein this layered approach (driven by meteorological observations not co-located with the stacks) showed a relatively modest impact on predicted plume heights. ~~As in our companion paper (Gordon *et al*., 2018), we found these algorithms resulted in under-predictions of plume rise, with 39 to 60% of predicted plume heights falling below half of the observed plume heights. However, we found here that a layered buoyancy approach for stable to neutral atmospheres, coupled with the assumption of free rise in convectively unstable atmospheres, resulted in much better model performance, both for atmospheric constituent concentrations and the predicted height of the plumes.~~ Persistent issues with over-fumigation of

Formatted: Width: 21.59 cm, Height: 27.94 cm

Formatted: Not Superscript/ Subscript

Formatted: Font: Italic

plumes in the model were linked to ~~positive biases in the predicted temperatures between the surface and 1km elevation~~a more rapid decrease in simulated temperature with increasing height than was observed. Thi~~s~~~~e~~ in turn may have le~~d~~~~d~~ to overestimates of near-surface diffusivity, resulting in excessive fumigation.

## 1 Introduction

Forecast ensembles of regional air-quality models tend to have relatively poor performance in their predictions of sulphur dioxide ($SO_2$), with normalized mean biases in the range +/-40%, Pearson's correlation coefficients (R) of less than 0.21, and normalized mean errors of more than 75% (Makar *et al.*, 2015b). These scores may be contrasted with those for atmospheric ozone ($O_3$) of +/- 13% for normalized mean bias, R more than 0.6, and normalized mean errors less than 37%. $SO_2$ is a primary emitted pollutant (it is not created by ~~chemistry~~photochemical reactions in the atmosphere), with the majority of anthropogenic $SO_2$ emissions in the study region coming from large smokestacks (Zhang *et al.*, 2018). In North America, such "major point sources" are often outfitted with Continuous Emissions Monitoring System (CEMS) instrumentation, which provides accurate hourly estimates of the emitted mass of $SO_2$, as well as estimates of parameters that govern the buoyancy- or momentum-driven rise of the resulting plumes, such as the temperature of the emissions, and their volume flow rate (volume emitted / unit time). Anthropogenic $SO_2$ emissions are the main source of most atmospheric sulphur deposition (Mylona, 1996) (reacting in the gas-phase with the OH radical to create sulphuric acid, and in cloud water and rain via aqueous chemistry to create bisulphate and sulphate ions). The poor performance of $SO_2$ predictions in air-quality models is therefore a matter of concern, and drives the need to better understand its causes. Some of the potential reasons for this poor performance include (in-)accuracy of the; (i) emissions information (less likely in cases where CEMS data are available), (ii) plume rise parameterization algorithms (which describe the vertical redistribution of the emitted mass according to the stack parameters and meteorological conditions: e.g., Briggs (1984)), (iii) errors in meteorological forecast ~~meteorological~~ variables (including wind speed and direction, etc., as well as those used in calculating plume rise), and (iv) $SO_2$ loss processes, such as oxidation (as noted above) and the deposition algorithms and meteorological inputs used for calculating the $SO_2$ deposition rate. Furthermore, a combination of these factors may drive the relative difference in model performance between $SO_2$ and $O_3$; we note, for example, that tropospheric $O_3$ is a secondary pollutant (driven by chemical formation and loss rather than direct emissions of ozone), and hence will be more spatially homogeneous than $SO_2$, with the implication that forecast accuracy for very local conditions will play more of a role in setting the ambient concentrations of $SO_2$ than $O_3$.

The prevalence of CEMS for $SO_2$ observations in both Canada and the U.S. (https://www.epa.gov/emc/emc-continuous-emission-monitoring-systems) implies that the CEMS-derived emissions inputs available for model simulations will be well characterized. However, reporting requirements vary between the countries. In Canada, emitting facilities are required to report estimates of their total annual emissions, as well as typical stack parameters, to the federal National Pollutant Release Inventory (NPRI, 2018), although individual Canadian provinces may require more detailed reporting. In the U.S., CEMS $SO_2$ data are reported at the national level to the U.S. EPA (EPA, 2015, 2018~~(a,b)).~~). In both countries, estimates of the typical stack volume flow rate (and/or the stack exit flow velocity) and effluent stack exit temperatures are reported and used for modelling, instead of hourly estimates recorded by CEMS. In the Canadian province of Alberta, regulatory reporting requirements include CEMS hourly observations of $SO_2$ and $NO_2$ emissions from selected large stacks, as well as hourly information on the stack effluent temperature and volume flow rate.

In our companion paper (Gordon *et al*., 2018) we note that past and current regional air-quality transport models (Im *et al*., ~~2015~~2015a; Byun and Ching, 1999; Holmes *et al*., 2006; Emery *et al*., 2010) and emissions processing models (CMAS, 2017; Bieser *et al*., 2011) describe the buoyancy- and/or momentum-driven vertical redistribution of emitted mass from stacks using variations on the work of Briggs (1969, 1975, 1984). In the latter work, observations of plume rise, stack parameter information, and meteorological conditions were used to generate parameterizations, linking these data to the height gained by the centerline of atmospheric plumes (the plume height), as well as the vertical extent of the bulk of the emitted mass about that centerline. However, subsequent early evaluations of the accuracy of these parameterizations (cf. VDI, 1985) have had mixed results, including parameterization estimates averaging 50% higher than observations (~~Giebel~~Gielbel, 1979); within 12 and 50% of observations (Ritmann, 1982); 30% higher than observations (England *et al*., 1976); 50% higher than observations (Hamilton, 1967). Recent studies using Reynolds averaged Navier-Stokes and large eddy simulation (~~RAND~~RANS-LES) modelling have shown that the integral model of Briggs overestimates the plume rise and its overestimation error increases as the role of atmospheric turbulence increases (Ashrafi *et al*., 2017), ~~and~~ as well as underestimates of plume rise, inferred from excessively high predicted surface concentrations (Webster and Thompson, 2002). Our companion paper~~, making~~ made use of different sources of meteorological observations~~,~~ to drive the Briggs (1984) plume rise algorithms, as well as CEMS data~~,~~ and aircraft observations of $SO_2$ plumes from multiple sources over a 29-day period~~,~~. There we found that the Briggs (1984) plume rise parameterization ~~tended~~ significantly ~~to~~ underpredicted plume heights in the vicinity of the multiple large $SO_2$ emissions sources in the Canadian Athabasca oil sands, with 34 to 52% of the parameterized heights falling below half of the observed height, compared to 0 to 11% of predicted plume heights being above twice the observed height, over conditions ranging from neutral, through stable to unstable. ~~.~~

However, in our companion paper we also noted the presence of considerable spatial heterogeneity in the meteorological observations used for the algorithm tests. Temperature profiles and other data used to define the input parameters for the Briggs algorithms were taken from two tall meteorological towers, a windRASS, and a research aircraft, and showed a substantial variation in the resulting plume height predictions, despite relatively close physical proximity of these sources of meteorological data (e.g. 8 km distance between the two meteorological towers). The region under study is subject to complex meteorological conditions due to the nature of the terrain (a river valley with up to 800 m of vertical relief, and open pit mines and settling ponds which may each be tens of $km^2$ in spatial extent). This heterogeneity cast some uncertainty on the results of the companion paper, in that the best application of the plume rise algorithms would be driven by the meteorology at the location of the stacks, rather than the location of the available meteorological instruments, and the latter suggested substantial local changes in meteorological conditions. As we show in the sections which follow, the spatial heterogeneity of meteorological conditions has a controlling factor on the predicted plume rise, and, in contrast to our companion paper, an approach making use of local temperature gradients between individual model layers has greatly improved accuracy in comparison to those inferring atmospheric stability conditions from the conditions at the top of the emitting stacks.

These underpredictions of plume rise are a potential source of concern, given that they imply that the underlying algorithms will bias $SO_2$ towards lower ~~elevations.~~concentrations. This will lead to more local rather than long-range sulphur deposition. ~~Sulpher~~Sulphur deposition is the focus of other work examining acidifying deposition associated with emissions sources in Alberta (Makar *et al*., 2018).

The work reported here has four main foci, driven by the need to evaluate and if possible improve the performance of both the algorithms governing plume rise and our air-quality model (Global Environmental Multiscale – Modelling Air-quality and Chemistry; GEM-MACH) which employs those algorithms. The main objectives of this study include: (1) an evaluation of the impacts of the plume rise algorithms on model performance, with the introduction of a new approach to calculate plume rise being compared to the standard Briggs (1984) approach; (2) estimation of the impact of hourly major point stack information on model results; and (3) an overall evaluation of the model performance using different configurations for the representation of plume rise and particle size distributions.

## 2 Model Description

### 2.1 Model Overview

Global Environmental Multiscale – Modelling Air-quality and Chemistry (GEM-MACH) is Environment and Climate Change Canada's comprehensive online air quality and chemical transport modelling system, currently in its second major revision. The model consists of an atmospheric chemistry module (Moran *et al.*, 2010), tightly coupled with the dynamical core and residing within the physics module of the Global Environmental Multiscale (GEM) weather forecast model (Cote *et al.*, ~~1998).~~1998a,b; Girard *et al*., 2014). Emissions for the model are provided using an emissions processing system based on SMOKE - Sparse Matrix Operator Kernel Emissions (~~SMOKE,~~ Coats, 1996~~, https://www.cmascenter.org/smoke~~). GEM-MACHv2 is a multiscale model, designed and exercised in a wide range of scales, from global chemical transport modelling, to regional air quality modelling with direct and indirect feedbacks between chemistry and meteorology (Makar *et al.*, 2015a,b), and urban scale air quality modelling (~~Stroud, 2016). The~~Munoz-Alpizar *et al*., 2017). The physical and chemical processes represented in the model regional air-quality prediction system ~~includes as its~~are summarized in Table 1. The main chemical components are the ADOM-II ~~mechanism~~ gas-phase chemistry mechanism ~~with 42 species (Lurmann *et al.*, 1986; Stroud, 2008);~~using the ~~Canadian Aerosol Module (Gong et al., 2003a,b), a size resolved, sectional approach with either 2 or 12 size bins, multi component, aerosol microphysics module , including~~ Young and Boris solver, and the aerosol module which includes process representation for particle nucleation, condensation, and coagulation (Gong et al, 2003)~~;as phase deposition based on the work of Jarvis (1976), Wesely *et al.* (1989), and Zhang *et al.* (2002, 2003); and particle~~, and deposition (Zhang *et al*., 2001). Additional aerosol processes include cloud scavenging, and in-cloud aqueous phase chemistry (Gong *et al.*, 2006), as well as equilibrium inorganic gas-aerosol partitioning (HETV scheme; Makar *et al*., 2003). Eight aerosol species are included in GEM-MACH: particle sulfate, nitrate, ammonium, primary organic carbon, secondary organic carbon, elemental (aka "black") carbon, sea-salt and crustal material. The model also features experimental options for feedback between weather and air-quality in 12-bin mode (Makar *et al*.,2015a,b). More detailed descriptions of GEM-MACH may be found in Makar *et al.*, 2015(a,b) and Im *et al*., 2015(a,b). We discuss elsewhere in this special issue the use of GEM-MACH for acid deposition estimates (Makar *et al*., 2018), bi-directional fluxes of ammonia to the boreal forest (Whaley *et al.*, 2018), the impact of updated emissions of volatile organic compounds and organic particulate matter (Zhang *et al.*, 2018) on model performance for these species (Stroud *et al*., 2018). The vertical transport of heat, moisture and momentum by turbulent eddies are represented by enhanced vertical diffusion is based on the turbulent kinetic energy (TKE) closure scheme (Mailhot and Benoit 1982) in the GEM model physics module. The same numerical scheme and coefficients of vertical diffusivity are used for the diffusive transport of chemical species within GEM-MACH. Table 1 provides an overview of the main processes represented in the atmospheric physics components of the GEM weather forecast model upon which GEM-MACH is based.

Formatted: English (U.S.)

Formatted: English (U.S.)

**Table 1**: A summary of the main physical and chemical processes represented in GEM-MACH regional model

| Condensation Scheme | Convective Parameterization | Aerosol Scheme | Radiative transfer | Boundary layer Parameterization | Surface processes | Gas phase mechanism | Gas-phase deposition |
|---|---|---|---|---|---|---|---|
| 2-moment cloud microphysics, (Milbrandt and Yau 2005a,b). | Kain and Fritsch (1990) and Kain (2004) | Canadian Aerosol Module (CAM); Sectional 2 or 12 bins; (Gong *et al*, 2003a,b) | Li and Barker (2005) | Moist-TKE closure scheme (Mailhot and Benoit, 1982) | ISBA2, (Belair *et al*., 2003a,b) | 42-species ADOM-II mechanism (Lurmann *et al*., 1986) | Jarvis (1976), Wesely *et al*. (1989), Zhang *et al*. (2002, 2003); |

Formatted Table

## 2.2 Model Setup and Configurations

A 2-bin simulation of GEM-MACH running in a nested configuration from a North American 10km resolution forecast to a 2.5km Alberta/Saskatchewan domain has been in continuous experimental forecast mode since October 2012, and this configuration is also used for operational forecasts by Environment and Climate Change Canada. While the 2-bin simulation reduces computational processing time by 25% in the current version of GEM-MACH, we investigate here the effect of this configuration on model accuracy relative to observations, employing the GEM-MACHv2 model in the Oil Sands 2.5-km nested system using the more detailed 12-bin aerosol size distribution configuration. We have carried out a set of retrospective simulations targeting the JOSM (the Governments of Canada and Alberta Joint Oil Sands Monitoring program) summer 2013 intensive campaign period (JOSM, 2011). The outer 10-km horizontal resolution domain, which covers most of continental Canada and United States, was configured with 82 vertical levels and a 5-min physics/15-min chemistry time step, with the chemical boundary and initial conditions provided by MOZART-4 climatology (Emmons *et al.*, 2010), and meteorological boundary/initial conditions provided by the GEM-based~~'s~~ Regional Deterministic Prediction System (RDPS, Caron *et al.*, 2014). The RDPS itself was driven by ~~data-assimilated~~ meteorological analyses generated using data assimilation. The RDPS was also used to drive a 2.5-km horizontal resolution regional weather-only simulation, using a modified GEM High Resolution Deterministic Prediction System configuration (HRDPS, Charron *et al.*, 2012). Both the modified HRDPS and the 10-km resolution GEM-MACH produced 36-hour simulations, the last 24 hours of which were used to provide the respective meteorological and chemical boundary conditions for a 24-hour GEM-MACH 2.5-km resolution simulation (which was configured with 64 vertical levels, and 1-min physics/2 min-minute chemistry model time steps). The use of the HRDPS in this fashion allowed each GEM-MACH 2.5km simulation to commence from a "spun-up" state for its cloud variables. For the chemical species, the last hour of each 24-hour simulation was used to provide initial conditions for the subsequent GEM-MACH simulation. This provided continuity of the chemical fields across subsequent 24-hour simulations. The GEM-MACH 10-km simulation and the HRDPS 2.5-km simulations, updated every 24 hours, provided ongoing boundary conditions and hence continuity with the meteorological analysis, thus preventing the high resolution meteorology from drifting chaotically from the analyses. The GEM-MACH 10km, HRDPS, and GEM-MACH 2.5km domains are shown in Figure 1. The retrospective simulations were carried out for the period August 1$^{st}$, 2013 to September 10$^{th}$, 2013, with the first 7 days results discarded as model spin-up time.

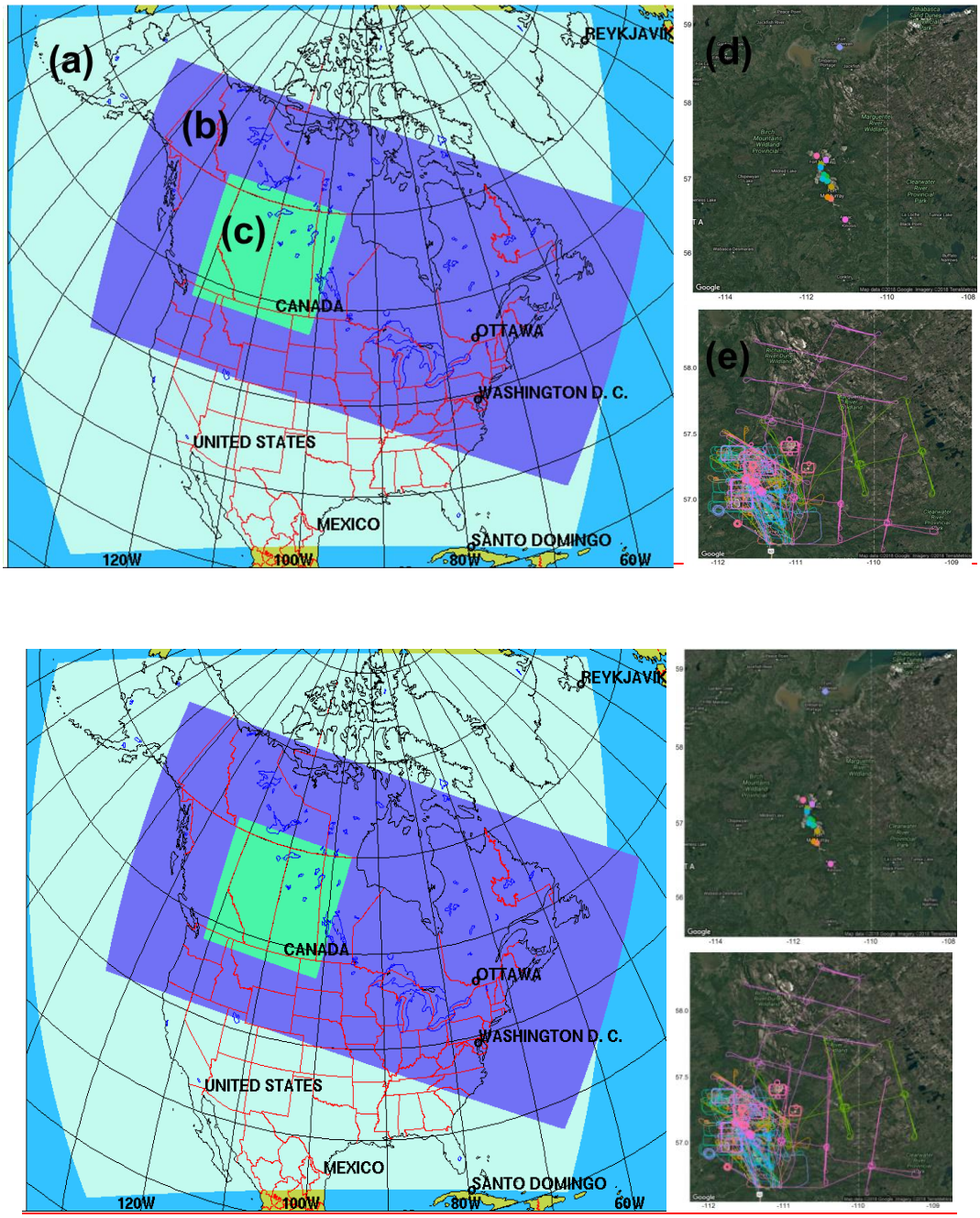



Figure 1: Schematic diagram showing the model simulation domains in the nested 2.5km resolution setup~~.~~ in the right panel:. (a) Light blue outermost domain: GEM-MACH 10km resolution North American forecast. (b) Dark blue domain: HRDPS 2.5km weather forecast. (c) Green innermost domain: GEM-MACH 2.5km forecast. ~~(d)~~The upper left panel is a GoogleEarth-referenced image showing the locations of the surface observations used in the study are shown in colored dots. ~~(e)~~And the lower left panel is a GoogleEarth-referenced image showing all 22 flight paths covered during the JOSM 2013 flight campaign.

The emissions used in our simulations were processed from inventory data from different sources, including the Canadian National Pollutant Release Inventory (NPRI) and Air Pollutant Emissions Inventory (APEI) data for 2013, within-facility specific 2010 data from the Cumulative Environmental Management Association (CEMA), and hourly Continuous Emissions Monitoring observations for hourly major point emissions of $SO_2$ and $NO_2$ for the province of Alberta (Alberta Environment and Parks). The latter sources account for 77% and 43% of total $SO_2$ and $NO_x$ emissions, respectively, from all NPRI point sources in Alberta, and 99% and 39% respectively for sources of these compounds solely within the Athabasca oil sands area (Zhang *et al*., 2018). The same set of emissions was used for all the simulation scenarios carried out for this study. The emissions set is discussed in detail in Zhang *et al* (2016, 2018). In the emissions processing, aerosols were chemically speciated for the 12-bin size distribution; the resulting emissions files were summed to the 2-bin distribution for the 2-bin simulations discussed below.

For the purpose of this study we have carried out 4 sets of model simulations, in order to evaluate the impact of (operational) 2-bin versus 12-bin aerosol size distribution, and of different algorithms for plume rise, on model performance.

**2.2.1 2-bin versus 12-bin model scenarios**

Gong *et al. (*2003) showed that a 12-bin sectional model is sufficient to accurately predict both aerosol number concentration and mass size distributions for most prevalent atmospheric conditions. However, because of the high computational cost and the requirement for a fast turn-around demanded in operational systems, the operational forecast configuration of GEM-MACH employs a 2-bin aerosol size distribution (bins bordered by diameter size cuts 0.01, 2.56, and 10.24 μm), with sub-binning used for those aerosol microphysical processes requiring more detailed aerosol sizing, such as nucleation (Moran *et al*., 2010). The 12-bin configuration (bins bordered by diameter size cuts 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28, 2.56, 5.12, 10.24, 20.48 and 40.96 μm) has been used for research purposes such as investigating aerosol–weather feedbacks (Makar *et al*., 2015a,b). ~~However, because of the high computational cost and the requirement for a fast turn-around demanded in operational systems, the operational forecast configuration of GEM-MACH employs a 2-bin aerosol size distribution, with sub-binning used for those aerosol microphysical processes requiring more detailed aerosol sizing, such as nucleation (Moran *et al*., 2010). The 12-bin configuration has been used for research purposes such as investigating aerosol–weather feedbacks (Makar *et al*., 2015a,b).~~ Here, both aerosol size distributions were used for 10-km and 2.5-km resolution nested simulations. The first two of our simulations are thus referred to as "2-bin" and "12-bin", and both make use of the original plume rise algorithms, employed in GEM-MACH (and described below). These simulations were compared to determine the relative impact of the more detailed size distribution on model performance relative to observations.

Formatted: English (U.K.)

**2.2.2 Plume rise algorithms: two alternative approaches**

As noted earlier, the set of empirical formulations and algorithms developed by Briggs (1984) for evaluating the plume rise height of major point source emissions has been the basis of plume rise calculations in several chemistry transport models such as GEM-MACH (Moran *et al*., 2010) and CMAQ/CMAx (Byun and Schere, 2006), as well as in regulatory air dispersion models such as AEROPOL (Kaasik and Kimmel, 2003) and CALPUFF (Levy *et al*., 2002). However, the details of how Brigg's algorithms were implemented may vary – we therefore provide the details of the GEM-MACH implementation, below. We follow with a revised plume-rise calculation procedure which in our subsequent evaluation is demonstrated to provide a more accurate estimation of final plume height. Our "12-bin" simulation noted above makes use of the original algorithm, while we refer to the revised algorithm as "Plume Rise" in our subsequent discussion.

The original implementation of the plume rise algorithm in GEM-MACH is based on the set of equations in Briggs (1984) which calculate the plume rise height above the top of the emitting stack, $\Delta h$, based on the atmospheric turbulence characteristics at the stack location. The formulae rely on a local estimation of the state of the atmosphere in the vertical at that location; the atmospheric stability, temperature gradients, and resulting formulae for plume height are predicated on the assumption that these stack-height conditions will continue throughout the atmospheric column until the maximum plume height is reached. However, in cases of more complex atmospheric conditions, where these conditions change significantly with height, the formulae may become inaccurate.

The equations depend on atmospheric stability parameters calculated in the meteorological module of the air-quality model, and include the boundary layer height ($H$), the Monin-Obhukov length ($L$), the surface wind friction velocity ($u_*$), the atmospheric temperature ($T_a$) and its gradient (~~$dT_a/dz$~~$\Delta T_a/\Delta z$), and the wind speed (U) at the stack height. An important parameter in the plume rise formulations is the emitted plume's initial buoyancy flux ($F_b$), which is dependent on the stack ~~height ($h_s$), the stack~~ exit temperature ($T_s$~~),~~) and the stack's exit volume flow rate ($V$), and is given by;

$$F_b = \frac{g}{\pi} V \frac{(T_s - T_a)}{T_s} \qquad (1)$$

Where g is the acceleration due to gravity. The emitted plume is buoyant and rises if $T_s > T_a$; $F_b$ is set to zero if $T_s < T_a$. If the stack height is within the predicted boundary layer depth ($h_s < h$), the plume rise is calculated based on the stability regimes at the stack height model level by the following equations:

For unstable conditions ($-0.25\ h_s < L < 0$),

$$\Delta h = \min\left[3\left(\frac{F_b}{U}\right)^{\frac{3}{5}} H_*^{-\frac{2}{5}},\ 30\left(\frac{F_b}{U}\right)^{\frac{3}{5}}\right]. \qquad (2)$$

For stable conditions ($0 < L < 2\ h_s$)

$$\Delta h = 2.6\left(\frac{F_b}{Us}\right)^{\frac{1}{3}}. \qquad (3)$$

And for neutral conditions ($L > 2\ h_s$ *and* $L < -0.25\ h_s$),

$$\Delta h = \min\left[39\frac{F_b^{3/5}}{U},\ 1.2\left(\frac{F_b}{u_*^2 U}\right)^{3/5}\left(h_s + 1.3\frac{F_b}{u_* U}\right)^{2/5}\right]. \qquad (4)$$

Where $H_* = -2.5u_*^3/L$ is the convective scale ~~velocity~~parameter, and $s$ is the stability parameter approximated by:

$$s = \frac{g}{T_a}\left(\frac{~~dT_a~~}{~~dz~~} + \frac{~~g~~}{~~c_p~~}\right)\left(\frac{\Delta T_a}{\Delta z} + \frac{g}{c_p}\right). \qquad (5)$$

Where ~~$dT_a/dz$~~$\Delta T_a/\Delta z$ is the vertical temperature gradient between the atmospheric temperature at the top of the stack and the temperature at the top of the model layer. We note here that some air-quality model implementations make use of one or the other formula of equations (2) and (4), as opposed to the minimum chosen here. In our companion paper (Gordon *et al*~~.,~~ 2018) we show that these differences have little impact on the calculated plume height.

The model also incorporates the potential for the buoyant plume to penetrate the top of the boundary layer (Hanna and Paine, 1988), which is accounted for by calculating the penetration parameter $P$ and using it to further adjust the plume rise $\Delta h$ calculated through the above formulae as:

Formatted: Font: +Body (Times New Roman)

$$P = \begin{cases} 1, & \frac{H - h_s}{\Delta h} \leq 0.5 \\ 1.5 - \frac{H - h_s}{\Delta h}, & 0.5 < \frac{H - h_s}{\Delta h} \leq 1.5 \\ 0, & \frac{H - h_s}{\Delta h} \geq 1.5 \end{cases} \quad (6)$$

Where H is the height of the boundary layer. The plume rise calculated earlier is then reset via;

$$\Delta h = min[(0.62 + 0.38P)(H - h_s), \Delta h] \quad (7)$$

Once the final value of the plume rise $\Delta h$ is calculated, the vertical spread of the plume and the emitted mass is then evaluated by using a common method from Briggs (1975) to specify the height of the top and bottom of the plume as;

$$\begin{matrix} h_{top} = h_s + 1.5\Delta h \\ h_{bottom} = h_s - 0.5\Delta h \end{matrix} \quad \begin{matrix} h_t = h_s + 1.5\Delta h \\ h_b = h_s - 0.5\Delta h \end{matrix} \quad (8)$$

In GEM-MACH, the plume top is further limited to the height of the boundary layer ($H$), if the penetration $P > 0$. During unstable conditions, the plume bottom is set to zero (the surface); that is, the plume is assumed to mix uniformly betweenthroughout the top of the atmosphere and the surface.boundary layer. We also note that the mass emitted into the plume is assumed to mix uniformly between $h_{top}$ $h_t$ and $h_{bottom}$ $h_b$; this is in contrast to the approach of Turner (1991), wherein a "top-hat" distribution centered on the value of hs $h_s$ was assumed, or the Gaussian distribution based on unpublished observations described in Byun and Ching (1999).

As described above, the original plume rise algorithm implemented in GEM-MACH does not account for potential changes in plume rise associated with the vertical variation in the atmospheric temperature and stability, which could be important for plume buoyancy especially during unstable conditions where the boundary layer depth could be much higher than the stack height. Similarly, changes in stability with height will affect plume rise. As reported in Gordon *et al.* (2018), when meteorological *observations* collected at oil sands sites are used to drive equations (1) through (5), the estimated plume heights were often underestimated, with between 37 to 52 percent of calculated values being less than ½ the observed height.

However, other approaches, which take into account the variation in height associated with atmospheric conditions in the vertical profile above the emitting stack, are available. Briggs proposed equations which would make use of changes in stability between layers and calculate the residual buoyancy flux between layers in the atmosphere – these are particularly amenable to the layered structure of atmospheric models (Briggs, 19851984, equations 8.84 and 8.85). This new algorithm is similar to other layer-by-layer approaches available in CMAQ (Byun and Ching, 1999), based on the hesitant-plume algorithm described in Turner (1991) and in dispersion modelling work by Erbrink (1994). In the new algorithm (hereafter referred to as the revised Briggs Plume rise or simply "plume rise") we utilized the model's calculated vertical profile of atmospheric temperature and wind speed to estimate the plume height as the height at which the emitted plume buoyancy flux dissipates totally. The initial plume buoyancy flux ($F_b$) at the top of the stack is calculated using equation (1) above, by using linear interpolation to evaluate the air temperature ($T_a$) and wind speed ($U$) at the stack height from the model's vertical profile. Under (locally) neutral and stable conditions, the buoyant plume is assumed to rise freely, and the residual buoyancy flux ($F_r$), remaining after it as it crosses the next atmospheric layer is given by:

$$F_{j+1} = \begin{cases} F_j - 0.015 s_j F_{j-1}^{\frac{1}{3}} (z_{j+1}^{\frac{8}{3}} - z_j^{\frac{8}{3}}), & \text{vertical plumes} \\ F_j - 0.053 s_j U_j (z_{j+1}^{3} - z_j^{3}), & \text{bent plumes} \end{cases} \quad (9)$$

For vertical plumes,

$$F_{j+1} = F_j - 0.015 s_j F_{j-1}^{\frac{1}{3}} \left( z_{j+1}^{\frac{8}{3}} - z_j^{\frac{8}{3}} \right), \qquad (9a)$$

and for bent plumes

$$F_{j+1} = F_j - 0.053 s_j U_j \left( z_{j+1}^3 - z_j^3 \right), \qquad (9b)$$

Here, ~~$s_j$~~$s_j$ is the local stability parameter for a given layer, calculated using (5) and layer-specific temperature values, and ~~$z_j$~~$z_j$ is the plume rise height when the plume reaches the bottom of the model's ~~j~~j'th layer. Briggs (1984) recommended the use of *both* formulae of (9), with the formula with greatest decrease in flux being used as the final value. Briggs also noted that the transition to bent plumes happens at a relatively low height above the stack, implying that that the residual buoyancy between layers is lost faster under windy conditions. At the stack height, ~~$F_{j=0} = F_b$~~$F_{j=0} = F_b$, and ~~$= h_s$~~$z_j = 0$ (that is, the vertical distances are relative to the top of the stack).~~. When the residual buoyancy flux becomes negative in (9), indicating that the plume height has been surpassed, the calculation is repeated to find the value of ~~$z$~~$z$ for which ~~$F=0$~~$F = 0$; the sum of this and the layer thicknesses transitioned to this height becomes the predicted plume rise. In our companion paper (Gordon *et al.*, 2018), this approach was found to provide similar results to the original Briggs' algorithms when driven by observations not co-located with the stacks. Our work here indicates that this algorithm has the potential to provide a more accurate estimate of plume rise, subject to caveats described below.

We note that the numerical coefficients in (9); 0.015 and 0.053, stem from two parameters; the entrainment constant for vertical rise conditions ($\alpha$, the entrainment coefficient for vertical plumes, nominally set to 0.08 by Briggs based on observations published in 1975 - the parameter in the first equation of (9) is a non-linear function of this $\alpha$ term; and $\beta$', the entrainment coefficient internal radius for bent-over plumes, set ~~to~~by Briggs ~~as~~to 0.4, though ranges from 0.45 to 0.52 were quoted elsewhere in Briggs, 1984). The choice of these parameters are based on data which are now over 40 years old, and may present an opportunity for future improvement of this revised plume rise approach.

The above formula (9) was recommended by Briggs for conditions which are stable to neutral at the stack height. We have defined stability in this case by comparing the dry adiabatic lapse rate to the local temperature lapse rate predicted by the model at the stack height and above. Briggs (1984) provided no equivalent formula for unstable conditions at the stack height, followed by stable profiles at higher elevations. The approach taken here has been to assume under convectively unstable conditions, the plume rises without loss of energy (that is, an assumption of zero entrainment) until the predicted temperature profile once again falls below the dry adiabatic lapse rate. Our first order approximation is thus to assume that under unstable conditions, there is minimal mixing entrainment of the rising plume with the surrounding atmosphere. This approach differs from that of Turner (1991), and the layered approach described in Byun and Ching (1999) where the residual buoyancy flux between layers is determined using different formulae based on the model-determined local atmospheric stability.

As in the original algorithm, the plume top and plume bottom are evaluated using equation (8) after the final plume rise has been evaluated. We do not apply the penetration equations (Eq.6 and Eq.7) since these corrections should be unnecessary in an approach making use of local changes in residual buoyancy. In our companion paper, this algorithm wa~~i~~s referred to as the "layered approach".

### 2.2.3 Hourly Emission Stack Temperature and Volume Flow Rate

We turn next to the available emissions data for driving the plume rise algorithms. Under Canadian federal reporting requirements to the National Pollutant Release Inventory (NPRI), annual total emissions of $SO_2$ and $NO_x$ from facilities are reported, along with a single set of stack parameters (stack height, stack diameters, average exit temperature, and average exit

velocity) to represent emissions throughout the year. In addition, hourly Continuous Emissions Monitoring data from large stacks are reported to the government of Alberta. These data include the hourly mass of emissions of $SO_2$ and $NO_2$, as well as hourly estimates of the time-varying stack parameters (volume flow rates and temperatures).

### 2.2.4 Simulation Scenarios

Our first two simulations use the "standard" annual NPRI reported stack parameters and the original plume rise algorithm for the 2-bin and 12-bin aerosol size distributions, while our second two simulations use the modified plume rise algorithm, first with the NPRI stack parameters, and second with emissions information derived from a combination of CEMS hourly stack parameters as well as engineering estimates of emissions during "upset conditions" in which the effluent is redirected to flare stacks (the latter estimates are considerably more uncertain than the CEMS information, but are nevertheless included here since they result in substantial changes in pollutant emissions and plume characteristics, see Zhang et al, 2018). ~~and second with CEMS derived hourly stack parameters.~~ The four scenarios examined are thus:

(1) A "2-bin" simulation: NPRI stack parameters, 2-bin aerosol size distribution, and the original plume rise

(2) A "12-bin" simulation: As in (1), but employing the 12-bin aerosol size distribution. Differences between (1) and (2) thus show the impact of the aerosol size distribution on performance.

(3) A "Plume rise" simulation: employing the layered plume rise algorithm, with emissions as in (2) Differences between (2) and (3) thus show the impact of the revised plume rise algorithm alone.

(4) An "Hourly" simulation: employing the layered plume rise algorithm, with volume flow rates and temperatures taken from the hourly CEMS data along with upset conditions. Differences between (3) and (4) thus show the impact of the initial buoyancy flux on the resulting plume rise, using the revised algorithm.

All of these simulations make use of the CEMS-derived mass of emitted $SO_2$ and $NO_x$.

## 3 Observations

The comparative statistics presented through this study were computed using the '~~modstat~~'modStat' function from the ~~openair~~'openair' R package (Carslaw and Ropkins, 2012), for complete pairs of valid model and observation data. The set of statistical measures and their formulas are presented in Table 2. Both surface monitoring network and aircraft observations have been used for model evaluation.

### 3.1 WBEA Surface Monitoring Networks

For the purpose of model evaluation, we have used hourly measurements of surface concentrations of PM2.5, $SO_2$, $NO_2$, and $O_3$ from a network of 10 air quality monitoring stations in the province of Alberta managed by the Wood Buffalo Environmental Association (WBEA) (see Figure 1(~~d~~upper-left panel)). The observation data have been filtered to remove extreme single-hour measurements that are greater than 150ppbv for $SO_2$, $NO_2$, and $O_3$, and 150 µg $m^{-3}$ for PM2.5 (single-hour spikes of this nature in hourly records are assumed to correspond to instrumentation errors or calibration times for the instruments). ~~The observation data have been filtered to remove extreme single-hour measurements that are greater than 150ppbv for $SO_2$, $NO_2$, and $O_3$, and 150 µg $m^{-3}$ for PM2.5.~~ Observations from August 10th, 2013 to September 10th, 2013 were selected for comparison to the model results, to align to with the period covered by the JOSM 2013 intensive aircraft measurement campaign.

Formatted: Font: Bold

Formatted: Font: 10 pt, Not Italic

Formatted: Not Superscript/ Subscript

Formatted: Not Superscript/ Subscript

### 3.2 JOSM Summer 2013 Intensive Campaign

From August 10th to September 10th, 2013, the National Research Council of Canada Convair aircraft was used as a mobile measurement platform to sample atmospheric constituents in the region of the Athabasca oil sands, with twenty-two flights taking place during the given time period (Figure ~~1c~~1, lower-left panel). These flights included flight paths designed for emission estimation, for the study of downwind transport and chemical transformation, and for satellite validation. Emission estimation flights took place around individual facilities at multiple altitudes, with the concentration and meteorological information gathered subsequently used to estimate fluxes entering and leaving the facility, and hence estimate emissions directly from aircraft observations (Gordon *et al*., 2015; Li *et al.*, 2017) Transformation flights were designed to follow plumes downwind, with observations taken in cross-sections at set distances downwind perpendicular to the plume direction, in order to study chemical transformations between point of emission and downwind receptors (cf. Liggio *et al*., 2016). Satellite validation flights incorporated aircraft vertical spirals at satellite overpass times, in order to improve satellite data retrieval algorithms (Whaley *et al*., 2018; Sheppard *et al.*, 2015). Here, we compare model predictions for our different simulations for $SO_2$, $NO_2$ and for PM1 sulfate, ammonium, and total organics to observations taken on-board the Convair using TS43, TS42 and Aerodyne Aerosol Mass Spectrometers (AMS) instruments, respectively. In order to allow for comparisons to the results from GEM-MACHv2 2.5km ~~oilsands~~oil sands model domain simulations, 10-second averages of the aircraft's positional data (latitude, longitude, elevation, and time) were created for all 22 flights. These data were in turn used to extract the ~~corresponding time and spatial linearly interpolated model values at the model's chemistry time resolution of 2 minutes~~corresponding linearly interpolated values in time and space from the model's 2 minute time step and 2.5km resolution, for each of the ~~instruments~~ species observed aboard the aircraft which~~that~~ were used for the model comparison. The nominal cruise speed of the National Research Council Convair 580 used in the experiment is 550 km/hour; a 10 second time interval thus represents an observation integration distance of 1.528 km, and a two minute time interval an observation integration distance of 18.3 km

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

## 4 Results and Discussion

### 4.1 Spatial Heterogeneity of Meteorological Conditions

We noted in our companion paper (Gordon *et al.*, 2018) that meteorological observations varied substantially in the study region depending on location, citing this as a possible confounding factor on the results of tests of the plume rise algorithms. This spatial heterogeneity was well captured by the high resolution GEM-MACH simulations, as is demonstrated by the example depicted in Figure 2, which shows the typical local variation in planetary boundary layer height (Figure 2(a)), ranging from about 1200m to 400m, the lower values corresponding to the main cleared areas (open pit mines, settling ponds) of the industrial facilities. The corresponding temperature profiles in several locations marked in Figure 2(a) are given in Figure 2(b): These show a substantial difference in model predicted stability at the three meteorological observation locations of Gordon *et al.* (2018) (windRASS, AMS03, and AMS05), and substantial differences between these and the locations of the main stacks of some of the facilities (Syncrude 1, CNRL, and Suncor). The temperature profiles show that the height and strength of the inversion may vary by over 100m in the vertical, and that the profiles do not merge with the larger scale flow until an elevation of 750m asl (450m agl) is reached. Given this level of variation, we might expect potential errors in calculated plume heights when applying the meteorological observations to plume rise at the stack locations, in turn suggesting that a re-examination of plume rise using the model results is worthwhile.

Formatted: Font: Italic

Formatted: Font: Italic

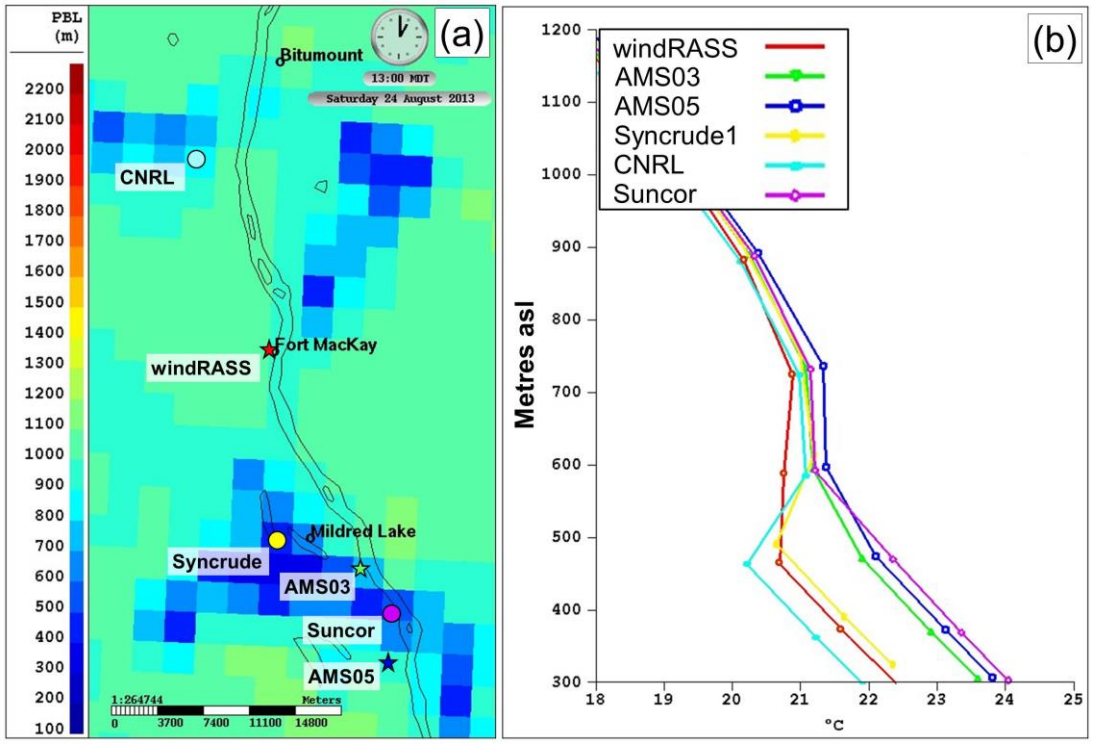



Figure 2: Examination of meteorological heterogeneity in the study area. (a) PBL heights (locations of large emitting stacks shown as circles, meteorological observation sites as stars). (b) Predicted temperature profiles at these locations and times; symbols indicate locations of model levels (a terrain following coordinate system is used in GEM-MACH).

## 4.2 Two-bin versus Twelve-bin Evaluation

We begin our evaluation by comparing the 2-bin and 12-bin particle size distribution simulations using identical emissions against Wood Buffalo Environmental Association's surface monitoring network PM2.5 measurements. The statistical comparison between these observations and all the 4 model scenarios is shown in Table ~~1~~3, and the corresponding histograms of observations (blue), 2-bin model simulated values (red) and 12-bin model simulation values (purple) is shown in Figure ~~2~~3. The statistics of Table ~~1~~3 show that the 12-bin simulation provides an overall improvement over the 2-bin model results across all metrics. For example, the magnitude of the ~~negative~~mean bias has decreased ~~by~~from -2.623 to -1.725 μg $m^{-3}$, a reduction of 34%, indicating that a sizeable fraction of particulate under-predictions in 2-bin simulations may be due to poor representation of particle microphysics through the use of the 2-bin distribution, despite sub-binning being used in some microphysics processes. The largest improvement in correlation coefficient and fraction of predictions within a factor of two also takes place going from the 2 to the 12 bin distribution. Figure ~~2~~3 shows that the model simulations are biased high for ~~particles~~PM2.5 concentrations less than 5 μ~~m diameter~~g $m^{-3}$, and are biased low for ~~the larger particle sizes.~~higher concentrations. ~~However, t~~The use of the 12-bin size distribution (purple histogram bars, Figure ~~2~~3) improves the fit to the observations (blue histogram bars), in comparison to the 2-bin distribution results (red histogram bars), though significant over-predictions of the frequency of low concentration events and under-prediction of high concentration events, remain.~~.~~

Formatted: Not Superscript/ Subscript
Formatted: Font: Symbol
Formatted: Superscript
Formatted: Superscript

### 4.3 Plume Rise Algorithm Evaluation

The simulation with the largest number of highest scores (bold-face numbers in Table ~~1~~3) is the "Plume rise" algorithm , which made use of the revised plume rise formulation, though the differences in performance between the "12-bin", "Plume rise" and "Hourly" simulations are relatively small. The latter small increment is expected, given that the observations are relatively close to the sources of primary particulate emissions, largely from surface sources of fugitive dust (see Zhang *et al.*, 2018). However, an increment of PM2.5 will be from secondary sources; about 99% of the anthropogenic $SO_2$ and $NH_3$ emissions, and about 40% of the $NO_x$ emissions in the Athabasca oil sands region originate in "major point source" stacks. The concentrations of these precursor species will therefore be influenced by the plume rise algorithm employed in model simulations, and hence secondary particulate species originating from these primary emissions may also be affected by plume rise. The small improvements in PM2.5 associated with the revised plume rise algorithm may thus represent the impact of secondary formation of particulate sulphate, ammonium and nitrate from $SO_2$, $NH_3$, and NOx, the latter having been influenced by the plume rise treatment. We examine this possibility using observations of PM1 particle sulphate, and ammonium taken with an aerosol mass spectrometer (AMS) aboard the NRCan Convair aircraft.

The aircraft's AMS instrument measures speciated atmospheric particle concentrations for particles less than 1µm size, and therefore cannot be compared with the 2-bin model results ~~(~~because the ~~two size bins are from 0 to 2.56 µm and 2.56 to 10.24 µm, hence the~~ smaller size bin (with upper diameter size cut 2.56 µm) will be biased high relative to the 1 µm size cut of the AMS~~).~~. While the modelled PM1 organic aerosols (OA) compared similarly to the AMS measurements for all the three~~3~~ model scenarios employing 12 particle bins, the PM1 sulphate and ammonium simulations with revised plume rise algorithm ("Plumerise" and "Hourly" simulations) produced better scores for most statistics than the ~~"12 bin"~~ simulation ~~compared~~ employing the ~~to the~~ original plume rise algorithm~~.~~, as shown in Table 4. Particulate sulphate largely originates in atmospheric oxidation of $SO_2$ by the OH radical in these flights – relatively little sulphate is emitted directly, and aqueous oxidation is largely absent due to the flights being cloud-free. Particle ammonium levels are closely linked to the sulfate through inorganic chemistry as well as being emitted by stacks in this region, and hence the ammonia results are consistent with the sulphate. The organic aerosols are at this distance downwind largely due to formation from area emissions sources of primary organic aerosol and of precursor volatile organic compounds to secondary organic aerosol formation, rather than large stack emissions, and hence are less affected by the plume rise treatment. A larger influence of plume rise on model results is expected for $SO_2$ and $NO_2$, due to the large fraction of their emissions originating in the large stacks of the Athabasca oil sands facilities.

**Formatted:** Not Superscript/ Subscript

**Formatted:** Not Superscript/ Subscript

**Formatted:** Not Superscript/ Subscript

**Formatted:** Not Superscript/ Subscript

**Formatted:** Not Superscript/ Subscript

**Formatted:** English (U.S.)

~~**Table 1**~~

**Table 2**: Statistical measures used in comparing model results with observations.

| Statistic | Formula |
|---|---|
| | $M_i$ = model time series; $O_i$ = observation time series |
| Number of complete data pair, | $n$ |
| Fraction of predictions within a factor of two, | $FAC2 = 0.5 \leq \frac{M_i}{O_i} \leq 2.0$ |
| Mean bias, | $MB = \frac{1}{n}\sum_{i=1}^{N} M_i - O_i$ |
| Mean Gross Error, | $MGE = \frac{1}{n}\sum_{i=1}^{N} \lvert M_i - O_i \rvert$ |
| Normalised mean bias, | $NMB = \frac{\sum_{i=1}^{n} M_i - O_i}{\sum_{i=1}^{n} O_i}$ |
| Normalised mean gross error, | $NMGE = \frac{\sum_{i=1}^{n} \lvert M_i - O_i \rvert}{\sum_{i=1}^{n} O_i}$ |
| Root mean squared error, | $RMSE = \left(\frac{\sum_{i=1}^{n}(M_i - O_i)^2}{n}\right)$ |
| Correlation coefficient, | $r = \frac{1}{(n-1)}\sum_{i=1}^{n}\left(\frac{M_i - \bar{M}}{\sigma_M}\right)\left(\frac{O_i - \bar{O}}{\sigma_O}\right)$ |
| Coefficient of Efficiency, | $COE = 1.0 - \frac{\sum_{i=1}^{n} \lvert M_i - O_i \rvert}{\sum_{i=1}^{n} \lvert O_i - \bar{O} \rvert}$ |
| Index of Agreement, | $IOA = \begin{cases} 1.0 - \frac{\sum_{i=1}^{n} \lvert M_i - O_i \rvert}{c\sum_{i=1}^{n} \lvert O_i - \bar{O} \rvert}, when \sum_{i=1}^{n} \lvert M_i - O_i \rvert \leq c\sum_{i=1}^{n} \lvert O_i - \bar{O} \rvert \\ \frac{c\sum_{i=1}^{n} \lvert O_i - \bar{O} \rvert}{\sum_{i=1}^{n} \lvert M_i - O_i \rvert} - 1.0, when \sum_{i=1}^{n} \lvert M_i - O_i \rvert > c\sum_{i=1}^{n} \lvert O_i - \bar{O} \rvert \end{cases}$ |

**Table 3**: Statistical comparison of GEM-MACH model simulation of surface PM2.5 with measurements from the WBEA observations between August 10th and September 10th, 2013. **Bold** face: best score. *Italics*: second best score.

| Statistic | PM2.5(μg/m$^3$) | | | |
|---|---|---|---|---|
| | **2-bin** | **12-bin** | **Plumerise** | **Hourly** |
| ~~Number of complete data pair, **nn**~~ | 6815 | 6815 | 6815 | 6815 |
| ~~Fraction of predictions within a factor of two, ***FAC2***~~***FAC2*** | 0.386 | 0.454 | **0.456** | *0.455* |
| ~~Mean bias, ***MB***~~ | -2.623 | **-1.725** | -1.813 | *-1.807* |
| ~~Mean Gross Error, ***MGE***~~ | 4.852 | 4.742 | **4.690** | *4.696* |
| ~~Normalised mean bias, ***NMB***~~ | -0.390 | **-0.257** | -0.270 | *-0.269* |
| ~~Normalised mean gross error, ***NMGE***~~ | 0.722 | 0.705 | **0.698** | *0.699* |
| ~~Root mean squared error, ***RMSE***~~ | 8.447 | 8.442 | **8.359** | *8.363* |
| ~~Correlation coefficient, ***rr***~~ | 0.122 | 0.151 | *0.154* | **0.155** |
| ~~Coefficient of Efficiency, ***COE***~~ | -0.213 | -0.185 | **-0.172** | *-0.174* |

| | | | | |
|---|---|---|---|---|
| ~~Index of Agreement,~~ ***IOA*** | 0.394 | 0.407 | **0.414** | *0.413* |

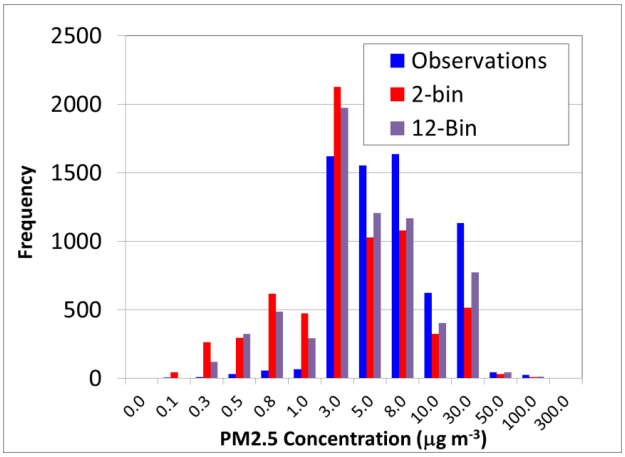



Figure ~~2~~3. Histogram of surface PM2.5 using Wood Buffalo Environmental Association surface monitoring data (blue), and the 2-bin (red) and 12-bin (purple) configurations of GEM-MACH. Both simulations make use of the original Briggs (1984) plume rise formulation.

**Table ~~2~~4**: Statistical comparison of PM1 sulphate, OA and ammonium atmospheric concentration from the aircraft AMS with the 2.5km resolution GEM-MACH simulations between August 13th and September 9th, 2013. **Bold** face: best score. *Italics*: second best score.

| | $SO_4$(µg/m$^3$) | | | OA(µg/m$^3$) | | | $NH_4$(µg/m$^3$) | | |
|---|---|---|---|---|---|---|---|---|---|
| **Statistic** | **12-bin** | **Plumerise** | **Hourly** | **12-bin** | **Plumerise** | **Hourly** | **12-bin** | **Plumerise** | **Hourly** |
| ~~**nn**~~ | 24523 | 24523 | 24523 | 24522 | 24522 | 24522 | 24523 | 24523 | 24523 |
| ***FAC2*** | **0.475** | *0.467* | 0.466 | **0.138** | *0.137* | *0.137* | **0.527** | **0.527** | *0.525* |
| ***MB*** | *-0.445* | -0.489 | **-0.435** | **-2.67** | *-2.669* | *-2.669* | **-0.014** | -0.034 | *-0.023* |
| ***MGE*** | 0.964 | **0.925** | *0.959* | **2.725** | *2.727* | *2.727* | 0.272 | **0.252** | *0.261* |
| ***NMB*** | *-0.397* | -0.436 | **-0.388** | -0.714 | *-0.713* | *-0.713* | **-0.051** | -0.121 | *-0.081* |
| ***NMGE*** | 0.861 | **0.826** | *0.856* | **0.728** | *0.729* | *0.729* | 0.976 | **0.904** | *0.937* |
| ***RMSE*** | 4.629 | **4.592** | *4.608* | 3.773 | 3.773 | 3.773 | 1.176 | **1.124** | *1.141* |
| ***r*** | 0.148 | *0.171* | **0.175** | **0.552** | *0.548* | *0.549* | 0.149 | *0.175* | **0.178** |
| ***COE*** | 0.203 | **0.235** | *0.207* | -0.111 | -0.111 | -0.111 | 0.061 | **0.13** | *0.099* |
| ***IOA*** | 0.601 | **0.618** | *0.603* | **0.445** | *0.444* | *0.444* | 0.53 | **0.565** | *0.549* |

The performance of the three model simulations using different plume rise algorithms, for surface mixing ratios of $SO_2$ observed at WBEA stations, is shown in Figure ~~3~~4. The model simulations are biased low for zero concentration levels (first bin, Figure ~~3~~4(a)), are biased high from 0.0 to 0.3 ppbv, biased low from 0.3 to 1.0 ppbv (Figure ~~3~~4(a)), and biased high for all

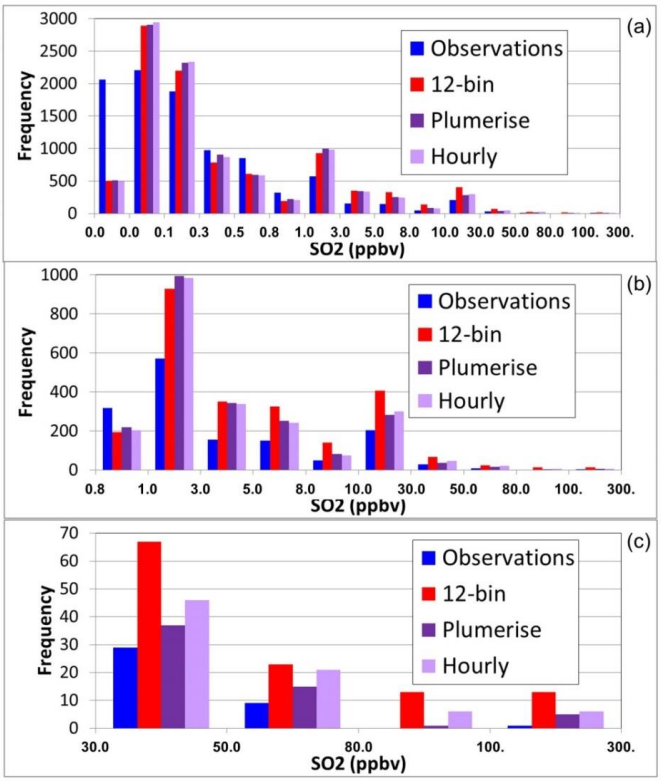
Formatted: Not Superscript/ Subscript
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic

concentrations above 1 ppbv (Figure ~~3~~4(b,c)). These last two ranges (Figure ~~3~~4(b,c)) result from surface fumigation of high concentration plumes in the region studied. While all model simulations are biased high for these fumigating plumes, the "Plumerise" and "Hourly" simulations have a reduced bias compared to the original plume rise algorithm "12-bin". The use of the hourly stack parameters derived from Continuous Emissions Monitoring ("Hourly") has somewhat worse performance than the same plume rise algorithm driven by annual reported stack parameters ("Plumerise").

The $SO_2$ statistics of Table ~~3~~5 show sometimes substantial improvements in model performance with the use of the revised plume rise algorithms, with the mean bias being reduced by 61%, the mean gross error by 27%, the correlation coefficient increasing by 26% and the index of agreement increasing by a factor of 2.26 between the 12-bin and "Plumerise" algorithms, and the "second best" score (italics) out of the three simulations being the "Hourly" simulation employing the revised volume flow rates and stack temperatures. For $NO_2$, the "Hourly" values (employing the hourly volume flow rates and temperatures) tend to have the best scores, though the differences between "Hourly" and "Plumerise" simulations, where the only difference in the plume treatment is in the source of data for the initial buoyancy flux, is relatively small. Both of the primary pollutants have shown a noticeable improvement in performance with the new plume rise treatment, with the pollutant for which most emissions are from stacks ($SO_2$) having the most noticeable changes.

Ozone, in contrast, is created or destroyed through secondary chemistry over relatively longer time-spans than the transport time from the sources in this comparison (spatial scales on the order of 10's of km). Accordingly, the impact of the plume rise of $NO_x$ on ozone formation is relatively minor, usually in the third decimal place (though first decimal place improvements occur for the mean bias with the use of the new plume rise algorithm).

**Formatted:** Justified


(a)
Frequency
3000
2500
2000
1500
1000
500
0
Observations
12-bin
Plumerise
Hourly
0.0 0.0 0.1 0.3 0.5 0.8 1.0 3.0 5.0 8.0 10.0 30.0 50.0 80.0 100. 300.
SO2 (ppbv)
(b)
Frequency
1000
800
600
400
200
0
Observations
12-bin
Plumerise
Hourly
0.8 1.0 3.0 5.0 8.0 10.0 30.0 50.0 80.0 100. 300.
SO2 (ppbv)
(c)
Frequency
70
60
50
40
30
20
10
0
Observations
12-bin
Plumerise
Hourly
30.0 50.0 80.0 100. 300.
SO2 (ppbv)

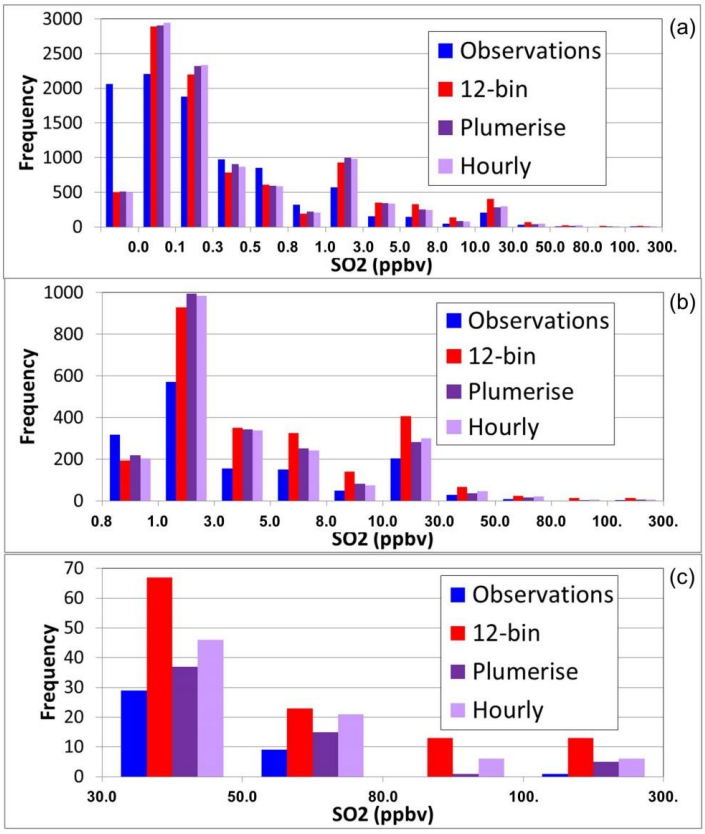



Figure ~~3~~4: Histograms of hourly surface $SO_2$ mixing ratios, in logarithmic mixing ratio bins, observations (blue), original plume rise algorithm (red), revised plume rise algorithm (dark purple), revised plume rise algorithm driven by hourly CEMS stack data (light purple). (a) All values. (b) 0.8 to 300 ppbv. (c) 30 to 300 ppbv.

**Table ~~3~~5**: Statistical comparison of $SO_2$, $NO_2$, and $O_3$ surface ~~concentration~~volume mixing ratio measurements from the WBEA surface observation network with the 2.5km resolution GEM-MACH simulations between August 10th and September 10th, 2013.

| **Statistic** | **$SO_2$(~~ppb~~ppbv)** | | | **$NO_2$(~~ppb~~ppbv)** | | | **$O_3$(~~ppb~~ppbv)** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **12-bin** | **Plumerise** | **Hourly** | **12-bin** | **Plumerise** | **Hourly** | **12-bin** | **Plumerise** | **Hourly** |
| **n** | 9457 | 9457 | 9457 | 6516 | 6516 | 6516 | 4384 | 4384 | 4384 |
| ***FAC2*** | 0.226 | **0.239** | *0.238* | 0.324 | *0.328* | **0.329** | **0.778** | *0.777* | *0.777* |
| ***MB*** | 1.215 | **0.476** | *0.631* | 1.006 | *0.919* | **0.910** | -0.977 | *-0.907* | **-0.898** |
| ***MGE*** | 2.41 | **1.756** | *1.906* | 4.456 | *4.388* | **4.384** | **7.239** | *7.269* | 7.271 |
| ***NMB*** | 1.133 | **0.444** | *0.588* | 0.262 | *0.239* | **0.237** | *-0.051* | **-0.047** | **-0.047** |
| ***NMGE*** | 2.247 | **1.637** | *1.777* | 1.159 | *1.141* | **1.140** | **0.375** | *0.376* | 0.377 |
| ***RMSE*** | 8.99 | **6.291** | *7.018* | 7.742 | *7.656* | **7.651** | **9.690** | *9.750* | 9.771 |
| ***r*** | 0.179 | **0.227** | *0.218* | **0.287** | **0.287** | *0.286* | **0.617** | *0.613* | 0.612 |
| ***COE*** | -0.648 | **-0.201** | *-0.304* | -0.264 | *-0.244* | **-0.243** | **0.220** | *0.217* | *0.217* |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ***IOA*** | 0.176 | **0.399** | *0.348* | *0.368* | **0.378** | **0.378** | **0.610** | *0.609* | 0.608 |

Overall, these results suggest that ~~(a)~~ the revised plume rise algorithm improves the model surface performance for primary pollutants largely emitted from stack sources ($SO_2$) or for which a large proportion of the emitted mass is via stack sources ($NO_2$). Also, the impact of the hourly volume flow rates and temperatures versus typical annual values is relatively small, though it results in a degradation of performance.

Statistical comparisons of model results computed against aircraft observations for $SO_2$, $NO_2$ and $O_3$, for all the flights in the aircraft campaign are shown in the Table ~~4~~6. Histograms of model performance for $SO_2$ aloft are shown in Figure ~~4~~5. With the exception of more negative biases, the two sets of atmospheric $SO_2$ concentrations calculated by the new plume rise algorithm driven using annual reported stack parameters again give the best results when compared to the aircraft measurements, for all statistical measures aside from the biases (the "Plume rise" and "Hourly" simulations are biased lower than the 12-bin simulation). The variation in the statistical performance between different plume rise algorithms aloft are larger than those noted above for the surface observation comparisons, for the model scenario with "Plume rise" and "Hourly" scenarios, with the former having the best overall performance. A more substantial improvement for $NO_2$ with the revised plume rise algorithm may be seen in comparison to the surface observation evaluation, with larger decreases in the mean bias, mean gross error, root mean square error, and increases in the scores for correlation coefficient, coefficient of error and index of agreement, between the "12-bin" and "Plumerise" simulations. The results for the two simulations using the new plume rise algorithm however remain similar for $NO_2$. It should be noted as well that the model generally performs better against the aircraft measurements than the comparisons to the surface observations across all the statistical measures for $NO_2$, reflecting the aircraft sampling a greater proportion of $NO_2$ mass originating from elevated plumes as opposed to surface sources. Similar to the surface observation comparisons, the atmospheric $O_3$ concentration calculated by the various model scenarios shows very minimal variation in the comparative statistics with the aircraft observation, with the exception of a marginally better correlation coefficient ($r = 0.477$) for original plume rise scenario compared to the result ($r = 0.6947$) for the new plume rise scenarios.

Figure ~~4~~5 shows the histograms comparing aircraft observations with the results of the three variations of plume rise algorithms for $SO_2$ (Figure ~~4~~5(a,b)) and $NO_2$ (Figure ~~4~~5(c,d)). In contrast to Figure ~~3~~4, all model simulations for $SO_2$ aloft are biased *low* between mixing ratios of 0.3 and 50 ppbv, and remain biased low above 50 ppbv for the Plume rise and Hourly simulations. Thus, model estimates of *surface* $SO_2$ mixing ratios (Figure 3) are biased high, while *aloft* (Figure ~~4~~5(a,b)), $SO_2$ mixing are biased low. A similar, though less pronounced, pattern may be seen for $NO_2$ (Figure ~~4~~5(c,d)), with model mixing ratios aloft biased low, for histogram bins between 0.1 and 10 ppbv. All versions of the model thus have a tendency to underpredict the *height* of the plumes, overestimating surface fumigation events, and underestimating occurrences when the plume remains aloft.

**Table ~~4~~6**: Comparison of statistical measures of $SO_2$, $NO_2$, and $O_3$ measurements from the aircraft campaign against the 2.5km resolution GEM-MACH simulations between August 13th and September 10th, 2013.

| **Statistic** | **TS43 - $SO_2$(~~ppb~~ppbv)** | | | **TS42 - $NO_2$(~~ppb~~ppbv)** | | | **TS49 - $O_3$(~~ppb~~ppbv)** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **12-bin** | **Plumerise** | **Hourly** | **12-bin** | **Plumerise** | **Hourly** | **12-bin** | **Plumerise** | **Hourly** |
| **n** | 29313 | 29313 | 29313 | 28114 | 28114 | 28114 | 29263 | 29263 | 29263 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ***FAC2*** | 0.233 | **0.249** | *0.243* | *0.300* | **0.306** | **0.306** | **0.953** | *0.950* | 0.949 |
| ***MB*** | **-0.186** | -0.795 | *-0.444* | *0.097* | **-0.007** | **-0.007** | -1.729 | *-1.539* | **-1.536** |
| ***MGE*** | 4.031 | **3.438** | *3.705* | 1.482 | **1.362** | *1.366* | **8.842** | *8.915* | 8.919 |
| ***NMB*** | **-0.057** | -0.244 | *-0.136* | 0.065 | **-0.005** | **-0.005** | *-0.056* | **-0.050** | **-0.050** |
| ***NMGE*** | 1.237 | **1.055** | *1.137* | 0.988 | **0.908** | *0.911* | **0.287** | *0.290* | *0.290* |
| ***RMSE*** | 14.332 | **11.153** | *12.345* | 3.521 | **2.951** | *2.957* | **11.757** | *11.961* | 11.982 |
| ***r*** | 0.234 | **0.34** | *0.317* | 0.416 | **0.493** | *0.491* | **0.477** | *0.469* | 0.468 |
| ***COE*** | 0.142 | **0.268** | *0.211* | 0.177 | **0.244** | *0.241* | **-0.171** | *-0.180* | -0.181 |
| ***IOA*** | 0.571 | **0.634** | *0.606* | 0.589 | **0.622** | *0.621* | **0.415** | *0.410* | *0.410* |

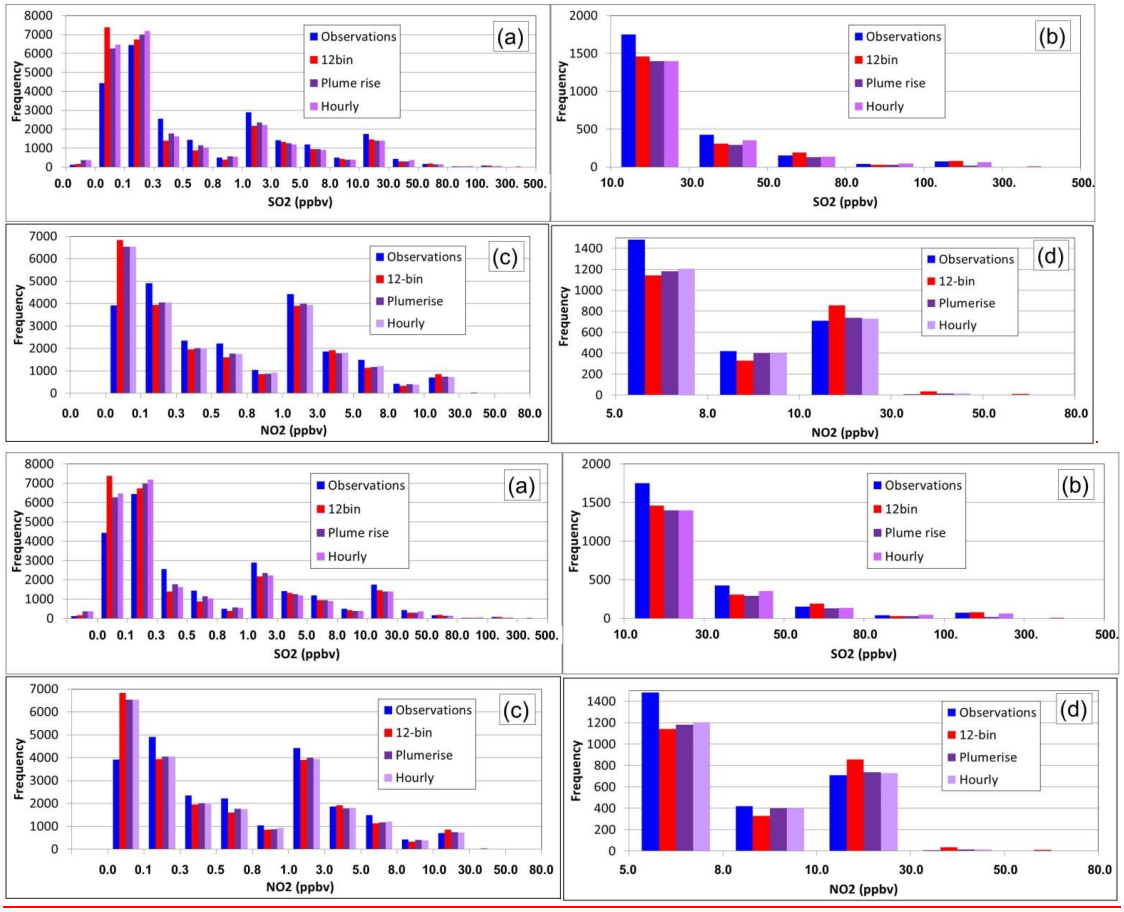
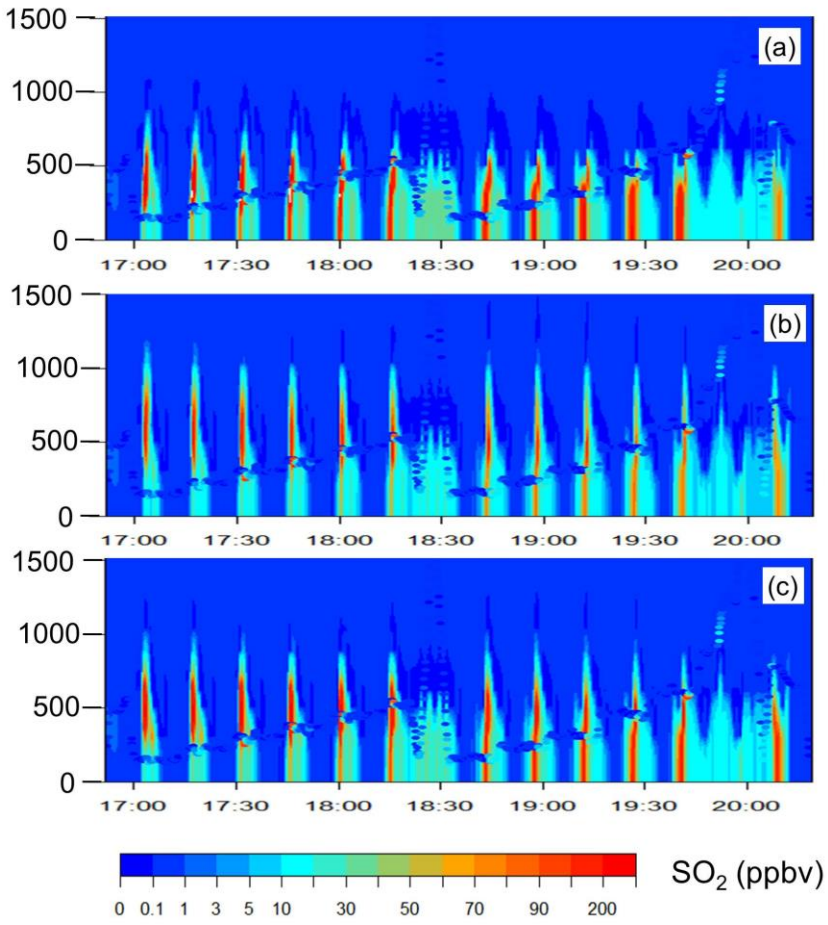


Figure ~~4~~5: Histograms comparing $SO_2$ and $NO_2$ simulations mixing ratios (ppbv) with aircraft observations. (a) All $SO_2$ values. (b) Higher $SO_2$ mixing ratios. (c) All $NO_2$ values. (d) Higher $NO_2$ mixing ratios. Left-most histogram bin in (a) and (c) correspond to values of zero.

Formatted: Line spacing: single

The results across the different simulations suggest that the overall model performance may be hampered by a tendency to place too much emitted mass close to the surface, and insufficient mass aloft. In order to determine possible causes for this behaviour, we carried out several additional analyses.

First, we examined the 12th flight of the observation study, which took place between 16:30 and 20:30 UTC (10:30 to 14:30 local time) on Aug 24th, as a case study to show the differences between the three simulations examining the impacts of the choice of plume rise algorithm and its input parameters. Flight 12 was an "emissions" flight, with the aircraft flying around the boundary of a single facility (Syncrude), with elevations gradually increasing in two successive sets of passes around the boundary. Data collected during flights of this nature were used to estimate emissions from the facility via calculation of the fluxes into and out of the facility from the collected data (Gordon *et al.*, 2015). During flight 12, the aircraft carried out two successive sequences circling the facility boundaries in gradual upward spirals (between 17:00 to 18:15 UTC, and 18:45 to 19:45 UTC), starting at the lowest aircraft altitude above the surface, and gradually increasing in elevation on each pass around the facility. The $SO_2$ plume was thus intersected at multiple times and multiple heights during each of these periods. Figure ~~5~~6(a,b,c) depicts the model-derived $SO_2$ mixing ratio profiles at 10 second intervals interpolated from the aircraft positions as a function of time, as mixing ratio contours, for the "12-bin", "Plume rise" and "Hourly" scenarios, respectively. The aircraft locations are shown as coloured dots over-plotted on the background model mixing ratio contours. Each high mixing ratio "spike" in the panels of Figure ~~5~~6 thus represents a successive pass through the model $SO_2$ plume – the change in these plumes as a function of time may be seen by following the changes in the plume cross-sections in each panel along the x-axis timeline, from left to right. Between 17:00 and 18:15 UTC, the simulated plumes are mostly aloft. The 12-bin simulation employing the original Briggs algorithm (Figure ~~5~~6(a)) begins to fumigate significantly by 17:30 UTC, with higher concentrations reaching the surface, while for the Plume rise simulation (Figure ~~5~~6(b)) the plume both reaches higher elevations, and experiences significantly less fumigation. The Hourly simulation (Figure ~~5~~6(c)) is intermediate between the other two simulations. In the second period (18:45 to 19:45 UTC), the fumigation behaviour becomes more pronounced for all three simulations, and once again is strongest for the 12-bin simulation (Figure ~~5~~6(a), weakest for the Plume rise simulation (Figure ~~5~~6(b)), and intermediate for the Hourly simulation (Figure ~~5~~6(c)).


1500
1000
500
0
(a)
17:00
17:30
18:00
18:30
19:00
19:30
20:00
(b)
(c)
0 0.1 1 3 5 10 30 50 70 90 200
SO2 (ppbv)

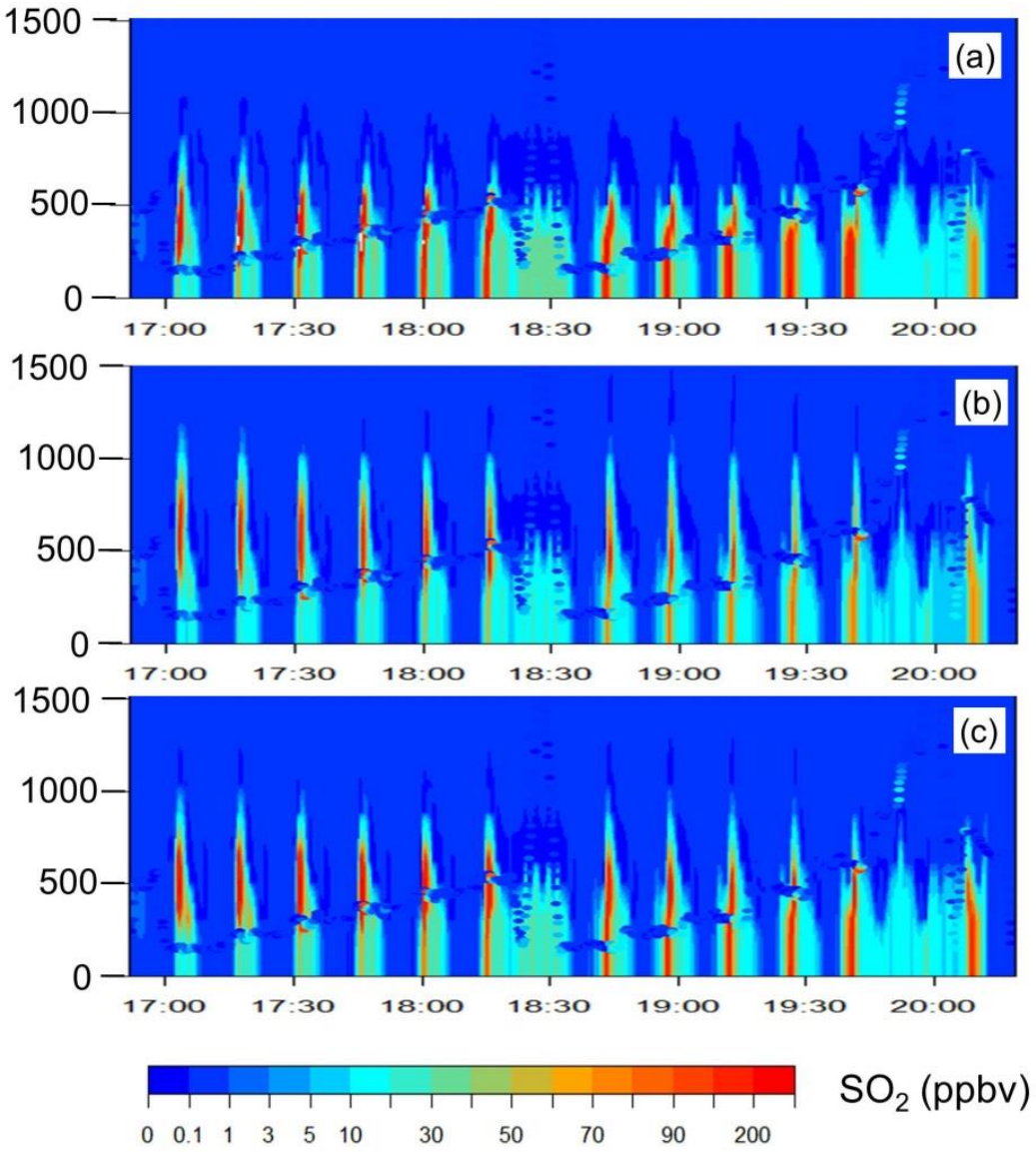
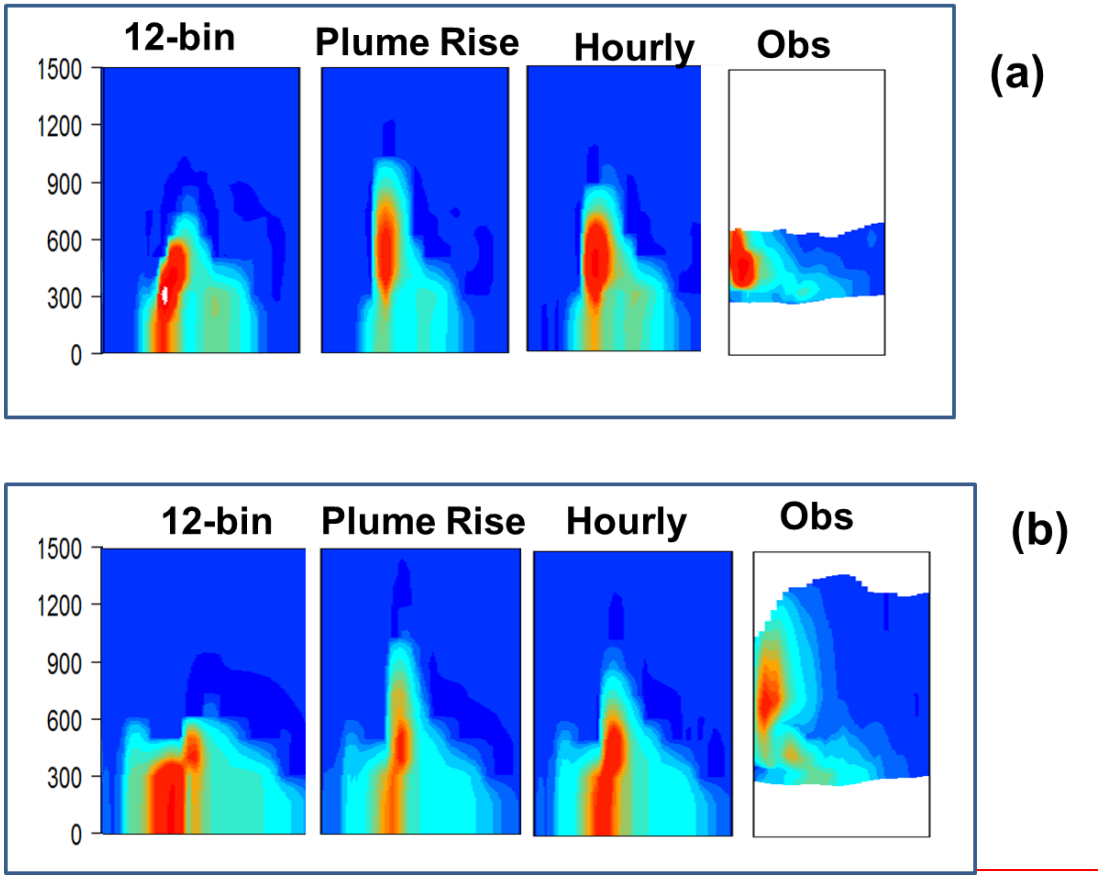



Figure ~~5~~6: Model $SO_2$ profile along the aircraft path for Flight 12 (a) "12-bin" simulation (original plume-rise algorithm), (b) "Plume rise" simulation (revised plume rise algorithm), and (c) "Hourly" simulation (revised plume rise algorithm combined with hourly data for volume flow rates and stack temperatures). Panels (a-c) show model predictions in the column of the aircraft trajectory as concentration contours – aircraft observed values at the aircraft locations at the given time (in UTC) are shown as coloured dots overplotting the background contours.

While the aircraft values are difficult to discern in Figure ~~5~~6, the collected aircraft $SO_2$ observation data at successive plume intersections during each of the two intervals were extracted from the data record, arranged so that "first plume intersection"

values were vertically aligned, and the vertical intervals between these successive aircraft passes were linearly interpolated in the vertical to yield observation-based cross-sections of $SO_2$ mixing ratios, for each of the two time intervals. These are compared to the model plumes between 17:42 and 17:54, and 19:08 and 19:25, in Figure ~~6~~7(a,b), respectively. In the first interval (Figure ~~6~~7(a)), the observed plume (far right profile) can be seen to be completely detached from the surface, with concentrations < 3 ppbv located below a > 100 ppbv region between 460 and 520m elevation. All three model plumes show more fumigation than the observations, with the "Plume rise" simulation showing the least fumigation of the three simulations, and the 12-bin simulation showing the most fumigation. In the second interval (Figure ~~6~~7(b)), the observed plume is located significantly higher than the model plumes (the "Plume rise" simulation plume is the closest of the three in terms of elevation, but all three model plumes underestimate the plume height by several hundred metres). While the observed plume during this second interval shows some signs of fumigation at the lowest elevation, the observed concentrations at lowest aircraft elevation are less than 30 ppbv, while the lowest model mixing ratios in the fumigation region are approximately 70 ppbv for the "Plume rise" simulation, and above 100 ppbv for the other two simulations. The case study thus echoes the statistical analysis of Figures ~~3~~4 and ~~4~~5: all model simulations tend to under-predict the plume top, and overpredict the extent of fumigation, for Flight 12.

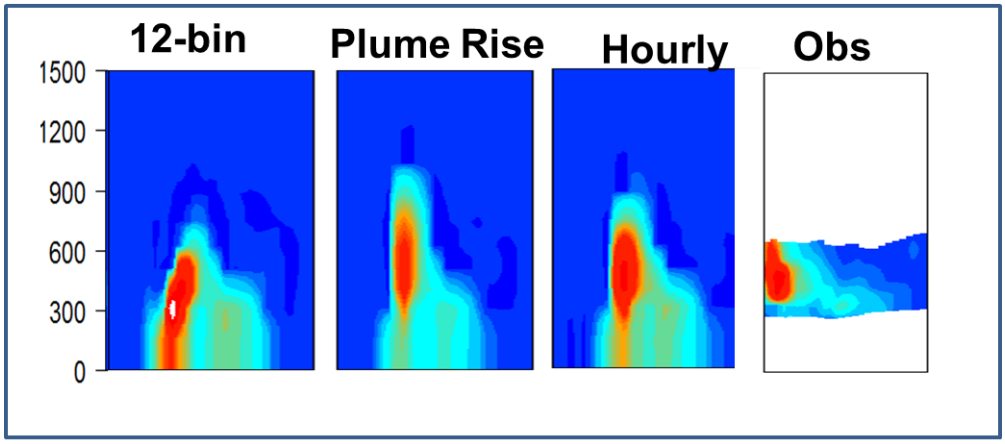



~~Figure 6. Zoomed view of Figure 5. (a) 17:42–17:54, observations interpolated from successive flight passes between 17:00 and 18:19. (b) 19:08–19:20, observations interpolated from successive flight passes between 18:42 and 19:45~~

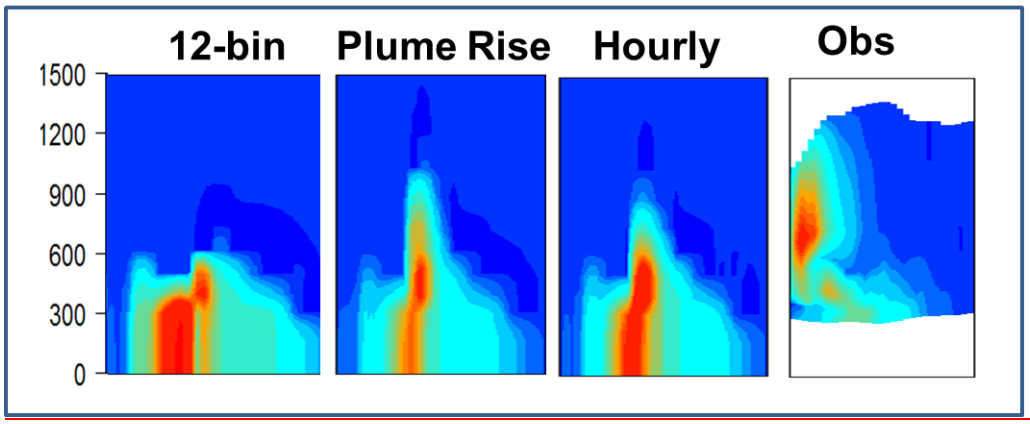
While the comparison is encouraging in that both of the simulations employing the new plume rise algorithm (Figure ~~6~~7(b,c)) out-perform the original (Figure ~~6~~7(a)) for most metrics (Figure ~~5~~6(f)), the use of the CEMS-observed volume flow rates and temperatures with the new algorithm result in a degradation of performance, relative to the simulation making use of annual averages for these parameters. That is, the believed-to-be-more-realistic stack parameters result in slightly worse performance: a cause for concern. The average of the hourly observed volume flow rates and temperatures for this facility's stack during flight 12 are 581.5 $m^3s^{-1}$ and 472.69K, respectively, while the corresponding annual reported values are 1174.5 $m^3s^{-1}$ and 513.2K. With respect to equation (1), the relative ratio of the buoyancy flux with these two sets of parameters will be:

$$R = \frac{V_r T_{s,o}(T_{s,r} - T_a)}{V_o T_{s,r}(T_{s,o} - T_a)} \qquad (10)$$

Where the subscripts *r* and *o* indicate the annual reported and hourly observed values of each quantity. Assuming an ambient temperature at stack height of 291K, the value of R is 2.28; that is, the initial buoyancy flux of the Plume rise simulation is over double that of the Hourly simulation. The hourly values are ~~known~~ believed to be more realistic during the period simulated (though include engineering emissions estimates during upset conditions) – the revised algorithm, while providing better results than the original, thus still has a tendency to under-predict the plume heights. In our companion paper, we found that the revised algorithm (therein referred to as the "layered approach") had no significant advantages over the original Briggs algorithms – here we have found this revised approach has considerable benefit, while showing the same overall tendency to under-predict plume heights as in our companion paper.




(a)




(b)

Figure 7. Zoomed-view of Figure 5. (a) 17:42-17:54 UTC, observations interpolated from successive flight passes between 17:00 and 18:19 UTC. (b) 19:08-19:20 UTC, observations interpolated from successive flight passes between 18:42 and 19:45 UTC

In order to demonstrate the extent to which the plume rise values themselves differ between flights, we have compared the calculated plume heights from each of the three algorithms examined here for 8 stacks (located at the Syncrude, Suncor, and CNRL facilities) against observations during the course of the study (Figure ~~7~~8). The observed plume rise values here were derived from estimates of the $SO_2$ plume centres from the aircraft campaign's emission box flights as estimated in our companion paper (Gordon *et al*. (2018)). Despite the differences visible in Figures ~~5~~6 and ~~6~~7, for flight 12, Figure ~~7~~8 shows that the revised algorithm has a significant impact on calculated plume heights, greatly increasing the number falling within a factor of two of the observations ~~(>70%)~~, while the original algorithm has the majority of calculated plume heights falling below the 1:1 line, in accord with Gordon *et al*. (2018)). However, the impact of the differences in volume flow rates and temperatures (Figure ~~7~~8(b) versus Figure ~~7~~8(c)) are usually relatively minor, with the exception of a few additional points falling below the 1:2 line for the Hourly (Figure ~~7~~8(c)) simulation. Table 7 shows the relative distribution of the 176 test cases compared in terms of their distribution about the 1:2, 1:1 and 2:1 lines of the scatterplots of Figure 8. The revised plumerise approach results in a

significant improvement in the distribution, and the use of CEMS data results in a slight further improvement in the average predicted plume height. We note that the relatively small differences between Figure 8(b) and 8(c), and between the last two columns of Table 7, imply that the residual buoyancy approach of equations 9 was relatively insensitive to the range of the initial buoyancy flux resulting from the two sets of emissions data used here, compared to the temperature gradients in equation (5). The large deviation between the annual reported and measured stack parameters for flight 12 may thus be an anomaly relative to the entire record across all 8 stacks examined here. Nevertheless, Figures ~~3~~4 to ~~6~~7 suggest that all of model simulations have a tendency to overestimate fumigation, so we continued our examination using Flight 12 as a case study.

The model concentrations of primary pollutants are also modified by vertical diffusion and advection. The use of a plume rise algorithm simultaneously with vertical diffusion implies the potential for "double-counting" of some proportion of the vertical mixing, in that the observation-based plume rise algorithms *de facto* incorporate vertical diffusion in their estimates of plume rise, while air-quality models must apply diffusion at all model grid-squares, including those in which plume rise algorithms have already distributed emitted mass in the vertical. If the relative impact of vertical diffusion versus buoyant plume rise is strong, this may result in excessive vertical mixing; the model effectively "double-counting" the vertical diffusion component of the net rise. The potential for overestimates of model diffusivity magnitudes resulting in excessive vertical mixing to the ground was investigated by carrying out a sensitivity run for Flight 12 in which diffusivities in the column were halved prior to their use in calculating vertical diffusion. This sensitivity run showed a minimal impact on model results – the magnitude alone of vertical diffusion did not influence the fumigation noted below. However, this test did not examine the potential changes associated with different magnitude changes in diffusivity as a function of height.

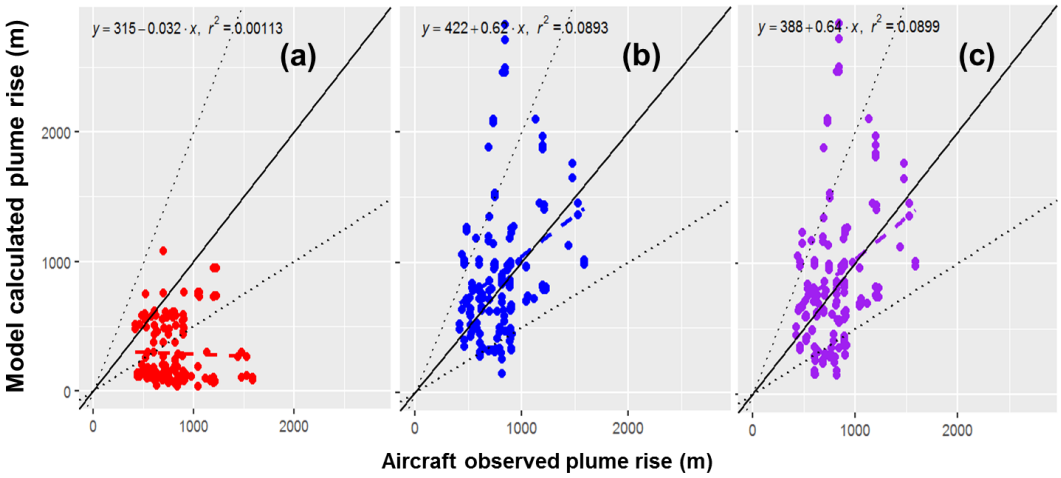



Figure ~~7~~8: Observed plume rise heights during aircraft emission box flights compared to model calculated plume rise using; (a) the original plume rise algorithm; (b) new plume rise algorithm; and (c) new plume rise algorithm and CEMS hourly stack temperature and volume flow rate.

**Table 7**: Comparison of model plume rise performance

| Number (model:observed) Total number of comparisons: 176 | Original Plume Rise Algorithm | New Plume Rise Algorithm (Layered Approach) | New Plume Rise Algorithm driven by CEMS data |
|---|---|---|---|
| Below 1:2 line | 116 | 24 | 29 |
| Between 1:2 and 1:1 line | 44 | 54 | 55 |
| Between 1:1 and 2:1 line | 15 | 75 | 69 |
| Above 2:1 line | 10 | 22 | 22 |
| Average plume height (m) (Observed: 822 m) | 289 | 935 | 914 |
| Ratio of average simulated to observed plume height | 0.35 | 1.13 | 1.11 |

Formatted: Line spacing: single
Formatted Table
Formatted: Line spacing: single
Formatted: Line spacing: single
Formatted: Line spacing: single
Formatted: Line spacing: single
Formatted: Line spacing: single
Formatted: Line spacing: single
Formatted: Line spacing: single

All of the plume rise algorithms are limited by the accuracy of the on-line model to accurately predict the meteorological quantities required in equations (1) through (9). We note that the original Briggs' algorithms (equations 1 through 8) are more strongly dependant on the model's ability to accurately predict meteorological conditions close to the surface, at stack height, as well as bulk parameters such as the Obukhov length, while the revised algorithm (equations 1,5 and 9) are more strongly dependent on the model's ability to accurately predict the temperature profile throughout the column.

We examined the model's temperature predictions, and compare to observations aboard the aircraft in Figure ~~8~~9. Figure ~~8~~9(a) shows the model-predicted temperatures in the columns around the Syncrude facility as colour contours in height versus time, similar to the mixing ratio cross-sections of Figure ~~5~~6. The corresponding aircraft temperatures are over-plotted on Figure ~~8~~9(a) as coloured dots employing the same temperature scale as the model values. The aircraft values, particularly in the first of the two emissions spiral periods (bracketed by vertical dashed lines in Figure ~~8~~9) suggest that the model temperatures are biased high in the lowest part of the atmosphere. Figure ~~8~~9(b) shows the temperature cross-sections interpolated from aircraft observations collected during the portion of the aircraft flight track crossing the $SO_2$ plume, to represent the average temperature profiles in each of the two regions. The first of these two cross-sections show an observed temperature inversion at the lowest aircraft altitudes, absent in the model temperature profile. In the second profile of Figure ~~8~~9(b), the inversion is no longer apparent. The model-predicted temperatures in the lowest part of the atmosphere are also biased high relative to both observation-based temperature cross-sections (compare Figure ~~8~~9(b) corresponding to the dashed line bordered regions of Figure ~~8~~9(a)). Figures ~~8~~9(c) and ~~8~~9(d) show the variation between model and observed temperatures in two other ways; as a pair of temperature time series during the model flight (Figure ~~8~~9(c)) and as a scatterplot showing the differences in temperature (observed – model) as a function of height (Figure ~~8~~9(d)).

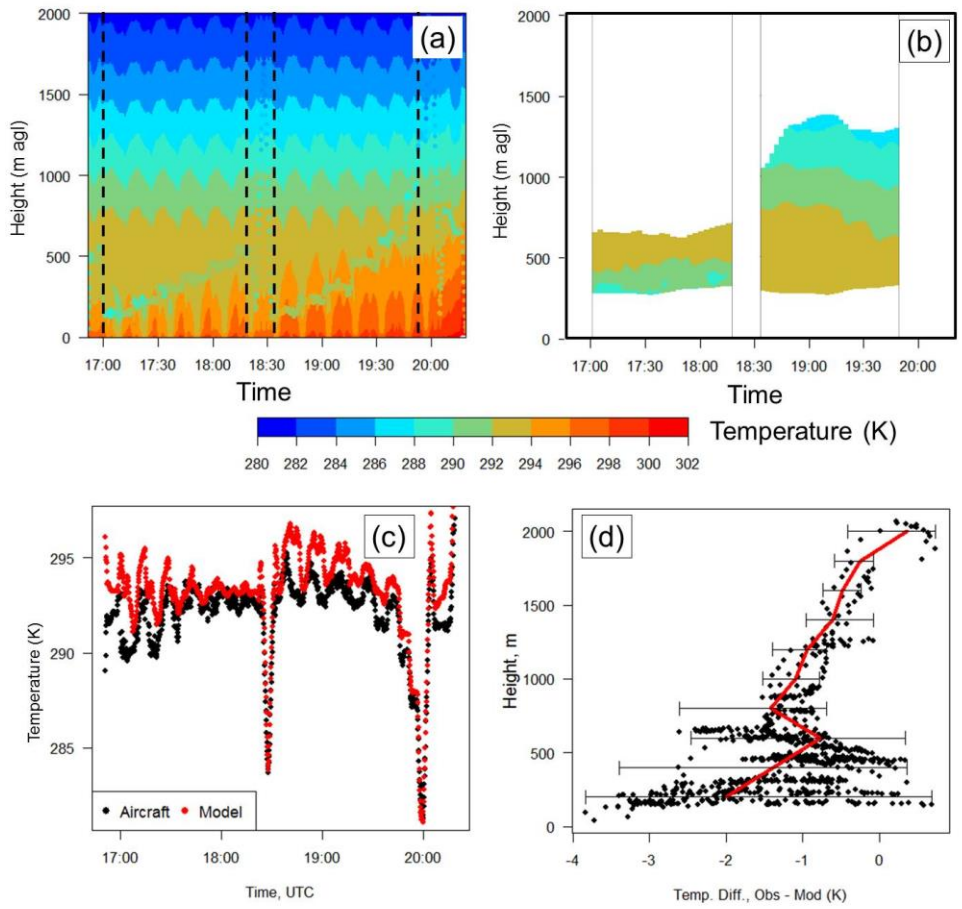



Figure ~~8~~9: Model versus observed temperatures, Flight 12. (a) Background model-predicted temperature profiles with observed temperatures overlaid (dots) with the same colour scale. (b) Observed temperatures along the portion of the transect containing the plumes, between 17:00 and 18:18 UTC, and 18:32 and 19:45 UTC. (c) Model (red) and observed (black) temperatures as a function of time; (d) Temperature deviation (observations – model) as a function of height, with the red line showing the mean deviation at every 100m.

All of these temperature comparisons suggest that, for Flight 12, the model tended to have positive temperature biases near the surface, biases which gradually decreased with height (Figure ~~8~~9(d)). The model atmosphere would thus be expected to be less stable than the observed atmosphere, with temperature gradients reduced in magnitude relative to observations. The model also reported positive values of the Obukhov length during the period (neutral to stable atmospheres; the Briggs' formula employed would be equations (3) or (4)), while the smaller magnitude temperature gradients in the model will drive parameter *s* (equation (5)) to smaller values. While *s* features in the stable atmosphere formula (3), it does not feature in the neutral atmosphere formula. That is, the original Briggs' formulae are relatively insensitive to errors in the temperature profile in near-neutral conditions, with only a weak influence via the $F_b$ term. However, the revised algorithms (equations (9), (1), and (5)) will be influenced by the accuracy of the temperature gradient at every point throughout the temperature profile. This analysis suggests that the original Briggs' algorithms (the "12-bin" simulations) will be less influenced by the temperature errors shown in Figure

~~8~~9(d), while the revised approach ("Plume rise" and "Hourly" will be more influenced by them, contributing to higher estimates of plume heights.

This particular case study thus places an important caveat on our results – while the revised plume rise approach provides better results, and a better estimate of plume rise relative to the observations, it may be doing so in part in response to a model overestimate of surface heating and the corresponding reduction in the magnitude of the temperature gradient, to which the latter algorithm is sensitive, and to which the original Briggs' algorithms are less sensitive.

Our final analysis examines the effects of the different plume rise algorithms on the broader region, through comparisons of multi-week average differences of surface and downwind vertical cross-section mixing ratios of $SO_2$ (Figure ~~9~~10). The change in $SO_2$ ("Plume Rise" – "12-bin") average surface mixing ratio and a representative cross-section are shown in Figure ~~9~~10(a,b), while the corresponding differences for the two simulations employing the revised algorithm ("Hourly" – "Plume Rise") are shown in Figure ~~9~~10(c,d). The first comparison (Figure ~~9~~10(a)) shows the substantial impact of the revised plume rise algorithm relative to the original Briggs' formulation; surface concentrations of $SO_2$ have decreased over most of the domain, often by by up to tens of percent. The corresponding cross-section (Figure ~~9~~10(b)) shows that most of the $SO_2$ removed from the surface is transported aloft, resulting in substantial relative increases in $SO_2$ mixing ratios throughout the lower troposphere. The second comparison (Figure ~~9~~10(c,d)) shows that the use of hourly CEMS stack parameter data results in substantial local increases and decreases – changes in plume height associated with the use of the hourly stack parameters are sometimes responsible for both positive and negative changes in the tens of percent, relative to the simulation driven by annual reported stack parameters. The $SO_2$ mass formerly being carried aloft now fumigates downwind, in the "Hourly" cross-section.

In similar evaluations for $NO_2$ (not shown), percentage differences of up to 10% in $NO_2$ surface mixing ratio and less than 1% maximum difference is surface ozone mixing ratio for the 30-day average period were found. The choice of a plume rise algorithm thus has a substantial impact on average surface and lower troposphere concentrations of those species predominantly emitted from large stacks.

Formatted: Line spacing: 1.5 lines

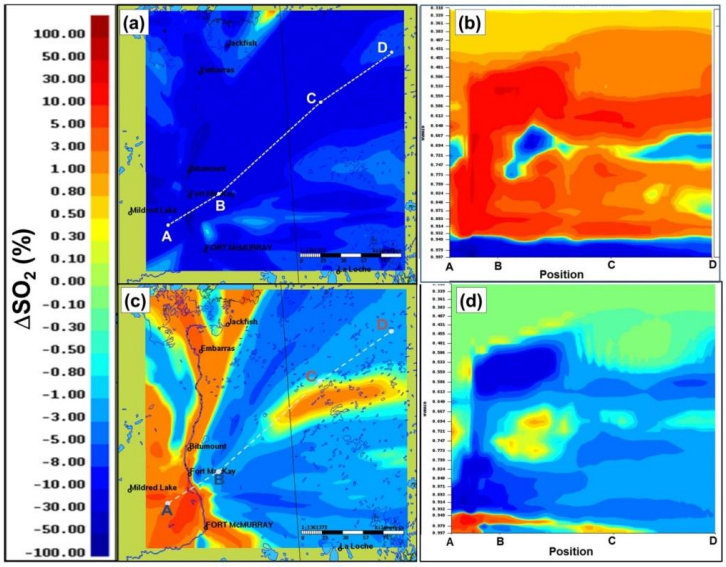



Figure ~~9~~10. Comparison of model-generated average mixing ratios: percent differences in multi-week averages. (a) Average surface mixing ratio percent differences for "Plume rise" – "12-bin". (b) Average cross-section percent differences along cross-section A→B→C→D, for "Plume rise" – "12-bin" . (c) Average surface mixing ratio percent differences for "Hourly" – "Plume rise". (d) Average cross-section percent differences along cross-section A→B→C→D, for "Hourly" – "Plume rise".

Formatted: Left, Line spacing: single

## 5 Conclusions

We have carried out a set of four model scenarios for a 2.5-km resolution nested domain using the GEM-MACH air quality forecast model for the Athabasca oil sands region of Alberta, Canada. These scenarios have allowed us to examine the relative impacts of aerosol size distribution and plume rise algorithms on model performance, relative to surface and aircraft observations of multiple chemical species.

While a 2-bin configuration with sub-binning of microphysical processes has been employed in the past for operational forecasting due to computational processing time constraints (Moran *et al.*, 2010), we find that the 12-bin configuration has better performance for all surface PM2.5 prediction metrics, including an overall 34% reduction in the magnitude of the bias of PM2.5, for a 25% increase in processing time.

Comparisons with the model and observed stack plumes showed that all algorithms tended to under-predict plume heights, in accord with our companion measurement-driven investigation of plume rise using the Briggs (1984) plume rise algorithm (Gordon et al., 2018). However, in contrast to that work, significant improvements to model performance were found with the adoption of a revised plume rise algorithm, also based on Briggs (1984), in which local changes in stability in individual model stable and neutral model layers are used to calculate the fractional reduction in buoyancy of the rising plume. Tests of the revised algorithm using ~~both~~ annually reported stack parameters, and hourly parameters from a combination of Continuous Emissions Monitoring and engineering estimates, both resulted in significant improvements in model performance in comparison to the original approach. However, the latter approach was also shown to be relatively insensitive to the range of initial buoyancy fluxes resulting from the two different emissions estimates, with the use of hourly observed (and presumed more accurate) stack parameters resulted in a slight degradation of performance relative to the use of annual reported values for these parameters. Further investigation using a specific case study suggested that the improvements associated with the revised algorithm may in part be due to model positive biases in lower atmospheric temperature, resulting in model underestimates in the magnitude of atmospheric temperature gradients. Nevertheless, the revised approach was found to correct much of the predominantly negative bias in predicted plume height seen for Briggs' original algorithms, correcting the biases in plume height noted in our companion paper, in which the algorithms were driven using observed meteorology.

Despite these improvements, and the tendency of the model to underestimate temperature gradients, the model still over-predicts the extent of fumigation for all plume rise algorithms tested, implying the need for further work. The revised approach found to be the most favorable in the current work is based on two key parameters; entrainment coefficients determined by Briggs from data collected in 1975 to be approximately 0.08 and 0.4 respectively; we recommend that these coefficients be re-estimated using more recent data.

Our simulations have shown that the choice of a plume rise parameterization has a very significant impact on downwind concentrations of $SO_2$ from the oil sands sources, with the approaches having the more accurate plume heights also resulting in significant reductions in surface $SO_2$, and increases in $SO_2$ aloft, helping to correct pre-existing positive and negative biases in the model at these elevations. Smaller impacts were found for $NO_2$, and minimal impacts for ozone.

*Code and data availability*

The aircraft observations used in this study are publicly available on the ECCC data portal (https://www.canada.ca/en/environment-climate-change/services/oil-sands-monitoring/monitoring-air-quality-alberta-oil-sands.html). The hourly surface monitoring network data are from the public website of the Wood Buffalo Environmental Monitoring Association (http://www.wbea.org/network-and-data/historical-monitoring-data). The model results are available

Formatted: Not Superscript/ Subscript
Formatted: Not Superscript/ Subscript
Formatted: Font: Not Bold, Italic
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Normal
Field Code Changed
Formatted: Font: (Default) Times New Roman, 10 pt

upon request to Ayodeji.Akingunola (ayodeji.akingunola@canada.ca). GEM-MACH, the atmospheric chemistry library for the GEM numerical atmospheric model (© 2007–2013, Air Quality Research Division and National Prediction Operations division, Environment and Climate Change Canada), is a free software which can be redistributed and/or modified under the terms of the GNU Lesser General Public License as published by the Free Software Foundation – either version 2.1 of the license or any later version. The specific GEM-MACH version used in this work may be obtained on request to ayodeji.akingunola@canada.ca. Much of the emissions data used in our model are available online: Executive Summary, Joint Oil Sands Monitoring Program Emissions Inventory report (https://www.canada.ca/en/environment-climate-change/services/science-technology/publications/joint-oil-sands-monitoring-emissions-report.html; Joint oil sands monitoring program emissions inventory report, 2018) and Joint Oil Sands Emissions Inventory Database (http://ec.gc.ca/data_donnees/SSB-OSM_Air/Air/Emissions_inventory_files/) and more recent updates may be obtained by contacting Junhua Zhang or Mike Moran (junhua.zhang@canada.ca; mike.moran@canada.ca).

*Author Contributions:* AA and PAM were responsible for the study design and methodology, model simulations, comparison to observations, and the writing of the manuscript and modifications of same. JZ, MDM and QZ contributed emissions data used in the modelling. AD and S-ML contributed aircraft observation data used for model evaluation. MG contributed aircraft plume height analyses, information on the companion paper, and contributed to the text and revisions of the manuscript.

*Competing Interests*: The authors declare that they have no conflict of interest.

*Special issue statement*: This article is part of the special issue "Atmospheric emissions from oil sands development and their transport, transformation and deposition (ACP/AMT inter-journal SI)". It is not associated with a conference.

*Acknowledgments:* This project was jointly supported by the Climate Change and Air Quality Program of Environment and Climate Change Canada, Alberta Environment and Parks, and the Oil Sands Monitoring program. The figures in this work were created using a combination of Environment Canada and Climate Change software and the R open-source programming language (R Core Team, 2017).

**~~6 6~~ References**

Ashrafi K., A. A. Orkomi, M.S. Motlagh, Direct effect of atmospheric turbulence on plume rise in a neutral atmosphere, *Atm. Poll. Res.*, 8 640-651, 2017.

Bieser, J., A. Aulinger, V. Matthias, M. Quante, H.A.C. Denier van der Gon, Vertical emission profiles for Europe based on plume rise calculations, *Environ. Pollut.*, 159, 2935-2946. doi:10.1016/j.envpol.2011.04.030, 2011.

Briggs, G.A.: Plume rise predictions, Lectures on air Pollution and environmental impact analyses. In: Workshop Proceedings, Sept. 29-Oct. 3, pp. 59-111, *American Meteorological Society*, Boston, MA, USA., 1975.

Briggs, G.A.: Plume rise and buoyancy effects, atmospheric sciences and power production. In: Randerson, D. (Ed.), DOE/TIC-27601 (DE84005177), TN. Technical Information Center, U.S. Dept. of Energy, Oak Ridge, USA,.pp 327-366, 1984.

Byun, D. W. and K.L. Schere: Review of the governing equations, computational algorithms, and other components of the Model-3 Community Multiscale Air Quality (CMAQ) Modeling System. *Appl. Mech. Rev.* *59*(2): 51-77. doi: 10.1115/1.2128636, 2006

**Formatted:** Font: (Default) Times New Roman, 10 pt
**Formatted:** Font: (Default) Times New Roman, 10 pt
**Formatted:** Font: (Default) Times New Roman, 10 pt
**Field Code Changed**
**Formatted:** Hyperlink, Font: (Default) Times New Roman, 10 pt, English (U.K.)
**Field Code Changed**
**Formatted:** Font: Italic
**Formatted:** Font: Italic
**Formatted:** Font: Italic
**Formatted:** Normal, Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers
**Formatted:** Font: (Default) Times New Roman, 10 pt
**Formatted:** Normal
**Formatted:** Font: Italic
**Formatted:** Normal, Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers
**Formatted:** Font: (Default) Times New Roman, 10 pt
**Formatted:** Font:
**Formatted:** Space After: 6 pt
**Formatted:** Font: Italic
**Formatted:** Space After: 6 pt

Byun, D.W. and J.K.S. Ching: Science algorithms of the EPA Models-3 community multiscale air quality (CMAQ) modeling system, US EPA, Office of Research and development, EPA/600/R-99/030, 1999

Carslaw, D.C. and K. Ropkins, openair — an R package for air quality data analysis. *Environmental Modelling & Software*. Volume 27-28, pp.52–61, 2012.

Caron, J-F. and co-authors: Changes to the Regional Deterministic Prediction System (RDPS) from version 3.2.1 to version 4.0.0. Canadian Meteorological Centre Technical Note. 2014 [Available on request from Environment Canada, Centre Météorologique Canadien, division du développement, 2121 route Transcanadienne, 4e étage, Dorval, Québec, H9P1J3 or via

http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/changes_e.html#20141118_rdps_4.0.0

Charron, M., S. Polavarapu, M. Buehner, P.A. Vaillancourt, C. Charette, M. Roch, J. Morneau, L. Garand, J.M. Aparicio, S. MacPherson, S. Pellerin, J. St-James, and S. Heilliette,: The Stratospheric Extension of the Canadian Global Deterministic Medium-Range Weather Forecasting System and Its Impact on Tropospheric Forecasts. *Mon. Wea. Rev.*, 140, 1924–1944, https://doi.org/10.1175/MWR-D-11-00097.1, 2012.

CMAS Website, https://www.cmascenter.org/smoke/, last accessed Sept., 2017.

Coats, C.J., 1996. High-performance algorithms in the sparse matrix operator kernel emissions (SMOKE) modeling system. *Proceedings of the Ninth AMS Joint Conference on Applications of Air Pollution Meteorology with AWMA.* American Meteorological Society, Atlanta, GA, pp. 584-588.

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998a. The operational CMC/MRB global environmental multiscale (GEM) model. Part 1: design considerations and formulation. *Mon. Wea. Rev*. 126, 1373-1395.

Côté, J., J.-G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998b. The operational CMC-MRB global environment multiscale (GEM) model. Part II: results. *Mon. Wea. Rev*. 126, 1397-1418.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), *Geosci. Model Dev.*, 3, 43-67, https://doi.org/10.5194/gmd-3-43-2010, 2010.

Emery C., Jung, K. Yarwood, G., Implementation of an Alternative Plume Rise Methodology in CAMx. Final Report, Work Order No. 582-7-84005-FY10-20, Environ International Corporation, 41 pp., 2010.

England, W.G., Teuscher, L.H., and Synder, R.B., A measurement program to determine plume configurations at the Beaver Gas Turbine Facility, Port Westward, Oregon. *J. Air. Poll. Contr. Assoc*., 986-989, 1976.

~~EPA. State Level Hourly Sulfur Dioxide (SO2) Data. https://www.epa.gov/air-emissions-modeling/state-level-hourly-sulfur-dioxide-so2-data~~

~~EPA, 2018(a): https://www.epa.gov/emc/emc-continuous-emission-monitoring-systems, last accessed February 1, 2018.~~

~~EPA, 2018(b):~~ EPA, 2015, Plain English Guide to the Part 75 Rule, https://www.epa.gov/sites/production/files/2015-05/documents/plain_english_guide_to_the_part_75_rule.pdf

Formatted: Font: Italic
Formatted: Space After: 6 pt
Field Code Changed
Formatted: Font: Italic
Formatted: Space After: 6 pt
Formatted: Line spacing: single
Formatted: English (U.K.)

EPA, 2018, EMC: Continuous Emission Monitoring Systems, https://www.epa.gov/emc/emc-continuous-emission-monitoring-systems, last accessed February 1, 2018.

Erbrink, Hans J., Plume rise in different atmospheres: A practical scheme and some comparisons with LIDAR measurements, *Atm. Env.* 28, 3625 -3636, 1994.

Gielbel, J., Messungen der Abgasfahnenüberhöhung eines Steinkohlekraftwerkes mit Hilfe von LIDAR (Plume Rise measurements of a pit coal power plant by means of LIDAR). (German) Schriftenreihe der Landesanstalt fur Immissionsschutz des Landes NRW, Heft 47, S. 42/59, 1979.

Girard, C., Plante, A., Desgagné, M., Mctaggart-Cowan, R., Côté, J., Charron, M., Gravel, S., Lee, V., Patoine, A., Qaddouri, A., Roch, M., Spacek, L., Tanguay, M., Vaillancourt, P.A., Zadra, A., Staggered vertical discretization of the canadian environmental multiscale (GEM) model using a coordinate of the log-hydrostatic-pressure type, *Mon. Wea. Rev.*, 142 (3), pp. 1183-1196, 2014.

Gong, S. L., L. A. Barrie, J.-P. Blanchet, K. von Salzen, U. Lohmann, G. Lesins, L. Spacek, L. M. Zhang, E. Girard, H. Lin, R. Leaitch, H. Leighton, P. Chylek, and P. Huang, Canadian Aerosol Module: A size-segregated simulation of atmospheric aerosol processes for climate and air quality models, 1, Module development, *J. Geophys. Res.*,108(D1), 4007, doi:10.1029/2001JD002002, 2003.

Gong, W., A.P. Dastoor, V.S. Bouchet, S.L. Gong, P.A. Makar, M.D. Moran, B. Pabla, S. Menard, L-P. Crevier, S. Cousineau and S. Venkatesh,: Cloud processing of gases and aerosols in a regional air quality model (AURAMS), *Atm. Res.* 82 (1-2), 248-275, 2006.

Gordon, M., ~~P.~~ Makar, ~~R.~~P. A., Staebler, ~~J.~~R. M., Zhang, ~~A.~~J., Akingunola, ~~W.~~A., Gong, ~~S-M.~~W., and Li~~.~~, S.-M.: A Comparison of Plume Rise Algorithms to Stack Plume Measurements in the Athabasca Oil Sands, *Atmos.* Chem. Phys~~*under*~~. *Discuss.*, https://doi.org/10.5194/acp-2017-1093, in review, ~~*ACP special issue on the oil sands,* 2018~~2017.

Gordon, M., S.-M. Li, R. Staebler, A. Darlington, K. Hayden, J. O'Brien, M. Wolde: 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, doi:10.5194/amt-8-3745-2015, 2015

Hamilton, P.M., Plume height measurements at Northfleet and Tilbury power stations, *Atm. Env.* 1, 379-387, 1967.

~~Hann~~Hanna, S. R. and R.J. Paine, Hybrid Plume Dispersion Model (HPDM) Development and Evaluation, *J. of App. Met.*, 28, 206-224, 1988.

Holmes, N.S., L. Morawska, A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available. *Atm. Env.* 40 (30), 5902-5928. doi: 10.1016/j.atmosenv.2006.06.003, 2006.

Im, U., Roberto Bianconi, R., Efisio Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., Baró, R., Bellasio, R., Brunner, D., Chemel, C., Curci, G., Flemming, J., Forkel, R., Giordano, L., Jiménez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote, C., Kuenen, J.J.P, Makar, P.A., Manders-Groot, A., Neal, L, Perez, J.L., Pirovano, G., Pouliot, G., San Jose, R., Savage, N., Schroder, W., Sokhi, R.S., Syrakov, D., Torian, A., Tuccella, P., Werhahn, J., Wolke, R., Yahya, K., Zabkar, R., Zhang, Y., Zhang, J., Hogrefe, C., Galmarini, S., Evaluation of operational on-line-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone, *Atm. Env.*, 115, 404-420, 2015(a).

Formatted: Space After: 6 pt
Formatted: Font: Italic
Formatted: Space After: 6 pt
Formatted: pb_toc_link1
Formatted: pb_toc_link1
Formatted: pb_toc_link1
Formatted: pb_toc_link1
Formatted: pb_toc_link1
Formatted: pb_toc_link1
Formatted: pb_toc_link1
Formatted: pb_toc_link1
Formatted: pb_toc_link1
Formatted: pb_toc_link1, Font: Times New Roman, Not Italic,
Formatted: pb_toc_link1, Font: Not Italic,
Formatted: Space After: 6 pt

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., Baró, R.., Bellasio, R., Brunner, D., Chemel, C., Curci, G., van der Gon, H.D., Flemming, J., Forkel, R., Giordano, L., Jiménez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L, Jorba, O., Knote, C., Makar, P.A., Manders-Groot, A., Neal, L, Perez, J.L, Pirovano, G., Pouliot, G., San Jose, R., Savage, N., Schroder, W., Sokhi, R.S., Syrakov, D., Torian, A, Tuccella, P., Wang, K., Werhahn, J., Wolke, R., Zabkar, R., Zhang, Y., Zhang, J., Hogrefe, C., Galmarini, S., Evaluation of operational on-line-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part II: Particulate Matter, *Atm. Env.*, 115, 421-411, 2015(b).

Joint Oil Sands Monitoring Plan~~:~~ (JOSM): Integrated Monitoring Plan for the Oil Sands: Air Quality Component, ISBN 978-1-100-18813-3, 72 pp., 2011. Available at http://publications.gc.ca/site/eng/394253/publication.html (last accessed Jan. 26, 2018).

Jarvis, P.G., The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field, *Phil. Trans. R. Soc. Lond., B.*, 273, 593-610, 1976.

Kaasik M., and Kimmel, V.: Validation of the improved AEROPOL model against the Copenhagen data set, *Int. J. Env. and Poll.* , 20, No. 1/2/3/4/5/6, DOI: 10.1504/IJEP.2003.004256, 2003.

Kain, J.S. and J.M. Fritsch, A one-dimensional entraining/detraining plume model and its application in convective parameterizations, *J. Atmos. Sci.*,47, 2784-2802, 1990.

Kain, J. S., The Kain-Fritsch convective parameterization: an update,. *J. Appl. Meteorol.*, 43, 170–181, 2004.

Levy, J.I.; Spengler, J.D.; Hlinka, D; Sullivan, D; Moon, D: Using CALPUFF to evaluate the impacts of power plant emissions in Illinois: mode sensitivity and implications. *Atm.. Env*., 36, 1063-1075, 2002.

Li, S.-M., Leithead, A., Moussa, S. G., Liggio, J., Moran, M. D., Wang, D., Hayden, K., Darlington, A., Gordon, M., Staebler, R., Makar, P. A., Stroud, C. A., McLaren, R., Liu, P. S. K., O'Brien, J., Mittermeier, R. L., Zhang, J., Marson, G., Cober, S. G., Wolde, M., and Wentzell, J. J. B.: Differences between measured and reported volatile organic compound emissions from oil sands facilities in Alberta, Canada, *Proc. Nat. Acad. Sci*., 114, E3756–E3765, 2017.

Liggio, J., ~~Moussa, S.G., Wentzell, J., Darlington, A., Liu, P., Leithead, A., Hayden, K., O'Brien, J., Mittermeier, R.L., Staebler, R., Wolde, M., and Li, S. M., Understanding the primary emissions and secondary formation of gaseous organic acids in the oil sands region of Alberta, Canada, *Atmos.* Chem. Phys., 17, 8411-8427, 2017.~~

~~Liggio, J.,~~ Li, S. M., Hayden, K., Taha, Y. M., Stroud, C., Darlington, A., Drollette, B. D., Gordon, M., Lee, P., Liu, P., Leithead, A., Moussa, S. G., Wang, D., O'Brien, J., Mittermeier, R. L., Brook, J. R., Lu, G., Staebler, R. M., Han, Y., Tokarek, T. W., Osthoff, H. D., Makar, P. A., Zhang, J., Plata, D. L., and Gentner, D. R.: Oil sands operations as a large source of secondary organic aerosols, *Nature*, 534, 91–94, https://doi.org/10.1038/nature17646, 2016.

Lurmann, F.W.,Lloyd, A.C., Atkinson, R., A chemical mechanism for use in long range transport/acid deposition computer modeling, *J. Geophys. Res*., 91, 10905-10936, 1986.

Makar, P.A., Bouchet, V.S., and Nenes, A., Inorganic Chemistry Calculations using HETV – A Vectorized Solver for the $SO_4^{2-}$-$NO_3^-$-$NH_4^+$ system based on the ISORROPIA Algorithms, *Atm. Env.*, 37, 2279-2294, 2003.

Makar, P.A., Gong, W., Milbrandt, J., Hogrefe, C., Zhang, Y., Curci, G., Zabkar, R., Im, U., Balzarini, A., Baro, R., Bianconi, R., Cheung, P., Forkel, R., Gravel, S., Hirtl, H., Honzak, L., Hou, A., Jimenz-Guerrero, P., Langer, M., Moran, M.D.,

Formatted: Space After: 6 pt

Formatted: Space After: 6 pt

Formatted: pb_toc_link1, Font: Times New Roman, Not Italic,

Formatted: Space After: 6 pt

Pabla, B., Perez, J.L., Pirovano, G., San Jose, R., Tuccella, P., Werhahn, J., Zhang, J., Galmarini, S., Feedbacks between air pollution and weather, part 1: Effects on weather. *Atm. Env.*, 115, 442-469, 2015 (a).

Makar, P.A., Gong, W., Hogrefe, C., Zhang, Y., Curci, G., Zabkar, R., Milbrandt, J., Im, U., Balzarini, A., Baro, R., Bianconi, R., Cheung, P., Forkel, R., Gravel, S., Hirtl, H., Honzak, L., Hou, A., Jimenz-Guerrero, P., Langer, M., Moran, M.D., Pabla, B., Perez, J.L., Pirovano, G., San Jose, R., Tuccella, P., Werhahn, J., Zhang, J., Galmarini, S., Feedbacks between air pollution and weather, part 2: Effects on chemistry. *Atm. Env.*, 115, 499-526, 2015 (b).

Makar, ~~Paul~~P., A., ~~A.~~ Akingunola, ~~J.~~A., Aherne, ~~A.S.~~J., Cole, ~~Y.~~A. S., Aklilu, ~~J.~~Y.-A., Zhang, ~~L.~~J., Wong, ~~K.~~J., Hayden, ~~S.M.~~K., Li, ~~J.~~S.-M., Kirk, ~~K.~~J., Scott, ~~M.D.~~K., Moran, ~~A.~~M. D., Robichaud, ~~H.~~A., Cathcart, ~~P.~~H., Baratzedah, ~~B.~~P., Pabla, ~~P.~~B., Cheung, P., Zheng, Q. ~~Zheng,~~, and Jeffries, D. S.: Estimates of Exceedances of Critical Loads for Acidifying Deposition in Alberta and ~~Saskatchewan, *under*~~Saskatchewan, *Atmos. Chem. Phys. Discuss.*, https://doi.org/10.5194/acp-2017-1094, in review~~, *ACPD*~~, 2018.

Mailhot, J. and R. Benoit, A finite-element model of the atmospheric boundary layer suitable for use with numerical weather prediction models, *J. Atm. Sci.*,39, 2249–2266, 1982.

Milbrandt, J.A., Yau, M.K., A multimoment bulk microphysics parameterization. Part I: analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, 62, 3051-3064, 2005(a).

Milbrandt, J.A., Yau, M.K., 2005b. A multimoment bulk microphysics parameterization. Part II: a proposed three-moment closure and scheme. *J. Atmos. Sci*. 62, 3065-3081, 2005(b).

Moran M.D., S. Ménard, D. Talbot, P. Huang, P.A. Makar, W. Gong, H. Landry, S. Gravel, S. Gong, L-P. Crevier, A. Kallaur, M. Sassi, Particulate-matter forecasting with GEM-MACH15, a new Canadian air-quality forecast model. In: Steyn DG, Rao ST (eds) *Air Pollution Modelling and Its Application XX*, Springer, Dordrecht, pp. 289-292, 2010.

Mylona, S. Sulphur dioxide emissions in Europe 1880–1991 and their effect on sulphur concentrations and depositions. Tellus B, 48: 662-689. doi:10.1034/j.1600-0889.1996.t01-2-00005.x, 1996.

NPRI, 2018: National Pollutant Release Inventory, https://www.canada.ca/en/environment-climate-change/services/national-pollutant-release-inventory/report.html, last accessed February 1, 2018.

Munoz-Alpizar, R., C. Stroud, S. Ren, S. Belair, S. Leroyer, V. Souvanlasy, L. Spacek, R. Pavlovic, D. Davignon, M. Moran, Towards an operational high-resolution air quality forecasting at ECCC, *19th EGU General Assembly, EGU2017*, proceedings from the conference held 23-28 April, 2017 in Vienna, Austria., p.3063, 2017

Rittmann, B.E., Application of two-thirds law to plume rise from industrial-sized sources, *Atm. Env.*, 16, 2575-2579, 1982.

Shephard, M. W., McLinden, C. A., Cady-Pereira, K. E., Luo, M., Moussa, S. G., Leithead, A., Liggio, J., Staebler, R. M., Akingunola, A., Makar, P., Lehr, P., Zhang, J., Henze, D. K., Millet, D. B., Bash, J. O., Zhu, L., Wells, K. C., Capps, S. L., Chaliyakunnel, S., Gordon, M., Hayden, K., Brook, J. R., Wolde, M., and Li, S.-M.: Tropospheric Emission Spectrometer (TES) satellite observations of ammonia, methanol, formic acid, and carbon monoxide over the Canadian oil sands: validation and model evaluation, *Atmos. Meas. Tech*., 8,5189–5211, doi:10.5194/amt-8-5189-2015, 2015.

Stroud, C. A., Morneau, G., Makar, P. A., Moran, M. D., Gong, W., Pabla, B., Zhang, J., Bouchet, V. S., Fox, D., Venkatesh, S., Wang, D., and Dann, T. OH-reactivity of volatile organic compounds at urban and rural sites across Canada: Evaluation of air quality model predictions using speciated VOC measurements, *Atm. Env*., 42, 7746–7756, 2008.



**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1, Font: Not Italic,
**Formatted:** pb_toc_link1, Font: Not Italic,
**Formatted:** Font: Italic
**Formatted:** Space After: 6 pt
**Formatted:** Space After: 6 pt
**Formatted:** Space After: 6 pt

Stroud, C. A., ~~P. A.~~ Makar, ~~J.~~P. A., Zhang, ~~M.~~J., Moran,~~-~~ M. D., Akingunola, A~~. Akinqunola.~~, Li, S.-M., ~~Li, A.~~ Leithead, ~~K.~~A., Hayden, ~~D. Wang: Impact of New~~K., and Siu, M.: Air Quality Predictions using Measurement-Derived Organic Gaseous and Particle Emissions ~~on Air Quality Predictions~~ in ~~the Athabasca Oil Sands~~a Petrochemical-Dominated Region, ~~*submitted to ACP/ACPD special issue,*~~ *Atmos. Chem. Phys. Discuss.*, https://doi.org/10.5194/acp-2018-93, in review, 2018.

Turner, D. B., L. W. Bender, J. O. Paumier, and P. F. Boone, Evaluation Of The TUPOS Air Quality Dispersion Model Using Data From The EPRI Kincaid Field Study, *Atm. Env.*, 25A, No. 10, 2187 – 2201, 1991.

VDI , 1985: VDI, Ausbreitung von Luftverunreinigungen in der Atmosphäre; Berechnung der Abgasfahnen-überhöhung. (Dispersion of air pollutants in the atmosphere; determination of plume rise) 1985-06 (German/English), Kommission Reinhaltung der Luft (KRdL) im VDI und DIN, 1985 – Normenausschuss, Available from: http://www.vdi.de.

~~Vet, R., Artz, R.S., Carou, S., Shaw, M., Ro, C-U., Aas, W., Baker, A., Bowersox, V.C., Dentener, F., Galy-Lacaux, C., Hou, A., Pienaar, J.J., Gillet, R., Forti, M.C., Gromov, S., Hara, H., Khodzher, T., Mahowald, N.M., Nickovic, S., Rao, P.S.P, Reid, N.W., A global assessment of precipitation chemistry and deposition of sulphur, nitrogen, sea salt, base cations, organic acids, acidity and pH, and phosphorus, *Atm. Env.*, 93, 3-100, 2014.~~

Webster H.N., D.J. Thomson, Validation of a Lagrangian model plume rise scheme using the Kincaid data set. *Atm. Env.*, 36, 5031–5042, 2002

Wesely, M.L., Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, *Atm. Env.*, 23, 1293-1304, 1989.

Whaley, C~~., P. A.~~. H., Makar, ~~M.W.~~P. A., Shephard, M., W., Zhang, L~~.~~., Zhang, J~~. Zhang, Q.~~., Zheng, ~~A.~~Q., Akingunola, ~~G.~~A., Wentworth, ~~J.~~G. R., Murphy, ~~S.~~J. G., Kharol, S. K., and ~~K. E.~~ Cady-Pereira~~:~~, K. E.: Contributions of natural and anthropogenic sources to ambient ammonia in the Athabasca Oil Sands and north-western Canada, ~~*under review, ACP,*~~ ~~2018~~*Atmos. Chem. Phys.*, 18, 2011-2034, https://doi.org/10.5194/acp-18-2011-2018, 2018.

Zhang, J~~.~~., Moran, M. D., Zheng, Q., Makar, P. A., Baratzadeh, P., Marson, G., Liu, P., and Li, S~~.~~-M~~.~~.: Emissions ~~preparation~~Preparation and ~~analysis~~Analysis for ~~air quality modelling for the JOSM project~~Multiscale Air Quality Modelling over the Athabasca ~~oil sands region~~Oil Sands Region of Alberta, Canada, ~~*under*~~*Atmos. Chem. Phys. Discuss.*, https://doi.org/10.5194/acp-2017-1215, in review ~~*ACP/ACPD special issue on the oil sands*~~, 2018.

Zhang, L., Moran, M.D., Makar, P.A., Brook, J.R., Gong, S., Modelling gaseous dry deposition in AURAMS: a unified regional air-quality modelling system, *Atm Env.*, 36, 537-560, 2002.

Zhang, L., Brook, J.R., and Vet, R., A revised parameterization for gaseous dry deposition in air-quality models, *Atm. Chem. Phys.*, 3, 2067-2082, 2003.

Zhang, L., Gong, S., Padro, J., and Barrie, L., A size-segregated particle dry deposition scheme for an atmospheric aerosol module, *Atm. Env.*, 35, 549-560, 2001

**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1, Font color: Auto,
**Formatted:** pb_toc_link1, Font color: Auto,
**Formatted:** pb_toc_link1, Font color: Auto,
**Formatted:** pb_toc_link1
**Formatted:** Space After: 6 pt
**Formatted:** Font: +Body (Times New Roman),
**Formatted:** Font: Italic
**Formatted:** Space After: 6 pt
**Formatted**
**Formatted**
**Formatted**
**Formatted** ...
**Formatted**
**Formatted**
**Formatted**
**Formatted**
**Formatted**
**Formatted**
**Formatted**
**Formatted**
**Formatted**
**Formatted**
**Formatted**
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted:** pb_toc_link1
**Formatted** ...
**Formatted:** pb_toc_link1