Review of Modelling below-cloud scavenging of size resolved particles in GEM-MACHv3.1 by Ghahreman et al. submitted to Atmos. Chem. Phys.

In this paper Ghahreman et al describe changes to the treatment of below cloud scavenging in the GEM-MACHv3.1 local area model. Specifically, they test the addition of multi-phase precipitation wherein both snow and liquid rain can scavenge aerosol separately, and the replacement of the Slinn (1984) scavenging coefficients with Wang et al (2014) scavenging coefficients. They run simulations with these different models in a domain over Canada for April 2018 and July 2018 months and then compared simulated aerosol deposition rates and near-surface concentrations with 3 observation networks. They find that the addition of multi-phase scavenging and the Wang BCS model leads to large changes in SO4 deposition rates and concentrations.

At the outset, I really like the study design – a simple sensitivity test using a detailed aerosol/cloud scheme embedded in a detailed LAM and with a clean comparison against observations. I think the study design is appropriate for this investigation. Where I am less impressed with the manuscript is in the interpretation and presentation of the results. I think that the statistical methods utilised by the authors are mostly appropriate (bias, R^2, etc) but the statistics are somewhat misused to support a preordained conclusion that the model is improved by the new scheme. The changes to aerosol deposition rates and concentrations are on the most part marginal and therefore don't support certain statements in the abstract or conclusion. I would like to see a more balanced conclusion which reflects the insignificance of impacts on NH4 and NO3 and how some SO4 metrics are improved while others are made worse by the inclusion of the multi-phase/Wang model.

I think the paper would benefit from structural re-organization and re-writing to remove repetition and add clarification, which is reflected in the sheer number of specific comments listed below. I recommend that the paper be revised according to the many comments below before being considered for publication.

General comments

The abstract is too wordy and does not follow the convention of describing the problem, then the method of solving the problem, and then the key results

Why was the Wang et al (2014) model chosen to replace Slinn (1984) rather than explicitly adding the missing processes of phoreses, electric charge and rear capture which has been shown to better represent observations and empirical models than both Slinn and Wang (see e.g., Fig 3 in Jones et al., 2022)?

Figures presenting the results in LAM domain (Figs 5-11 and S1-S3) would significantly benefit from having domain-average values also presented. At present, the colour scale and the significant inner-domain variability means it is very difficult to identify the overall impact of changes – e.g., whether the net difference between Wang2014 and multi-phase in Fig. 10 b-d is negative or positive

The figure and table captions are in general not descriptive enough (in some places, outright confusing) and should at the least include the specific metric plotted (e.g., accumulated aerosol deposition in Figs 5 and 6 which are currently erroneously labelled as concentrations). This should also be done for figures in the supplement.

Please choose informative names for the simulations and then use these names consistently throughout the manuscript.

My biggest issue with the paper is that it does not convincingly show that replacing the single-phase Slinn BCS scheme with the multi-phase Wang BCS scheme significantly improves aerosol concentrations relative to observations, which is explicitly stated as the main result of the paper in both the conclusions and the abstract. Case in point, the abstract includes the line: “Improvements in model performance (via scores for correlation coefficient, normalized mean bias, and/or fractional number of model values within a factor of two of observations) could also be seen, between the base case and the two simulations based on multiphase partitioning for NO3-, NH4+, and SO42.” This is patently not true – the paper shows no improvements for NO3 or NH4 (e.g., no areas with significant changes in Fig 9c,d and insignificant improvements to only some skill scores in Tables 1 and 2). I think it is fair to say there is a significant difference in SO4 concentrations between the simulations, but in terms of improvement, SO4 is improved against APQMP observations and made worse against CAPMoN observations in terms of bias, while changes to R2 and FAC2 are marginal. I recommend that the authors be clearer about the significance of their results and include less generalised comments about how the model is improved.

Specific comments

[L1] The title is different to the title in the Supplement. I think the more succinct title here is the more appropriate

[L9] It is unclear what you mean by distributions – aerosol size distributions? Please be more descriptive

[L11] References to Makar et al (2018) are unnecessary in the abstract. Consider also removing references to Slinn (1984) and Wang et al (2014).

[L13] Please give the long-name for GEM-MACH here

[L16] “GEM-MACH simulations…” – please be a little bit more specific about the experiment design (e.g., Regional GEM-MACH simulations in a local-area domain over Canada)

[L17] It would help to define the chemical formulae at the outset, e.g., sulfate (SO42=) and then use the chemical formulae or the chemical name consistently throughout the paper. Bouncing between formulae and names is confusing to the reader who might think that one refers to particulates and the other is in reference to gaseous component

[L31] Sentences beginning “The aerosol scavenging rates…” and ending “…bigger differences for aerosols larger that 1um” – these are very technical statements for an abstract and should probably be moved to the model description or results. They do not add anything to the abstract. Instead, earlier in the abstract you could describe the main difference between Wang and Slinn in one sentence or at least the motivation for moving from Slinn to Wang.

[L47] – “though the relative importance of aerosol dry deposition may have increased as a result of new observational studies” – the statement is rather awkward and could be interpreted that aerosol dry deposition was directly influenced by the observational studies. Please reword

[L49] – by ‘precipitation chemistry’ do you mean the chemical reactions within the hydrometeor. This phrase is rather ambiguous

[L50] – when you list the microphysical processes involved in BCS you could include the rear-capture effect. I think it is well understood that rear-capture in the wake of an oblate droplet is an important BCS process (e.g., https://acp.copernicus.org/articles/17/4159/2017/) and the reason it is not often considered is that the original BCS studies considered droplets to be spherical

[L60] You say “below-cloud wet scavenging” or “below-cloud scavenging” a lot, consider acronymising this to make the manuscript easier to read

[L67] “while rain droplets are usually assumed to be spherical in shape” – it is true that most of the early BCS studies assumed spherical raindrops. It would be worth noting that this assumption is erroneous, particularly for larger raindrops, which are much more oblate

[L79 – L87] – this really doesn’t belong in the abstract, please move it to the methodology section (2.2). This describes the underlying BCS scheme and does not fit well in the abstract

[L87] I find it curious that the extinction efficiency $E(dp, Dd)$ which dominates the variation in the scavenging coefficient is not described in more detail here. Potentially you could say “the Slinn (1984) extinction efficiency is prescribed as a linear combination of the extinction efficiencies due to Brownian motion, interception, and impaction. Therefore, the Slinn model misses important processes in the Greenfield gap such as thermophoresis which are included implicitly in the Wang et al (2014) model. This omission forms the motivation for testing the Wang scheme in this study”.

[L87] – Are the extinction efficiencies calculated offline or online? If offline, what parameters and assumptions were made to generate the extinction efficiency (temperature, pressure, etc)? What model was used for the raindrop terminal velocity? These are important assumptions which will affect the results

[L94] – What is the motivation for testing Wang et al (2014) over Slinn? At present, it seems arbitrary

[L95] – referring to GEM-MACH, typically the acronym rather than the long name is in parentheses. Also, this should be defined at the first use of the model’s acronym

[L112] “fully-coupled” with reference to GLMs or LAMs typically refers to atmosphere-ocean coupling unless otherwise specified, please be more specific about what is fully coupled to what

[L130] Sentence beginning “The default GEM-MACH model includes…” - This seems like a very comprehensive aerosol and chemistry model. Has it been used for other simulations - is it validated against observations? Why the significant complexity? If used operationally, will your changes (multi-phase, Wang) be incorporated operationally? This should probably be added to the results section.

[L184] “GEM-MACH simulations were carried out on a limited area model (LAM) domain with 2.5 km × 2.5 km (red) resolution, nested from a 10 km × 10 km (blue) horizontal resolution, for the months of April and July, 2018” – refer to figure 1 here

[L200-L265] – I like the fact that you name the simulation here: “base-case”, “multi-phase”, “Wang2014”. Please use these names consistently throughout the manuscript to identify which simulations you refer to. For example, in the supplement in the caption for Fig. S2 you say “partitioned and base experiments (e.g. rain/snow – base)” which is confusing and unnecessary when you have well defined names for these simulations

[L212-L238] Sentences beginning “In both the base-case…” and ending “solid phase precipitation at high altitudes” – This should not be in the description of the simulation but in a separate paragraph after to enhance readability

[L241] “similar to case 2 above” – this is an example of where the simulation name should be used. “case 2” is ambiguous, please instead say “multi-phase”. Capitalization of the simulation names (“BASE”, “MULTI”, “WANG2014”) may also increase readability. The actual BCS model could then be referred to as wang2014 to distinguish from the simulation which is a separate entity

[L242-L267] – Sentences beginning “Figure 3 compares the Slinn1984 and Wang2014 …” and ending “where the differences are up to 30% (Wang et al. 2014)”. Similar to above, this should not be in the description of the simulation but in a separate paragraph, and probably in the BCS model description (section 2.2) rather than the simulation description (section 2.3)

[L254] – “Jones et al (2022) showed that the thermophoresis mostly enhances the collection of accumulation mode particles (0.1 – 1 µm)” – Jones et al (2022) primarily showed that rear-capture was the dominant BCS process for much of the raindrop size distribution. Similar to the results of this study, Jones et al found little difference in aerosol burdens when using Wang and Slinn, but a large difference between explicit Slinn + phoresis + rear capture and Wang or Slinn. Would your results have been different if you had used the Slinn+ph+rc scheme rather than Wang as the improved model?

[L264] – “The assumptions result in a smaller range of changes for both rain and snow scavenging values as a function of size, generally within 10% for all particle sizes except for particle within 0.1 µm - 2.0 µm for rain scavenging, where the differences are up to 30% (Wang et al. 2014)” – I’m not sure what you mean - the scavenging coefficients seem to differ by orders of magnitude with size not 10-30%?

[L313] – “HSO3- deposition mostly occurs close to the emission sources, while the wet deposition of the oxidized form, SO4= is the dominant and more efficient in downwind regions” – You have not presented maps of emissions and so it is unclear to the uninformed where in the domain is a source and where is a downwind region. I would suggest as a matter of course that in the supplement you plot the accumulated emissions for each of the species for April / July and then refer to these figures when you mention sources/downwind regions etc.

[L328] “Figure 8 indicates enhancement of the scavenged sulphate particles …” – this is certainly true in most but not all regions. Consider saying “overall enhancement”

[L337] “The 90% confidence intervals show …” Please be more specific about the exact statistical test used. I assume a t-test but this is not mentioned

[L349] “Overall, the Wang2014 scheme has slightly lower HSO3 - caused by the feedback in the model, and mixed changes of SO4=, NO3 - and NH4 +. July” – are these overall changes significant? Please call quantify the domain mean change in these components between the different simulations

[L352] Sentence beginning “Here, the regions where the differences are significant…” - is there a correlation between the deposition anomalies and the precipitation rate? surely this would be easy to determine? Perhaps calculating an R^2 between the spatial maps (see here for a method https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1998WR900018)

[L358] "The lower scavenging of the Slinn's scheme can be explained by its lack of processes such as thermophoresis, which may increase the collection efficiency for particles in the size range of 0.01−1 µm (Jones et al., 2022). This may also explain the underestimation of scavenging coefficient from the Slinn (1984) scheme, and the differences between two schemes for particles below 1 µm (refer to section 2.3)." – there is a lot of repetition here about Slinn missing important processes in the Greenfield gap. Please consider condensing this

[L362] – "Figure S4b indicates the difference between two scavenging schemes." Replace 'indicates' with 'shows' or 'highlights' and note this is for snow rather than rain. Please be clearer and more specific with your descriptions of e.g., figures and results as currently it is rather ambiguous throughout the manuscript

[L364] – "Given the fact that the solid precipitation is dominating in the April precipitation" – specifically refer to Figure 2 here

[L376] – "Comparison of the observed SO4= data with the simulation results (Fig. 12d-f), suggests a better agreement with observations by including the Multiphase rain-snow partitioning, and further improvement in agreement associated with the use of the Wang et al. (2014) scavenging scheme" – to the blind eye, the changes are absolutely minimal. Please include goodness of fit metrics in Fig. 12 (NMB, R2, RMSE, etc) to quantify if there is any improvement as without including these metrics, it is difficult to validate this assertion. If the goodness of fit metrics are in the tables, then refer to them in this sentence as well as Fig 12. Fig 12 alone does not show better agreement

[L380] "(from 0.46 to -0.05)" - for SO2! Please be more careful with what you refer to. The lack of description is jarring

[L381] "Wang2014 experiment has a better correlation (R = 0.86) and better factor 2 (Fac2 = 0.64) values compared to the other two GEM-MACH experiments (Table 1)" – without the baseline correlation and factor 2 scores, it is difficult to gauge whether Wang2014 is an improvement. Include the scores for the other 2 experiments here. Are the differences significant or is it in the noise? If you had run a different case study, would you expect to see the same results? See also [L402] for a similar lack of values in the base case which would aid comparability. I would argue that your concluding remark "Overall, the Wang2014 simulation has superior performance to the base case and multiphase Slinn1984 simulations" is only valid if you directly compare the goodness of fit metrics between the simulations. Additionally, I would argue against your assertion given that the NMB is much worse for Wang2014 for SO4 than for the base case for CAPMON!

[L407] – "The impacts of partitioning and Wang2014 scavenging on modelled ambient concentration of speciated PM2.5" – you now move from ions in rain water (deposition) to near-surface concentrations. I assume they are near-surface, please can you clarify this. Please be more careful when describing the metrics, especially in the figures and their captions. It is difficult to determine at present whether the figures show deposition rates or near surface concentrations as this is lacking from the captions

[L412] – "Corresponding 90% confidence interval scores for the difference plots are shown in the lower panels." – I really like the fact that you include the goodness of fit metrics in Figure 16 – I would prefer that a similar thing was done for Fig 12 to aid visual presentation rather than having the values separately in a table but this is just a suggestion. I also like that in Figs 13-15 that the 90 % CI spatial maps are included alongside the anomalies and wonder why you did not do this for Fig 10 and 11 to aid visual presentation

[L438] "For example, the multi-phase approach resulted in the most significant improvement in modelled SO4= wet deposition flux over Alberta (at APQMP sites, and in comparison to previously published work which had wet sulphate positive biases of +200% across combined CAPMoN and APQMP sites, Makar et al., 2018), as well as improvement in modelled ambient particulate sulfate concentration at NAPS sites" – this is my biggest contention with this paper, I don't think that the conclusions are well supported by the results. For example, from table 1, the SO4 NMB is better compared to APQMP but significantly worse compared to CAPMoN. The changes in R and FAC2 are marginal at best. The conclusion should be more conservative I feel – are any of the results actually significant?

[L447] "The Wang et al., (2014) scheme is based on a semi-empirical approach, providing an overall best fit to an ensemble of existing parameterization and observations." – the phraseology is wrong here. Wang certainly performed a best fit optimisation to some existing models but not to observations. They fit their model to the 90% of the parameterizations – an arbitrary choice meant to emphasize that the upper end of the models best fit with observations. However, when you actually compare the Wang-derived scattering coefficients against Laakso "observations" (see Fig. 3 in Jones et al 2022) there remains a significant disparity between Wang and observations. I think this should be highlighted.

[Figure 1] Longitude coordinates are included for the grid cells but latitude coordinates are not – please include these

[Figure 2] why do you say "at the model hybrid level of 0.98"? This is ambiguous – is it near the surface or high in the atmosphere? You also refer to (2c) when I think you mean (2d)

[Figure 5 and 6] – The captions are not informative enough. Do you mean concentration in rainwater? Are these accumulated totals over the month (mass per area) or fluxes (mass per area per time)? Is it normalised to nitrogen and sulfur totals (e.g., mass[S] per area) or in units of substance mass? In the text ([L310]) it is implied that these are contributions to mean wet deposition which is rather ambiguous. Additionally, why do you use µmoles for deposition and then µg for concentration – I would stick to one or the other?

[Figure 7] I am confused as to why you use the units of m rather than mm for the precipitation accumulations? The values range from ~5e-5 to ~7e-3 so mm would be more appropriate. I assume this is an accumulation rather than a rate (which would have units of per time)

[Figure 8] as with figures 5 and 6, the metric being plotted is not described (i.e., difference in accumulated aerosol deposition). Also, please label the subplots (a,b,c, etc) in all of the figures

[Figure 12] Please include goodness of fit metrics in Fig. 12 (NMB, R2, RMSE, etc) to quantify if there is any improvement as without including these metrics, it is difficult to validate this assertion

[Figure 13-15] – I like these figures. The caption is clear and goodness of fit is included in the lower panels

[Tables 1 and 2] Please say what metric is being evaluated in these tables (accumulated deposition for 1, near-surface concentration for 2)