

Review of “Simulating Lightning NOx Production in CMAQv5.2: Evaluations of Scientific Updates” by Kang et al.

General comments:

This paper uses the CMAQ air quality model, driven by WRF meteorology to present various approaches to simulating lightning NOx over the USA. These approaches utilise the NLDN lightning data to: 1) directly determine Lightning NOx (LNOx) in the model, 2) to combine with simulated convective precipitation to determine LNOx, or 3) develop relationships between CP and lightning, that could then be applied to CP simulation when NLDN observations are not available. These 3 approaches are all valid approaches to consider and provide a useful comparison in this work. The authors describe the approaches and show how the models perform in different years in order to determine their robustness. The authors state that applying the NLDN observations directly will provide the highest fidelity LNOx simulation. However, when observations are not available they conclude that the third option of parametrisation based on CP is appropriate, despite some issues which have been well-described.

In general, I am pleased to see such a paper. Very often lightning parametrisations are not well documented so this paper is welcomed. Furthermore, the thorough description of the 3 approaches will serve other modelling groups well should they be trying to decide how best to develop a lightning scheme for their own models. However, the paper is not ready for publication in its current form. Firstly, data description sections for the NLDN data, and the WRF driving model, and how it couples to CMAQ, need to be added. Secondly, there is no discussion of appropriateness for chemical transport and reactions of applying a LNOx scheme that does not depend on the underlying simulated convection (through use of CP), as in the case with one scheme presented (hNLDN). This point is not a hindrance for the results of this paper which only looks at the LNOx itself, but it may be an issue should one look at ozone or OH for example. Thirdly, the parametrisation developed has relationships that are highly spatially dependent, and therefore the caveats to applying it to different climates must be discussed. In summary, I can see that all the approaches presented here can have their uses, but that better data description and discussion of caveats is needed.

Specific comments:

L33. “future climate studies”: I think it is debatable whether the model can do this because it does not seem necessary to me that the spatial dependency of relationships developed for fig5 will hold in different climates. The authors must at least include discussion of this in the main text and justify their opinion.

L33. “simulations focused outside the NLDN region”: Given the spatially dependent relationships of the model produced, and that there have only been produced over the NLDN region, I don’t see how a model has been developed that simulate anywhere else. Please can authors clarify how their model can be applied elsewhere in the main text. Otherwise, I can see that a method has been developed that could be applied elsewhere where lightning observations exist, so a statement to this effect could still be included.

L65. I think the Murray (2016) paper on lightning and air quality would be nice reference to include here <https://link.springer.com/article/10.1007%2Fs40726-016-0031-7>

L84. This is just one paper that has provided extra evidence on LNOx per flash, and it looks only at the gulf of mexico. For this statement you need to reference at least a selection of the raft of studies that have added to this estimate since the 2007 schumann and huntrieser review. Here’s some

clues, they are not all required (it's not a review) but hopefully you can find some of them (Huntrieser et al., 2008; Cooray et al., 2009; Huntrieser et al., 2009; Bucsela et al., 2010; Ott et al., 2010; Huntrieser et al., 2011; Miyazaki et al., 2014; Pollack et al., 2016)

Before section 2. Need a description of the NLDN since it is integral to this paper. Notably, is it cloud-to-ground or total lightning? You also need some basic description of the WRF model and its version used here, since that is driving the convective precipitation variable that is integral to the paper.

L132. "A local adjustment is applied...": Is this done at the end of each month simulation. Or maybe CMAQ is not run at the same time WRF? You need to clarify these details of the model setup, for it to make sense how the LNOx scheme is being applied.

Fig1. Ideally all the starting points (inputs) would be on the left, leading then to the outputs on the right. I believe these starting points are "NLDN raw data", "ICCG climatology", and "Gridded met data". This would make the flow clearer.

L141. "yield": I find this term confusing. DO you mean NOx per flash? Flashes per CP? Or something else? Please clarify.

Eq3. I'm not entirely clear why Ratio_NLDN2CP and LTratio are both needed. Can the same thing not be achieved with $LTratio = \frac{\text{sum}(\text{NLDN})}{\text{sum}(\text{CP})}$? i.e. apart from the cap of 50, you are fitting a gradient with zero-intercept to each grid cell? Anyway, this is an existing scheme that's already been used so I guess it is what is. There's certainly no problem, I just think it would be clearer not to have two parameters where one would do. If there is a good reason, then it would be worth adding it to the text.

Eq 3. I think it's kind of being implied throughout that NLDN is just CG flashes. THIS is the only way this equation would make sense to me. See earlier comment, please add a description of NLDN early on where you make this explicit.

L164 onwards: this could be a separate subsection in order to break up the various components a bit more clearly.

L164-173. Would be helpful to show a figure of an example distribution, or better an example IC distribution and CG distribution, and the combined distribution.

Eq7. Doesn't the ICCG_ratio need to come into the vertical profile equations somewhere? Possibly this equation. Otherwise the final vertical weightings of the LNOx column will assume equal numbers of each flash type?

Eq7. Why multiply by 0.2? Isn't that something to do with the ocean grid cells, but isn't the lower distribution for CG flashes?

Section 2.2. This updated lightning scheme no longer depends on where convection occurs in WRF. This is often thought to be problematic because in some cases LNOx will not be transported and react as though in a convective environment. Please acknowledge this aspect of the update and give arguments for why it is appropriate.

Fig2. Could add an extra bar for all month correlation.

Fig2. Are all the bars significant? There's only 12 points for each, so worth checking. Could just add a horizontal line at the correlation needed for significance at 5% level.

Fig3. Please use grey for where there is no data, and a different colour for where values are close to zero. Also, rainbow colour bars are unappealing for several reasons <https://www.climate-lab-book.ac.uk/2014/end-of-the-rainbow/> Please consider changing it.

Fig3 scatter. You could add lines of best fit for each colour, and for all, to make things a little clearer.

L236. "East and west". R1 and R5? Or all regions and R5? Please clarify what east and west refers to.

L241. "...and the log-linear is stronger in the upper value range.....". There doesn't look to be much in it to me, the spread around the log-linear best fit line is similar all the way along. I would just remove this end part of the sentence.

Fig5. Could add a panel for the NLDN lightning climatology. This will help interpret the relevance of each location.

Fig5. These fits by location make it questionable to apply in a different climate. It is quite possible the response of lightning to CP for a location could change in a different climate, e.g. updraught strength could feasibly reduce but CP increase. This could affect the lightning production. You need to discuss this point if you want to include any claim that the model can be applied to different climates.

Fig6. Are the log-linear slopes and the intercepts also stable over time? Either add the plots to the figure5 or describe in the text

L280. Is the same version of WRF not used for the whole time period? This must be explained in a WRF data description section that needs to be added before section 2.

L314. "...dynamic cutoff values are used...": please show the resultant column LNOx annual cycle with this approach on fig8

Fig9. Why only these 2 years. Would be fine to have many panels of all available years. Or if these years demonstrate a particular point then fine, but it would be good to add an extra panel with a climatology of each model, with standard deviation bars of each month to show interannual variability

L336. "...poor simulation of 2011 precip...": I think this is too strong. You have not shown the precip is poorly simulated, you have shown that the model based on CP doesn't work as well 2011. Lightning depends on many factors that may not be captured by CP variability. One of these factors may have varied in 2011 leading to poor model performance. I think you can say that one possible explanation for poor model performance is if CP was poorly simulated. If you want to say any more then you need to get precip observations and compare to the simulation.

Fig10. Variance of pNLDN looks too low compared to hNLDN. I think it's worth mentioning this in the text.

Fig11. Colour scale on this figure is not helpful. There needs to be a much larger upper value in order to see the detail. Or a logscale is often useful for such plots.

L347. "...agree with each other for both years...": I can't tell if this is true because the colour scale lose so much of the detail through saturating.

L352. "...without including observations": although reanalysis is driving the WRF model? This is something that needs to be clarified in WRF data description section.

L363. Worth commenting here what other schemes are also available in WRF to do the same thing. E.g. I presume a cloud-top height scheme exists?

L386. Perhaps worth adding a bit more positivity regarding your paper along the lines of: "In this paper we have developed and demonstrated a method that can now be applied to new observations as they become available."

Technical comments:

L120. "inline": Do you mean "online"? or maybe "interactive"?

L139. "...in that...": "in which"?

Fig1. Can the quality of the image be increased? The text isn't as clear as it could be. Increase the font size too?

L167 "as in Wang...": "drawing from Wang who..."?

Fig3 caption. "... for other months". "...for other months (not shown)"

Fig5. Can the image quality be increased.

Figures in generally are of poor image quality. Please can the dpi be increased.

L372. "2018": we are now in 2019. Rephrase the sentence