Responses to Referee #1

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.

We thank the reviewer for the overall positive assessment of the manuscript and for the constructive comments. Incorporation of the reviewer's suggestions and revisions in response to the comments has greatly improved the revised manuscript.

Firstly, data description sections for the NLDN data, and the WRF driving model, and how it couples to CMAQ, need to be added.

In this application, WRF and CMAQ were run separately (offline). CMAQ used the meteorological fields output by WRF and prepared using MCIP as input files. We have now added a data description and model configuration section after the introduction section to address the NLDN data source and model configuration issues.

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.

In all the schemes, the lightning produced NO is treated as an emitted species and its atmospheric flux due to lightning is added to the existing NO from other emissions. This NO then undergoes the same chemical and physical processes as any other emission species does. In terms of the potential mismatch between LNO production and convective transport of other ozone precursors that could occur with the hNLDN scheme, we would recommend running WRF with lightning assimilation (Heath et al., 2016), then convection will be forced to occur at the correct times and locations, and will consequently eliminate any such possible mismatch.

We have provided additional information and recommended the use of lightning assimilation in WRF simulations in Section 3.3: Updates to the lightning module and the LNO production scheme.

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.

Yes, we agree. We have now revised our manuscript to provide the caveat and removed the obscure description regarding the application in areas outside the study domain.

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.

LNOx emission estimates for future climate scenarios are needed to adequately assess both air quality and atmospheric deposition amounts under these scenarios. We include the description here to help serve this need through the development, application and evaluation of the pNLDN approach that attempts to parameterize LNOx emissions as a function of convective rain using historical data. We nevertheless acknowledge the relationship may not hold if the climate changes dramatically in a such that the range of change is outside that of the magnitude of the historical data. We have now added the caveat in the summary and conclusions.

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.

We agree with the reviewer that the current discussion is confusing. We have removed this description in the revised manuscript.

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

We already reference this publication at multiple locations in the manuscript text.

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)

We have now added Bucsela et al., 2010; Huntrieser et al., 2009 and 2011; Ott et al., 2010, in addition to Pickering 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.

Thank you for the suggestion. We have now added the necessary information in the data source and model configuration section.

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.

CMAQ is run offline, i.e., WRF fields were processed using MCIP to provide hourly input for CMAQ simulations. The local adjustment factor varies monthly and is applied at each hour to the input LNOx emissions. We have now provided the clarification in the “data source and model configuration” section.

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.

Thanks for the suggestion. In the revised manuscript, we have revised the figure as suggested.

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

Yes, it means the lightning flash yield per unit of CP. We have now provided this clarification.

Eq3. I’m not entirely clear why Ratio_NLDN2CP and LTratio are both needed. Can the same thing not be achieved with LTratio=sum(NLDN)/sum(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.

Thanks for the comments. It is true that this is an existing scheme that is in the process of being replaced with the updated scheme.
Yes, it would be possible to combine Ratio_NLDN2CP and LTratio into one term that varies with location and month. However, we chose to break it into a term that varies just with month (Ratio_NLDN2CP that converts from mm of convective precipitation per hour to flashes per hour) and a term that varies with location and month (LTratio, unitless). The rationale was that LTratio could be set to 1 for some applications and allowed to vary for others.

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

Thanks. We have now separated this part as Section 3.2.

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.

An example vertical lightning NO distribution is available in Allen et al. (2012), Fig. 1. In all the schemes as shown by the equations in the manuscript, the total column of lightning NO emissions is generated based on lightning flashes, then the column NO is distributed vertically. We don't vertically distribute either IC or CG flashes. Additional clarification is provided in the new data source and model configuration section.

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?

The total number of CG and IC flashes are determined first, the vertical profile equations are applied separately, and then the total emissions from each source are summed.

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?

We agree that equation 7 in its current form could be confusing. To address the reviewer's question and avoid confusion for other readers, in the revised manuscript we have recast equation 7 as follows:

$$W = (Bottom_{Frac} - Top_{Frac}) \times F1 + (Bottom2_{Frac} - Top2_{Frac}) \times F2$$

In Equation 7, the weight (W) at each layer is the combination of two distribution density. The sum of each density through all the model column (all vertical layers) should be 1. However, the wider distribution (WMU=350 hPa, and WSIGMA = 200 hPa) extends beyond the top of the model domain, as such the sum of the first distribution is less than 1 (~0.93), while the second distribution (WMU=600 hPa and WSIGMA = 50 hPa) does add up to 1. Thus, in order to ensure that the sum of W through all the layers is equal to 1 and while also resembling the vertical distribution shown in Allen et al. (2012), each distribution needs to be scaled. The scaling factors F1 and F2 in the revised equation thus control the relative contributions of the two distributions to the vertical allocation of lightning emission. In the current CMAQ configuration, F1 = 1, and F2 = 0.2 (default). Also note that the vertical distribution obtained by this methodology is insensitive to the particular IC/CG ratio present in a given grid cell. When information regarding the vertical distributions for IC/CG flashes becomes available (for example, Lightning Mapping Array data could be used to obtain nominal distributions for IC and CG flashes), the factors F1 and F2 could be derived based on the IC/CG ratio at a particular grid cell to possibly represent variability in LNOx vertical distribution more accurately in time and space.

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.

As we mentioned earlier, in all the schemes, the lightning produced NO is treated as an emission species and it is added to the existing NO from other emissions, then it undergoes the same chemical/photochemical and physical processes as any other emission species do. In terms of the potential mismatch between LNO production and convective transport of other ozone precursors that

could occur with the hNLDN scheme, we would recommend running WRF with lightning assimilation (Heath et al., 2016), so that convection will be forced to occur at the correct times and locations, which will eliminate any such mismatch. We have provided additional information and recommended the use of lightning assimilation in WRF simulations in Section 3.3: Updates to the lightning module and the LNO production scheme.

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

It has been added in the revised Figure 2.

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.

Yes, this Figure has modified as suggested.

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.

We have revised the figure as suggested.

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

It has been revised as suggested.

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

We have modified both the text and the figure caption to convey that R1 constitutes the West region while R2-R5 constitute the East region.

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

The sentence is revised as suggested.

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

The NLDN Lightning flashes over the same modeling domain for July 2012 and January and July 2013 were presented in our earlier publication (Heath et al., 2016).

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.

Thanks, we have revised the manuscript as suggested.

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

Additional panel for the log-linear parameters has been added in the revised Figure 6.

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.

The clarification has been added as suggested in the data source and model configuration section.

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

In Figure 9a, the red line is the resultant column LNOx using the dynamic cutoff values. This same line is now also added to the revised Figure 8 as suggested.

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 stardard deviation bars of each month to show interannual variability

We are able to add more panels for additional years. Since we have been using 2011 and 2013 as the representative years through the manuscript, we would like to stick to these two years to make it consistent. We are working to make additional analysis on all the available years to also assess trends and spatial variability – these will be reported in a future contribution.

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.

Thanks for the suggestion, we have revised the description to remove the obscureness.

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

We have added the description.

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.

This Figure has been revised as suggested. Thanks.

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.

The revised figure should make it clearer and we have revised the description accordingly.

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

By the observations here we mean that the lightning flash observations. We have added the clarification in the data source and model configuration 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?

The WRF model itself doesn’t simulate lightning and thus it doesn’t provide lightning NO_X production schemes. The WRF-Chem model does contain the cloud-top-height lightning prediction scheme, as well as LNOx schemes.

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.”

Thanks, this point is well taken.

Technical comments:
L120. “inline”: Do you mean “online”? or maybe “interactive”?

All these words are interchangeably used in the modeling community. To add a little more clarification, we described it as “based on simulated parameters at run time” versus using static emission inputs.

L139. “…in that…”: “in which”?

Thanks, it has been revised as suggested.

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

Yes, in the revised manuscript, we have improved the quality of all the figures.

L167 “as in Wang…”: “drawing from Wang who…”?

Thanks, this part of the sentence has been removed as suggested by Reviewer 2.

Fig3 caption. “.. for other months”. “…for other months (not shown)”

Thanks, it has been revised as suggested.

Fig5. Can the image quality be increased.
Figures in generally are of poor image quality. Please can the dpi be increased.

Y**es, all the figure quality has been improved by separating them from the main text (the resolution was degraded when the figures were inserted into the main text).**

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

Thanks, we have revised the description with updated reference.

Responses to Referee #2

The manuscript introduces an updated scheme for calculating the lightning NOx emissions by using the gridded hourly NLDN flash data. The updated scheme has improvements in simulating the NOx emissions compared with the previous scheme using the monthly NLDN flash data, which also requires two different scale factors in determining the lightning flash. The study also developed another scheme using linear and log-linear parameters, which is suitable to use when the hourly lightning flash data are not available, or the air quality simulations are set up to run real-time forecast, or future climate simulations. I personally appreciate the content and scope this study introduced. Natural source, such as lightning, will play an important role in determining the O_3 attainment, especially in the western U.S. I think the manuscript is acceptable to be published by the journal. I have some comments that need the authors to address.

We thank the reviewer for his/her positive comments and recommendation for publication.

Major comments:
Allen et al. (2010, 2012) developed the lightning scheme also using the flash rates from the NLDN, the same as the author proposed. I did not see what updates or advances the authors made considering that. Is that mainly because Allen et al. used monthly flash rates, while the authors used hours? Please elaborate.

The lightning scheme in CMAQ5.0 was developed by Allen et al. (2010, 2012) and was based on monthly NLDN data. In order to redistribute the monthly data into the modeling domain for hourly simulations, additional factors are applied based on the meteorological model predicted convective precipitation. Yes, the use of hourly NLDN data is one of the advances over the existing scheme in CMAQ5.0. When the hourly NLDN data are used, the lightning flashes are directly converted into lightning NO without dependence on the quality of the meteorological fields. In addition, the averaging over neighboring grid boxes and use of both linear and non-linear fits in the pNLDN scheme results in a better fit than in Allen et al.

The quality of the figures embedded in the manuscript are really low. I suggest the authors prepare clear plots when they submitted the manuscripts for review.

In the revised manuscript, the plots are separate from the main text and the resolution is improved.

Minor comments:
Line 28: suggest to remove "scheme and associated LNOx"

Following the reviewer's suggestion, we have modified the sentence to read: "We first document the existing LNOX production and vertical distribution algorithm".

Line 64-65: the authors should add some references listing how the previous studies about lightning NOx affect surface ozone, before the authors could make the conclusion of the importance of LNOx.

Many of the references already cited earlier in this paragraph (Murphy, 2016; Ott et al., 2010; and Schumann and Huntrieser, 2007) discuss and summarize estimated impacts of

lightning NOx on surface ozone. We believe these studies provide adequate justification for the importance of LNOx on atmospheric chemistry and resultant air quality.

Line 82: remove "For instance"

Thanks. It is revised as suggested.

Line 94: use abbreviations for "could-to-ground" since it was defined before

Thanks. Change has made.

Line 97-106: I suggest moving this parts into methodology.

Thanks for the suggestion, but we think that this description better fits in the introduction section.

Line 108: what is old and new scheme? It is confusing since the manuscript mentioned at least 4 schemes: previous parameterizations; Allen et al. 2010; hourly NLDN, and the newly developed parametrization scheme.

Following the reviewer's suggestion, we have now modified the sentence as "a preliminary assessment of the spatial and temporal distribution of LNO columns in the existing (mNLDN), updated (hNLDN), and newly developed (pNLDN) schemes."

Line 120: I suggest to remove this paragraph since this lightning NOx option was not discussed later in the manuscript any more.

There are two purposes for developing this manuscript: 1) to update the existing schemes using the most up-to-date information and develop a generic scheme for use without observed lightning data, 2) to document the schemes used in previous CMAQ releases. Even though the preliminary parameterization scheme wasn't discussed later in the manuscript (due to production of unrealistic high LNO rates), we still want to keep it because it existed in earlier CMAQ versions. We feel that inclusion of this brief discussion would be useful to model users who have also used previous versions of the modeling system and these earlier LNOx parameterizations.

Line 131: to convert "what"?

To address the reviewer's question, we have reworded this discussion to (line 164-166 of the revised manuscript as "First, a global factor (lightning yield) is applied at each grid cell to produce lightning flashes from model CP. Then, a local adjustment (LTratio) is applied at each grid cell to ensure that the local CP- and NLDN-based flash rates match."

Line 167-168: remove "Wang et al (1998) …."

Thank you for catching the typo. It is now removed.

Line 170: change to "and NO produced by CG flashes at a lower layer of the atmosphere (600hPa)"

Thanks. Change is made as suggested.

Line 211: CP was already defined in previous content

Thanks. It has been revised.

Line 211: In section 4, the authors used the different version and configurations of WRF to explain the performances of different years LNOx simulations. So here it would be helpful to show the verions/configuration of the WRF from 2002 to 2014.

To address the reviewer's suggestion, we have reworded the discussion on lines 251-253 of the revised manuscript as: "We analyzed meteorological fields generated by the WRF model simulations from 2002 to 2014 over the continental United States to examine the relationship between the observed lightning flashes and the predicted CP. Though the WRF model has evolved over a few versions (from version 3.4 to 3.8), the Kain-Fritsch (KF) convective scheme (Kain and Fritsch, 1990) was used consistently in simulations for all years." We have also added a data description and model configuration section after introduction to address the model configuration issue.

Line 245: convective precipitation was defined earlier

It is now revised.

Line 276-277: rewrite this sentence.

This sentence has been rewritten and now reads (lines 305-309)"As indicated in Figure 6, the spatial patterns of slopes generated using data from different time periods for both linear (upper panel) and log-linear regressions (lower panel) are similar except that larger values are created over the Great Plains east of the mountains when the most recent years' data (2009-2014) were used to perform the linear regression .

Line 280: see comments earlier. Please list the differences for the WRF versions.

Thanks, we have now added the information in the new data source and model configuration section.

Line 335-337: how the authors make the conclusions that the poor relations of NLDN flashes and model predicted CP was associated with the poor simulations of precipitation by WRF?

We agree with both reviewers that poor correlation between NLDN flashes and modeled convective precipitation may not completely be attributable to poor precipitation simulation by WRF. We have revised the discussion to now discuss other factors (in addition to WRF simulation errors) than can influence this poor correlation.

Simulating Lightning NO_x Production in CMAQv5.2: Evolution of Scientific Updates

Daiwen Kang[1]*, Kenneth Pickering[2], Dale Allen[2], Kristen Foley[1], David Wong[1], Rohit Mathur[1], and Shawn Roselle[1]

[1]National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA

[2]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, USA

*Corresponding author: Daiwen Kang, US EPA, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, USA. Tel.: 919-541-4587; fax: 919-541-1379; e-mail: kang.daiwen@epa.gov

Abstract

This work describes the lightning NO_x (LNO_x) production schemes in the Community Multiscale Air Quality (CMAQ) model. We first document the existing LNO_x production scheme and ~~associated LNOx~~ vertical distribution algorithm. We then describe updates that were made to the scheme originally based on monthly National Lightning Detection Network (mNLDN) observations. The updated scheme uses hourly NLDN (hNLDN) observations. These NLDN-based schemes are good for retrospective model applications when historical lightning data are available. For applications when observed data are not available (i.e., air quality forecasts and~~, future~~ climate studies that assume~~ under~~ similar climate conditions~~, and simulations focused outside the NLDN~~), we have developed a scheme that is based on linear and log-linear parameters derived from regression of multiyear historical NLDN (pNLDN) observations and meteorological model simulations. Preliminary assessment for total column LNO_x production reveals that the mNLDN scheme overestimates LNO_x by over 40% during summer months compared with the updated hNLDN scheme that reflects the observed lightning activity more faithfully in time and space. The pNLDN performance varies with year, but it generally produced LNO_x columns that are comparable to hNLDN and mNLDN, and in most cases, it outperformed mNLDN. Thus~~Nevertheless~~, when no observed lightning data are available, pNLDN can provide reasonable estimates of LNO_x emissions over time and space for this important natural NO_x source that influences air quality regulations.

1. Introduction

Lightning nitrogen oxide~~s~~ (LNO$_X$~~; NO$_X$ = NO + NO$_2$~~) is~~are~~ produced by the intense heating of air molecules during a lightning discharge and subsequent rapid cooling of the hot lightning channel (Chameides, 1986). Since NO and NO_2 are often coexistent in equilibrium after immediate release, they~~it~~ are~~is~~ often collectively referred to as nitrogen oxide~~s~~ (NO$_X$; NO$_X$ = NO + NO$_2$). ~~And for~~ NO$_X$ produced by lightning flashes ~~, it~~ is ~~also~~ referred to as lightning NO$_X$ (LNO$_X$) in the literature. As one of the major natural sources of NO~~$_x$~~, LNO~~$_x$~~ is mainly produced in the middle and upper troposphere. It plays an essential role in regulating ozone (O_3) mixing ratios and influences the oxidizing capacity of the troposphere (Murray, 2016). Despite much effort in both observing and modeling LNO~~$_x$~~ during the past decade, considerable uncertainties still exist with the quantification of LNO~~$_x$~~ production and distribution in the troposphere (Ott et al., 2010). Most~~ t studies~~ estimates of global LNO$_X$ production range~~ing~~ from 2 to 8 Tg (N) yr^{-1}, which is ~~or about~~ 10-15% of the total NO$_x$ budget (Schumann and Huntrieser, 2007). However, owing to the concerted efforts to reduce anthropogenic NO$_X$~~$_x$~~ emissions within the U.S. in recent decades, it is expected that the relative burden of LNO$_X$ and its associated impact on atmospheric chemistry will increase. As a result, it is important to include LNO$_X$ even when modeling ground-level air quality and the interaction of air-surface exchange processes.

To simulate the amount of LNO~~$_x$~~ production in space and time in a chemical transport model (CTM), it is important to know: 1) where and when lightning flashes occur, 2) the amount of LNO~~$_x$~~ produced per flash, and 3) how LNO~~$_x$~~ is vertically distributed. Historically, the lightning flash rates are derived with the aid of parameterizations in CTMs (Price and Rind, 1992; Allen et al.,2000, 2010, 2012; Barthe et al., 2007; Miyazaki et al., 2014). Various schemes have been developed for determining LNO~~$_x$~~ production per flash based on assumptions regarding LNO~~$_x$~~ production efficiency per flash or the energy ratio of cloud-to-ground (CG) flashes to intra-cloud (IC) flashes (Schumann and Huntrieser, 2007). The parameterizations, derived based on theoretical analysis (e.g., Price et al. 1997), laboratory studies (Wang et al., 1998), limited aircraft or satellite observations, or a combination of these methods, are generally too simplified and have large uncertainties (Miyazaki et. al., 2014) and cannot represent well the regional and temporal variability of lightning activity (Boccippio, 2001; Medici et al., 2017).

Over the past decades, our understanding of the production and distribution of LNO_x has been greatly improved with the aid of ground-based lightning detection networks (e.g., Nag et al., 2014; Rodger et al., 2006), aircraft measurements for specific storms (e.g., Huntrieser et al., 2011), satellite observations (Pickering et al., 2016; Medici et al., 2017; Boersma et. al., 2005), and modeling studies (e.g. Zoghzoghy et al., 2015; Cummings et al., 2013). ~~For instance, e~~Even though there are still substantial sources of uncertainty, the LNO_x production rate per flash is now more robust than earlier literature estimates (Bucsela et al., 2010; Huntrieser et al., 2009 and 2011; Pickering et al., 2016; Ott et al., 2010).

A LNO_x production module, based on the lightning flash rate and LNO_x parameterizations of Allen et al. (201~~0~~2), was first introduced in the Community Multiscale Air Quality (CMAQ) (Byun and Schere, 2006) model Version 5.0 (CMAQv5.0) that was released in 2012. That scheme, like the schemes used in previous works (Kaynak et al., 2008; Smith and Mueller, 2010, and Koo et al., 2010), uses flash rates from the National Lightning Detection Network (NLDN) (Orville et al., 2002) to constrain LNO_x. Specifically, LNO_x production is proportional to convective precipitation and is scaled locally so that the monthly average convective-precipitation based flash rate in each grid cell matches the average of monthly total NLDN flash rate, where the latter is obtained by multiplying the detection-efficiency adjusted ~~cloud-to-ground~~CG flash rate by Z+1, where Z is the climatological IC/CG ratio from Boccippio et al. (2002). This scheme, even though it is constrained by NLDN data, depends on the upstream convective precipitation predicted by the meteorological model, that may be resolution dependent and ~~itself~~ generally shows low skill and large regional variations (e.g., Casati et al., 2008). With the availability of NLDN lightning flash data, an algorithm is implemented to estimate hourly LNO_x production from NLDN lightning flash data, avoiding the dependence on the presence of convective precipitation in the model. For modeling exercises where the observed lightning flashes are not available (e.g., real-time air quality forecasts~~,~~ and past- or future-year projection studies~~, and air quality simulations focused outside the NLDN~~), different options are needed to provide the LNO_x estimates. A LNO_x parameterization scheme is developed based on the relationship between the observed NLDN lightning flashes and modeled convective precipitation from a set of Weather Research and Forecasting (WRF) model simulations (the model used to create meteorological inputs for CMAQ) of~~from~~ 2002 to 2014 over the continental United States.

In this manuscript, we present the updates/development of the LNO_x module that was released in CMAQ version 5.2 in June 2017 and a preliminary assessment of the spatial and temporal distribution of LNO columns in the existing (mNLDN), updated (hNLDN), and newly developed (pNLDN) ~~previous and~~ schemes. ~~in their production of total LNO_x columns in space and time.~~ In a follow-on manuscript, a comprehensive evaluation of model performance with the various schemes will be presented.

Section 2 of this paper provides the data description and model configurations. Section 3 ~~2 of this paper~~ describes the existing and updated LNO_x schemes in CMAQ that are based on the NDLN data. Section 4~~3~~ presents an analysis of the historical relationship between NLDN lightning flashes and model-predicted convective precipitation. Section 5~~4~~ provides the derivation of parameterization scheme based on the analysis in Section 4~~3~~. Section 6~~5~~ is the assessment of the ~~old and new~~mNLDN, hNLDN, and pNLDN schemes on their production of total LNO_x columns. With discussions, we conclude this study in Section 7~~6.~~

2. Data source and model configuration

2.1 NLDN data

The observed lightning activity data ~~with CG flashes~~ were obtained from the National Lightning Detection Network (NLDN) (Orville, 2008). The raw CG flashes ~~from the raw data~~ were gridded onto the model~~ing~~ horizontal grid cells ~~with either~~ hourly for use in the hNLDN scheme and then ~~values or being~~ aggregated into monthly mean ~~(hourly)~~ values for use in the mNLDN scheme~~depending on the scheme used~~. The NLDN CG flashes have a detection efficiency of 90%-95% and a location accuracy of approximately 500 m. The detection efficiency for NLDN IC flashes is ~~generally~~ lower and more variable (Zhu et al., 2016), so the climatological IC/CG ratio developed by Boccippio et al. (2001) is used to quantify LNO production by IC flashes.

2.2 Model configurations

The meteorological fields used in developing the LNO schemes are provided by WRF (Stamarock and Klemp, 2008). The WRF output fields were processed using the Meteorology-Chemistry Interface Processor (MCIP) to provide input for CMAQ modeling system (Otte and Pleim, 2010). We leveraged on the archived WRF simulations from 2002 to 2014 to derive the regression-based scheme (pNLDN). The archived meteorological outputs were generated from three WRF versions: version 3.4 for 2002-2005, version 3.7 for 2006-2013, and version 3.8 for 2014.

NO is the direct product of lightning flashes, and after release, a large portion of it can be quickly turned into NO_2 by reaction with O_3 and other species in the atmosphere. Under most circumstances, NO and NO_2 coexist in~~under the~~ chemical/photochemical equilibrium, so ~~traditionally the~~ lightning produced nitrogen oxides are generally referred to as LNO_X. But only NO is involved in the actual implementation of the schemes in CMAQ.~~;~~ ~~w~~We, hereafter, refer to all the schemes as LNO schemes. All the LNO schemes include three steps: 1) derive or use observed lightning flashes at a grid cell, 2) translate the lightning flashes into total column lightning NO at the grid cell, and 3) distribute the total column NO among~~into~~ model layers based on vertical distribution algorithms. After the lightning NO is injected into the vertical layers, it is then combined with (added to) the existing NO from other emissions (both anthropogenic and biogenic sources). From there, it undergoes the same chemical/photochemical and physical processes as any other species do.

~~2.~~3. Description of the LNO~~x~~ module in CMAQ: existing schemes and updates

3~~2~~.1 Lightning module and the existing LNO~~x~~ schemes

Beginning with CMAQv5.0, the LNO~~x~~ module contains two options for inline (based on model simulated parameters at the run time) LNO~~x~~ production. The first option is an over-simplified parameterization that assumes that~~ any~~ 1 mm hour^{-1} of convective precipitation (CP) corresponds to 147 lightning flashes for a 36 x 36 km^2 horizontal grid cell (which should be scaled for other resolutions). A preliminary analysis indicated that this scheme produced unrealistically excessive LNO~~x~~ during summer months (not shown). This option was removed from CMAQ in version 5.2.

The second option in CMAQv5.0 was developed by Allen et. al. (2010; 2012) and utilized monthly National Lightning Detection Network (hereafter referred to as mNLDN) flash data. In this scheme, flashes are assumed to be proportional to CP with the relationship varying locally with a two-step adjustment so that monthly average CP-based flash rates match the NLDN observations. First, a global factor (lightning yield) is applied at each grid cell to ~~convert~~ produce lightning flashes from model CP~~ to flashes~~. Then, a local adjustment (LTratio) is applied at each grid cell to ensure that the local CP- and NLDN-based flash rates match. Figure 1 shows the data preprocessing for LNO_x production using mNLDN data in CMAQ. First, CG flashes are gridded onto the modeling grid that is specified in the model input meteorological file using the Fortran program, NLDN_2D. The output (GRIDDED NLDN IOAPI) is the monthly mean lightning flash density (LFD) over the model domain in IOAPI format. Ocean_factor, Calc_strike_factor, and ICCG are R scripts that are used to convert NLDN CG flashes to quantities that are proportional to LNO production. The ~~in whichthat the~~ Ocean_factor script ingests the land-ocean mask and indicates values of 1 for grid cells that contain land and 0.2 for grid cells that only contain ocean. A value of 0.2 is used for oceanic-grid cells because the amount of lightning produced per unit of convective rain is ~~very~~ ~~lightning flash yield from CP of marine convection is~~ approximately five times less for marine convection than for~~that of~~ continental convection (Christian, et al., 2003). The Calc_strike_factor script ingests the gridded NLDN CG lightning flash data and the CP values predicted by the upstream meteorological model WRF to calculate the Ratio_NLDN2CP according to the following equation:

$$Ratio\_NLDN2CP = \frac{\sum_{i=1}^{nT}\sum_{j=1}^{nC} NLDNflashes}{\sum_{i=1}^{nT}\sum_{j=1}^{nC} CP} \quad (1)$$

where nT is the total time steps, and nC is the total grid cells. Ratio_NLDN2CP is the ratio of the monthly average total flashes over the domain to the monthly average CP over the domain, and it is used to convert the CP values to flash rates. The ICCG script interpolates the climatological IC/CG ratio (Boccippio et al., 2001) onto the model grid cells according to their geographical location and month of the year. Then the Fortran program, LTNG_2D_DATA, collects all the information generated in the prior steps plus the LNO_X production rate: moles NO per CG (MOSLN) and IC (MOLSNIC) flash to generate one input file (one file for each month of the year) that contains all the lightning parameters needed by the CMAQ lightning module. An additional local adjustment factor LTratio (monthly value at each grid cell) is needed to ensure that the local CP- and NLDN-based CG flash rates match.

$$LTratio = \frac{\sum_{i=1}^{nT} NLDNflashes}{\sum_{i=1}^{nT} CP \times Ratio\_NLDN2CP} \qquad (2)$$

This value is capped at 50 to avoid placing ~~estimating~~ excessive amounts of lightning-NO emissions in model grid cells with much less ~~minimal~~ CP than observed in an attempt to match observed monthly flash rates. Finally, the moles of NO produced per hour and grid cell is calculated in the lightning module in CMAQ as:

$$CLNO = CP \times Ratio\_NLDN2CP \times LTratio \times Ocean\_factor \times (MOLSN + MOLSNIC \times ICCG) \quad (3)$$

where CLNO is the moles of NO, and Ratio_NLDN2CP x LTratio x Ocean_factor is the lightning yield per unit CP.

3.2 Vertical distribution algorithm

The moles of LNO~~x~~ are then distributed vertically using the two-peak algorithm described in Allen et al. (2012), which is a preliminary version of the segment-altitude distributions (SADs) of flash channel segments derived from Northern Alabama Lightning Mapping Array data by Koshak et al (2014) convolved with pressure~~, as in Wang et al. (1998) found LNOx was proportional to pressure in laboratory experiments~~. A two-peak distribution is used because NO produced by IC flashes ~~occurs~~ is centered at a higher layer of the atmosphere (350 hPa) than ~~and than~~ NO produc~~tion~~ed by CG flashes ~~at a lower layer of the atmosphere~~ (600 hPa). Accordingly, LNO~~x~~ is distributed with two Gaussian normal distributions: the upper distribution has a mean pressure of 350 hPa and a standard deviation of 200 hPa, and the lower distribution has a mean pressure of 600 hPa and a standard deviation of 50 hPa. For each CMAQ layer, the pressure (p) is calculated as following:

$$p = \sigma \times (psfc - ptop) + ptop \qquad (4)$$

where σ is the sigma value of the layer, psfc is the surface pressure, and ptop is the pressure at the top of the model domain.

At each pressure level (p), the ~~cumulative distribution function (CDF) parameter for a~~standardized Gaussian parameter ~~normal distribution~~ (x) is calculated as:

$$x = (p - WMU)/(\sqrt{2} \times WSIGMA) \qquad (5)$$

where WMU is the mean value of the distribution (either 600 hPa or 350 hPa), and WSIGMA is the standard deviation of the distribution (either 50 hPa or 200 hPa).

Then the fraction of the column emissions at the pressure p is calculated by the following distribution function:

$$Frac(x) = 0.5 \times \{1.0 + SIGN(1.0, x) \times \sqrt{1.0 - e^{\left(-4.0 \times \frac{x^2}{\pi}\right)}}\,\} \qquad (6)$$

where SIGN is a function that produces 1.0 if x >= 0, and -1.0 otherwise.

At each model layer, the weighted contribution is:

$$W = (Bottom_{Frac} - Top_{Frac}) \times F1 + (Bottom2_{Frac} - Top2_{Frac}) \times F2 \qquad (7)$$ ~~0.2~~

where W is the weight at a model layer, $Bottom_{Frac}$ and Top_{Frac} are the fractional contribution calculated by Equation (6) at the bottom and top of the model layer, respectively, for the upper distribution peak (WMU = 350 hPa, and WSIGMA = 200 hPa), and $Bottom2_{Frac}$ and $Top2_{Frac}$ are for the lower distribution peak (WMU=600 hPa and WSIGMA = 50 hPa). F1 and F2 are scaling factors that control the relative contributions to W from the top and the bottom distributions, respectively. Ideally, W would match the vertical profile presented in Figure 1 by Allen et al. (2012) and the sum of W at all the layers is equal to 1. In the current CMAQ configuration, F1 = 1 and F2 = 0.2.

Finally, the LNO_x at each layer is:

$$LTEMIS(L) = W(L) \times CLNO \qquad (8)$$

where LTEMIS(L) is the LNO_x at layer L, W(L) is the weight at layer L as calculated by Equation (7), and CLNO is the total column LNO_x.

3.3 Updates to the lightning module and the LNO_x production scheme

As described above, the LNO_x production scheme, mNLDN, calculates CLNO using scaled values of the convective precipitation. To simplify the procedure to generate LNO_x, in

CMAQv5.2 we used the gridded hourly NLDN (hNLDN) flash data in the lightning module, which reduces Equation 3 to:

$$CLNO = NLDNCGflashes \times Ocean\_factor \times (MOLSN + MOLSNIC \times ICCG) \quad (9)$$

NLDNCG flashes are generated using a Fortran program adapted from NLDN_2D by reading in the raw NLDN CG flashes, Ocean_factor and ICCG are the same as in Equation 3, but the R scripts are replaced by a Fortran program to put all these parameters (including the parameters associated with regression analysis described in the next two sections) into one file as parameter input file for CMAQ. MOLSN and MOLSNIC have default values of 350 moles flash^{-1}, but they can be modified in the CMAQ run script via environment variables.

Since the hNLDN scheme directly injects LNO into the modeling grid cells based on observed CG flashes, there is a possibility that a disconnection exists between LNO and other convectively transported precursor species for O_3 production. However, when the lightning assimilation technique (Heath et al., 2016) based on the same observed lightning flashes is applied in WRF simulations, other precursor species will be forced to occur at the correct times and locations. Therefore, it is recommended that lightning assimilation be applied in WRF simulations when hNLDN scheme is used in CMAQ to produce LNO emissions.

~~3.~~4. Examining the relationship between NLDN flashes and modeled CP

The existing LNO_X production schemes in CMAQ depend heavily on ~~convective precipitation (~~CP~~)~~ amounts predicted by WRF. We analyzed meteorological fields generated by the WRF model simulations from 2002 to 2014 over the continental United States to examine the relationship between the observed lightning flashes and the predicted CP. Though the WRF model has evolved over a few versions (from version 3.4~~1~~ to 3.8~~7~~), the Kain-Fritsch (KF) convective scheme (Kain and Fritsch, 1990) was used consistently in simulations for all years. We first examined the relationship between lightning flashes, which were aggregated into hourly flash counts and gridded onto the modeling grid cells and the modeled hourly CP from WRF over the continental US (12 km horizontal grid spacing). The results (not shown) showed little to no correlation between the observed lightning flashes and the predicted CP, regardless of the time period examined. However, when the lightning flashes and CP were each aggregated to mean values over geographical regions (the entire modeling domain as the extreme) for each month in the time series, as shown in Figure 2, the correlation between the two quantities was obvious. This suggests that although the model-predicted CP is not a good predictor of lightning events in space and time, it does show the skill to predict cumulative lightning activity across geographic regions for a given month. Further analysis of the relationship indicates unique distribution patterns in space over the contiguous United States through the years. As shown in Figures 3a and 3b, lightning yields per unit CP are smaller in the eastern US than in other areas confirming that the lightning yield varies regionally. The original scheme used a universal lightning yield for the entire modeling domain, while Allen et al. (2012) allowed the yield to vary locally. This ~~As shown in Equations 1 and 2, the original scheme and Allen et al. (2012) used a universal lightning yield for the entire modeling domain; however, this~~ analysis indicates that the yield is lowest in the east (Region 1) but similar in regions 2–5, which could be combined. Figure 4a shows the scatter plots and the corresponding linear regression equations, as well as the correlation coefficients (r). Again, the data points over the two regions (East: Region 1 and West: Regions 2-5 in Figure 3a) are distinct, and the slope (0.05) associated with the linear regression equation over the East is less than half of the value over the West (0.13), meaning that the lightning yield over the west is more than twice that over the eastern U.S. Further analysis reveals that better relationships exist when logarithmic translation is taken for both NLDN flashes and CP as shown in Figure 4b; i.e., after applying the translation, the correlation coeficients increased for both the West and East regions ~~and the log-linear relationship is stronger at the upper value range than that at the lower value range~~.

4~~.~~5. LNO$_X$ scheme based on the relationship between NLDN flashes and CP

Statistically, the relationship between CP~~convective precipitation~~ rate and NLDN lightning flash rate over large regions suggests similar yields within each region. But considerable scatter still exists within each region and the overall statistics may be dictated by certain large values. As an estimate, the most direct approach would be to use regression equations to determine LNO$_X$ from CP for western U.S. grid cells and regression equations for eastern U.S. grid cells as shown in Figures 4a and 4b. However, in addition to the concern associated with variations within a region ~~mentioned earlier,~~ this direct application would also cause some practical problems: 1) the analysis regions are arbitrary; and 2) the LNO$_X$ production would be spatially inconsistent with abrupt changes along the bordering grid cells separating regions. Therefore, instead of deriving regression equations using the regional data, linear (log-linear) regression equations are derived using data averaged over an area of adjacent grid cells (analogous to the derivative concept to cut regions into small areas that cover adjacent model grid cells). In areas that lack enough data points to perform the regression, data are filled using the inverse-distance weighting (IDW) spatial interpolation technique (Lu and Wong, 2008). Figure 5 shows the spatial linear (upper panel) and log-linear (lower panel) regression parameters and the correlation coefficients over patches of 3 x 3 grid cells (36 x 36 km^2 in area) using the data from 2002 to 2014, respectively. As shown in Figure 5, significantly larger slope values appear over the Mountain West and Central Plains states indicating a greater lightning yield per unit CP over these regions than in other regions. Comparison of the two correlation coefficient maps reveals that the log-linear relationship has higher correlations over larger areas than the simple linear relationship. However, both approaches have correlations >0.5 in regions with frequent lightning activity.

4.1 Stability over time

A robust parameterization scheme should be relatively insensitive to the training time period. In order to test this, the lightning yield (slope of the linear and log-linear regression was re-calculated using data from 2002-2012 (P02-12), 2002-2014 but excluding 2011 and 2013 (P02-14sb2), and 2009-2014 (P09-14). The r~~R~~esults are shown in Figure 6. ~~Cross-examination of~~ As indicated in Figure 6~~s 6a–c and Figure 5 (upper left)~~, ~~indicates that~~ the spatial patterns of slopes generated using data from different time periods for both linear (upper panel) and log-linear regressions (lower panel) are ~~very~~ similar except that larger values are created ~~except~~ over the Great Plains east of the mountains when the most recent years' data (2009-2014) were used to perform the linear regression. This difference may be attributable to the evolution of the WRF model and the NLDN data (Nag et al., 2014) through the years, and it also indicates that the parameters need to be updated to include the most recent data available.

To test the sensitivity of LNO_x to the parameters derived from different time periods, Figure 7 shows the total monthly column LNO_x for 2011 and 2013 generated using different set of parameters derived using linear regression from different time periods, and for comparison, the LNO_x produced by the updated NLDN based scheme, hNLDN, described in Section 2 is also included. As shown in Figure 7a, in 2011 the parameter schemes (pNLDN) (except for P09-14) tend to underestimate LNO_x during summer months (June, July, and August, JJA) compared with hNLDN scheme, but in 2013 (Figure 7b), the pNLDN schemes are mixed in producing LNO_x with both over- and under- estimate during the summer months. In both years, very small differences are observed with the pNLDN scheme with parameters from different time periods except P09-14. P09-14 parameters seem to produce the most LNO_x during summer months in both years making it the best to match LNO_x produced by hNLDN scheme in 2011 but it yields more overestimation in June and July of 2013.

4.25.2 Sensitivity to logarithmic scales

As discussed earlier, the log-linear regression between NLDN lightning flashes and CP produced better correlation coefficients than the simple linear regression. We also noticed, however, that if the log scale parameters are applied to all the data, too much LNO_x is produced relative to the hNLDN scheme, especially during winter months when both lightning activity and convective precipitation occur less frequently. This high bias exists because the log scale tends to inflate contributions from small values when linear regression is performed after the log transformation. To test the impact of log scale on the production of LNO_x, we choose the summer months (JJA) in 2011 and specify a series of cutoff values for CP (cm), that is, linear regression parameters are applied if CP is smaller than a specific cutoff value, and log-linear regression parameters are applied if otherwise. Figure 8 shows the monthly total column LNO_x produced with CP cutoff values from 0.1 (P01) to 0.6 (P06) cm. As indicated in Figure 8, the smaller the cutoff value is, the more LNO_x produced. When the cutoff value of 0.2 is applied, LNO_x production best matched those produced by hNLDN; however, the summer months in 2011 are different from other years, in that significantly more lightning flashes and convective precipitation were observed in the continental US, especially in the east and southeast US. When the same cutoff value (0.2) is applied to other years, LNO_x is overestimated compared with that produced by hNLDN scheme. For generalized application to all years, dynamic cutoff values are used with this scheme (the result is also shown in Figure 8). Specifically, if CP is greater than the intercept value at a location from linear regression, the log-linear regression parameters are used; otherwise, the linear regression parameters are applied. This technique demonstrates acceptable results for all the years studied.

5.6. Assessment of LNO_x production schemes

As a preliminary assessment of these LNO_x production schemes, we only investigate the distribution of column LNO_x in time and space; a more detailed evaluation of the impact of these schemes on air quality will be presented in a subsequent study.

Figure 9 shows the monthly total column LNO_x produced by the different schemes for the years 2011 and 2013. For both years, mNLDN scheme tends to generate significantly more LNO_x during warm months (May–September) than hNLDN and pNLDN schemes. Collectively during May–September, mNLDN produced about 40% (39% in 2011 and 42% in 2013) more LNO_x than hNLDN. The regression parameter-based scheme, pNLDN, underestimated LNO_x during summer months (JJA) in 2011 compared to hNLDN, but the two schemes generally agree well in 2013. As mentioned earlier, the significant underestimate of LNO_x by pNLDN may be attributed to underestimated convective precipitation in WRF, which reduced the count of lightning flashes during this period. There were about 17% more lightning flashes during JJA in 2011 than the same period in 2013 over the continental US. The relatively poor ~~simulation of 2011 precipitation is also evident in Figure 2 as the~~ correlation coefficient between NLDN flashes and model predicted CP values in 2011 is also evident in Figure 2 which was the second least ~~in 2011~~ among the 13 years studied. The daily total column LNO_x produced by these schemes for July 2011 and July 2013 is presented in Figure 10. Among the schemes, mNLDN produced the most LNO_x on most of the days in July for both years. Except for a few days, pNLDN underestimated LNO_x in 2011 relative to the other approaches, but in 2013 it produced comparable results to hNLDN except that for the first few days of the month, LNO_x was overestimated by pNLDN. In addition, the day-to-day variance genera~~e~~ted by pNLDN seems smaller compared with hNLDN for both years.

The spatial distributions of monthly total column LNO_x produced by each of the three schemes over the contiguous United States for July 2011 and July 2013 are presented in Figure 11. Overall, the spatial patterns generally agree with each other for both years with pNLDN producing relatively smaller values ~~a~~, especially ~~but the patterns produced by pNLDN deviate~~ along the edges or over locations where LNO_x amounts are relatively small. Note that both hNLDN and mNLDN are based on the same monthly observed data, so consequently they produced similar spatial patterns. The pNLDN is derived based on the linear and log-linear regression parameters using multiple years' historical observed data and model simulations with different versions, and it is applied to a specific period without including observations. Nevertheless, as the main intention for pNLDN to be applied is when there are no observed lightning data available (such as air quality forecasts and past or future climate simulations with similar climate conditions), it can provide the reasonable estimate for LNO_x comparable to hNLDN and mNLDN.

~~6.~~7. Summary and discussions

In this study, we described the LNO_x production schemes in the CMAQ model's lightning module and updated the existing monthly NLDN observation-based scheme with the current understanding and resources. For retrospective model applications, the hourly NLDN observation-based scheme, hNLDN, is expected to provide the highest-fidelity spatial-temporal LNO_x. If observations are not available, such as in air quality forecasts and future climate studies, the linear and log-linear regression parameter-based scheme, pNLDN, provides a spatial-temporal estimate of LNO_x. Note that even though the pNLDN scheme can provide LNO estimates for past or future climate studies, the spatial dependency of the relationship presented here may not hold under ~~with significant~~ changinge ~~of~~ climate conditions.

Large uncertainties are still associated with each of these schemes resulting from the various assumptions common to all the LNO_x production schemes, e.g., the uniform NO_x production rate per flash, the IC/CG ratios, the difference of LNO_x production rates over land and ocean, and uniform vertical profiles in time and space. The regression parameter-based scheme suffers additional uncertainties resulting from the way the parameters are derived. First, the CP values were only produced by the KF convective scheme in this regression analysis. If other convective schemes are used in the upstream meteorological model, the regression relationship will differ. Spatially this scheme is only applicable to the area over which the regression analysis was performed (here, the contiguous United States). In addition, the parameters may need to be reproduced when the model resolution or version is changed or when updated observational data ~~and model simulations~~ become available.

Lightning and LNO_x will remain an active research area in atmospheric sciences for the forseeable future. For example, lightning data from ~~especially when the~~ Geostationary Lightning

Mapper (GLM) instruments on ~~on board~~ the Geostationary Operational Environment Satellite (GOES) 16 and 17 ~~R (GOES-R) series~~ (Goodman et al., 2013; Rudlos~~h~~ky et al., 2019) are now~~becomes~~ ~~fully operational in 2018~~publicly available. With more observations (both at surface and in space) available, the assumptions associated with the LNO_x schemes will be updated to reflect the evolving understanding of LNO_x production in time and space. For example, Medici et al. (2017) recently updated IC/CG ratios over the contiguous United States based on the relative occurrence of CG and IC flashes over an 18.5-year period. Their study updates the Boccippio et al. (2001) climatology used in this study that employed 4-year datasets. In addition, NASA George C. Marshall Space Flight Center is updating the vertical distributions of lightning channel segments (SAD) based on 9-year North Alabama Lightning Mapping Array (NALMA) datasets (W. Koshak, personal communication, 2018). In addition, the Lightning Mapping Array data could be used to obtain nominal distributions of IC and CG flashes and that information could be used to derive the scaling factors (F1 and F2) associated with the vertical LNO distribution algorithm in Equation 7, thus the vertical LNO distribution could be represented more accurately in time and space. When all these data are available, we will examine and adapt these updates to the lightning parameterizations and make them available in future CMAQ releases. In this paper we have developed and demonstrated a method that can now be applied to new observations as they become available.

Code and data availability

CMAQ model documentation and released versions of the source code, including all model code used in his study, are available at https://www.epa.gov/cmaq. The data processing and analysis scripts are available upon request. The WRF model is available for download through the WRF website (http://www.wrf-model.org/index.php).
The raw lightning flash observation data used are not available to the public but can be purchased through Vaisala Inc. (https:// www.vaisala.com/en/products/systems/lightning-detection). The immediate data except the lightning flash data behind the figures are available from https://zenodo.org/record/2590452 (Kang, et al., 2019). Additional input/output data for CMAQ model utilized for this analysis are available upon request as well.

Disclaimer: The views expressed in this paper are those of the authors and do not necessarily represent the views or policies of the U.S. EPA.

Author Contribution

Daiwen Kang: data collection, algorithm design, model simulation, analysis, and manuscript writing.

Kenneth Pickering: algorithm formation and manuscript writing.

Dale Allen: algorithm formation and manuscript writing.

Kristen Foley: algorithm formation, data analysis, and manuscript writing.

David Wong: code update.

Rohit Mathur: manuscript writing.

Shawn Roselle: manuscript writing.

Acknowledgement:

The authors thank Brian Eder, Golam Sarwar, and Tanya Spero (U.S. /EPA) for their constructive comments and suggestions during the internal review process.

References

Allen, D. J., Pickering, K. E., Stenchikov, G., Thompson, A., and Kondo, Y.: A three-dimensional total odd nitrogen (NOy) simulation during SONEX using a stretched-grid chemical transport model, J. *Geophys. Res.*, 105, doi:10.1029/2010JD014062, 2000.

Allen, D. J., Pickering, K. E., Duncan, B., and Damon, M.: Impact of lightning NO emissions on North American photochemistry as determined using the Global Modeling Initiative (GMI) model, *J. Geophys. Res.*, 115, doi:10.1029/2010JD014062, http://dx.doi. org/10.1029/2010JD014062, 2010.

Allen, D. J., Pickering, K. E., Pinder, R. W., Henderson, B. H., Appel, K. W., and Prados, A.: Impact of lightning-NO on eastern United States photochemistry during the summer of 2006 as determined using the CMAQ model, *Atmos. Chem. Phys*., 12, 1737–1758, doi:10.5194/acp-12-1737-2012, 2012.

Barthe, C., Pinty, J. -P., and Mari, C.: Lightning-produced NOx in an explicit electrical scheme tested in a Stratosphere-Troposphere Experiment: Radiation, Aerosols, and Ozone case study, *J. Geophys. Res.,* 112, D04302, doi:10.1029/2006JD007402, 2007.

Boccippio, D. J., Cummins, K. L., Christian, H. J., and Goodman, S. J.: Combined Satellite- and Surface-Based Estimation of the Intracloud–Cloud-to-Ground Lightning Ratio over the Continental United States, *Mon. Weather Rev*., 129, 108–122, 2001.

Boersma, K. F., Eskes, H. J., Meijer, E. W., and Kelder, H. M.: Estimates of lightning NOx production from GOME satellite observations, *Atmos. Chem. Phys*., 5, 2311–2331, 2005, http://www.atmos-chem-phys.net/5/2311/2005/.

Bucsela, E. J., Pickering, K. E., Huntemann, T. L., Cohen, R. C., Perring, A., Gleason, J. F., Blakeslee, R. J., and Albrecht, R. I.: Lightning-generated NOX seen by the ozone monitoring instrument during NASA's Tropical Composition, Cloud and Climate Coupling Experiment (TC4). *J. Geophys. Res*., 115, D00J10, doi:10.1029/2009JD013118, 2010.

Byun, D. W. and Schere, K. L.: Rewiew of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system, *Appl. Mech. Rev*., 59, 51-77, 2006.

Casati, B., Wilson, L., Stephenson, D., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U.,Ebert, E., Brown, B., and Mason, S.: Forecast verification: current status and future directions, Meteorol. Appl., 15, 3–18, 2008.

Chameides, W. L.: The role of lightning in the chemistry of the atmosphere. In The Earth's Electrical Environment, Chapter 6, National Academy Press, Washington, D. C., ISBN 0-309-03680-1, 1986.

Christian, H. J., Blakeslee, R. J., Boccippio, D. J., Boeck, W. L., Buechler, D. E., Driscoll, K. T., Goodman, S. J., Hall, J. M., Koshak, W. J., Mach, D. M., and Stewart, M. F.: Global frequency and distribution of lightning as observed from space by the Optical Transient Detector, *J. Geophys. Res*., 108(D1), 4005, doi:10.1029/2002JD002347, 2003.

Cummings, K. A., Huntemann, T. L., Pickering, K. E., Barth, M. C., Skamarock, W. C., Holler, H., Betz, H. -D., Volz-Thomas, A., and Schlager, H.: Cloud-resolving chemistry simulation of a Hector thunderstorm, *Atmos. Chem. Phys*., 13, 2737–2777, doi:10.5194/acp-13-2757-2013, 2013.

Goodman, S. J., Blakeslee, R. J., Koshak, W. J., Mach, D., Bailey, J., Buechler, D. Carey, J., Schultz, C., Bateman, M., McCaul Jr., E., and Stano, G.: The GOES-R Geostationary Lightning Mapper (GLM), *Atmos. Res*., 125-126, 34-39, doi:10.1016/j.atmosres.2013.01.006, 2013.

Heath, N. K., Pleim, J. E., Gilliam, R. C., and Kang, D.: A simple lightning assimilation technique for improving retrospective WRF simulations. *J. Adv. Model. Earth Syst*., 8, 1-19, doi:10.1002/2016MS000735, 2016.

Huntrieser, H., Schlager, H., Lichtenstern, M., Roiger, A., Stock, P., Minikin, A., Höller, A., Schmidt, K., Betz, H.-D., Allen, G., Viciani, S., Ulanovsky, A., Ravegnani, F., and Brunner, D.: NOX production by lightning in Hector: first airborne measurements during SCOUT-O3/ACTIVE. *Atmos. Chem. Phys*., 9, 8377–8412, doi:10.5194/acp-9-8377-2009, 2009.

Huntrieser, H., Schlager, H., Lichtenstern, M., Stock, P., Hamburger, T., Holler, H., Schmidt, K., Betz, H. D., Ulanovsky, A., and Ravegnani, F.: Mesoscale convective systems observed during AMMA and their impact on the NO*x* and O_3 budget over West Africa. *Atmos. Chem. Phys*., 11, (6):2503–2536,. doi:10.5194/acp-11-2503-2011, 2011.

Kang, D., Pickering, K., Allen, D., Foley, K., Wong, D., Mathur, R., and Roselle, S.: data set, https://doi.org/10.5281/zenod.2590452, 2019.

Kaynak, B., Hu, Y., Martin, R. V., Russell, A. G., Choi, Y., and Wang, Y.: The effect of lightning NO*x* production on surface ozone in the continental United States. *Atmos Chem Phys*. 8(17):5151–5159. doi:10.5194/acp-8-5151-2008, 2008.

Koo, B., Chien, C. J., Tonnesen, G., Morris, R., Johnson, J., Sakulyanontvittaya T., Piyachaturawat, P., and Yarwood, G.: Natural emissions for regional modeling of background ozone and particulate matter and impacts on emissions control strategies. *Atmos. Environ*.,44(19):2372–2382. doi:10.1016/j.atmosenv.2010.02.041, 2010. .

Lu, G. Y., and Wong, D. W.: An adaptive inverse-distance weighting spatial interpolation technique, *Computers & Geosciences*, 34, 1044-1055, 2008.

Medici, G., Cummins, K. L., Cecil, D. J., Koshak, W. J., and Rudlosky, S. D.: The intracloud lightning fraction in the contiguous United States, *Mon. Wea. Rev*., 145, 4481–4499, doi:10.1175/MWR-D-16-0426.s1, 2017

Miyazaki, K., Eskes, H. J., Sudo, K., and Zhang, C.: Global lightning NOx production estimated by an assimilation of multiple satellite data sets, *Atmos. Chem. Phys*., 14, 3277-3305, doi:10.5194/acp-14-3277-2014, 2014.

Murray, L. T.: Lightning NO_x and Impacts on Air Quality, *Curr Pollution Rep.*, doi: 10.1007/s40726-016-0031-7, 2016.

Nag, A., Murphy, M. J., Cummins, K. L., Pifer, A. E., and Cramer, J. A.: Recent Evolution of the U.S. National Lightning Detection Network, 23rd Intl. Lightning Detection Conference, Tucson, Arizona, USA,18-19 March 2014. http://www.vaisala.com/en/events/ildcilmc/Pages/ILDC-2014-archive.aspx

Novak, J. H. and Pierce, T. E.: Natural emissions of oxidant precursors, *Water Air Soil Poll*., 67, 57-77, 1993.

Orville, R. E., Huffines, G. R., Burrows, W. R., Holle, R. L., and Cummins, K. L.: The North American Lightning Detection Network (NALDN) – first results: 1998-2000, *Mon. Wea. Rev*., 130, 2098–2109, 2002.

Orville, R. E.: Development of the National Lightning Detection Network, *Bull. Am, Meteorol. Soc*., 89, 180–190, doi:10.1175/BAMS-89-2-180, 2008.

Ott, L. E., Pickering, K. E., Stenchikov, G. L., Allen, D. J., DeCaria, A. J., Ridley, B., Lin, R.-F., Lang, S., and Tao, W.-K.: Production of lightning NO_X and tis vertical distribution calculated from three-dimensional cloud-scale chemical transport model simulations, *J. Geophys. Res.*, 115, D04301, doi:10.1029/2009JD011880, 2010.

Otte, T. L., and Pleim, J. E.: The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system: updates through MCIPv.3.4.1. Geosci. Model Dev., 3, 243-256, doi:10.5194/gmd-3-243-2010, 2010.

Pickering, K. E., Bucsela, E., Allen, D., Ring, A., Holzworth, R., and Krotkov, N.: Estimates of lightning NO_X production based on OMI NO_2 observations over the Gulf of Mexico, *J. Geophys. Res. Atmos.*, 121, 8668-8691, doi:10.1002/2015JD024179, 2016.

Price, C., Penner, J., and Prather, M.: NO_X from lightning 1. Global distribution based on lightning physics, *J. Geophys. Res.*, 102, 5929-5941, 1997.

Price, C., and Rind, D.: A simple lightning parameterization for calculating global lightning distributions. *J. Geophys. Res.*, 97, 9919-9933, doi:10.1029/92JD00719, 1992.

Rodger, C. J., Werner, S., Brundell, J. B., Lay, E. H., Thomson, N. R., Holzworth, R. H., and Dowden, R. L.: Detection efficiencyof the VLF World-Wide Lightning Location Network (WWLLN): Initial case study. *Ann. Geophys*., 24, 3197–3214, doi:10.5194/angeo-24-3197-2006, 2006.

Rudlosky, S. D., Goodman, S. J., Virts, K. S., and Bruning, E. C.: Initial Geostationary Lightning Mapper Observations. Geophys. Res. Lett., 46, 1097-1104, doi:10.1029/2018GL081052, 2019.

Schumann, U. and Huntrieser, H.: The global lightning-induced nitrogen oxides source, *Atmos. Chem. Phys*., 7, 3823-3907, doi:10.5194/acp-7-3823-2007, 2007.

Smith, S. N., and Mueller, S. F.: Modeling natural emissions in the Community Multiscale Air Quality (CMAQ) Model-I: building an emissions data base. *Atmos Chem Phys*., 10(10):4931–4952. doi:10.5194/acp-10-4931-2010, 2010.

Skamrock, W. C., and Klemp, J. B.: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, 227, 3465-3485, doi:10.1016/j.jcp.2007.01.037, 2008.

Zhu, Y., Rakov, V. A., Tran, M. D., and Nag, A.: A study of National Lightning Detection Network responses to natural lightning based on ground truth data acquired at LOG with emphasis on cloud discharge activity. *J. Geophys. Res.*, 121, 14,651-14,660, doi:10.1002/2016JD025574, 2016.

Zoghzoghy, F. G., Cohen, M. B., Said, R. K., Lehtinen, N. G., and Inan, U. S.: Ship-borne LF-VF oceanic lightning observations and modeling, *J. Geophys. Res. Atmos*., 120, 10890-10902, doi:10.1002/2015JD023226, 2015.

Wang, Y., DeSilva, A. W., Goldenbaum, G. C., and Dickerson, D. D.: Nitric oxide production by simulated lightning: Dependence on current, energy and pressure, *J. Geophys. Res.*, 103, 19,149-19,159, 1998.