Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

Geoscientific Model Development
Discussions
Open Access

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

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

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, future climate studies, 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. 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.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

Geoscientific Model Development Discussions

1. Introduction

Lightning nitrogen oxides (LNO_X; $NO_X = NO + NO_2$) are produced by the intense heating of air molecules during a lightning discharge and subsequent rapid cooling of the hot lightning channel (Chameides, 1986). 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 studies estimate global LNO_X production ranging from 2 to 8 Tg (N) yr^{-1} 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 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, even though there are still substantial sources of uncertainty, the LNO_X production rate per flash is now more robust than earlier literature estimates (Pickering et al., 2016).

A LNO_X production module, based on the lightning flash rate and LNOx parameterizations of Allen et al. (2010), 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 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 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, 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) 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 old and new 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 describes the existing and updated LNO_X schemes in CMAQ that are based on the NDLN data. Section 3 presents an analysis of the historical relationship between NLDN lightning flashes and model-predicted convective precipitation. Section 4 provides the derivation of parameterization scheme based on the analysis in Section 3. Section 5 is the assessment of the old and new schemes on their production of total LNO_X columns. With discussions, we conclude this study in Section 6.

2. Description of the LNO_X module in CMAQ: existing schemes and updates

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 $hour^{-1}$ convective precipitation (CP) corresponds to 147 lightning flashes for a 36x36 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 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

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

Geoscientific Model Development Discussions

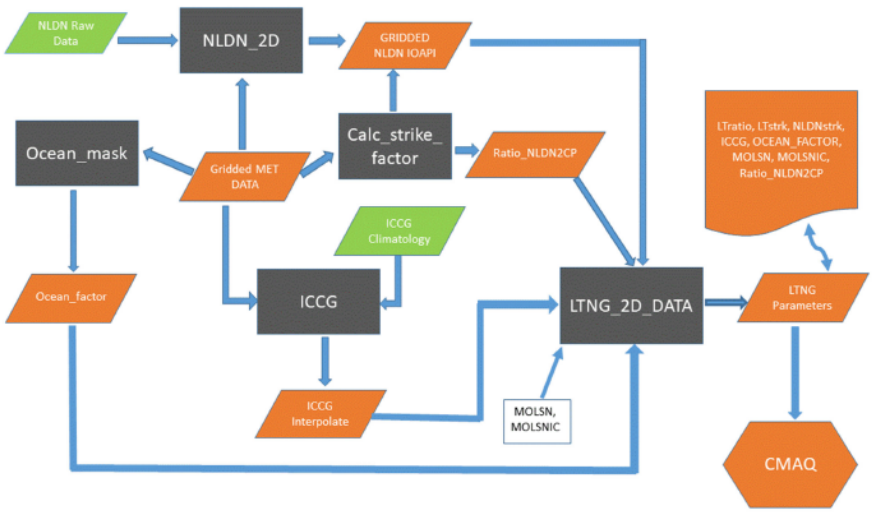

Figure 1. Flowchart of data preprocessing for LNOx production using mNLDN scheme in CMAQ. The black diamonds are the R scripts or Fortran programs and the texts within the diamonds are the names of the scripts/programs used with CMAQ release, the green parallelograms are external data files and the orange parallelograms are output files, and MOLSN and MOLSNIC are two constant values.

preprocessing for LNOx 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 in that the Ocean_factor 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 lightning yield of marine convection is approximately five times less than 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:

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

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

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

Geoscientific Model Development Discussions
Open Access

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 estimating excessive amounts of lightning-NO emissions in grid cells with minimal CP. 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.

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 at a higher layer of the atmosphere (350 hPa) than NO production by CG flashes (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 Gaussian 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:

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

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).

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.

2.2 Updates to the lightning module and the LNOx 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:

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

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

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.

3. 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 3.1 to 3.7), 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 lighting 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

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

Geoscientific Model Development Discussions
Open Access
EGU

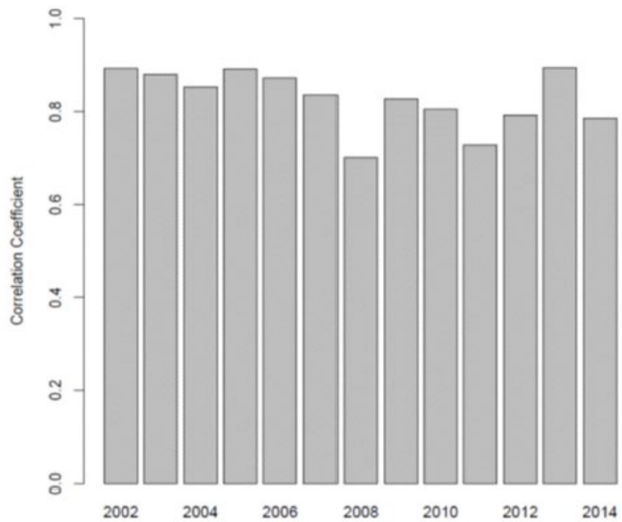

Figure 2. Correlation coefficients between 12 monthly mean NLDN lightning flash density and mean convective precipitation from 2002 to 2014 over the model domain.

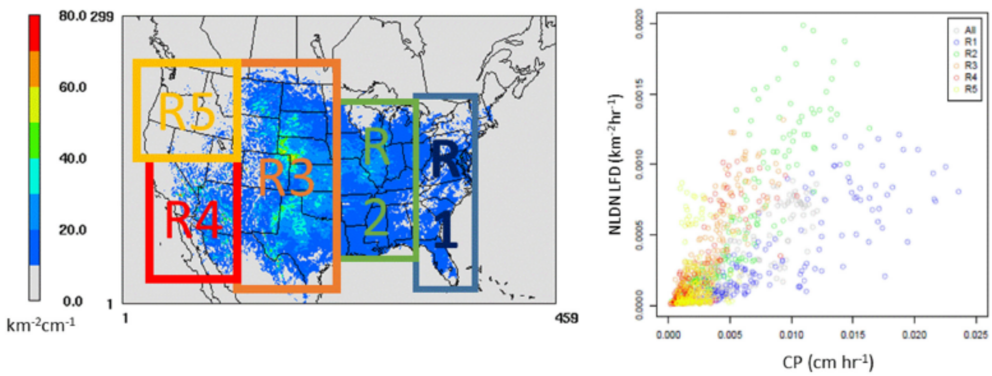

Figure 3. a. The ratio (background) between lightning flash density and modeled convective precipitation (CP) in July (2002-2014; similar patterns for other months) and the analysis regions (R1 to R5). b. Comparison of monthly mean NLDN lightning flash density (km^{-2} hr^{-1}) and modeled convective precipitation for the domain (All) and regions (R1 to R5) from 2002-2014. Each plotted pixel represents the monthly mean value (13 (years) x 12 (months) total pixels) over each region.

confirming that the lightning yield varies regionally. 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

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

Geoscientific Model Development Discussions
Open Access
EGU

linear regression equations, as well as the correlation coefficients (r). Again, the data points over

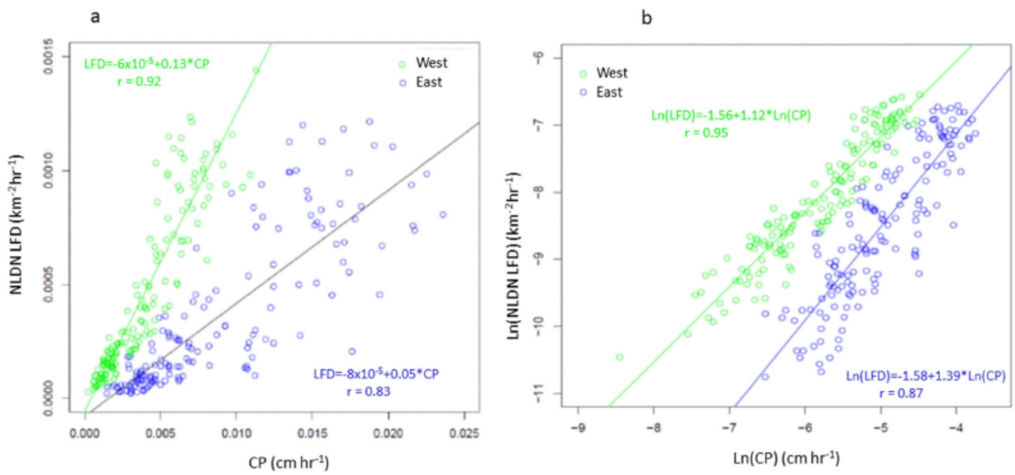

Figure 4. Comparison of monthly mean NLDN lightning flash density (km^{-2} hr^{-1}) and modeled convective precipitation for the West (green) and East (blue) from 2002-2014: a. linear scale, b. logarithmic scale. Each plotted pixel represents the monthly mean value (13 (years) x 12 (months) total pixels) over each region.

the two regions (East and West) 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; 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. LNO_X scheme based on the relationship between NLDN flashes and CP

Statistically, the relationship between 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

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

analysis regions are arbitrary; and 2) the LNOx 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).

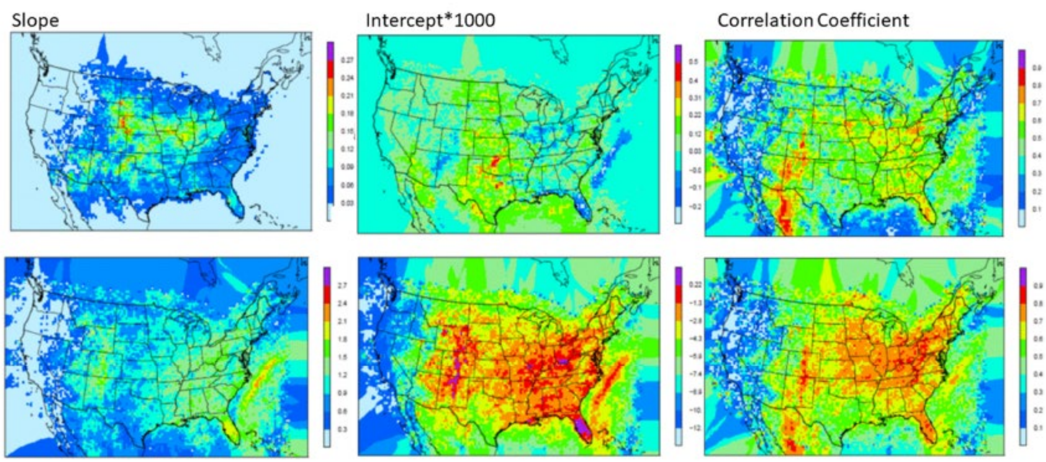

Figure 5. Parameters of linear (upper frame) and logarithmic linear (lower frame) regression parameters generated using all the data from 2002-2014: left column: Slope, middle column: Intercept, and right column: Correlation coefficient.

Figure 5 shows the spatial linear (upper panel) and log-linear (lower panel) regression parameters and the correlation coefficients over patches of 3x3 grid cells (36x36 km^2 in area) using the data from 2002 to 2014, respectively. As shown in Figure 5, significantly large 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.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

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 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). Results are shown in Figure 6. Cross-examination of Figures 6a-c and Figure 5 (upper left) indicates that the spatial patterns of slopes generated using data from different time periods 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

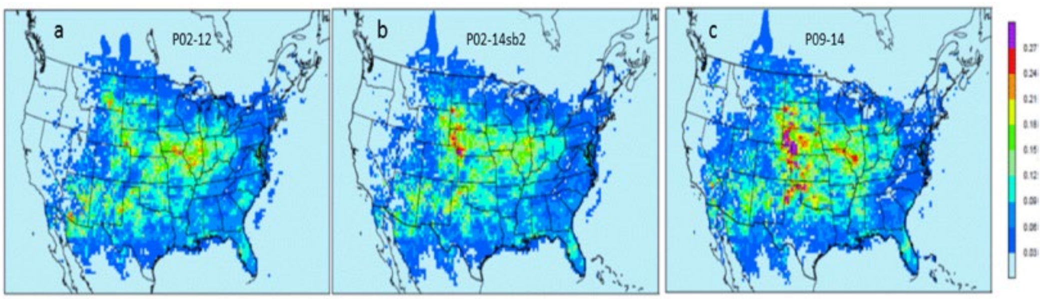

Figure 6. The slope maps from linear regression using data from different time period. a. Data from 2002-2012, b. Data from 2002-2014 excluding 2011 and 2013, c. Data from 2009-2014.

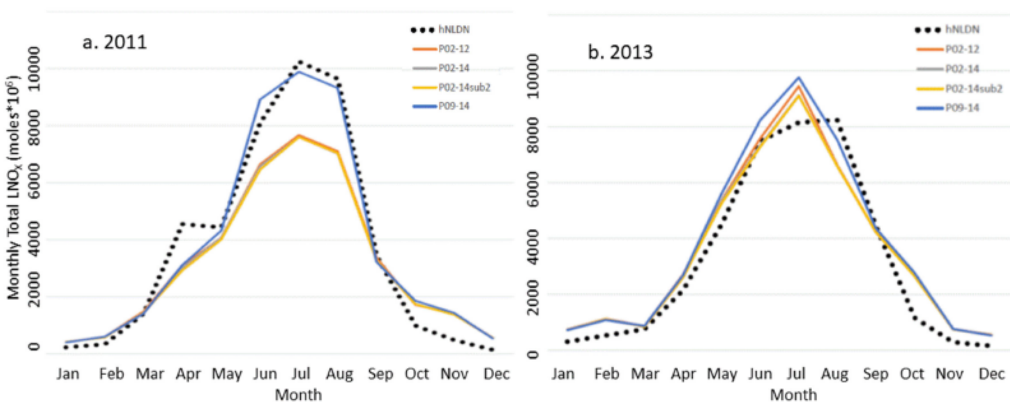

Figure 7. Total monthly column LNOx over the model domain using parameters derived from different time periods for a. 2011 and b. 2013. NLDN: LNOx is produced by the hourly NLDN lightning flashes, P02-12: parameters derived using data from 2002-2012, P02-14: parameters derived using data from 2002-2014, P02-14sb2: parameters derived using data from 2002-2014 excluding 2011 and 2013, P09-14: parameters derived using data from 2009-2014.

regression. This difference may be attributable to the evolution of the WRF model and the

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

Geoscientific Model Development Discussions

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.

4.2 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

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

Geoscientific Model Development
Discussions
Open Access

regression parameters are applied if CP is smaller than a specific cutoff value, and log-linear

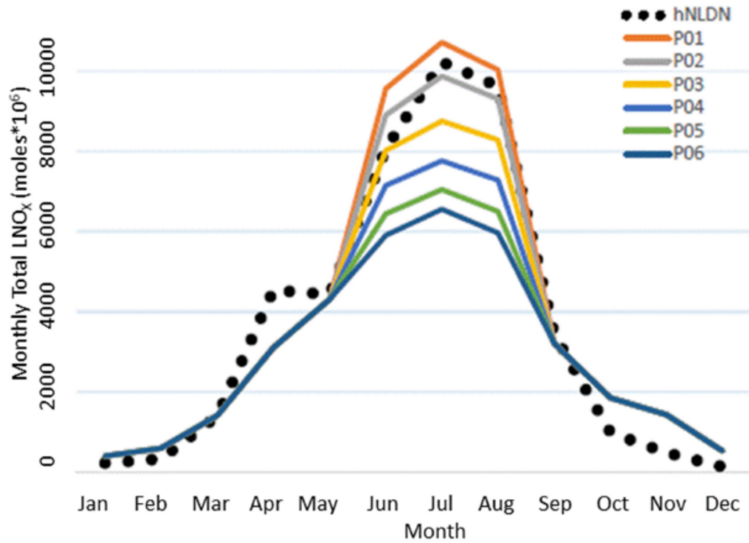

Figure 8. Total monthly column LNOx over the model domain using different CP cutoff values during summer months in 2011. hNLDN: LNOx produced by the hNLDN scheme, P01-P06: CP (cm) cutoff values from 0.01 (P01), 0.02 (P02), to 0.06 (P06). Linear regression parameters are applied when CP is less than the cutoff value, and log-linear regression parameters are used if otherwise.

regression parameters are applied if otherwise. Figure 8 shows the monthly total column LNOx 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 LNOx produced. When the cutoff value of 0.2 is applied, LNOx 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, LNOx is overestimated compared with that produced by hNLDN scheme. For generalized application to all years, dynamic cutoff values are used with this scheme. 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. 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.

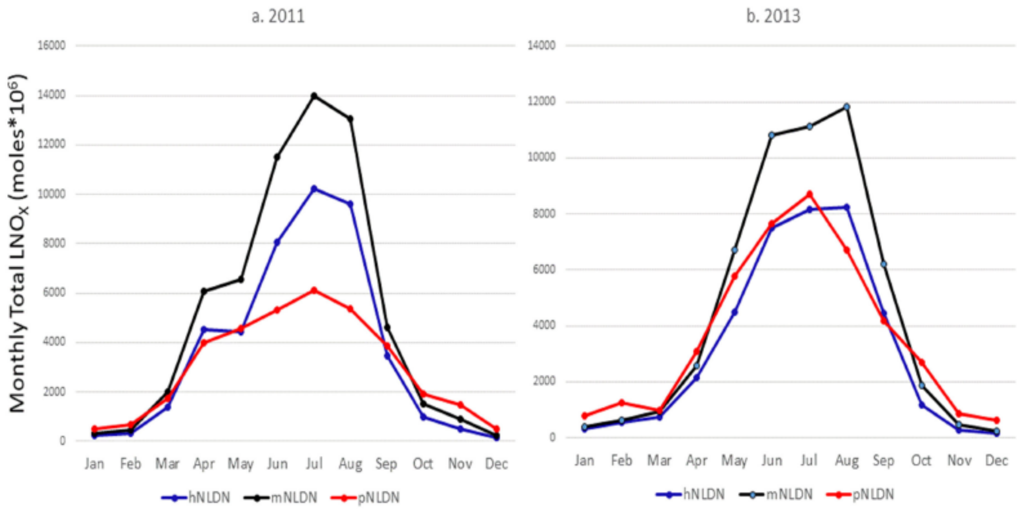

Figure 9. Total monthly column LNOx over the model domain with different LNOx production schemes for 2011 and 2013

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 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

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

Geoscientific Model Development Discussions

presented in Figure 10. Among the schemes, mNLDN produced the most LNOX on most of the days in July for both years. Except for a few days, pNLDN underestimated LNOX 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, LNOX was overestimated by pNLDN.

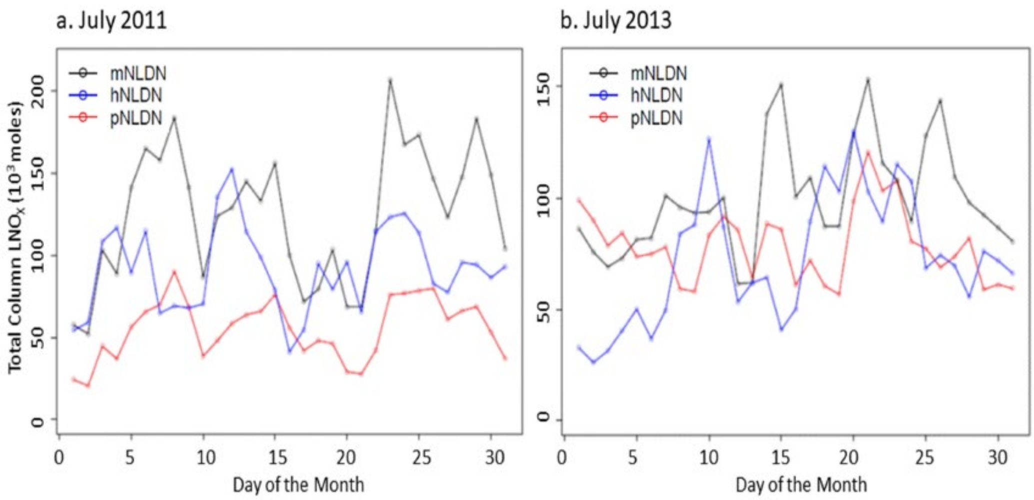

Figure 10. Total daily column LNOx over the model domain with different LNOx production schemes for 2011 and 2013

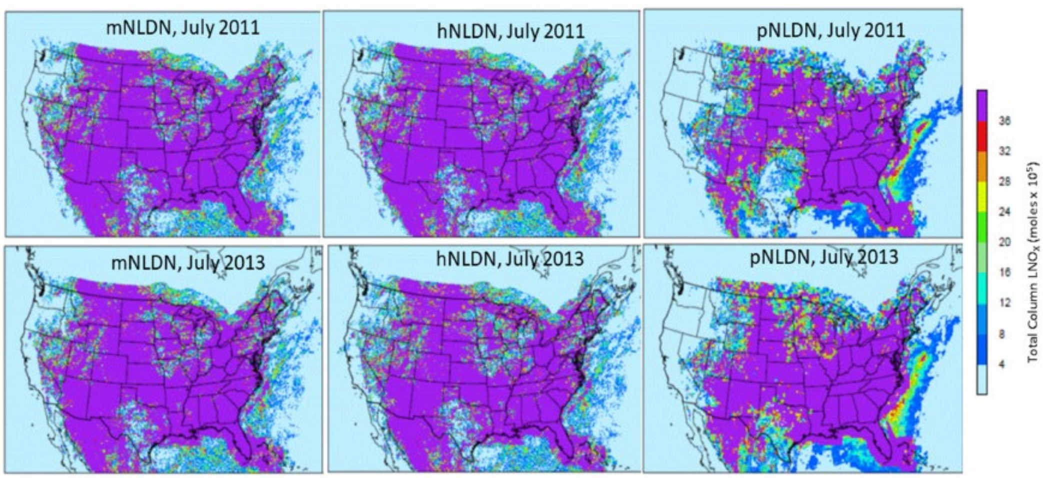

Figure 11. Spatial distribution of monthly column LNOx with different LNOx production schemes for July 2011 (upper frame) and July 2013 (lower frame)

The spatial distributions of monthly total column LNOX 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, 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 future climate simulations), it can provide the reasonable estimate for LNO_X comparable to hNLDN and mNLDN.

6. 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.

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 updated observational data and model simulations become available.

Lightning and LNO_X will remain an active research area in atmospheric sciences, especially when the Geostationary Lightning Mapper (GLM) on board the Geostationary Operational

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-33
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 March 2019

Environment Satellite R (GOES-R) series (Goodman et al., 2013) becomes fully operational in 2018. 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). 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.

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/.

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 Lighting Mapper (GLM), *Atmos. Res*., 125-126, 34-39, doi:10.1016/j.atmosres.2013.01.006, 2013.

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 O3 budget overWest 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.

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.

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.

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.

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.