

JGR Atmospheres

RESEARCH ARTICLE

10.1029/2020JD033990

Key Points:

- CAPE, LCL, and column-integrated humidity are the most important environmental variables for predicting lightning in the tropics & subtropics
- A logistic regression using only large-scale environmental variables and interactions accurately predicts lightning up to 86% of the time
- A parameterization based on the logistic regression captures the general pattern of lightning occurrence over land and ocean

Correspondence to:

M. Etten-Bohm,
montana.etten-bohm@tamu.edu

Citation:

Etten-Bohm, M., Yang, J., Schumacher, C., & Jun, M. (2021). Evaluating the relationship between lightning and the large-scale environment and its use for lightning prediction in global climate models. *Journal of Geophysical Research: Atmospheres*, 126, e2020JD033990. <https://doi.org/10.1029/2020JD033990>

Received 2 OCT 2020

Accepted 4 FEB 2021

Evaluating the Relationship Between Lightning and the Large-Scale Environment and its Use for Lightning Prediction in Global Climate Models

Montana Etten-Bohm¹ , Junho Yang², Courtney Schumacher¹ , and Mikyoung Jun³ 

¹Department of Atmospheric Sciences, Texas A&M University, College Station, TX, USA, ²Department of Statistics, Texas A&M University, College Station, TX, USA, ³Department of Mathematics, University of Houston, Houston, TX, USA

Abstract The objective of this study is to determine the relationship between lightning observed by the Tropical Rainfall Measuring Mission (TRMM) Lightning Imaging Sensor (LIS) and seven large-scale environmental variables obtained from the 3-hourly Modern-Era Retrospective analysis for Research and Application version 2 (MERRA-2) reanalysis in the tropics and subtropics. The large-scale environmental variables used are: convective available potential energy (CAPE), normalized CAPE (nCAPE), lifting condensation level (LCL), column saturation fraction (r), 700-hPa omega, low-level wind shear (LS) from 900 to 700 hPa, and deep wind shear (DS) from 900 to 300 hPa. All environmental variables show a significant shift toward larger values when lightning is present except for shear. DS decreases when lightning is present, while LS shows little mean change. However, strong geographical differences exist in the relationship between the environmental variables and lightning occurrence, particularly between land and ocean and the tropics and subtropics. Using a logistic regression, a lightning parameterization for global climate models (GCMs) is created using the above environmental variables as predictors while also adding geographic indicators (coast, slope, and latitude) and terms representing interactions between the predictors. The logistic regression predicts lightning occurrence accurately up to 86% of the time and is further applied to MERRA-2. While there are regions of overprediction and underprediction, the lightning parameterization performance shows promising potential for use in GCMs.

1. Introduction

Predicting lightning in global climate models (GCMs) is important for understanding not only how lightning will vary with climate change, but also how upper tropospheric chemistry and wildfires associated with lightning (e.g., Krause et al., 2014) will be impacted. One of the earliest and simplest GCM lightning parameterizations used only cloud top height to simulate global lightning flash rate densities (Price & Rind, 1992). Complexity was added to this framework in ensuing years by either including different cloud top height/lightning relationships for land and ocean or utilizing different cloud variables, like cloud droplet number concentration, convective precipitation rate, convective mass flux, and cold cloud depth (see Clark et al., 2017 for a comprehensive list of more recent parameterizations). However, there is a high level of uncertainty in the representation of convective cloud properties in GCMs, which impacts the magnitude and even the sign of the resulting predicted lightning trends in future climate scenarios (e.g., Clark et al., 2017; Finney et al., 2018; Tost et al., 2007). Because it is generally accepted that GCMs are more proficient at simulating the large-scale environment than convective cloud properties, this study focuses exclusively on analyzing the relationship between lightning and large-scale environmental parameters to evaluate how well lightning can be explained and predicted on a global scale.

The production of graupel and ice within a cumulonimbus cloud can lead to the development of lightning, and thunderstorm evolution can be highly impacted by changes in large-scale environmental parameters such as convective available potential energy (CAPE), humidity, and wind shear (Williams et al., 2002; Zipser & Lutz, 1994). Thus, studies sometimes incorporate environmental variables in GCM lightning parameterizations, in addition to convective cloud or other properties produced by a GCM (e.g., Romps et al., 2014; Stolz et al., 2017). For example, CAPE is often used as a predictor of lightning occurrence as it relates to updraft intensity, and therefore graupel and ice production (Williams & Stanfill, 2002). Romps

et al. (2014) used CAPE from radiosonde data over the continental United States (CONUS) and observations of twelve-hourly precipitation from radar and rain-gauge data from the National Oceanic and Atmospheric Administration River Forecast Centers for 1 year to create a lightning flash rate parameterization. Evaluating the parameterization on the output from multiple GCMs, Romps et al. (2014) found that all models produced large increases in CAPE (with a mean of $+11\%/^{\circ}\text{C}$) over CONUS between the current climate and the late-21st century, and therefore an overall increase in lightning flash rate. The mean precipitation increase in the models was $+1.5\%/^{\circ}\text{C}$, but ranged from -1.8% to $+4.2\%$, thus rainfall is a weaker constraint in their lightning prediction and can change sign depending on the model used. In addition, although this parameterization worked well over CONUS, it does not translate well to a global scale, particularly over the ocean (Romps et al., 2018).

Stolz et al. (2017) had better success in differentiating land/ocean patterns with their lightning parameterization, which is based on 6-hourly reanalysis data and lightning observations from space, by using normalized CAPE (nCAPE) and warm cloud depth (WCD), in addition to other variables. The thermodynamics of a thunderstorm may be better quantified using nCAPE compared to CAPE because it takes the depth of the CAPE into account, where “fat CAPE” is typically found over land and “skinny CAPE” over ocean (Stolz et al., 2015). This is relevant when considering lightning occurrence, as lightning tends to occur in fatter CAPE cases because of the associated stronger updrafts (Orville & Henderson, 1986; Stolz et al., 2015, 2017; Zipser et al., 2006). Lifting condensation level (LCL), or related proxies like cloud base height and WCD (Stolz et al., 2015; Williams & Stanfill, 2002), have also been shown to help distinguish land and ocean lightning occurrence because moister areas, like over the ocean, tend to have lower LCLs, and therefore lower cloud bases, which has been shown to be linked to less lightning.

Stolz et al. (2017) also included measures of humidity and shear in their lightning parameterization: they represented moisture by using the average relative humidity (RH) between 850 and 500 hPa and calculated wind shear between 1,000 and 500 hPa. Low-level wind shear (LS) has been linked to the longevity and intensity of squall lines (Rotunno et al., 1988) and more generally plays an integral role in downdrafts and cold pools and the triggering of new deep convection (Tompkins, 2001; Weisman & Klemp, 1982), which could also influence lightning production. The final parameter used in their lightning parameterization is cloud condensation nuclei (CCN) derived from a global atmospheric model with an included microphysics module. Using a multiple-linear regression model with their five chosen variables, Stolz et al. (2017) found that total lightning density (i.e., the total lightning flash rate normalized by the area of the associated convective feature) was positively correlated with nCAPE, CCN, and wind shear and negatively correlated with WCD and RH. Overall, they found that wind shear and RH were of secondary importance when compared to the other parameters except when looking at localized regions, like the Amazon and the Congo, where RH becomes a primary factor, but was still inversely related with lightning. They postulated that the negative relationship with humidity was related to the entrainment of drier air, which causes stronger downdrafts and therefore stronger subsequent convection.

In many ways, our study follows that of Stolz et al. (2017) closely. However, there are three notable differences between our methodologies that produce differing results on the relationship between lightning and the large-scale environment. The first is that we use a different set of environmental parameters. We do not use CCN because it is also subject to large uncertainties, akin to the convective cloud properties from GCMs. We also chose to investigate the column-integrated humidity because of its strong relationship to tropical rain production. Column saturation fraction (r), which is a measure of how humid a column is relative to its saturation specific humidity and is analogous to column water vapor, can help to better quantify the moisture profile within a region as opposed to using isolated levels. Precipitation has been shown to increase for higher r (e.g., Bretherton et al., 2004), and more recent studies have found that precipitation- r curves differ for varying precipitation types, with a more exponential pick-up observed for stratiform rain and a more linear relationship observed for convective rain (Ahmed & Schumacher, 2015, 2017). However, the relationship of r with lightning has not yet been studied in the literature, and a potentially predictive link may lie within. We further chose to investigate large-scale wind parameters beyond 1,000–500 hPa wind shear. Deep wind shear (DS) has been shown to increase hydrometeor detrainment in thunderstorms and increase the stratiform rain area in deep convection (Li & Schumacher, 2011) and could increase the area in which lightning occurs. Large-scale vertical motion and its relationship with lightning has been mini-

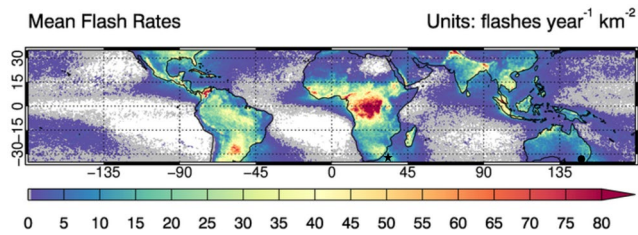


Figure 1. Mean lightning flash rates (flashes year⁻¹ km⁻²) observed by TRMM LIS over the tropics and subtropics from 1998 to 2013. The locations of the peak instantaneous LIS flash rates over land (301 flashes day⁻¹ km⁻²) and ocean (307 flashes day⁻¹ km⁻²) from the subset data are represented by the circle and star, respectively.

mally investigated, and only on a regional scale (Bang & Zipser, 2016), despite many studies showing its correlation with precipitation (e.g., Bony et al., 2004; Davies et al., 2013), particularly at 500 hPa.

The second difference between our methodology and that of Stolz et al. (2017) is that we extend our logistic regression to include interactions between variables. The use of interaction terms helps to explain the relationships between the variables themselves and tends to yield a better prediction. The third difference, which is a feature that differentiates our study from most lightning parameterization efforts, is that we focus on lightning occurrence rather than flash rate density (i.e., frequency over intensity) since we feel that predicting lightning occurrence is more strongly motivated by our observational analysis and statistical methodology. Ultimately, this study has two main objectives: (1) analyze the relationship between the large-scale environment and lightning with

previously unstudied parameters and (2) create a simple lightning parameterization for GCMs using only environmental variables and a logistic regression.

2. Data and Methods

2.1. Observational Data Sets

The TRMM satellite was a joint project between the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA) (Kummerow et al., 1998) and LIS was an instrument onboard TRMM that detected lightning at storm-scale resolution ($\sim 3\text{--}6$ km). Although the detection efficiency of the LIS instrument can vary based on latitude and land/ocean differences (Erdmann et al., 2020), the TRMM LIS improved upon its predecessor, the OTD onboard Microlab-1 (later renamed Orbview-1), by detecting approximately 90% of all lightning in TRMM's swath (Cecil et al., 2014). TRMM orbited from the end of 1997 to early 2015. Partial years (i.e., 1997 and 2015) are excluded from our analysis, as well as 2014 since TRMM was being prepared to come down from orbit and its altitude varied frequently, thus affecting the lightning retrievals (Albrecht et al., 2016).

Figure 1 shows mean flash rates for the TRMM LIS climatology. A sharp contrast in flash rates between land and ocean is observed, consistent with many previous studies (e.g., Albrecht et al., 2016; Orville & Spencer, 1979; Virts et al., 2013; Williams & Stanfill, 2002). Hotspots of lightning occurrence are observed over places like the southeastern United States, Argentina, and the Congo in central Africa, with mean flash rates often exceeding 40 flashes year⁻¹ km⁻², whereas mean flash rates over the open ocean never exceed three flashes year⁻¹ km⁻². The peak flash rate for land and ocean are denoted by the circle and star, respectively. Since the two points are close to the LIS southern data region boundary, the data quality of LIS could be called into question. However, previous studies (i.e., Albrecht et al., 2016; Cecil et al., 2014) have also found these regions to have high lightning intensity.

Seven environmental variables are used in this study to investigate the relationship between the large-scale environment and lightning occurrence and intensity: CAPE (J kg⁻¹), nCAPE (J kg⁻¹ m⁻¹), LCL (hPa), r (unitless), omega at 700 hPa (Pa s⁻¹), LS from 900 to 700 hPa (m s⁻¹), and DS from 900 to 300 hPa (m s⁻¹). These variables are obtained or derived from the Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2) dataset produced by NASA's Global Modeling and Assimilation Office (Gelaro et al., 2017). The reanalysis data are available globally at 3-hourly temporal resolution and $0.5^\circ \times 0.625^\circ$ spatial resolution with 72 vertical pressure levels from the surface to 0.01 hPa. For use in a lightning parameterization, MERRA-2 variables were extrapolated to 1,000 hPa over missing data regions for LS, DS, and omega to give better predictions over higher elevation areas and to better match what's done in GCMs. The variables are matched temporally and spatially with the TRMM LIS lightning dataset at $0.5^\circ \times 0.5^\circ$ resolution between 35°N and 35°S and then classified into either a lightning environment or nonlightning environment.

A lightning environment is classified as the time period 30 min prior to a lightning occurrence within a grid box. This choice is consistent with the use of individual grid values to parameterize subgrid-scale processes

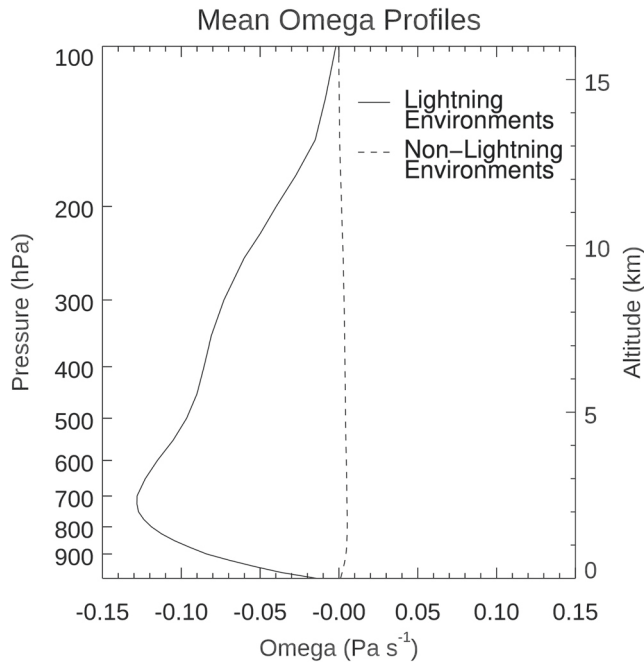


Figure 2. MERRA-2 omega profiles when lightning is present and not. Profiles are averaged from 35°S to 35°N for 1998.

in GCMs and a typical GCM time step of 15–30 min. A nonlightning environment is classified as any grid box that did not contain lightning. The 30-min requirement effectively discards ~75% of lightning occurrences and only grid boxes overflowed by TRMM are considered for both lightning and nonlightning environments. However, over 600,000 lightning flashes were analyzed during the 16-year TRMM LIS observational record so sample size is not a concern. We also note that the sampling duration of TRMM LIS is about 90 s (Cecil et al., 2014), so lightning could occur outside of the overpass time but within the 30-min window and thus bias the nonlightning environments, particularly over the ocean where storms have lower flash rates compared to land. However, lightning is so rare in general that any misclassifications would have only a very minor impact on the nonlightning environment statistics.

CAPE and nCAPE are calculated from the MERRA-2 temperature and specific humidity variables using the `cape_sound.pro` Interactive Data Language (IDL) script created by Dominik Brunner (http://en.verysource.com/code/3056618_1/cape_sound.pro.html), based on Emanuel's "calcsound" (Emanuel, 1994). For CAPE, parcels are lifted from multiple different levels to find the largest value of CAPE (i.e., to find the most unstable CAPE) and for nCAPE, the positive area of CAPE is divided by the depth of CAPE, which is defined as the vertical distance (in meters) between the level of free convection (LFC) and the equilibrium level (EL). LCL is calculated following Romps (2017) and as done in Ahmed and Schumacher (2015), r is calculated by dividing the column water vapor by the saturated column water vapor. Omega is obtained directly from the

MERRA-2 data set. In many previous studies (e.g., Bony et al., 2004), omega at 500 hPa is used to differentiate convectively active regions from less convectively active regions, and therefore was the pressure level initially used in this study. However, mean omega profiles for lightning and nonlightning environments from the MERRA-2 data set show that the maximum vertical motion for lightning environments occurs at 700 hPa (Figure 2). This is likely where the effects of convergence and orographic lift just below are maximized, both of which play significant roles in lightning production. Therefore, 700 hPa is chosen for the analysis to best represent the large-scale vertical motion within lightning environments. LS and DS are estimated as following: $LS = \sqrt{(u_{900} - u_{700})^2 + (v_{900} - v_{700})^2}$ and $DS = \sqrt{(u_{900} - u_{300})^2 + (v_{900} - v_{300})^2}$ where u_i and v_i are the zonal and meridional winds at the i th pressure level, respectively.

2.2. Logistic Regression

A logistic regression (McCullagh & Nelder, 1989) can be used to quantify the predictive ability and relative importance of particular environmental variables for parameters such as total lightning density and rain type occurrence (e.g., Stolz et al., 2017; Yang et al., 2019). In this study, the logistic regression determines whether lightning will occur or not. The model was trained using data from TRMM LIS and MERRA-2 for 2003 and tested with data from 2004. These years were chosen because they had similar sea surface temperatures in the tropical Pacific, with near-neutral or slightly warmer than normal conditions (i.e., weak El Niño) present throughout each year.

For the logistic regression model used in this study, the logit transformation of the probability of a lightning occurrence at grid point s , denoted by $\text{logit}(p(s))$, is expressed as a linear combination of the predictors:

$$\text{logit}(p(s)) = \log \frac{p(s)}{(1 - p(s))} = \beta_0 + \beta_1 X_1(s) + \cdots + \beta_p X_p(s), \quad (1)$$

where $X_i(s)$ denotes the predictor (i.e., CAPE, r , etc.) value at grid point s and β_i represents the coefficient for the covariate X_i . The coefficients are estimated using a statistical inference method, Maximum Likelihood

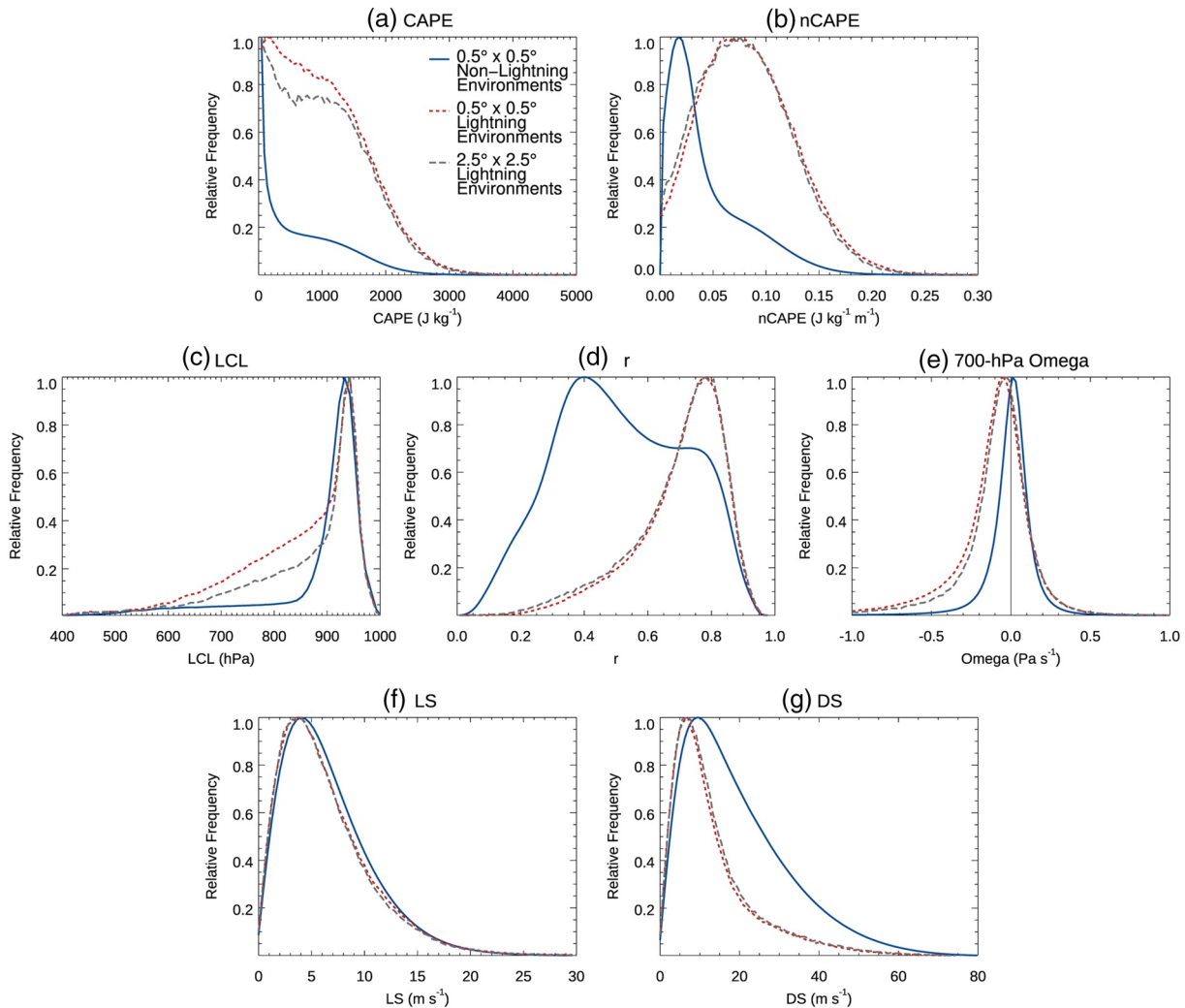


Figure 3. Distributions of MERRA-2 variables for differing environments (1998–2013). Solid blue lines show distributions for 0.5° nonlightning environments, dashed red lines show distributions for 0.5° lightning environments and dashed gray lines show distributions for 2.5° lightning environments.

Estimation (MLE) method, and then the fitted model outputs the predicted probability of lightning occurrence at grid point s , $\hat{p}(s)$, from zero (0% chance of lightning occurrence) to one (100% chance of lightning occurrence).

To predict if lightning will occur in a specific grid box, a cutoff probability (p_c) is specified so that if $\hat{p}(s) > p_c$, lightning is predicted in that grid box. In this study, p_c was chosen to equal the sample proportion of lightning occurrences as compared to all grid points. Lightning occurs $\sim 0.48\%$ of the time (as observed by the TRMM LIS), which is why p_c was chosen to be 4.8×10^{-3} . The result of the model is compared to observations from TRMM LIS for 2003 to test the accuracy of the regression before implementing in a lightning parameterization. Another focus of this study was to quantify the importance of each variable in predicting lightning, so the relative importance of each term is also analyzed following the methods of Pratt (1987) and Thomas et al. (2008).

3. Results

3.1. Observed Relationships Between Lightning and Large-Scale Environmental Variables

Histograms of CAPE, nCAPE, LCL, r , 700-hPa omega, LS, and DS values for lightning environments and nonlightning environments from MERRA-2 for 1998 to 2013 are shown in Figure 3. To examine sensitivity

to spatial resolution, 2.5° and 0.5° grids are separated for lightning environments. The axes are normalized by the maximum count value for each of the variables with the exception of CAPE and nCAPE in which the second largest value was used as the normalizing factor because of the very large number of values at zero. The large lightning counts at CAPE and nCAPE = 0 may be due to the depletion of CAPE from already precipitating storms. In addition, gridded reanalysis fields may not always capture the true storm environment. Apart from LS, large differences exist between lightning environments and nonlightning environments for all variables. CAPE, nCAPE, and r show large shifts to higher values for lightning occurrence. Previous work has shown a strong relationship between CAPE and nCAPE and lightning (Romps et al., 2014, 2018; Stolz et al., 2015, 2017; Williams et al., 2002). However, the change in the mode of r from 0.4 for nonlightning environments to 0.8 for lightning environments provides justification for the use of r as a lightning predictor, even though Stolz et al. (2017) argued that humidity was a secondary factor. Most lightning environment LCLs peak at about the same height as in nonlightning environments (950 hPa); however, lightning becomes relatively more common when LCLs are above 850 hPa. Omega at 700 hPa shifts negative when lightning occurs, indicating enhanced large-scale upward motion. Despite many previous studies showing a relationship between lightning and shear (i.e., Allen et al., 2011; Púčik et al., 2015; Taszarek et al., 2017; Weisman & Klemp, 1982; Westermayer et al., 2017), LS shows very little change between lightning and nonlightning environments, while DS shows a shift toward weaker shear for lightning environments, most likely attributable to the overall low DS in the tropics where most lightning production occurs. However, many of the previously mentioned studies focused on midlatitude environments, and could help explain the difference in findings. CAPE and LCL lightning environment values are most sensitive to grid size, with a larger shift away from nonlightning environment values at 0.5°. Therefore, the 0.5° grid was chosen for the remainder of the analysis but applying these results to models with coarser resolution remains appropriate since the distributions of lightning environment variables do not generally change when resolution changes.

Figure 4 shows maps of the mean value of each variable for nonlightning environments and the difference between the lightning and nonlightning environments. CAPE in nonlightning environments is generally highest over the Amazon, the Congo and warm ocean regions (Figure 4a) and increases for lightning environments globally (Figure 4b). The largest increases in CAPE occur over cooler ocean waters, indicating that large CAPE is necessary for lightning occurrence in these regions. Figure 4c shows that nCAPE for nonlightning environments has relatively lower values than CAPE over the Maritime Continent and West Pacific warm pool, but relatively higher values than CAPE over land regions where “fat CAPE” cases are more prevalent (e.g., Argentina, eastern Brazil, the Sahara, southern Africa, Saudi Arabia, and Australia). These land areas generally show a decrease in nCAPE when lightning occurs (Figure 4d), especially over arid regions in Saudi Arabia and Australia, suggesting invasions of moist tropical airmasses are more important for lightning generation there than steep lapse rates. However, over high lightning areas like the southeastern US and Congo, nCAPE slightly increases during lightning events. LCLs are closer to the surface over oceanic and moist land regions (Figure 4e), and these are the regions that show the smallest change in LCL when lightning occurs (Figure 4f). The biggest change for LCL in lightning environments occurs over the Sahel where LCL values increase (i.e., the LCL becomes lower) by about 100 hPa, suggesting that low-level moisture plays a significant role in lightning production in this typically dry environment. LCLs generally lower over land for most lightning environments with notable exceptions over the Amazon, northern Africa, and southern Australia—three very disparate regions in terms of land surface and climate.

Figures 4g and 4h show that r increases almost everywhere for lightning environments, but primarily over subtropical areas where r is relatively low. Very little change exists in r during lightning events over high-humidity locations like the Amazon, Congo, and Maritime Continent. Figure 4i shows strong large-scale upward motion at 700 hPa in these high-humidity land areas, in addition to oceanic Intertropical Convergence Zones (ITCZs). Note that areas of high-elevation are removed only for visual purposes because there is often an increase in omega along higher terrain due to rising air on the windward side of mountains, not representative of the large-scale, nonterrain driven rising motion. More upward motion (i.e., stronger negative values) is present almost everywhere for lightning environments (Figure 4j), with the most extreme differences over Argentina, Saudi Arabia, and Australia, perhaps indicating that 700-hPa omega is more important in areas where moisture is not abundant.

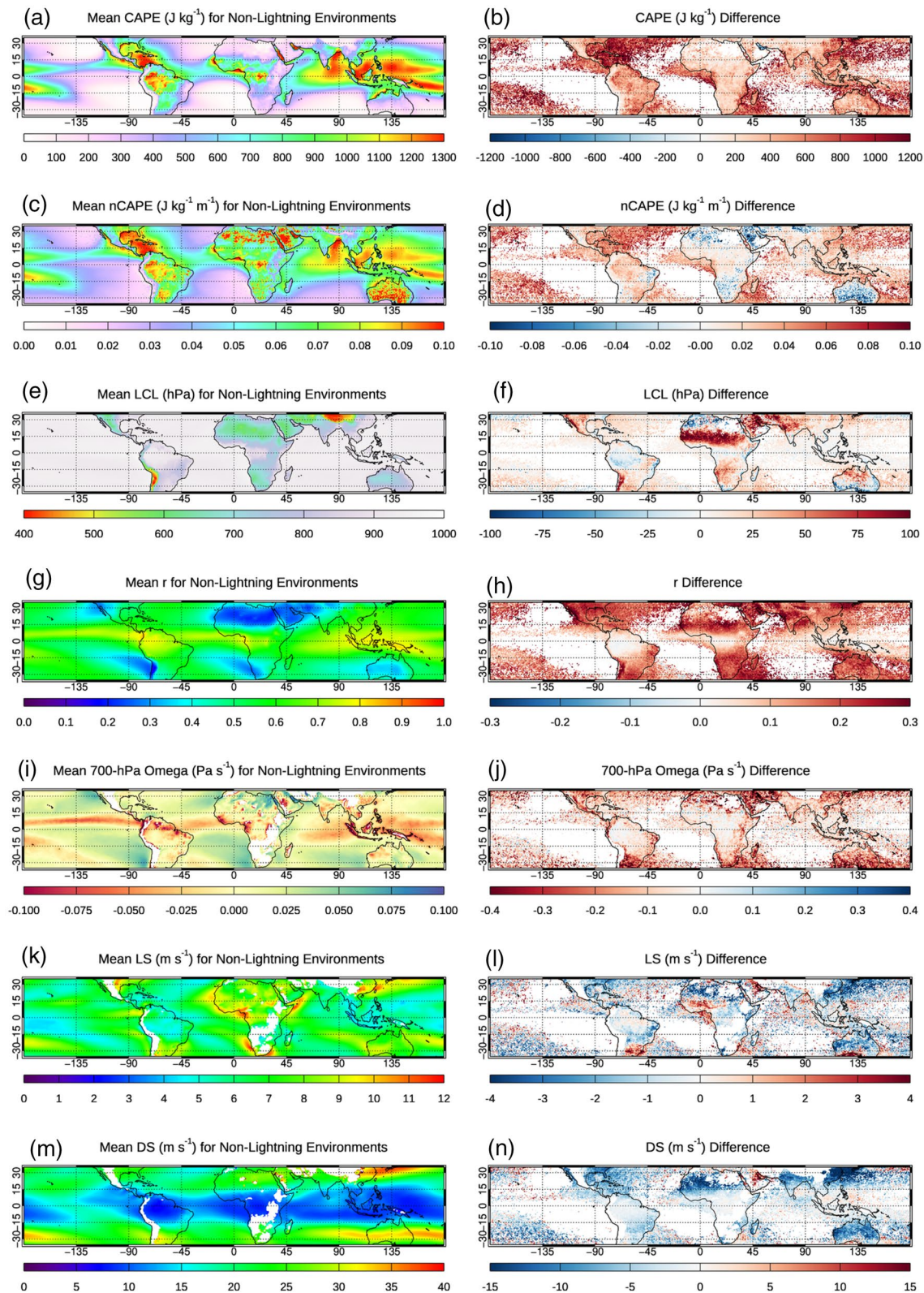


Figure 4. Left column shows mean values of MERRA-2 variables for nonlightning environments between 1998 and 2013. Right column shows difference maps between mean MERRA-2 variables for lightning environments and nonlightning environments. Warm colors show an increase for lightning environments and cool colors show an increase for nonlightning environments.

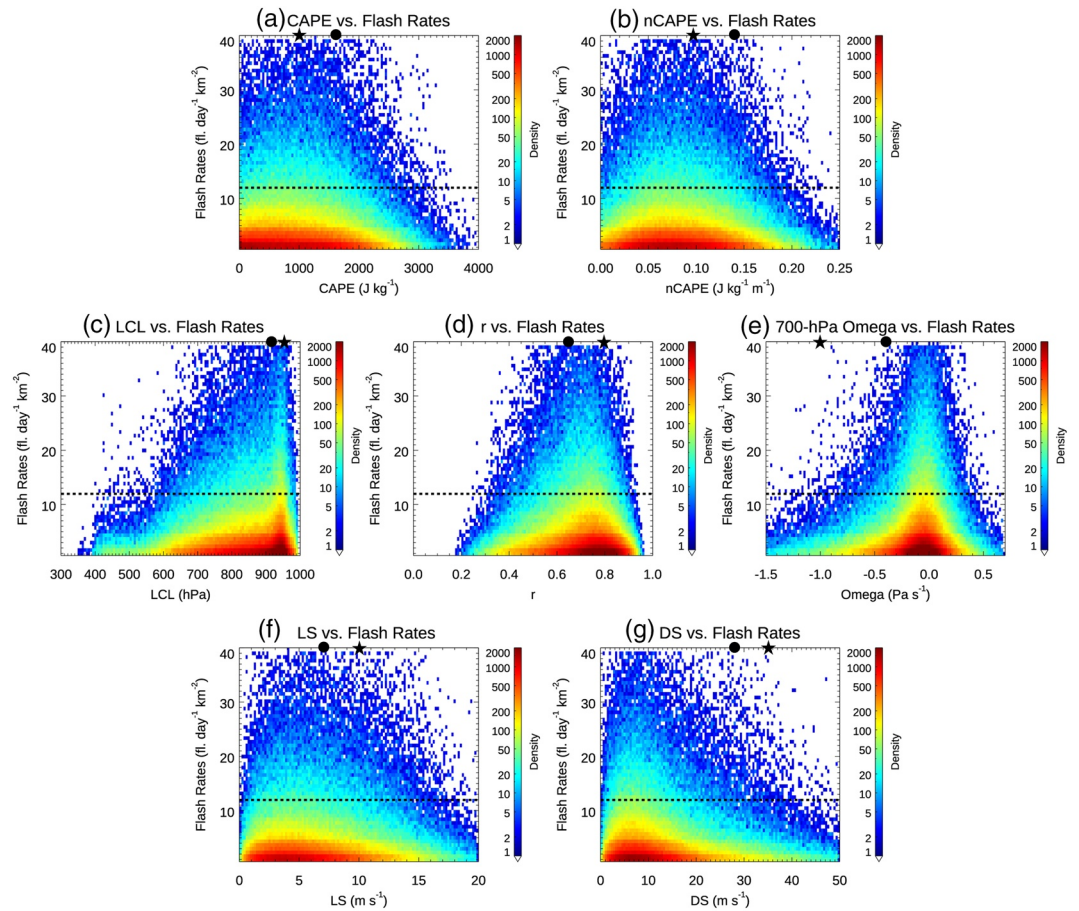


Figure 5. Histogram density plots of MERRA-2 large-scale environmental variables and TRMM LIS flash rates (1998–2013). The density color bar is logarithmic and the horizontal, dashed, black line represents the 95th percentile of flash rates. The symbols at the top of each panel represent the value of the environmental parameter for which the peak LIS flash rate was associated with, for land (circle) and ocean (star).

LS is generally strongest over coastal regions throughout the tropics and subtropics, while DS is weak in the tropics and strengthens into the subtropics (Figures 4k and 4m). Both LS and DS show regionally varying patterns of positive and negative changes in lightning environments (Figures 4l and 4n). For example, LS increases over Argentina and West Africa but decreases over Southeast Asia and Australia during lightning events. All of these locations tend to exhibit naturally high low-level shear, so it is unclear why there would be an opposite trend when lightning is present. DS generally decreases across much of the tropics and subtropics in lightning environments, although there are pockets of increased deep shear when lightning occurs, such as over the Pacific ITCZ and Saudi Arabia. Overall, Figure 4 presents a complex view of how lightning relates to its environment, but this complexity has coherence that we hope to exploit for lightning prediction.

To examine how each of these variables affects not only lightning occurrence, but also lightning intensity, Figures 5 and 6 present histogram density plots of the instantaneous 0.5° grid values from 1998 to 2013 for each MERRA-2 variable compared to the lightning flash rate. The binning interval is 0.690 flashes $\text{day}^{-1} \text{km}^{-2}$ for the ordinate variable, 35 J kg^{-1} for CAPE, $0.002 \text{ J kg}^{-1} \text{m}^{-1}$ for nCAPE, 6.125 hPa for LCL, 0.009 for r , 0.019 Pa s^{-1} for 700-hPa omega, 0.175 m s^{-1} for LS, and 0.438 m s^{-1} for DS. While the histograms are only plotted for flash rates up to 40 and 30 flashes $\text{day}^{-1} \text{km}^{-2}$, the peak instantaneous LIS flash rate for the 0.5° climatology over land was ~ 301 flashes $\text{day}^{-1} \text{km}^{-2}$ and occurred in Southeast Australia (32.75°S , 147.25°E ; depicted in Figure 1) on 7 November 2005 at 09:13 UTC (19:13 LT). Over ocean, the peak flash rate was ~ 307 flashes $\text{day}^{-1} \text{km}^{-2}$ and occurred off the east coast of South Africa (32.25°S , 33.25°E ; also depicted in Figure 1) on December 18, 2006 at 03:10 UTC (05:10 LT). The associated environmental parameter value for the peak flash rates for land and ocean are denoted by a star and circle, respectively, on each figure.

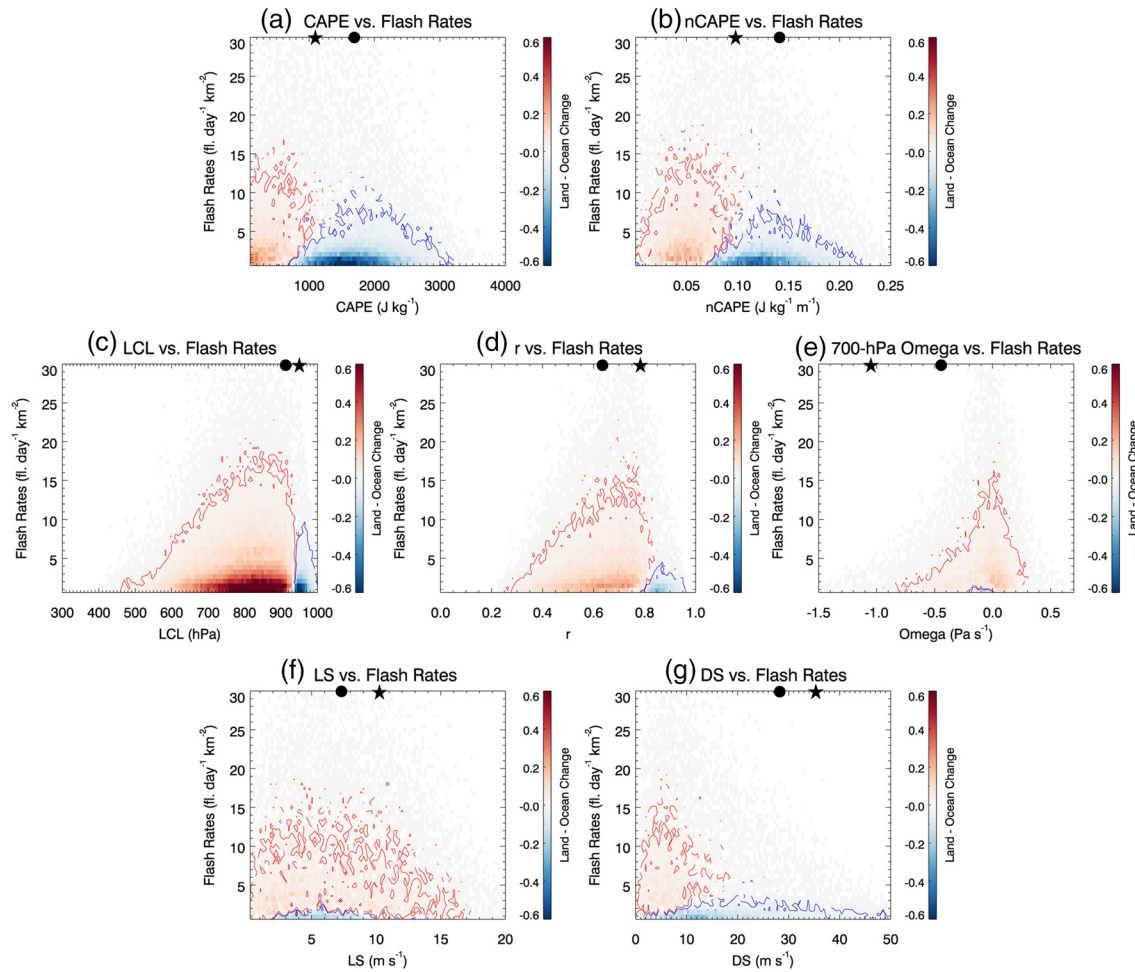


Figure 6. Land-ocean differences for environmental variables and flash rates from Figure 5. Values of -0.015 and 0.015 are shown in the blue and red contours. The symbols at the top of each panel represent the value of the environmental parameter for which the peak LIS flash rate was associated with, for land (circle) and ocean (star).

Consistent with Figures 3, Figure 5 shows that most lightning occurs for low-to-moderate CAPE, moderate nCAPE, high LCL values (i.e., low LCL heights), high r , slightly negative 700-hPa omega (i.e., rising motion), low-to-moderate LS, and low DS with large spread observed for all variables. Figure 5 further shows that most lightning occurs at rates of less than two flashes $\text{day}^{-1} \text{km}^{-2}$, with the 95th percentile at about 12 flashes $\text{day}^{-1} \text{km}^{-2}$ (denoted by dashed black line). An important result from this figure is that the highest flash rates do not occur at the most extreme environmental values, but rather in a wide range around the environmental variable mode. For example, high (i.e., 95th percentile) lightning flash rates can occur with values of CAPE ranging from 0 to 3,000 J/kg, r ranging from 0.3 to 0.9, and 700-hPa omega ranging from -1.0 to 0.4 Pa/s . The most extreme cases over land and ocean (denoted by the circle and star) also generally do not occur at the environmental parameter mode and sometimes occur at the edge of the observed ranges, most notably in 700-hPa omega and DS, which may be due to the subtropical location of the peak flash rates. As shown in Figures 4, 700-hPa omega decreases the most in the subtropics when lightning is present, perhaps from weather front activity. Also, the climatological DS is much smaller in the tropics, so lightning in the subtropics will almost always be associated with much higher DS values. Because extreme lightning environments are ensconced in the regular range of lightning environments, Figure 5 suggests that it would be difficult to parameterize extreme lightning events using just environmental variables.

As shown in Figures 1 and 4, strong contrasts in lightning flash rates and large-scale environmental variables exist between land and ocean. To further explore possible relationships between the environment and

extreme lightning events, density plots similar to Figure 5 were created by subtracting the normalized histograms for land and ocean (Figure 6). Positive values indicate that there is a stronger relationship between the environmental variable and flash rate over land, and negative values indicate a stronger relationship over ocean. Figures 6a and 6b show that similar magnitude flash rates over ocean typically occur at higher values of both CAPE and nCAPE compared to over land. This is most likely because more CAPE/nCAPE naturally exists over the ocean (Figures 4a and 4c). Over land, there is a trend toward higher flash rates as CAPE/nCAPE increases, although more so for nCAPE. Over ocean, this trend is much weaker.

LCL and r show similar relationships with flash rate parsed by land/ocean (Figure 6c and 6d). Flash rates over land tend to increase as LCLs become lower and r increases in magnitude, while lightning over ocean tends to only occur at very low LCLs (below 925 hPa) and high r (greater than 0.8). Thus, more moisture yields more intense lightning over land, while little difference in flash rate magnitude is observed over ocean because of the smaller observed dynamic range in LCL and r . Surprisingly, the largest flash rates over land occur at near zero or slightly positive values of 700-hPa omega (Figure 6e), suggesting that the sub-grid-scale triggering of convection in large-scale subsiding environments is an important factor in lightning production. LS shows negligible differences between land and ocean lightning environments, and essentially no relationship with flash rate in either region (Figure 6f). Lightning over ocean occurs for a broad range of DS values and shows no relationship with flash rate (Figure 6g). Higher flash rates are more prevalent over land when DS increases to 5 and 10 m s⁻¹, so a small to moderate amount of deep shear appears conducive to convection that produces high flash rates.

Figure 6 shows that more favorable convective conditions (i.e., moderate-to-high CAPE and nCAPE, low LCLs, high r , and moderate DS) have to be present for lightning to occur over ocean compared to land, and a pick-up of higher flash rates over land occurs for lower LCLs and higher values of CAPE, nCAPE, r , and DS, with mostly nonexistent relationships between these variables and flash rate increase over ocean. We also note that the most extreme lightning events over land and ocean do not always follow the land-ocean relationships in Figure 6. For example, the most extreme ocean event had lower CAPE and nCAPE and more negative 700-hPa omega than the most extreme land event. This is in part due to the fact that the peak ocean event occurred in a region strongly influenced by Africa, so is not representative of true open ocean conditions.

Analyzing relationships *between* environmental variables may also provide insight on conditions conducive to lightning production. Because of the importance of these three variables in the logistic regression (see next section), Figure 7 shows the joint distributions of lightning occurrence for CAPE/ r , LCL/CAPE, and LCL/ r . As r increases, lightning tends to occur at larger CAPE (Figure 7a). The highest lightning densities exist for r between 0.65 and 0.85 and for CAPE between 500 and 2,000 J kg⁻¹. Thus, CAPE can help predict when lightning occurs at various values of r . Figure 7b shows how this relationship changes for land versus ocean. Over land, lightning occurs at low-to-moderate CAPE and moderate-to-high r with a positive relationship between the two variables (i.e., as r increases, CAPE also increases). Over ocean, lightning typically occurs at moderate-to-high CAPE and high r , but the variables exhibit a negative relationship (i.e., as r increases, CAPE decreases). Taking into account interactions between these two variables would help lightning predictions over both land and ocean.

Figure 7c shows that lightning occurrence is most prevalent at low LCLs (between 1,000 and 900 hPa) and low-to-moderate CAPE (less than 3,000 J kg⁻¹), but that as CAPE increases lightning occurs at lower LCLs (or higher pressure values). This negative relationship is predominant over land (Figure 7d), whereas no relationship between the two variables is observed over ocean (i.e., lightning at low LCLs is observed for all values of CAPE below 3,000 J kg⁻¹). Figure 7e shows a complicated relationship for lightning occurrence between r and LCL. Lightning occurrence maximizes at low LCLs and high r , but there is a broad range of r values (maximized between 0.4 and 0.9) that lightning occurs at when LCLs are low, as well as a broad range of LCLs (maximized between 1,000 and 650 hPa) that lightning occurs at when r is high. However, there is another part of the plot in which lightning occurs more often as LCLs become lower and r increases. This negative linear trend is evident for land environments (Figure 7f), whereas there is no evident relationship between LCL and r over ocean. This information, along with the relationships between the other environmental variables and lightning occurrence, will be formalized with a logistic regression and implemented in a lightning parameterization in the next section.

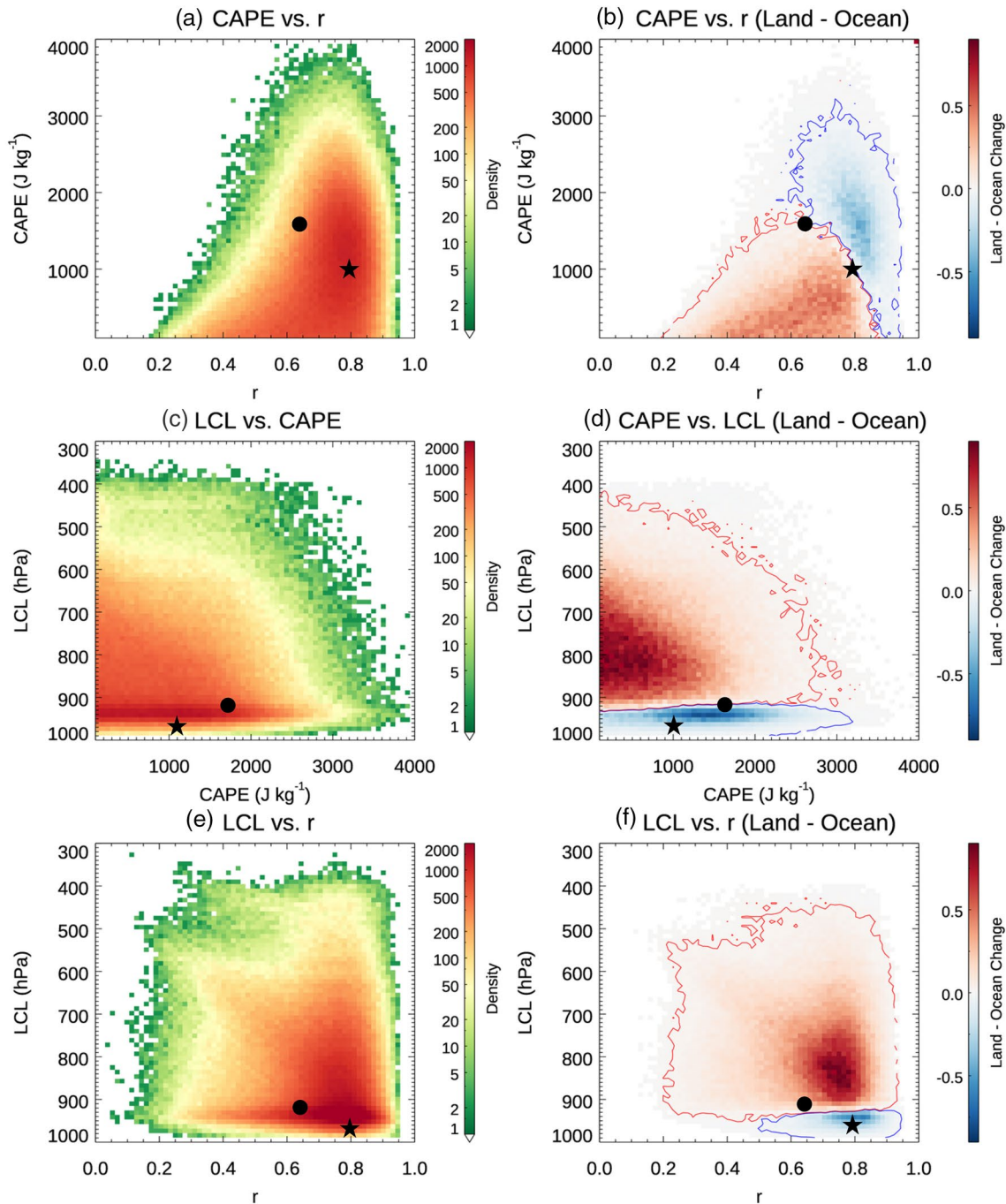


Figure 7. MERRA-2 CAPE versus r , LCL versus CAPE, and LCL versus r values for all lightning occurrences (left column) and land-ocean differences (right column) from 1998 to 2013 between 35°N and 35°S. Values of -0.015 and 0.015 are shown in the blue and red contours in the right column. The symbols on each panel represent the value of the environmental parameters for which the peak LIS flash rate was associated with, for land (circle) and ocean (star).

3.2. Statistical Model and Parameterization

3.2.1. Logistic Regression

The logistic regression model from Section 2.2 is used to quantify the results presented in Section 3.1. In creating possible covariate sets for the logistic regression, CAPE and nCAPE were found to have a high correlation (0.92), which causes serious multicollinearity problems. Therefore, only CAPE is used in the logistic regression analysis because of its simpler applicability in parameterizations in GCMs and sensitivity

Table 1
Probability of Detection (POD) for Nonlightning and Lightning Occurrence Predicted by the Logistic Regression Compared to TRMM LIS for Global, Land, and Ocean Cases

Model	p_c	Nonlightning Occurrence (%)	Lightning Occurrence (%)
Model a - CAPE, LCL, and r only			
All	4.8E-3	74.9	84.6
Land		54.4	89.6
Ocean		81.9	72.1
Model b - CAPE, LCL, r, and their interactions			
All	4.8E-3	77.5	84.9
Land		64.5	90.5
Ocean		82.2	70.6
Model c - All variables, geographic indicators, and interactions			
All	4.8E-3	78.9	86.0
Land		63.4	93.0
Ocean		83.7	71.2

The cutoff probability is denoted by p_c .

tests showed generally negligible differences in the output of the logistic regression model using one or the other variable. Three covariate sets are used to predict lightning occurrence via the logistic regression model: **a** CAPE, LCL, and r only, **b** CAPE, LCL, r , and their interactions, and **c** all environmental variables, three geographic indicators (i.e., coast, slope and latitude), and a subset of interaction terms.

These models were chosen to represent the simplest to most complicated parameterizations possible for the environmental variables analyzed in the previous section. Geographic indicators were included in the third model to assess the impact of using fixed, nonatmospheric variables. Coastal regions were defined as grid boxes that had between 10% and 90% land. Coast was added as a variable because previous studies have shown an increase in lightning along coastal regions, often as a result of frictional convergence, and some areas even see more lightning along the coast than further inland (Biswas & Hobbs, 1990). Land slope was calculated by differencing the maximum and minimum elevation between a central grid box and the highest and lowest surrounding grid box, and then dividing by the distance across the central grid box. Land slope was added because it was expected to help represent the growth of convection on the windward side of mountains, like the pick-up of lightning observed over the Himalayas in Asia and the Sierra Madre Occidental in Mexico shown in Figure 1. Latitude was added because of the strong latitudinal difference in flash rates.

The environmental variables and geographic indicators are considered “main effect” terms as they are independent variables that have an effect directly on the dependent variable, lightning, as compared to the interaction terms that consider the effect of each independent variable on each other in predicting the dependent variable. The interaction terms used in models **b** and **c** are listed in Table 1. The interaction terms were chosen following the results presented in Figures 4, 6, and 7. In addition, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Akaike, 1998; Schwarz, 1978) were used to find the optimal sets of covariates. For model **c**, the optimal covariate sets chosen based on the AIC and BIC differed slightly, where BIC eliminated the CAPE/coast, omega/LS, LS/latitude, and coast/latitude terms and AIC did not. However, after testing the prediction accuracy with and without the four terms, it was determined that not including the CAPE/coast, omega/LS, LS/latitude, and coast/latitude interaction terms improved the model's prediction accuracy, and therefore were not used in model **c**.

Regression tests were performed using different geographic data sets (i.e., land-only and ocean-only data) to potentially improve the regression compared to using all data. While the ocean-only regression improved nonlightning prediction for model **a** and the land-only regression improved lightning prediction for model **c**, all other cases worsened, and therefore the land-ocean distinctions were not considered further. A correction to account for rare lightning occurrence (King & Zeng, 2003) was also applied to potentially improve the regression, but was not deemed useful as its prediction accuracy worsened for all models.

Table 2 shows the probability of detection (POD) for the three sets of covariates (i.e., the different models) globally, for land, and for ocean. At each grid point, when the predicted probability from the logistic regression fit is less than p_c , the model returns a zero (no lightning occurrence predicted) and when the predicted probability is greater than p_c , the model returns unity (lightning occurrence predicted). A contingency table outputs the fraction of correctly and incorrectly predicted pixels in terms of lightning versus no lightning, from which the POD is obtained. Model **a** predicted no lightning occurrence correctly 74.9% of the time. Model **b** improved on model **a** by predicting no lightning occurrence correctly 77.5% of the time, and model **c** performed best, predicting no lightning occurrence correctly 78.9% of the time. For lightning prediction, each model improves on the previous, where the POD for lightning occurrence was 84.6% in model **a**, 84.9% in **b**, and 86.0% in **c**. When looking at prediction accuracy separately over land and ocean, model **b** does best over land for nonlightning occurrence. For lightning occurrence, model **a** has the best POD over ocean, while model **c** has the best POD globally and over land, showing that each model has its own strengths.

Table 2
Estimated Logistic Regression Coefficients for the Three Models

	Model a	Model b	Model c
	CAPE, LCL, and r only	CAPE, LCL, r, and their interactions	All variables, geographic indicators, and interactions
CAPE	0.632	0.779	0.856
LCL	−1.362	−1.303	−1.360
r	1.489	1.230	1.308
Omega	–	–	−0.427
LS	–	–	0.231
DS	–	–	0.051
Coast	–	–	0.672
Slope	–	–	0.308
Latitude	–	–	−0.238
CAPE/LCL	–	−0.360	−0.066
CAPE/r	–	−0.050	−0.093
CAPE/omega	–	–	0.043
CAPE/LS	–	–	−0.056
CAPE/DS	–	–	0.065
CAPE/slope	–	–	0.015
CAPE/latitude	–	–	−0.031
LCL/r	–	−0.167	−0.378
LCL/omega	–	–	−0.111
LCL/LS	–	–	0.140
LCL/DS	–	–	0.184
LCL/slope	–	–	0.188
LCL/latitude	–	–	−0.067
r/omega	–	–	0.209
r/LS	–	–	−0.042
r/DS	–	–	−0.081
r/slope	–	–	−0.117
r/latitude	–	–	0.183
LS/DS	–	–	−0.055
LS/coast	–	–	−0.108
DS/coast	–	–	−0.193
DS/latitude	–	–	−0.090
Coast/slope	–	–	−0.187
Latitude/slope	–	–	−0.022

All terms are statistically significant at the 0.1% level.

Overall, nonlightning predictions are better over ocean (~80%) versus land (~60%) and lightning predictions are better over land (~90%) versus ocean (~70%), which is consistent with the fact that much more lightning occurs over land than ocean.

Table 1 presents the estimated coefficients (β_i s) in Equation 1 for each model. The main effect and interaction terms are all considered statistically significant at the 0.001 level. The magnitude of the coefficient

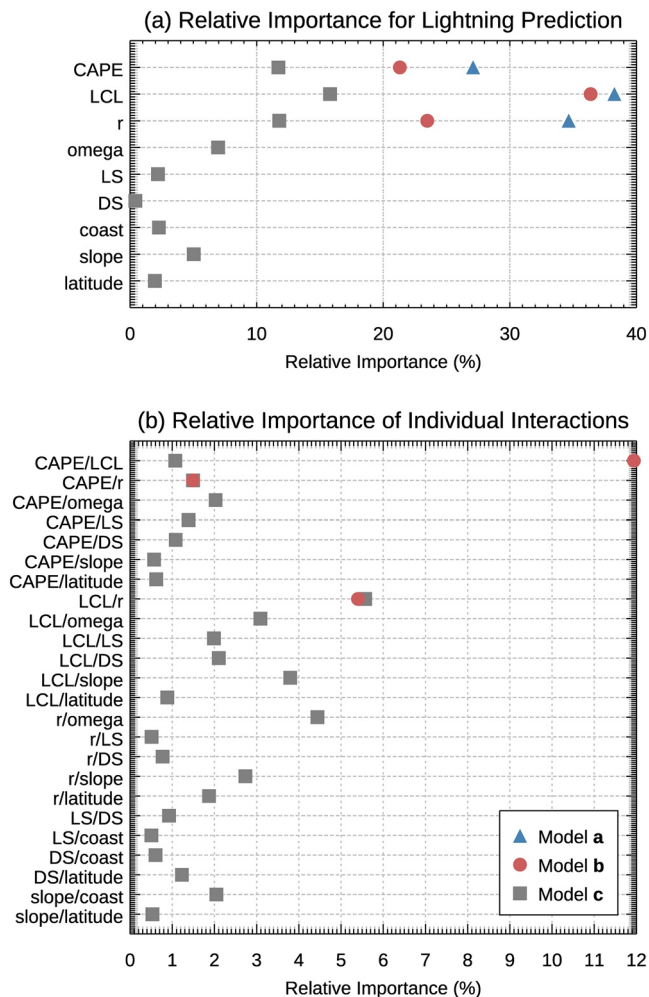


Figure 8. Relative importance of MERRA-2 variables using the logistic regression for model **a** CAPE, LCL, and r only (blue triangles), model **b** CAPE, LCL, r , and their interactions (red dots), and model **c** all, variables, geographic indicators, and interaction terms (gray squares). The sum of all variables for each model is 100%.

helps to explain the relative importance (discussed later) of each term and the sign of the coefficient indicates the correlation between the individual variable and lightning occurrence (i.e., direct [positive] or inverse [negative] correlations). Positive correlations between lightning occurrence and the terms are observed for CAPE, r , LS, DS, coast, and slope, while negative correlations are observed for LCL, 700-hPa omega, and latitude. Note that although the estimated coefficient value is negative for LCL and omega, pressure decreases with height and upward motion is negative, making them direct relationships. The sign of these correlations is consistent with the analysis in Section 2. However, the estimated coefficient for DS is negative for models without interactions (not shown). This is likely due to lightning occurring at only low-to-moderate values of DS (Figures 3f and 5f) and therefore a weak relationship between DS and lightning occurrence exists after accounting for interactions. The sign of the interaction term coefficients is more nuanced. For example, the negative coefficient for the CAPE/ r interaction term means that the effect of CAPE on lightning occurrence decreases as r increases.

While the coefficients in Table 1 change when additional main effects or interactions are considered, an analysis of relative importance provides more information on which terms have the greatest impact on the final lightning prediction. The importance of each variable is calculated by dividing the absolute value of the estimated coefficient by its standard error, then scaled so that the sum of all variables is 100%. The relative importance of each term for each logistic regression case is presented in Figure 8, where the top panel shows the relative importance of the main effects for each model, and the bottom panel shows the relative importance of the interactions in models **b** and **c**. In model **a**, LCL accounts for over 38% of the total importance, followed by r (35%) and CAPE (27%). In model **b**, while each of the main effects still have a relative importance exceeding 20%, CAPE/LCL plays a large role, accounting for almost 12% of the total importance, followed by LCL/ r (5%). CAPE/ r shows the least importance of the three interactions, perhaps because it does not discern land interactions (where most of the lightning occurs) as crisply as the other interactions, as shown in Figure 7. In model **c**, CAPE, LCL, and r are still the most important main effect terms, each accounting for over 12% of the total importance, followed by 700-hPa omega (7%) and slope (5%), with the shear parameters, coast, and latitude showing the least importance (<2% for each). The interactions in model **c** account for a little

over 40% of the total importance, where LCL/ r is the most important interaction, followed by r /omega, LCL/slope, LCL/omega, and r /slope. All other interactions account for less than 2.5% of the total relative importance in model **c**.

Overall, Figure 8 shows the importance of LCL, r , and CAPE in predicting lightning occurrence, with secondary importance placed on 700-hPa omega and slope, and the least importance placed on shear, coast, and latitude. Stolz et al. (2017) concluded that humidity was of secondary importance globally, but we have found that the use of column saturation fraction (r) provides highly useful information about lightning production. We further found a positive relationship between humidity and lightning occurrence (Table 1), whereas Stolz et al. (2017) found a negative relationship. Cumulonimbus generally prefers to form in moister environments and we saw the largest increase in r when lightning was present in the drier subtropics (Figures 4g and 4h). While it is possible that Stolz et al. (2017) found a different relationship because they analyzed midlevel humidity compared to our measure of column-integrated humidity, it is more likely because of the way they defined their lightning climatology in relation to the convective feature area observed by the other TRMM satellite instruments, such that they were more focused on internal storm interactions with humidity rather than a more generally conducive environment for convective occurrence.

While model **c** provides the best overall prediction statistics (Table 2), models **a** and **b** use fewer main effects and many fewer interactions with not much loss in predictability, so are arguably better candidates for a GCM parameterization, for which simplicity is desired. However, the use of the three interaction terms in model **b** notably improves the POD over land compared to model **a**, so we argue that interactions should be considered in future parameterization applications.

3.2.2. Lightning Parameterization

In this section, we apply each logistic regression case to MERRA-2 data from the test year (2004) to assess the geographical fidelity of the lightning predictions. Figure 9 shows the observed mean lightning occurrence from TRMM LIS (1998–2013) and the predicted mean lightning occurrence using MERRA-2 fields for each model for 2004. Note that Figure 9a represents lightning frequency rather than the mean intensity (shown in Figure 1) because the logistic regression predicts occurrence rather than intensity. Also, since the parameterization predicts lightning at every grid point and every time, and TRMM only observes one location at most twice a day, the full range of years for TRMM LIS was used to give more samples and a fairer comparison. Figures 1 and 9 show generally similar spatial patterns, such as a strong land/ocean

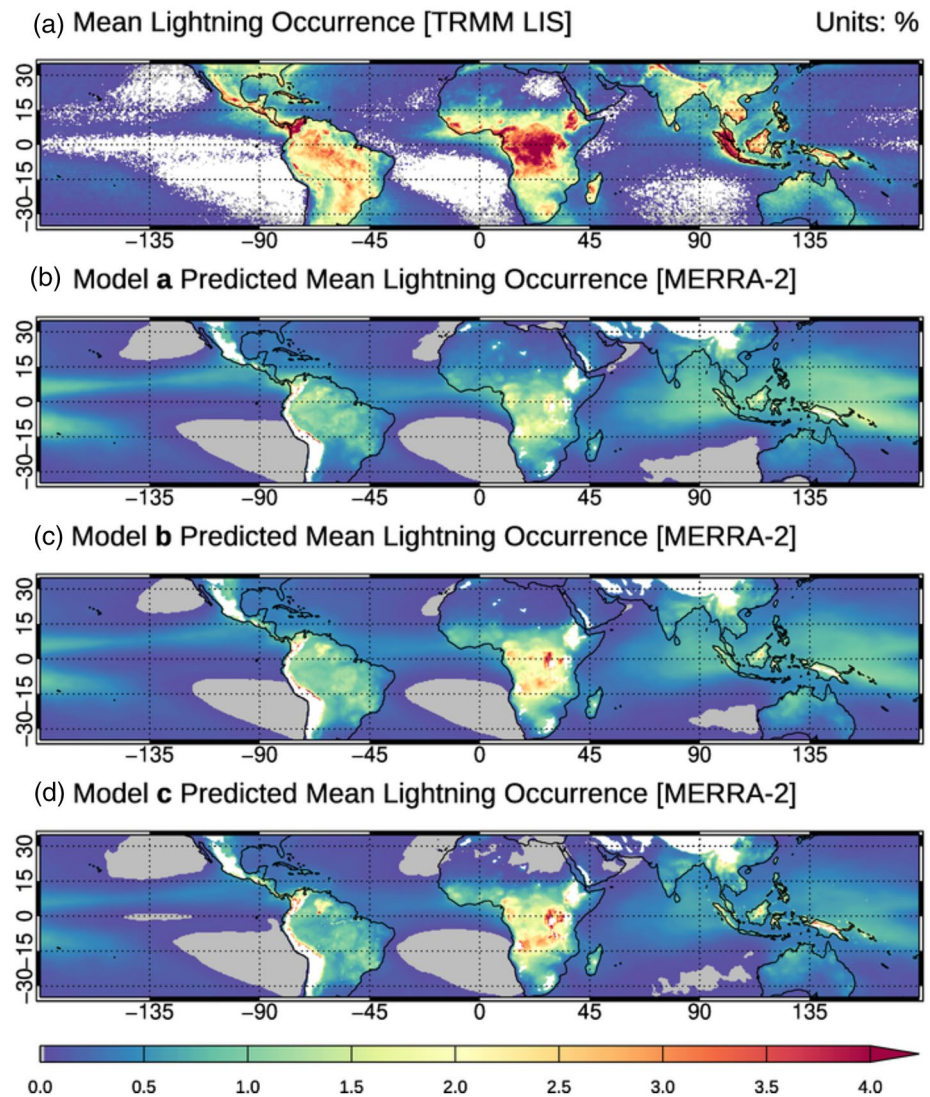


Figure 9. Mean lightning occurrence for 1998–2013 from TRMM LIS (a) and mean predicted lightning occurrence for 2004 from MERRA-2 variables using the logistic regression parameterization for model **a** CAPE, LCL, and r only, model **b** CAPE, LCL, r , and their interactions and model **c** all variables, geographic indicators, and interaction terms.

contrast and hot spots over Venezuela and the Congo, but other hot spot regions like Argentina become less pronounced in the occurrence map and lightning occurs at about the same frequency as over the Amazon. Figures 9b–9d use the logistic regression given in Equation 1 and the coefficients from Table 1. Elevation greater than 1,500 m is removed in the MERRA-2 prediction plots because of the inaccurate predictions from the logistic regression, likely due to the LCL term.

Figure 9 shows that each logistic regression underpredicts lightning occurrence over land and overpredicts it over ocean compared to LIS. Because there is significantly more lightning over land, this represents an overall underprediction by the logistic regression models. However, in places like the southeastern United States, for example, the severe underpredictions could be attributed to the known model bias that produces more precipitation than observed in MERRA-2 (Gelaro et al., 2017). Nevertheless, we chose not to perform a bias correction. Despite this global underprediction, the general spatial pattern of lightning occurrence is represented well by the logistic regression model applied to MERRA-2 data, with more lightning occurring over land than ocean and agreement in the relative difference in lightning occurrence over major land areas (e.g., Australia, the Amazon, and the Congo in increasing order of magnitude). As model complexity increases from model a to c, the underpredictions over land and overpredictions over ocean lessen and the spatial pattern more closely resembles observations from TRMM LIS. However, while model c represents the best prediction, model b still performs well and may be preferred in an operational parameterization because of its simplicity.

4. Conclusions

Seven environmental variables from MERRA-2 are used to investigate the relationship between the large-scale environment and lightning occurrence and intensity from TRMM LIS observations from 1998 to 2013. The variables used are CAPE, nCAPE, LCL, r , 700-hPa omega, wind shear from 900 to 700 hPa, and wind shear from 900 to 300 hPa. Large differences exist for these variables between lightning environments and nonlightning environments across the tropics and subtropics. CAPE, r , and 700-hPa omega increase for lightning environments globally, whereas nCAPE, LCL, LS, and DS show both increases and decreases. Strong land-ocean variations exist between lightning and the large-scale environment, with more favorable convective conditions (i.e., moderate-to-high CAPE and nCAPE, low LCLs, high r , and moderate DS) necessary for lightning to occur over ocean compared to land. Relationships between multiple environmental variables and lightning occurrence also exist. For example, lightning occurrence at low CAPE increases as r increases over land and when CAPE is moderate, lightning generally only occurs when r is large. Higher flash rates occur for slightly higher CAPE, nCAPE, and LCLs, lower r , slightly lower 700-hPa omega, and slightly higher LS and DS; however, these changes are well within the range of observed lightning environments for small flash rates making a parameterization for lightning intensity more difficult. Thus we focus on prediction of lightning occurrence.

Using a logistic regression technique, this study quantified the relationship between lightning occurrence and the large-scale environmental parameters on a global scale. LCL, column humidity (r), and CAPE (or nCAPE as they turned out to be generally interchangeable) were found to be the most important environmental predictors for lightning occurrence, with 700-hPa omega of secondary importance. Consistent with Stolz et al. (2017), the shear parameters were found to be of only minor importance when compared to the other large-scale environmental variables, but humidity was shown to be more beneficial in lightning prediction than found in Stolz et al. (2017). The addition of interaction terms and geographic indicators in the logistic regression was further explored and it was shown that models that included one or both types of terms improved upon the model that was only based on environmental variables.

Lightning parameterization studies to date show a wide range in lightning predictions. Using MERRA-2 fields, lightning parameterizations based on the logistic regression models presented previously are used to predict lightning occurrence and the resulting maps are compared to TRMM LIS observations. Overall, the general pattern of lightning occurrence is represented well for each model, with the parameterization including all the environmental variables, geographical indicators and many of their interactions performing best. However, a simpler formulation with just LCL, r , CAPE, and their interactions performed almost as well.

There remain regions of under and over predictions in even our best model, a problem that persists in other recent parameterization efforts (e.g., Magi, 2015; Stolz et al., 2017). However, our parameterization improves upon predictions from previous studies over the Congo and oceans. Our parameterization is also not dependent on cloud and precipitation variables, which need their own parameterization and which can vary significantly between GCMs, potentially propagating large uncertainty. Further, our parameterization does not include land-ocean scaling factors, but is still able to capture the land-ocean dichotomy that other studies have struggled with (such as Romps et al., 2018). However, further research is warranted to investigate why underpredictions still occur over certain land regions and to find which large-scale drivers of lightning are missing from current parameterizations.

Data Availability Statement

The TRMM LIS data were obtained from NASA GHRC (https://ghrc.nsstc.nasa.gov/lightning/data/data_lis_trmm.html) and the MERRA-2 data were obtained from NASA GES DISC (<https://goldsmr5.gesdisc.eosdis.nasa.gov/data/MERRA2/>).

Acknowledgments

This work was supported by NASA Grant NNX17AH66 G S003.

References

- Ahmed, F., & Schumacher, C. (2015). Convective and stratiform components of the precipitation-moisture relationship. *Geophysical Research Letters*, 42, 10453–10462. <https://doi.org/10.1002/2015GL066957>
- Ahmed, F., & Schumacher, C. (2017). Geographical differences in the tropical precipitation-moisture relationship and rain intensity onset. *Geophysical Research Letters*, 44, 1114–1122. <https://doi.org/10.1002/2016GL071980>
- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). New York: Springer. https://doi.org/10.1007/978-1-4612-1694-0_15
- Albrecht, R. I., Goodman, S. J., Buechler, D. E., Blakeslee, R. J., & Christian, H. J. (2016). Where Are the Lightning Hotspots on Earth?. *Bulletin of the American Meteorological Society*, 97(11), 2051–2068. <https://doi.org/10.1175/BAMS-D-14-00193.1>
- Allen, J., Karoly, D., & Mills, G. (2011). A severe thunderstorm climatology for Australia and associated thunderstorm environments. *Australian Meteorological and Oceanographic Journal*, 61(3), 143–158. <https://doi.org/10.22499/2.6103.001>
- Bang, S. D., & Zipser, E. J. (2016). Seeking reasons for the differences in size spectra of electrified storms over land and ocean. *Journal of Geophysical Research: Atmospheres*, 121, 9048–9068. <https://doi.org/10.1002/2016JD025150>
- Biswas, K. R., & Hobbs, P. V. (1990). Lightning over the Gulf Stream. *Geophysical Research Letters*, 17(7), 941–943. <https://doi.org/10.1029/GL017i007p00941>
- Bony, S., Dufresne, J.-L., Le Treut, H., Morcrette, J.-J., & Senior, C. (2004). On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, 22(2–3), 71–86. <https://doi.org/10.1007/s00382-003-0369-6>
- Bretherton, C. S., Peters, M. E., & Back, L. E. (2004). Relationships between Water Vapor Path and Precipitation over the Tropical Oceans. *Journal of Climate*, 17(7), 1517–1528. [https://doi.org/10.1175/1520-0442\(2004\)017<1517:RBWVPA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<1517:RBWVPA>2.0.CO;2)
- Cecil, D. J., Buechler, D. E., & Blakeslee, R. J. (2014). Gridded lightning climatology from TRMM-LIS and OTD: Dataset description. *Atmospheric Research*, 135–136, 404–414. <https://doi.org/10.1016/j.atmosres.2012.06.028>
- Clark, S. K., Ward, D. S., & Mahowald, N. M. (2017). Parameterization-based uncertainty in future lightning flash density. *Geophysical Research Letters*, 44, 2893–2901. <https://doi.org/10.1002/2017GL073017>
- Davies, L., Jakob, C., May, P., Kumar, V. V., & Xie, S. (2013). Relationships between the large-scale atmosphere and the small-scale convective state for Darwin, Australia. *Journal of Geophysical Research: Atmospheres*, 118, 11534–11545. <https://doi.org/10.1002/jgrd.50645>
- Emanuel, K. A. (1994). *Atmospheric convection*. New York, NY: Oxford University Press.
- Erdmann, F., Defer, E., Caumont, O., Blakeslee, R. J., Pédeboy, S., & Coquillat, S. (2020). Concurrent satellite and ground-based lightning observations from the Optical Lightning Imaging Sensor (ISS-LIS), the low-frequency network Meteorage and the SAETTA Lightning Mapping Array (LMA) in the northwestern Mediterranean region. *Atmospheric Measurement Techniques*, 13(2), 853–875. <https://doi.org/10.5194/amt-13-853-2020>
- Finney, D. L., Doherty, R. M., Wild, O., Stevenson, D. S., MacKenzie, I. A., & Blyth, A. M. (2018). A projected decrease in lightning under climate change. *Nature Climate Change*, 8(3), 210–213. <https://doi.org/10.1038/s41558-018-0072-6>
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., & Zhao, B. (2017). The Modern-Era retrospective analysis for research and applications, Version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>
- King, G., & Zeng, L. (2003). Logistic regression in rare events data. *Journal of Statistical Software*, 8(2), 137–163. <https://doi.org/10.18637/jss.v008.i02>
- Krause, A., Kloster, S., Wilkenskeld, S., & Paeth, H. (2014). The sensitivity of global wildfires to simulated past, present, and future lightning frequency. *Journal of Geophysical Research: Biogeosciences*, 119, 312–322. <https://doi.org/10.1002/2013JG002502>
- Kummerow, C., Barnes, W., Kozu, T., Shiue, J., & Simpson, J. (1998). The Tropical Rainfall Measuring Mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, 15(3), 809–817. [https://doi.org/10.1175/1520-0426\(1998\)015<0809:TTRMMT>2.0.CO;2](https://doi.org/10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2)
- Li, W., & Schumacher, C. (2011). Thick Anvils as viewed by the TRMM precipitation radar. *Journal of Climate*, 24(6), 1718–1735. <https://doi.org/10.1175/2010JCLI3793.1>
- Magi, B. I. (2015). Global lightning parameterization from CMIP5 Climate Model Output. *Journal of Atmospheric and Oceanic Technology*, 32(3), 434–452. <https://doi.org/10.1175/JTECH-D-13-00261.1>
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.
- Orville, R. E., & Henderson, R. W. (1986). Global distribution of midnight lightning: September 1977 to August 1978. *Monthly Weather Review*, 114(12), 2640–2653. [https://doi.org/10.1175/1520-0493\(1986\)114<2640:GDOMLS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<2640:GDOMLS>2.0.CO;2)

- Orville, R. E., & Spencer, D. W. (1979). Global lightning flash frequency. *Monthly Weather Review*, 107(7), 934–943. [https://doi.org/10.1175/1520-0493\(1979\)107<0934:GLFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1979)107<0934:GLFF>2.0.CO;2)
- Pratt, J. W. (1987). Dividing the indivisible: Using simple symmetry to partition variance explained. In *Proceedings of the 2nd International Conference in Statistics* (pp. 245–260). Tampere, Finland: Department of Mathematical Sciences, University of Tampere.
- Price, C., & Rind, D. (1992). A simple lightning parameterization for calculating global lightning distributions. *Journal of Geophysical Research*, 97(D9), 9919–9933. <https://doi.org/10.1029/92JD00719>
- Púčik, T., Groenemeijer, P., Rýva, D., & Kolář, M. (2015). Proximity soundings of severe and nonsevere thunderstorms in Central Europe. *Monthly Weather Review*, 143(12), 4805–4821. <https://doi.org/10.1175/MWR-D-15-0104.1>
- Romps, D. M. (2017). Exact expression for the lifting condensation level. *Journal of the Atmospheric Sciences*, 74(12), 3891–3900. <https://doi.org/10.1175/JAS-D-17-0102.1>
- Romps, D. M., Charn, A. B., Holzworth, R. H., Lawrence, W. E., Molinari, J., & Vollaro, D. (2018). CAPE times P explains lightning over land but not the land-ocean contrast. *Geophysical Research Letters*, 45, 12623–12630. <https://doi.org/10.1029/2018GL080267>
- Romps, D. M., Seeley, J. T., Vollaro, D., & Molinari, J. (2014). Projected increase in lightning strikes in the United States due to global warming. *Science*, 346(6211), 851–854. <https://doi.org/10.1126/science.1259100>
- Rotunno, R., Klemp, J. B., & Weisman, M. L. (1988). A Theory for Strong, Long-Lived Squall Lines. *Journal of the Atmospheric Sciences*, 45(3), 463–485. [https://doi.org/10.1175/1520-0469\(1988\)045<0463:ATFSL>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<0463:ATFSL>2.0.CO;2)
- Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Stolz, D. C., Rutledge, S. A., & Pierce, J. R. (2015). Simultaneous influences of thermodynamics and aerosols on deep convection and lightning in the tropics. *Journal of Geophysical Research: Atmospheres*, 120, 6207–6231. <https://doi.org/10.1002/2014JD023033>
- Stolz, D. C., Rutledge, S. A., Pierce, J. R., & van den Heever, S. C. (2017). A global lightning parameterization based on statistical relationships among environmental factors, aerosols, and convective clouds in the TRMM climatology: Lightning Parameterization From TRMM. *Journal of Geophysical Research: Atmospheres*, 122, 7461–7492. <https://doi.org/10.1002/2016JD026220>
- Taszarek, M., Brooks, H. E., & Czernecki, B. (2017). Sounding-derived parameters associated with convective hazards in Europe. *Monthly Weather Review*, 145(4), 1511–1528. <https://doi.org/10.1175/MWR-D-16-0384.1>
- Thomas, D. R., Zhu, P., Zumbo, B. D., & Dutta, S. (2008). On measuring the relative importance of explanatory variables in a logistic regression. *Journal of Modern Applied Statistical Methods*, 7(1), 21–38. <https://doi.org/10.22237/jmasm/1209614580>
- Tompkins, A. M. (2001). Organization of tropical convection in low vertical wind shears: The role of cold pools. *Journal of the Atmospheric Sciences*, 58(13), 1650–1672. [https://doi.org/10.1175/1520-0469\(2001\)058<1650:OOTCIL>2.0.CO;2](https://doi.org/10.1175/1520-0469(2001)058<1650:OOTCIL>2.0.CO;2)
- Tost, H., Jöckel, P., & Lelieveld, J. (2007). Lightning and convection parameterisations—Uncertainties in global modeling. *Atmospheric Chemistry and Physics*, 7(17), 4553–4568. <https://doi.org/10.5194/acp-7-4553-2007>
- Virts, K. S., Wallace, J. M., Hutchins, M. L., & Holzworth, R. H. (2013). Highlights of a new ground-based, hourly global lightning climatology. *Bulletin of the American Meteorological Society*, 94(9), 1381–1391. <https://doi.org/10.1175/BAMS-D-12-00082.1>
- Weisman, M. L., & Klemp, J. B. (1982). The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Monthly Weather Review*, 110(6), 504–520. [https://doi.org/10.1175/1520-0493\(1982\)110<0504:TDOESC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<0504:TDOESC>2.0.CO;2)
- Westermayer, A. T., Groenemeijer, P., Pistotnik, G., Sausen, R., & Faust, E. (2017). Identification of favorable environments for thunderstorms in reanalysis data. *Meteorologische Zeitschrift*, 26(1), 59–70. <https://doi.org/10.1127/metz/2016/0754>
- Williams, Rosenfeld, D., Madden, N., Gerlach, J., Gears, N., Atkinson, L., & Avelino, E. (2002). Contrasting convective regimes over the Amazon: Implications for cloud electrification. *Journal of Geophysical Research*, 107(D20), 8082. <https://doi.org/10.1029/2001JD000380>
- Williams, & Stanfill, S. (2002). The physical origin of the land-ocean contrast in lightning activity. *Comptes Rendus Physique*, 3(10), 1277–1292. [https://doi.org/10.1016/S1631-0705\(02\)01407-X](https://doi.org/10.1016/S1631-0705(02)01407-X)
- Yang, J., Jun, M., Schumacher, C., & Saravanan, R. (2019). Predictive statistical representations of observed and simulated rainfall using generalized linear models. *Journal of Climate*, 32(11), 3409–3427. <https://doi.org/10.1175/JCLI-D-18-0527.1>
- Zipser, E. D., & Lutz, K. R. (1994). The vertical profile of radar reflectivity of convective cells: A strong indicator of storm intensity and lightning probability? *Monthly Weather Review*, 122(8), 1751–1759. [https://doi.org/10.1175/1520-0493\(1994\)122<1751:TVPORR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<1751:TVPORR>2.0.CO;2)
- Zipser, E. J., Cecil, D. J., Liu, C., Nesbitt, S. W., & Yorty, D. P. (2006). Where are the most intense thunderstorms on Earth? *Bulletin of the American Meteorological Society*, 87(8), 1057–1071. <https://doi.org/10.1175/BAMS-87-8-1057>