EVALUATING THE RELATIONSHIP BETWEEN LIGHTNING AND LARGE-SCALE ENVIRONMENTAL VARIABLES

A Thesis

by

MONTANA ETTEN-BOHM

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Courtney Schumacher
Committee Members,	Mikyoung Jun
	Yangyang Xu
Head of Department,	R. Saravanan

December 2018

Major Subject: Atmospheric Sciences

ABSTRACT

The objective of this study is to determine the relationship between lightning and six large-scale environmental variables: convective available potential energy (CAPE), normalized CAPE (nCAPE), column saturation fraction (r), 700-hPa omega, 900-700 hPa low-level wind shear (LS) and 900-300 hPa deep wind shear (DS). Lightning data is obtained from the Tropical Rainfall Measuring Mission's (TRMM) Lightning Imaging Sensor (LIS) from 1998 to 2013 and large-scale environmental variables are derived from 3-hourly Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) data. Each dataset is binned at 0.5° x 0.5°. MERRA-2 data is considered to represent a lightning environment when lightning occurs within 30 minutes of the MERRA-2 time stamp.

CAPE, nCAPE and r show clear distinctions in lightning environments compared to non-lightning environments and the largest flash occurrences are associated with low-to-moderate CAPE, moderate nCAPE, slightly negative values of 700-hPa omega (i.e., rising motion), high r, low-to-moderate LS and low DS. Clear geographical distinctions for flash occurrences exist between land and ocean for CAPE, nCAPE and r and between tropical and sub-tropical areas for CAPE, nCAPE, r and DS. The relationship of r with other variables for lightning occurrences is evaluated and it is shown that CAPE and omega with r have the clearest relationships.

The association between large-scale environmental variables and lightning is analyzed globally for latitudes between 35°N and 35°S using two statistical models. Using a generalized linear model (GLM), nCAPE and r are the primary predictors for lightning prediction. Using a point-process model, nCAPE is the best predictor, with strong regional contrasts present. The GLM is used in a lightning parameterization to predict lightning from MERRA-2 and the Community Atmosphere Model version 5 (CAM5). Predicted lightning from both datasets generally agrees with observations from the TRMM LIS, which supports the use of a lightning parameterization in GCMs.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Professor Courtney Schumacher and Professor Yangyang Xu of the Department of Atmospheric Sciences as well as Professor Mikyoung Jun of the Department of Statistics at Texas A&M University.

Statistical analysis in sections 3.2.1 and 3.2.2 was conducted in part by Junho Yang of the Department of Statistics.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

This work was supported by Ramesh Kakar and NASA's Weather and Atmospheric Dynamics program.

NOMENCLATURE

CAM5	Community Atmosphere Model version 5
CAPE	Convective Available Potential Energy
DS	Deep (900-300hPa) Wind Shear
GCM	Global Climate Model
GLM	Generalized Linear Model
LGCP	Log-Gaussian Cox Process
LI	Land Indicator
LIS	Lightning Imaging Sensor
LS	Low-Level (900-700hPa) Wind Shear
MERRA-2	Modern-Era Retrospective analysis for Research and Applications, version 2
NASA	National Aeronautics and Space Administration
nCAPE	Normalized Convective Available Potential Energy
OTD	Optical Transient Detector
r	Column Saturation Fraction
ReLogit	Rare Event Logistic Regression Model
TRMM	Tropical Rainfall Measuring Mission

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

Lightning has many environmental and societal impacts including its modification of the NO_x cycle (Schumann and Huntrieser 2007), its ability to spark forest fires (Flannigan et al. 2009) and its threat as one of the leading causes of weather-related fatalities in the United States (Curran et al. 2000). Lightning is created by a difference in charge within a cloud itself (intra-cloud; IC), between a cloud and the ground (cloud-to-ground; CG) or between two separate clouds (cloud-to-cloud; CC). This charge separation often occurs due to the presence and collision of graupel and ice particles within clouds (Reynolds et al. 1957; Takahashi 1978; Churchill and Houze 1984; Saunders et al. 1991). The production of graupel and ice particles heavily relies on the influence of large-scale environmental variables (Zipser and Lutz 1994; Williams et al. 2002), therefore a potentially predictive link exists between the large-scale environment and lightning.

Various studies have evaluated the influence of large-scale environmental variables on lightning occurrence using both observational and model data. Convective Available Potential Energy (CAPE) has been long thought of as a major contributing factor to the electrification of clouds because it is a proxy for updraft strength (Rutledge et al. 1992; Williams et al. 1992; Zipser and Lutz 1994). For example, Williams et al. (2002) studied convection in four environmental regimes over Brazil using radiosonde data, the Brazil Lightning Detection Network and field charge antennae, and determined that higher peak flash rates occurred at moderate-to-high CAPE. The study area was primarily over land, however the westerly wind regime had convective properties similar

to the ocean and was weak in its lightning production even though a comparison of CAPE values between all regimes showed little variability.

More lightning occurs over land typically as a result of stronger updrafts and differing surface properties (Orville and Henderson 1986; Zipser et al. 2006). Stolz et al. (2015) used normalized CAPE (nCAPE) to better quantify the environmental conditions associated with the land-ocean lightning contrast. The "shape-of-the-CAPE" is considered when using nCAPE, and can differentiate between "fat CAPE," typically found over land, and "skinny CAPE," typically found over ocean. For example, if two regions both have CAPE of 1000 J kg^{-1} and the depth of CAPE is 2000 m over land and 4000 m over ocean, nCAPE would equal 0.5 J kg^{-1} m^{-1} and 0.25 J kg^{-1} m^{-1}, respectively. Thus, while regular CAPE calculations would not be able to discern the thermodynamic environment between the land and ocean regions, nCAPE could provide a more physically based distinction.

Stolz et al. (2015) used the 6-hourly European Reanalysis dataset (ERA-Interim), a global transport model (GEOS-Chem) with the online microphysics module TOMAS and the Tropical Rainfall Measuring Mission's (TRMM) Lightning Imaging Sensor (LIS) to evaluate the occurrence of total lightning density (TLD) over the tropics against nCAPE, warm cloud depth (WCD) – which is the vertical distance between the lifting condensation level (LCL) and the freezing level – and cloud condensation nuclei (CCN). More flashes occurred at moderate-to-high nCAPE, variable CCN and large WCD. They also found that nCAPE helped quantify the land-ocean lightning contrast, as more lightning occurred at lower nCAPE over the ocean.

Stolz et al. (2017) expanded upon this study by evaluating the occurrence of TLD against nCAPE, WCD, CCN, 1000-500 hPa wind shear and the average relative humidity (RH) between 850 and 500 hPa using ten statistical models. They created a lightning parameterization using a multiple-linear regression model, which was the best performer of the ten models, and found that TLD was positively correlated with nCAPE, CCN and wind shear, and was negatively correlated with WCD and RH. It was concluded that wind shear and RH were of secondary importance when compared to nCAPE, CCN and WCD.

Rather than evaluating RH at isolated levels as in Stolz et al. (2017), column saturation fraction (r) is a measure of how humid a column is relative to its saturation specific humidity (similar to column-integrated water vapor). Many studies have linked precipitation (P) to r (e.g., Bretherton et al. 2004) and have found that P – r curves vary for differing precipitation types and intensity (Ahmed and Schumacher 2015, 2017). However, studies relating lightning and r are absent in the literature. In addition, large-scale vertical motion (omega) at 500 hPa has also been used as a proxy for precipitation (e.g., Bony et al. 2004). However, like r, the link between large-scale omega and lightning in the literature is lacking. Furthermore, the relationship between mid-layer wind shear (1000-500 hPa) and lightning has been identified (e.g., Stolz et al. 2017) and was shown to have secondary importance for lightning production. However, the relationship between low-level wind shear (LS) and deep wind shear (DS) and lightning has not yet been studied. LS has been shown to increase the longevity of storms and a strong relationship has been observed between LS and squall line strength (Rotunno et al. 1988). DS has been shown to enhance hydrometeor detrainment and therefore increase the stratiform rain area

in deep convection (Li and Schumacher 2011), thus similar to omega and r, there may be an unidentified link between LS and lightning and DS and lightning because of the relationship of LS and DS to precipitation.

Increasing the understanding of lightning and large-scale environmental variables will also help when predicting lightning and creating lightning parameterizations; however very few studies have analyzed lightning in global climate models (GCMs). Lightning occurrence is rare and is considered to have a relatively weak signal and interaction with other model variables, and therefore is typically not parameterized and/or output. In addition to Stolz et al. (2017), Romps et al. (2014) and Magi (2015) have attempted to shows the significance and importance of a lightning parameterization.

Magi (2015) used a subset of Coupled Model Intercomparison Project phase 5 (CMIP5) models to evaluate monthly mean total precipitation, convective precipitation and convective mass flux from the GCMs in relation to observed global lightning flash rate distributions from a combined dataset of the Optical Transient Detector (OTD) and LIS. An empirical lightning parameterization from the GCM output and lightning observations was created using a polynomial fit. Over land, the lightning parameterization applied to the GCM output was found to under (over) predict flash rates in areas where higher (lower) occurrences of lightning were observed in satellite measurements. Generally, flash rates were over-predicted across all ocean areas compared to observations. Magi (2015) used monthly mean values in his calculations, but better correlations and statistical fits are more likely with higher temporal resolution data.

Romps et al. (2014) created a lightning parameterization from one year of twelve-hourly observations of precipitation and CAPE over the Continental United States (CONUS) and applied it to GCM outputs. The majority of GCMs predicted an increase in precipitation and they all predicted an increase in mean CAPE over CONUS, and therefore an overall increase in lightning. While this parameterization worked well over CONUS, it most likely would not translate over the oceans and/or on a global scale because relationships between lightning and the environment can exhibit strong regional differences and thus need to be accounted for, as other studies have observed (e.g., Williams et al. 2002; Zipser et al. 2006). In addition, precipitation is a highly derived, and often biased, quantity in models so focusing on large-scale environmental variables is warranted when examining the ability of GCMs to represent changes in lightning production.

Statistical models can be used to create lightning parameterizations. When evaluating the statistical relationship between meteorological variables and certain events, there are two ways to quantify the result: (1) determining whether the event occurs and (2) assessing the intensity of the event. Logistic regressions can be used to determine if an event occurs when the response is dichotomous, i.e., when there are only two possible outcomes (McCullagh and Nelder 1989). Previous studies (e.g., Stolz et al. 2017) have used linear logistic regressions to quantify the occurrence of lightning and have found this method to be effective. The spatial Log-Gaussian Cox Process (LGCP) is a point-process model first introduced by Moller et al. (1998). This model differs from a linear logistic regression by determining both the occurrence and intensity of an event; by extension, a

LGCP approach is more complex and its accuracy in determining the occurrence and intensity of an event is much more precise than a simple regression model. Unlike the linear logistic regression method, quantifying lightning flash rates and intensity using a LGCP model has not been directly addressed in the literature.

Motivated by this review of the current literature, the work presented here will have three distinct objectives: (1) Analyze the relationship, if any, between variables that characterize the large-scale environment with the occurrence of lightning between 35°N and 35°S latitude, (2) Evaluate the use of a LGCP statistical model by comparing results to those obtained using a linear regression model, and (3) Create a lightning parameterization using the linear regression model and use it to predict lightning in a GCM. In addition, this work will assess the robustness of the conclusions presented by Stolz et al. (2017) that RH and wind shear are of secondary importance when used as a predictor of lightning. This will be done by (a) using r as a proxy for column-integrated humidity, rather than using RH at isolated levels, and (b) distinguishing between a low-level and deep wind shear as opposed to using a mid-layer shear parameter.

2. DATA AND METHODS

2.1. Observational Datasets

Ground instrumentation, like lightning mapping arrays (LMAs), were some of the only sources of lightning data prior to space-borne instruments. Multiple studies have evaluated the influence of large-scale environmental variables on lightning using LMAs (e.g., Samsury and Orville 1994; Molinari et al. 1994; Williams et al. 2002) despite sparse to non-existent data on a global scale, particularly over the ocean. The introduction of the World-Wide Lightning Location Network in the late-1990s (WWLLN, Rodger et al. 2005) helped to populate these data sparse regions and ensured that LMAs remained a common source of lightning data during the satellite era.

In 1995, the OTD onboard Microlab-1 (later renamed Orbview-1) was the first space-borne lightning observing instrument to be launched into space (Christian et al. 2003). Between 1995 and 2000, the OTD provided previously unobtainable observations of lightning occurrences over oceanic regions but lacked in its detection efficiency (Cecil et al. 2014). In November of 1997, the TRMM satellite – a joint project between the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA) – was launched (Kummerow et al. 1998). The mission effectively ended in 2014 when the satellite ran out of fuel and began a slow drift downward. One of the instruments onboard TRMM was LIS, which improved on the OTD's detection efficiency by detecting approximately 90% of all lightning in TRMM's swath (Cecil et al. 2014). LIS detected lightning at storm-scale resolution (~3-6 km) and was able to observe a single point for approximately 80-100 seconds.

The OTD and LIS provide a valuable dataset of the occurrences and intensity of global lightning between 1995 and 2015 that has been utilized to evaluate the influence of large-scale environmental variables on lightning (e.g., Magi 2015). For this study, OTD data is excluded based on the suboptimal detection efficiency and resolution. LIS orbit data files are obtained from the NASA Earth Observing System Data and Information System Global Hydrology Resource Center (http://ghrc.nsstc.nasa.gov/) and a full TRMM LIS climatology is built to examine the interannual variability during TRMM's lifetime. TRMM orbited from the end of 1997 to early 2015, but the partial years, 1997 and 2015, are eliminated for consistent interannual analysis.

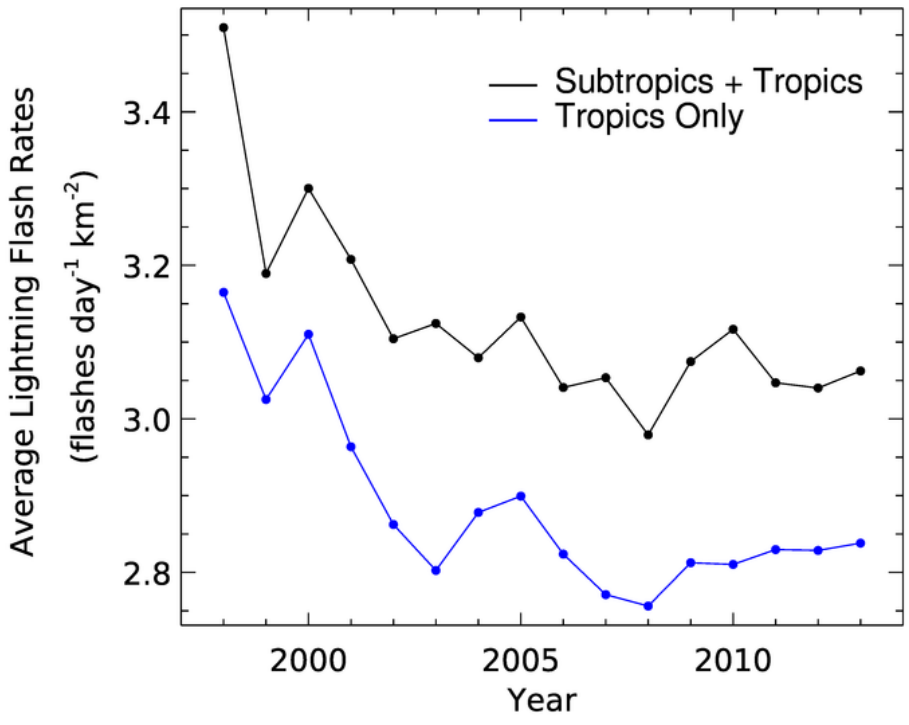

Figure 1. Climatology of flash rates from TRMM LIS. Black line shows average flash rates for latitudes between 35°N and 35°S and blue line shows average flash rates for tropical areas, between 20°N and 20°S.

Figure 1 shows the year-to-year variability in the LIS average lightning flash rates separated into tropics only (blue line) and subtropics + tropics (black line). Most of the lightning observed by LIS occurs in the tropics and there are some notable interannual variations associated with ENSO. For example, the strong El Niño in 1998 is hypothesized to be the cause of the large flash rates at the beginning of the climatology with lower values during the La Niña event the following year. After 2000, there appears to be a decline in average flash rates until 2008 after which the values begin to increase until 2013. It is unclear what may be causing this trend, but the overall range in average flash rates amount to only ~10% of the maximum observed value. Further, the tropics only and subtropics + tropics generally show a similar pattern, albeit with some differences year-to-year.

During 2014, TRMM was being prepared to come down from orbit and its altitude varied frequently. Following Albrecht et al. (2016), 2014 is also excluded for the remainder of the analysis. The TRMM satellite was boost to a higher altitude in 2001 in order to conserve fuel. Thus, flash rates (units of flashes s^{-1} km^{-2}) are used instead of flash counts (units of flashes s^{-1}) throughout the analysis to avoid artificial changes in counts because of the boost.

The global climatology of flash rates is analyzed within the context of the larger scale atmospheric conditions. Atmospheric observations are limited over large parts of the tropics, therefore reanalysis data, such as the ERA-Interim dataset used in Stolz et al. (2015, 2017), is necessary to describe the 3-D state of the large-scale environment. The Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-

2) is a reanalysis dataset produced by NASA's Global Modeling and Assimilation Office (Gelaro et al. 2017). MERRA-2 data is available globally at 3-hourly temporal resolution with a horizontal resolution of 0.5° x 0.625° and 72 vertical pressure levels from the surface to 0.01 hPa.

MERRA-2 reanalysis data files are obtained from the NASA's Goddard Earth Sciences Data and Information Services Center (https://goldsmr5.gesdisc.eosdis.nasa.gov/). MERRA-2 provides the pressure, temperature, specific humidity and omega fields as well as the zonal (u) and meridional (v) wind components that were used to calculated the six large-scale environmental variables used for analysis in this study: CAPE (J kg^{-1}), nCAPE (J kg^{-1} m^{-1}), r (unitless), omega (Pa s^{-1}) at 700 hPa, LS (m s^{-1}) from 900-700 hPa and DS (m s^{-1}) from 900-300 hPa. These environmental variables are calculated from the 3-hourly MERRA-2 data and matched with the TRMM LIS data between 35°N and 35°S. The data is classified as a lightning environment if a lightning flash occurs within 30 minutes of the MERRA-2 time stamp, otherwise it is classified as a non-lightning environment.

CAPE is calculated by taking the temperature and specific humidity data provided by MERRA-2 and using the cape_sound.pro Interactive Data Language (IDL) script created by Dominik Brunner (https://svn.ssec.wisc.edu/repos/bennartz_group/LIBRARY/idl/std_libs/atmos_phys/). Dividing CAPE by the depth of the positive CAPE area (i.e., the area between the Level of Free Convection (LFC) and the Equilibrium Level (EL)) gives an approximation of nCAPE.

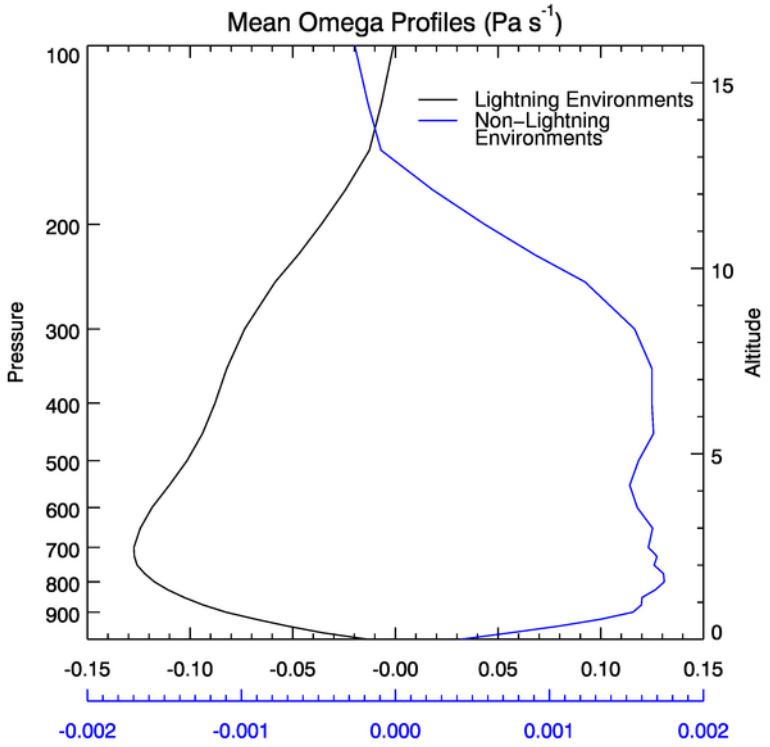

Figure 2. Mean omega profiles for differing environments. Lightning environments are shown in black and non-lightning environments are shown in blue (note differing x axis for differing environments). Profiles are averaged for the years 1998-2013.

Figure 2 shows mean omega profiles for lightning environments and non-lightning environments from the MERRA-2 data set. The non-lightning environments show very weak sinking motion (positive values) through most levels and the lightning environments show strong rising motion (negative values), particularly at lower levels, with the maximum occurring at 700 hPa. Because of this maximum separation, omega at 700 hPa is chosen to best represent the large-scale vertical motions for lightning environments.

As done in Ahmed and Schumacher (2015), r is calculated by first estimating the saturation vapor pressure (e_s), column water vapor (CWV) and the saturation CWV (CWV_s), following equations (1)-(3), respectively,

$$CWV = qv_0 * \frac{\Delta p_0}{g} + \cdots + qv_i * \frac{\Delta p_i}{g} \quad \textbf{(1)}$$

$$e_s = x_0 + x_1(T - T_r) + x_2(T - T_r)^2 + \cdots + x_i(T - T_r)^i \quad \textbf{(2)}$$

$$CWV_s = qv_{s_0} * \frac{\Delta p_0}{g} + \cdots + qv_{s_i} * \frac{\Delta p_i}{g} \quad \textbf{(3)}$$

where $qv_{0(i)}$ is the mixing ratio at the lowest (ith) pressure level, $\Delta p_{0(i)}$ is the difference in pressure between the lowest (ith) level and the level immediately above, g is the acceleration due to gravity, $x_0, \ldots, x_i$ are the coefficients of e_s, T is temperature in K and $T_r = 273.15$. The subscript *'s'* indicates the variable is at saturation. Dividing CWV by the CWV_s, gives an estimate of r. LS and DS are estimated using equations (4) and (5)

$$LS = \sqrt{(u_{900} - u_{700})^2 + (v_{900} - v_{700})^2} \quad \textbf{(4)}$$

$$DS = \sqrt{(u_{900} - u_{300})^2 + (v_{900} - v_{300})^2} \quad \textbf{(5)}$$

where u_i and v_i are the zonal and meridional winds at the ith pressure level.

The MERRA-2 and TRMM LIS data was initially binned at three different resolutions: 2.5°, 1.0° and 0.5°, and tests were performed to evaluate the sensitivity of lightning environment distributions to resolution. Figure 3 shows histograms for MERRA-2 derived variables for lightning occurrences between 1998 and 2013 for 0.5° x 0.5° resolution (solid blue lines) and 2.5° x 2.5° resolution (dashed red lines). The axes are normalized by the maximum count value, except for CAPE and nCAPE where the second largest count values were used as the normalizing factor to exclude the high counts at 0. CAPE and nCAPE show the biggest differences between the two grid resolutions, where more lightning occurs for higher CAPE and nCAPE at 0.5° x 0.5° resolution (Figure 3a

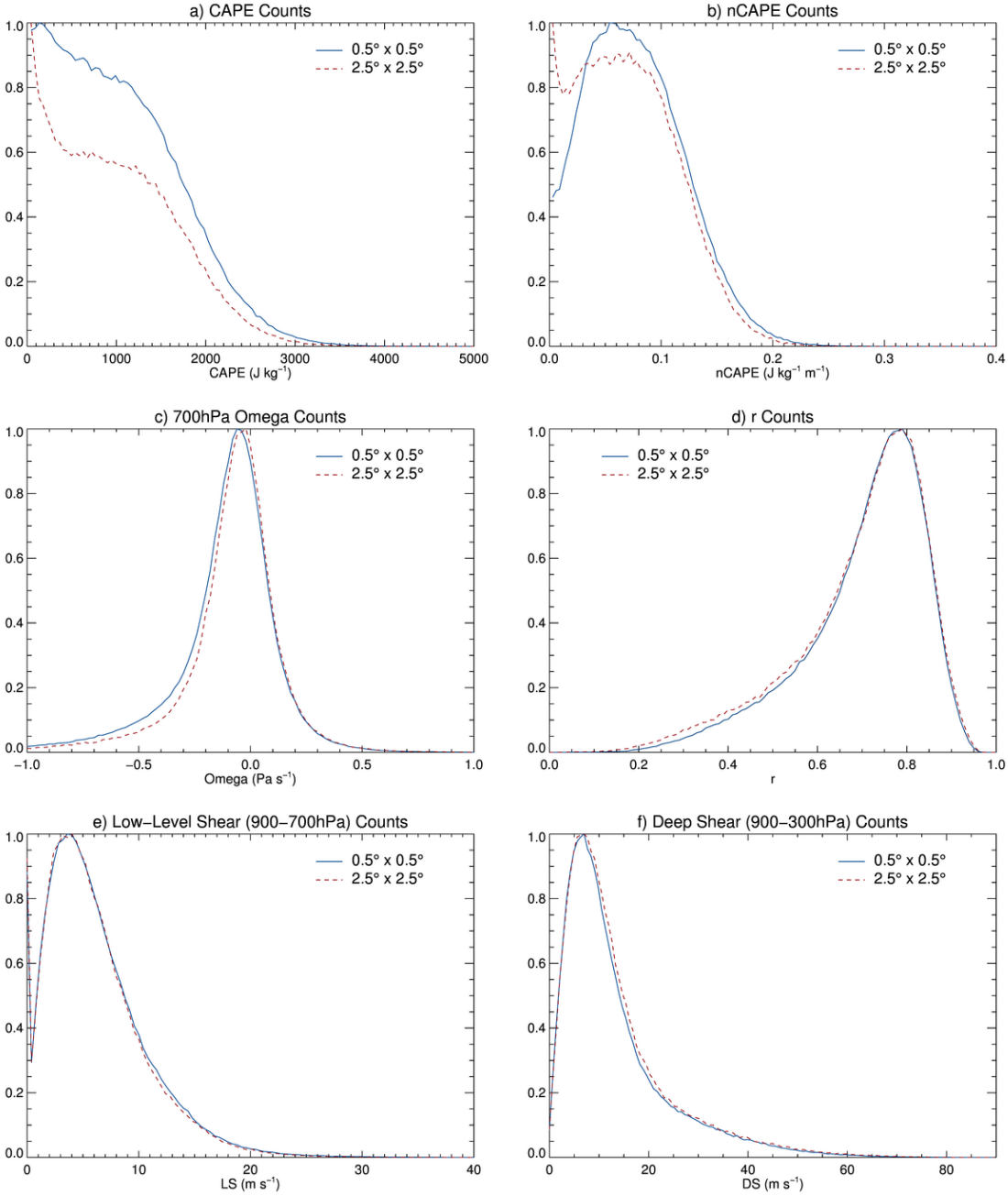

Figure 3. Distributions of MERRA-2 variables for differing grids (1998-2013). Solid blue lines show distributions for 0.5° x 0.5° grid and dashed red lines show distributions for 2.5° x 2.5° grid.

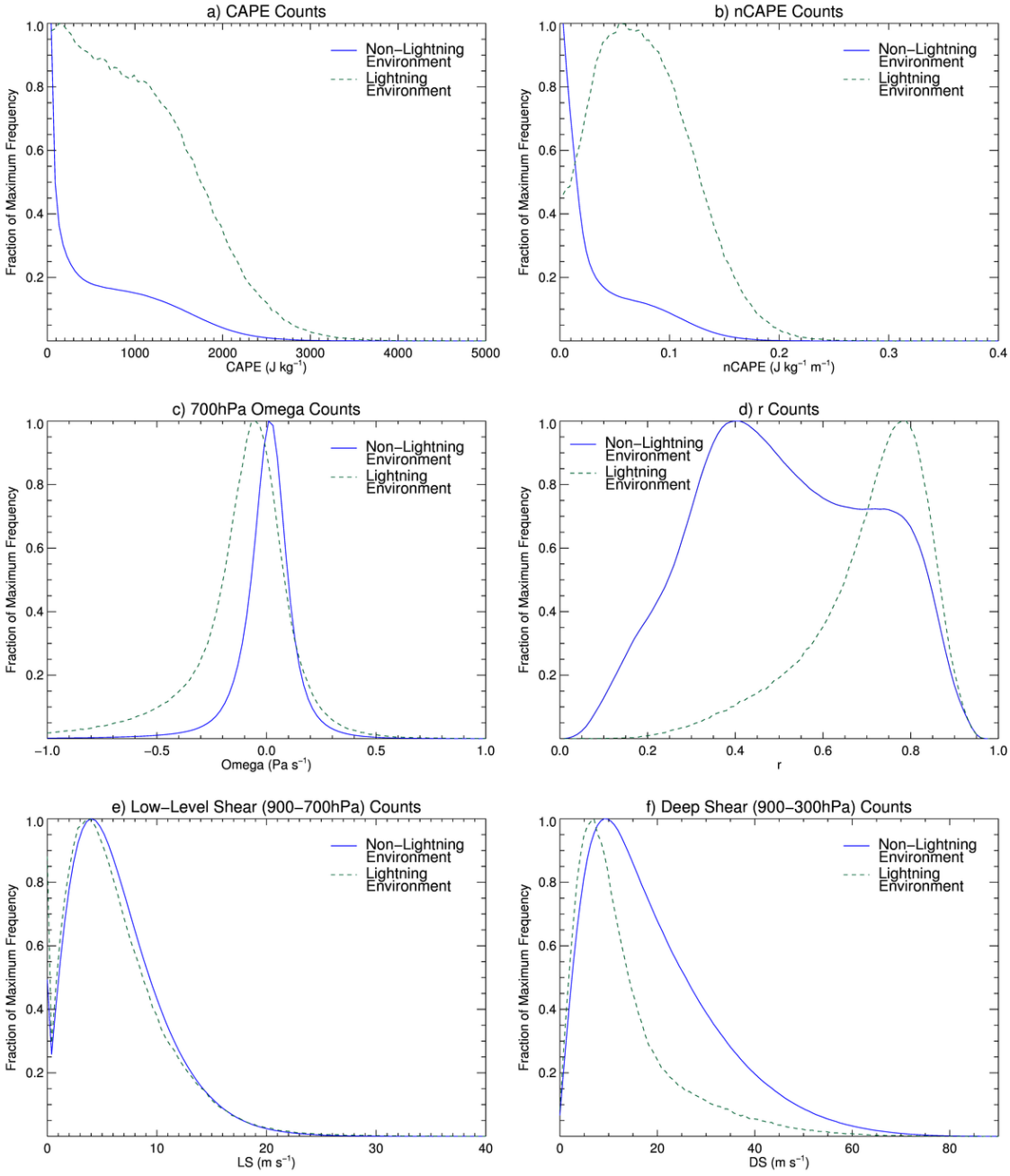

Figure 4. Distributions of MERRA-2 variables for differing environments (1998-2013). Solid blue lines show distributions for non-lightning environments and dashed green lines show distributions for lightning environments.

and b). Omega, r, LS and DS show minimal differences between the two grid resolutions (Figure 3c, d, e and f). Therefore, the 0.5° grid was chosen to more accurately represent CAPE and nCAPE for lightning environments.

Histograms of CAPE, nCAPE, omega, r, LS and DS observed within 0.5° grids with lightning between 1998 and 2013 are compared to environmental values from grids with no lightning (Figure 4). The axes are again normalized by the maximum count value for each of the variables, with the exception of CAPE and nCAPE where the second largest value was used as the normalizing factor. Figure 4 shows that the distributions of the large-scale parameters are clearly distinguishable between the lightning and non-lightning environments, with the exception of LS. The largest differences are reflected in CAPE, nCAPE and r. Figure 4a shows that 74% of lightning occurs at CAPE greater than 500 J kg^{-1} and only 56% of non-lightning environments have CAPE greater than 500 J kg^{-1}. For non-lightning (lightning-only) environments, 68 (30) % of nCAPE instances lie between 0 and 0.05 (Figure 4b). Figure 4d also presents justification for the use of r as a predictor of lightning, as the distribution of values for all environments is much broader than for lightning-only, with the peak in the latter shifted to higher r.

As expected, omega at 700 hPa shows a shift towards increased rising motion in lightning-only environments (Figure 4c), however the smaller difference between the distribution curves suggests that omega will be of secondary importance in predicting lightning compared to CAPE, nCAPE and r. The distribution of LS for lightning-only environments is almost indistinguishable from the non-lightning environments (Figure

4e). However, the difference in distributions becomes larger for the DS case (Figure 4f), motivating the inclusion of the alternate shear parameters in subsequent analysis.

Mean values of flash rates from TRMM LIS and atmospheric variables derived from MERRA-2 between 1998 and 2013 are shown in Figure 5 for all longitudes between

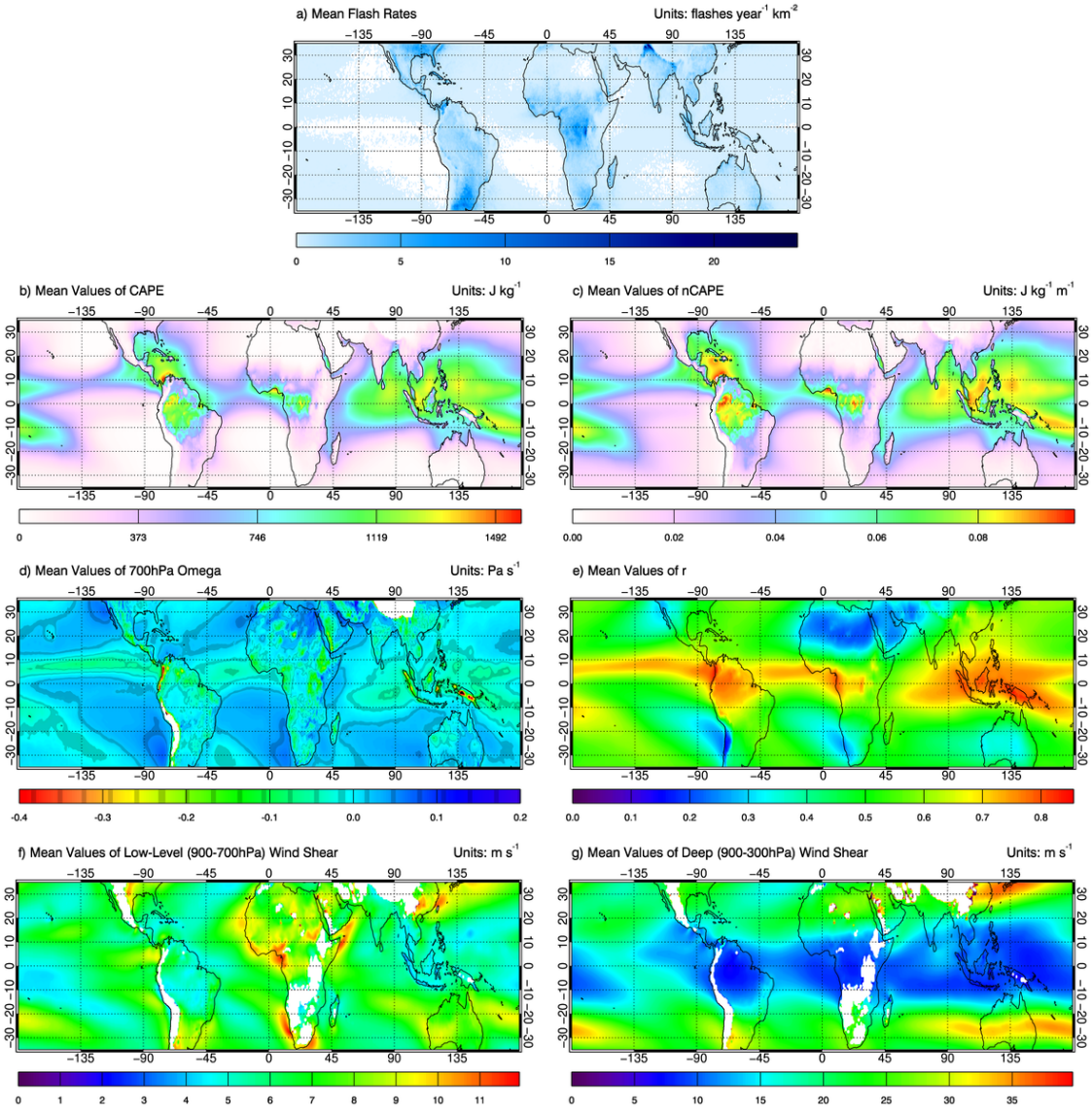

Figure 5. Mean maps for lightning and MERRA-2 variables for the years 1998-2013. Panel (a) shows flash rates (flashes year^{-1} km^{-2}) obtained from TRMM LIS and (b)-(g) shows environmental variables derived from MERRA-2.

35°N and 35°S. Higher flash rates are more prevalent over land compared to the ocean, consistent with many previous studies (e.g., Orville and Spencer 1979; Williams and Stanfill 2002; Virts et al. 2013; Albrecht et al. 2016). Flash rates over the ocean are never more than 3 flashes $year^{-1}$ km^{-2}, while flash rates exceed 15 flashes $year^{-1}$ km^{-2} over parts of the Congo region of Africa, the Himalayas in Southeast Asia, Argentina and the southeastern United States. Over land, higher flash rates are often correlated with moderate-to-high CAPE (>1000 J kg^{-1}), high nCAPE (>0.08 J kg^{-1} m^{-1}), moderate-to-high r (>0.8) and negative omega indicative of rising motion. Higher LS (>10 m s^{-1}) and DS (>30 m s^{-1}) values are associated with the higher flash rates in South America, but not in all flash intensive regions. This regionality could be caused by the low-level jet present near the Andes (Stensrud 1996). Regions of zero flash rates (e.g., the southeast Pacific and southern Atlantic Oceans) are reflected in near-zero mean values of CAPE, nCAPE and r, which helps motivate this study in understanding the geographical relationship with variables and lightning.

2.2. Statistical Models and Lightning Parameterization

Correlations between lightning and environmental variables found by two linear regression models and one LGCP model are evaluated in order to determine which model will work best in a lightning parameterization. The two generalized linear models (GLMs) utilized in this study determine whether lightning will occur using a logistic regression (McCullagh and Nelder 1989). The first model is trained using data obtained for 2003 from the TRMM LIS and MERRA-2 reanalysis (see section 2.1). The year 2003 is chosen

because it lies outside of an El Niño or La Niña period in which the data may not represent normal conditions. In this model, the logit of the transformation of the probability of a lightning occurrence, $logit\big(p(s)\big)$, where s is a grid point, is expressed as the linear combination of the environmental predictors:

$$logit\big(p(s)\big) = \log\left\{\frac{p(s)}{1-p(s)}\right\} = \beta_0 + \beta_1 X_1(s) + \cdots + \beta_p X_p(s) + \epsilon(s) \qquad \textbf{(6)}$$

where $X_i(s)$ denotes the predictor (e.g., nCAPE, r, etc.) at grid point s, β_i represents the coefficients found by the model, and $\epsilon(s)$ denotes the associated error term. To evaluate the differences, if any, between correlations (i.e., the magnitude and sign of the coefficients found in equation (6)) of lightning and large-scale variables over different surfaces, the model is trained using data over land and ocean, land-only, and ocean-only.

Lightning is detected by TRMM LIS just 0.01% of the time within a half hour of the MERRA-2 three hourly samples; therefore, lightning is considered a rare event. This situation can result in inaccurate coefficients and potentially large underestimation in the probability of lightning occurring. The extent of this bias is tested using a second linear regression model. The rare event logistic model (ReLogit) was first introduced by King and Zeng (2003) and utilizes the same logistical regression method after applying a correction that accounts for small sample sizes. Training the ReLogit model on the 2003 dataset constitutes the "rare event" experiment and the coefficients of each predictor are compared to those obtained from runs over each surface type.

In order to test the predictability of lightning using these linear regression models, the output of equation (6), which ranges between zero (no chance of lightning occurring)

to one (100% chance of lightning occurring), is examined. These probabilities of lightning flash occurrences for 2003 will be compared with observations obtained from TRMM LIS for the tropical and sub-tropical regions in order to determine a cutoff probability, p_c. Lightning is defined to occur over locations where $\hat{p}(s) \geq p_c$, where $\hat{p}(s)$ is the predicted probability of the lightning occurrence.

The outputs of these linear regression models will give a probability of lightning occurring in each bin, however there is no information regarding the flash rate (i.e., the intensity). The spatial LGCP was first introduced by Moller et al. (1998) and will give a probability and intensity for a region of interest. First, an intensity process is defined as $R(s) : D \rightarrow [0, \infty)$ where $D \subset \mathbb{R}^2$, i.e., an intensity process is a non-negative valued stochastic process. The point process $X(\cdot)$ is a nonnegative integer-valued process defined on the domain D and is called the spatial LGCP with the intensity function $R(\cdot)$ if (1) $X(\cdot)$ is a Cox process (Cox 1955) with intensity $R()$, and (2) $\log(R(\cdot))$ is a Gaussian process (MacKay 2003). In other words, $X(\cdot)$ is said to be a spatial LGCP with intensity R if X and R satisfy the following two conditions:

$$\begin{cases} X(s) \sim Possion(R(s)) \\ \log\big(R(s)\big) = \gamma_0 + \gamma_1 X_1(s) + \cdots + \gamma_p X_p(s) + Y(s) \end{cases} \quad \textbf{(7)}$$

where $X_i(s)$ is the ith predictor at a grid point s and $Y(\cdot)$ is assumed to be a mean zero stationary bivariate gaussian process. One of the most widely used covariance models for the stationary process $Y(\cdot)$ is the Matérn covariance, including the case when the smoothness parameter is held constant ($\nu = 0.5$), called the exponential covariance

(Matérn 1986), where K_ν is a modified Bessel function of the second kind, σ^2 is the variance and ϕ is the scale parameter.

$$cov\big(Y(s_1), Y(s_2)\big) = C(\|s_1 - s_2\|) = \begin{cases} \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\|s_1 - s_2\|}{\phi}\right)^{\nu} K_\nu \left(\frac{\|s_1 - s_2\|}{\phi}\right) & \text{Matérn} \\ \sigma^2 \exp\left(-\frac{\|s_1 - s_2\|}{\phi}\right) & \text{exponential} \end{cases} \quad \textbf{(8)}$$

Evaluating the covariance will give a measure of the dependence of the data.

2.3. Community Atmosphere Model Version 5 Data

GCMs rarely include a lightning parameterization due to the small spatial and temporal scales involved. However, it is beneficial to investigate whether GCMs are capable of predicting lightning using the statistical methods described in section 2.2. The Community Atmosphere Model version 5 (CAM5) does not currently include a lightning parameterization but can provide output at a similar temporal (three-hourly) and spatial (0.25° x 0.25°) resolution to the MERRA-2 observational dataset. Large-scale atmospheric variables from CAM5 (pressure, temperature, specific humidity, omega, and the u and v wind components) are interpolated to a 0.5° x 0.5° grid at three-hourly resolution (i.e., the same resolution as the MERRA-2 data) for 2003. Calculations of CAPE, nCAPE, r, LS and DS follow the same methods described in section 2.1 and are input into the statistical models described in section 2.2. The predicted flash occurrences using the linear regression model are compared to the occurrences predicted by the regression model for CAM5 environmental data from 2003 and further analyzed with respect to the observed flash rates from TRMM LIS.

3. RESULTS

3.1. Relationship of Lightning and Large-Scale Environmental Variables

3.1.1. Single Variable Relationships

Density plots from 1998 to 2013 for all longitudes and latitudes between 35°N and 35°S of CAPE, nCAPE, omega, r, LS and DS against observed lightning flash rates are shown in Figure 6. Greater lightning density is associated with low-to moderate CAPE, moderate nCAPE, slightly negative 700-hPa omega (i.e., rising motion), high r, low-to-moderate LS and low DS, although there is large spread for all variables. Comparing Figure 4 and Figure 6, the same overall distribution of environmental values during lightning occurrence can be seen; however, it is now clear that the density of flash rates is greatest below approximately 2 flashes day^{-1} km^{-2} and the highest flash rates are very rare with varying relationships to the environmental parameters.

While higher densities of lightning (e.g., greater than 10 flashes day^{-1} km^{-2}) occur at CAPE between 0 and 1500 J kg^{-1} (Figure 6a) and nCAPE between 0.02 and 0.11 J kg^{-1} m^{-1} (Figure 6b), the highest flash rates occur in more narrow ranges of CAPE and nCAPE centered around 1000 J kg^{-1} 0.08 J kg^{-1} m^{-1}, respectively. This analysis shows that the highest flash rates don't necessarily occur at the highest possible CAPE or nCAPE. Similarly, the highest flash rates occur at slightly negative 700-hPa omega rather than at extreme negative values (Figure 6c).

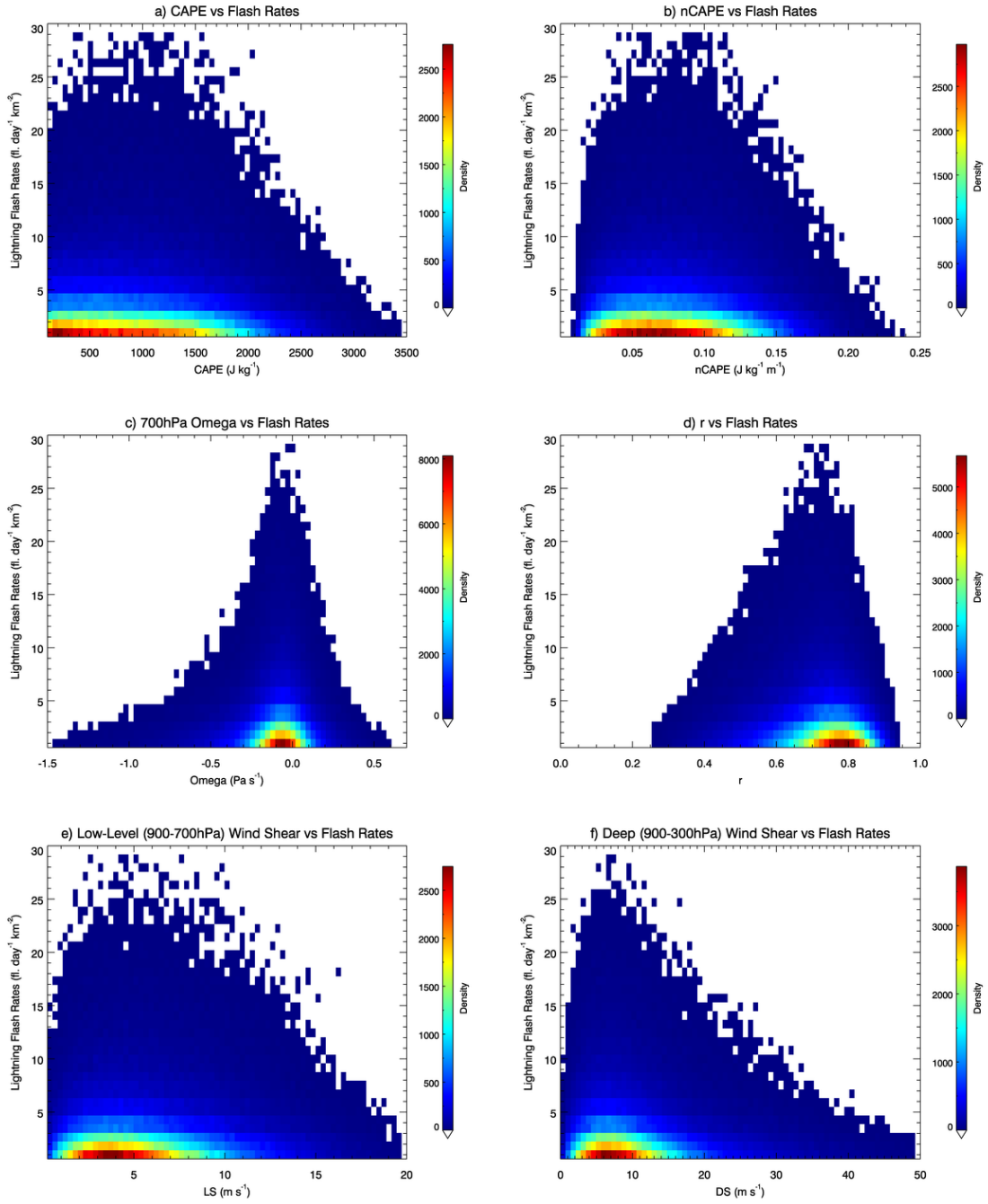

Figure 6. Density plots of large-scale environmental variables and flash rates (1998-2013).

Figure 6d shows that almost all lightning flashes occur at r equal to or greater than 0.25, with 59% of flashes occurring between 0.7 and 0.9, where a non-linear increase in flash rates occurs. This is indicative of the moisture source required to fuel the storms that produce lightning and unlike the other variables, the highest flash rates occur at the largest r. High flash rates also occur for LS primarily between 2 and 10 m s^{-1} and DS less than 15 m s^{-1} (Figure 6e and Figure 6f), thus high shear appears inimical to large lightning production.

Figure 7 shows the influence the underlying surface type has on the relationship between flash rates and the environmental conditions. Subtracting the normalized histograms for land and ocean produces positive (negative) values where there is a stronger relationship between the variables over land (ocean). Blue and red lines show contours of -0.03 and 0.03. Figure 7a and Figure 7b show a clear distinction between land and ocean, which makes sense in context of Figure 5, where a distinction in both flash rates and values of CAPE and nCAPE are observed. The histogram differences show that lightning is more likely to be associated with CAPE less than 700 J kg^{-1} and nCAPE less than 0.06 J kg^{-1} m^{-1} over land. However, these thresholds increase slightly as flash rates increase. Greater CAPE and nCAPE are more likely to be associated with lightning over ocean, potentially because (1) higher CAPE and nCAPE generally occur over the ocean (Figure 5) and/or (2) more CAPE, and therefore stronger updrafts, are required for lightning to occur over ocean.

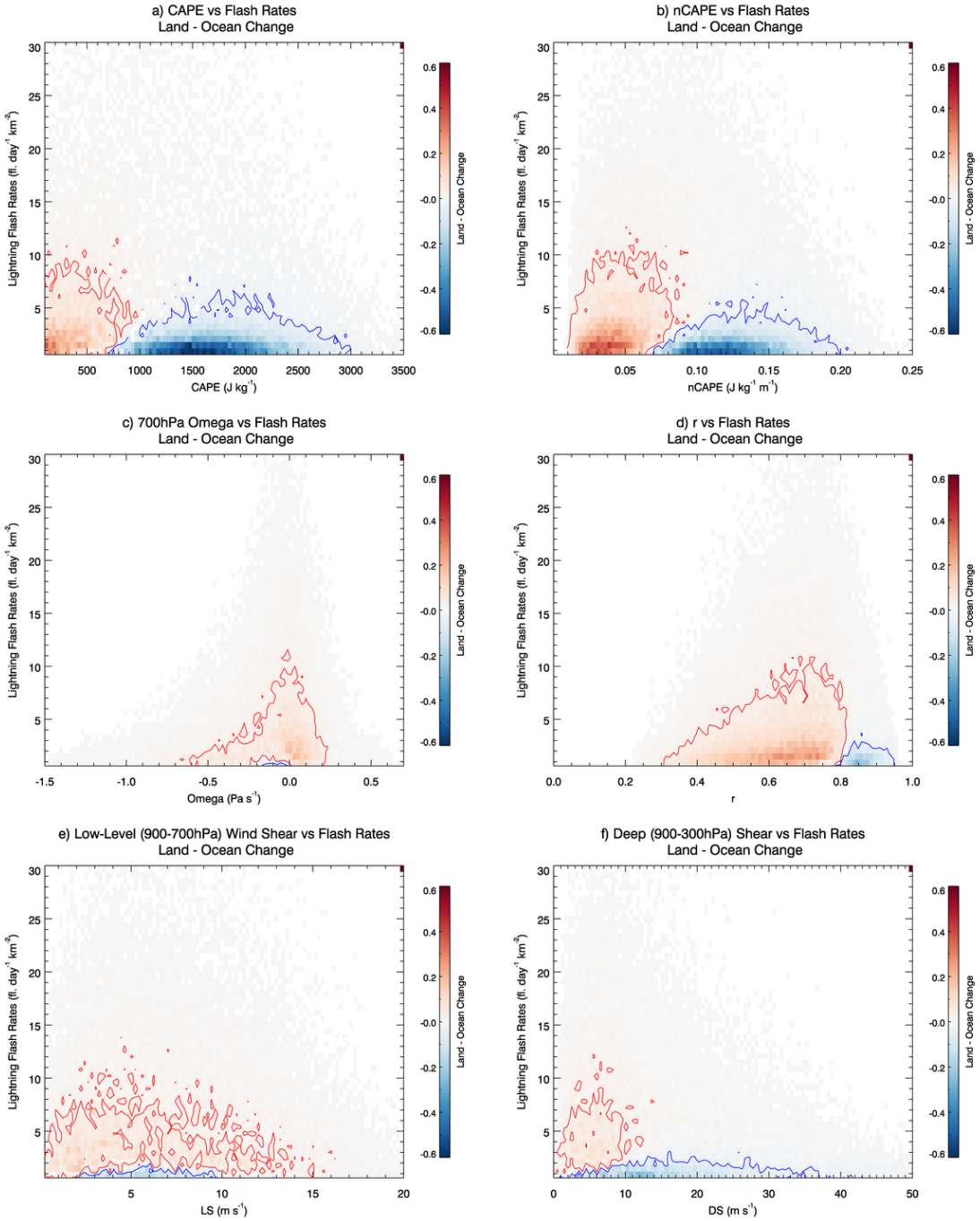

Figure 7. Land-ocean differences for environmental variables and flash rates (1998-2013). Values of -0.03 and 0.03 are shown in the blue and red contours.

Lightning is associated with higher nCAPE over ocean compared to over land (Figure 7b). This result appears inconsistent with Stolz et al. (2017), who suggested that lower nCAPE is expected over ocean and higher nCAPE is expected over land. However, recall that nCAPE uses CAPE and the "shape-of-the-CAPE" in its calculation (i.e., nCAPE = CAPE/z, where z is the depth of the CAPE layer). Since lightning occurs at significantly higher values of CAPE over ocean than over land (Figure 7a), the "skinny" nature of CAPE (and thus a larger z) only slightly weakens the land/ocean nCAPE relationship in Figure 7b (i.e., it is similar to the CAPE relationship, albeit less pronounced).

Omega at 700 hPa shows a shift towards positive values during lightning occurrences over land (although more highly negative omega values are also observed) and a more concentrated occurrence over ocean with values focused at -0.1 Pa s^{-1} (Figure 7c). The onset for lightning for r over ocean (land) is greater than (less than) 0.8 (Figure 7d), suggesting that lightning can occur in drier environments over land. Furthermore, Figure 7d shows that higher r is associated with higher flash rates over land at r values less than 0.8, while there is little difference in flash rate magnitude with changes in r over ocean compared to land. LS (Figure 7e) shows negligible differences between land and ocean and DS (Figure 7f) shows a shift towards lower DS for land at higher flash rates and higher DS for ocean at low flash rates. Figure 7 suggests that in order for lightning to occur over ocean, more favorable conditions (i.e., moderate-to-high CAPE and nCAPE, high r and moderate DS) must be present compared to land.

In addition to land/ocean differences, latitudinal differences are also evident between lightning and environmental parameters. Figure 8 shows histograms of flash rates by environmental variables for 1998-2013 in tropical areas (20°N to 20°S; 1st column), sub-tropical areas (20°N to 35°N and 20°S to 35°S; 2nd column), and the normalized difference between the tropics and subtropics (3rd column). Positive values (warm colors)

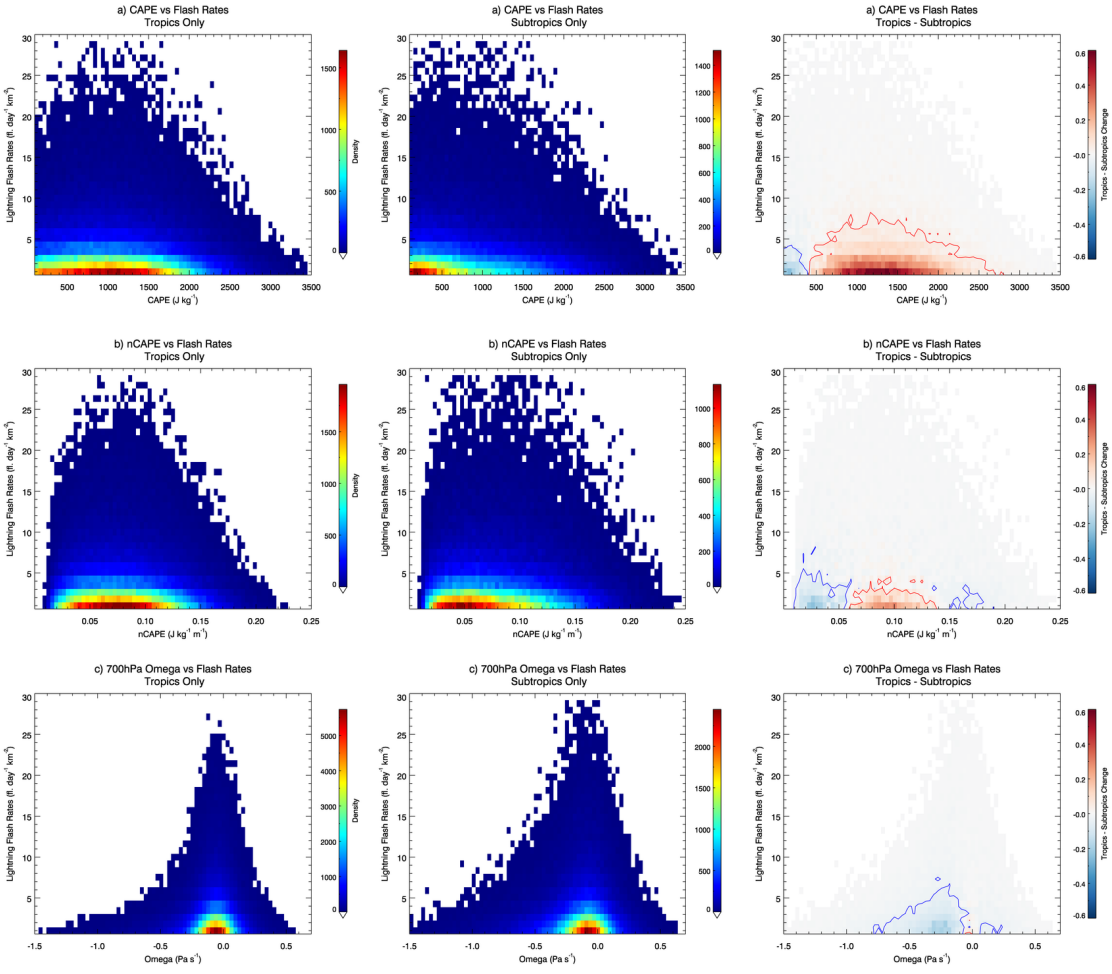

Figure 8. Tropics-subtropics density plots of environmental variables and flash rates (1998-2013). First column shows density plots for tropical areas, second column shows density plots for subtropical areas and third column shows the differences between the first two columns. Values of -0.03 and 0.03 are shown in the blue and red contours.

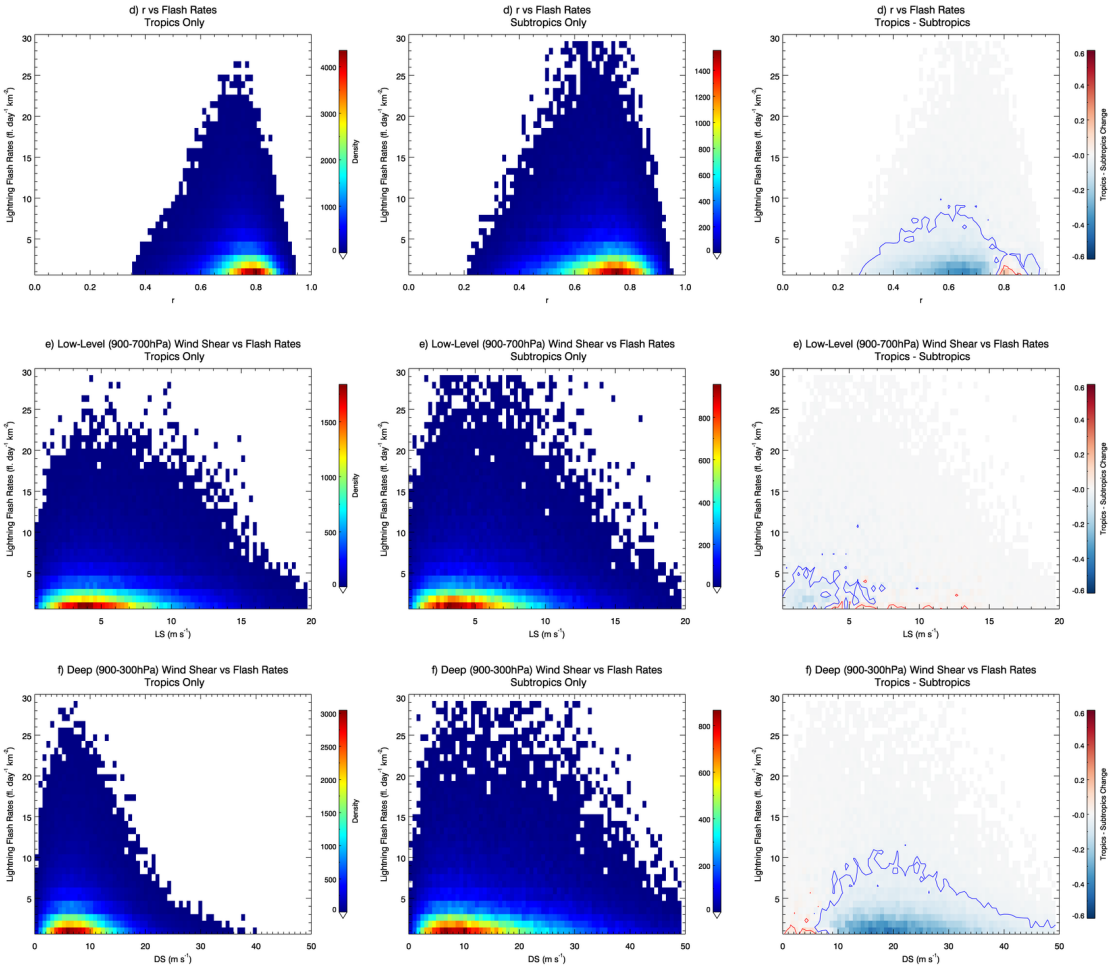

Figure 8. Continued

show stronger relationships for tropical areas and negative values (cool colors) show stronger relationships for sub-tropical areas. Lightning with flash rates less than 10 flashes year^{-1} km^{-2} occurs for higher CAPE in the tropics (Figure 8a), although there is also generally higher CAPE in the tropics versus the subtropics (Figure 5b). Between the tropics and subtropics, nCAPE associated with lightning doesn't shift as much as CAPE (Figure 8b). Subtle differences can be found at low, moderate and high nCAPE where a shift occurs towards the subtropics at low nCAPE, the tropics for moderate nCAPE and

slightly back to the subtropics for high nCAPE, most likely due to the land-ocean differences between the tropics and subtropics. While omega associated with lightning shows very little difference between the tropics and subtropics (Figure 8c), r shows a more distinct pattern with higher r in the tropics for low flash rates (Figure 8d). Again, this is likely due to the fact that the tropics are simply more moist (Figure 5e). LS shows very little latitudinal preference in relation to lightning production (Figure 8e), but DS shows a more dramatic tropics/subtropics comparison (Figure 8f). Lightning occurs for DS primarily less than 20 m s^{-1} in the tropics, but at all values between 0 and 50 m s^{-1} in the subtropics, which can be linked back to the different large-scale flow in both regions (Figure 5g).

Figure 7 and Figure 8 show that large geographical differences occur when looking at lightning production and its relationships with large-scale variables. Motivated by these results, maps of these variables only when lightning is occurring are examined (Figure 9). CAPE is higher over ocean during lightning occurrences, with some areas exceeding 2000 J kg^{-1}, particularly in the Caribbean Sea (Figure 9a). Little lightning production is expected in the Caribbean (Figure 5a), but Figure 9a shows that for lightning to occur, high CAPE must be present. Figure 9b shows minimal differences between CAPE and nCAPE when lightning is present except that some areas over land (ocean) increase (decrease) in magnitude when compared to CAPE, as expected for nCAPE where high values of nCAPE are expected over land and lower values over ocean. This is most distinguishable for land in Argentina, the southeastern United States and South Africa and for the Caribbean Sea and the Indian Ocean. Figure 9c gives some explanation for high flash rates in Argentina,

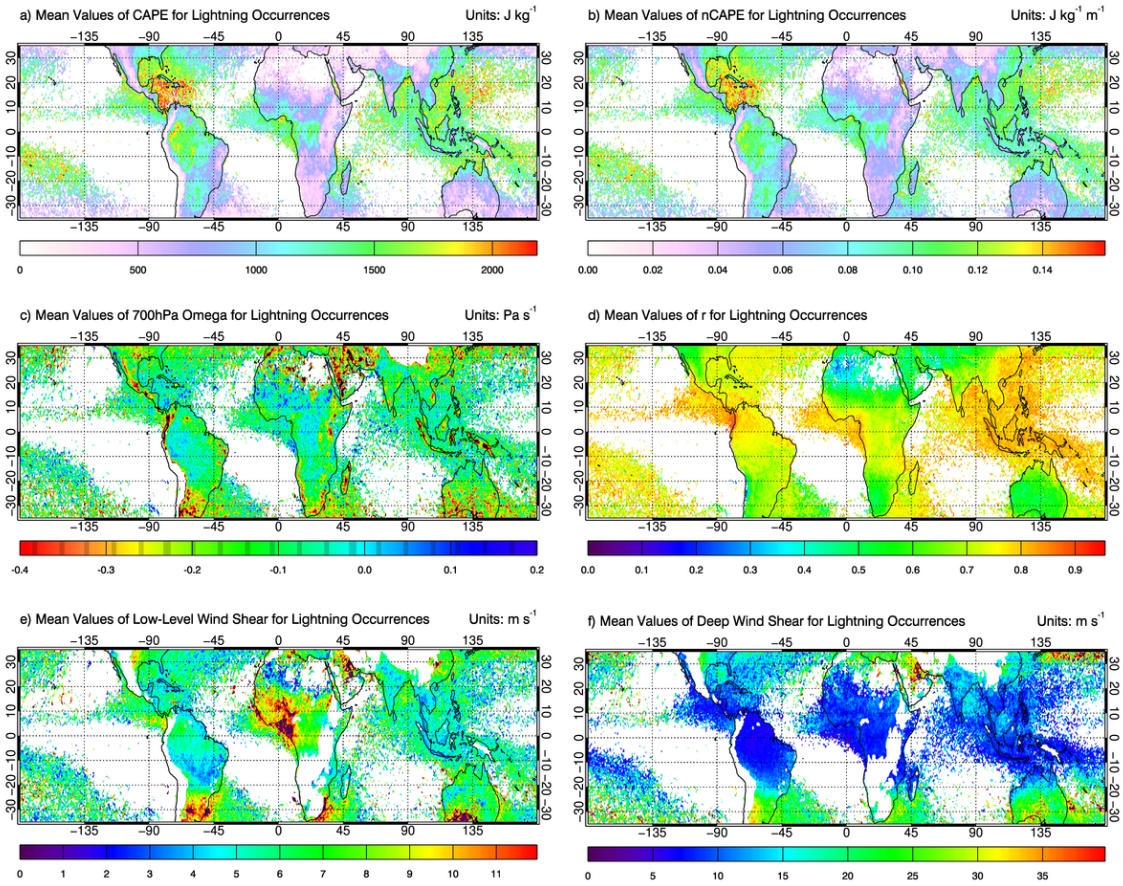

Figure 9. Mean environmental maps of MERRA-2 derived variables for lightning occurrences for the years 1998-2013.

where it can now be observed that large rising motion is associated with lightning occurrences. Figure 9d shows some regional differences and supports the conclusions of Figure 7d and Figure 8d where higher r is expected for lightning occurrences over the ocean and that the lightning occurs in the subtropics in less moist environments. LS and DS show enhanced mean maps of those presented in Figure 5f and g and can also help to explain large lightning production in areas like Argentina, where high LS and DS are observed (Figure 9e and f). Overall, these maps show that large regional contrasts are

present and that for lightning production to occur, variable thresholds may change from region to region.

3.1.2. Two-Variable Relationships

Figure 10 shows the two-parameter relationship between one of the best predictors, r, and each of the other variables for lightning occurrences. CAPE and r show the clearest relationship: as r increases from moderate-to-high values, lightning occurs at higher CAPE with the highest lightning density at high r/moderate CAPE (Figure 10a). A similar relationship is seen with nCAPE, although there is a less gradual "pick-up" in nCAPE at moderate r (Figure 10b). Figure 10c shows high lightning densities at high r and a focused range of omega between 0.0 and -2.0 Pa s^{-1}. At high r, LS between 1 and 7 m s^{-1} and DS between 2 and 13 m s^{-1} are associated with higher lightning occurrence. Figure 10 suggests that at high r, environments with moderate-to-high CAPE and nCAPE, negative omega and moderate LS and DS are conducive to lightning production, thus refining the predictive potential of each of these parameters.

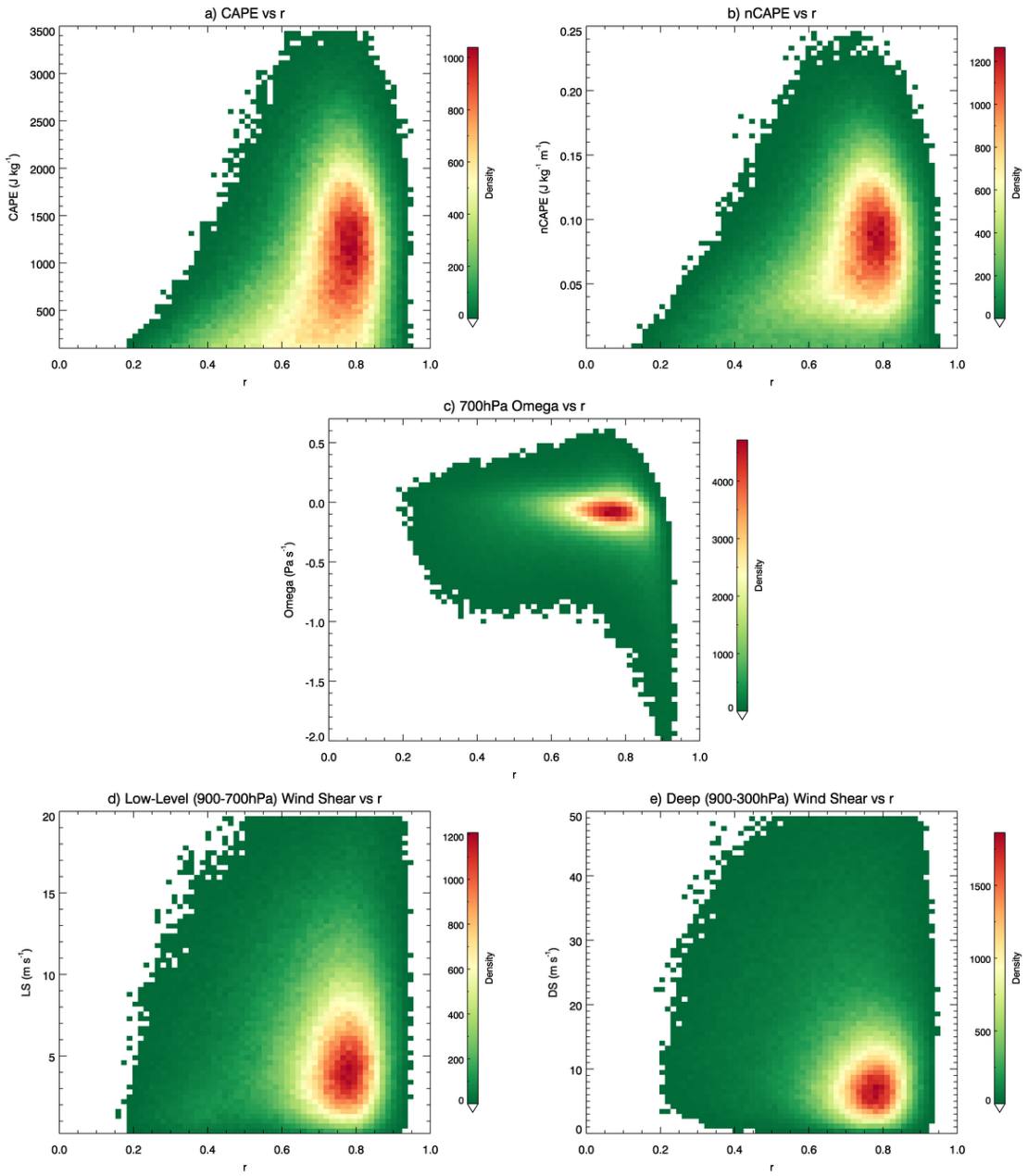

Figure 10. Two-variable density plots from MERRA-2 for all flash occurrences (1998-2013).

Following the same method as Figure 7, two-parameter density lightning difference plots between land and ocean are shown in Figure 11. Large differences between land and ocean are apparent. Lightning over ocean typically occurs at moderate-to-high CAPE and nCAPE and high r (shown as the blue areas in Figure 11a and b). Over land, there is a clear shift to lightning occurrence in environments with low-to-moderate CAPE and nCAPE and moderate r (shown as the red areas in Figure 11a and b). The pick-up for high r/moderate CAPE and nCAPE primarily occurs over land. Lightning over the ocean occurs at higher r for similar values of 700-hPa omega compared to over land (Figure 11c). Figure 11d shows that over ocean (land), lightning occurs at all values of LS when r is high (moderate-to-low). Unlike the LS case, Figure 11e shows that for land surfaces, lightning occurrence increases as DS decreases and r increases.

Figure 11 shows that there is a large land-ocean contrast in environmental characteristics during instances of lightning that can be more fully explained in a multi-parameter space. Overall, these plots help to explain this contrast, particularly over land, where pickups occur at increasing r/low-to-moderate CAPE, increasing r/low-to-moderate nCAPE and decreasing r/low-to-moderate DS. Figure 11c and d show similar results to those presented in Figure 7c and d, and don't provide significant additional information when combining the two parameters.

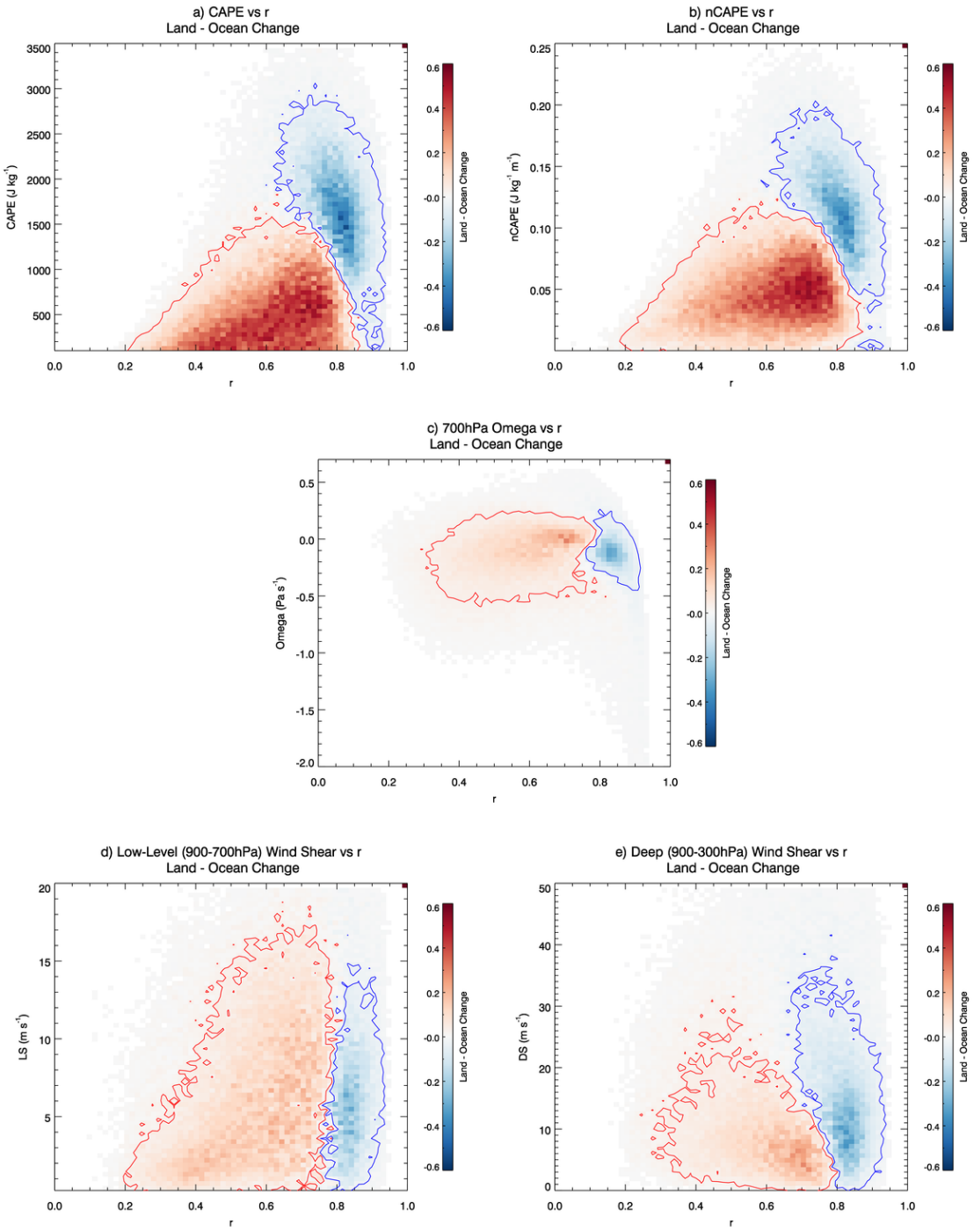

Figure 11. Land-ocean differences for two-variable density plots of MERRA-2 environmental variables (1998-2013). Values of -0.03 and 0.03 are shown in the blue and red contours.

As shown in Figure 6a, lightning primarily occurs for low-to-moderate CAPE, which is unexpected. Instead, most lightning is expected to occur for high CAPE. To investigate this further, omega profiles for lightning occurrences are presented in Figure 12a to see how they change for differing CAPE intensities. The plot shows mean omega profiles for lightning occurrences for all CAPE (black line), weak CAPE (<500 J kg^{-1}; blue line), moderate CAPE (500 J kg^{-1} ≤ CAPE ≤ 2000 J kg^{-1}; green line) and strong CAPE (>2000 J kg^{-1}; red line). Figure 12a shows that weak CAPE cases are associated with the strongest rising motion, particularly in the lower levels. This is most likely due

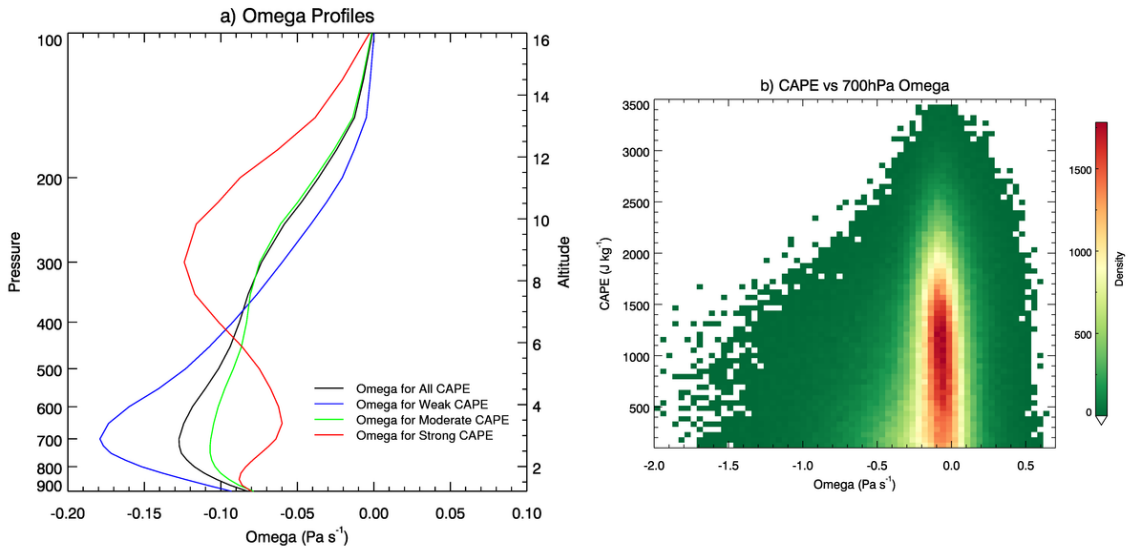

Figure 12. Omega-CAPE relationships. Left panel shows omega profiles for varying CAPE intensities averaged for years 1998-2013. CAPE thresholds are defined as weak CAPE: CAPE < 500 J kg^{-1}, moderate CAPE: 500 J kg^{-1} ≤ CAPE ≤ 2000 J kg^{-1} and strong CAPE: CAPE > 2000 J kg^{-1}. Right panel shows a density plot for CAPE vs. 700-hPa omega (1998-2013).

to strong CAPE being consumed by previous storms, and therefore only weak CAPE is observed at the time of lightning occurrence. Figure 12b shows the relationship between

CAPE and 700-hPa omega for lightning occurrences supporting the fact that low-to-moderate CAPE is associated with large rising motion in the lower troposphere.

3.1.3. Extreme Lightning

Investigating flash rates greater than a certain threshold, or extreme lightning, is beneficial because it can explain the conditions for which the extreme events will occur. Extreme lightning is defined here as a flash rate greater than approximately 12 flashes day^{-1} km^{-2}, which is the 95th percentile of the TRMM LIS climatology (Figure 13). This threshold is based on approximately 631,000 lightning cases. It is hypothesized that the environmental relationships observed in Figures 4 through 8 will be strengthened when looking at the extreme lightning cases only.

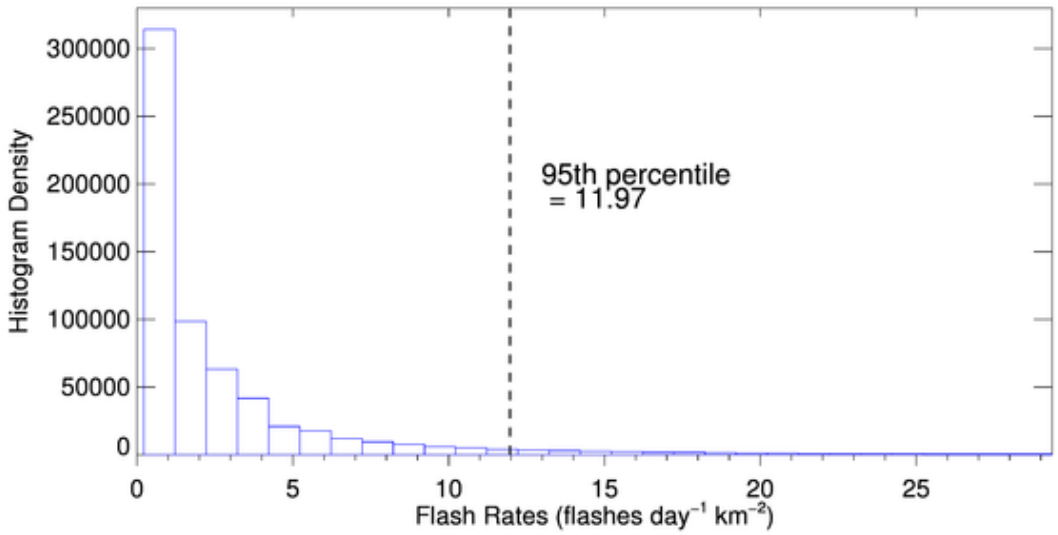

Figure 13. Distribution of lightning flash rates from TRMM LIS between 1998 and 2013. Dashed vertical line shows 95th percentile.

Figure 14 shows density plots for the same variable pairs as Figure 10 where extreme and non-extreme lightning cases are considered separately. The first column shows non-extreme lightning cases, the second column represents extreme lightning cases, and the third column shows the magnitude change between the first and second columns. Similar, but weakened relationships are shown in column two compared to column one. Surprisingly, the extreme lightning cases are associated with lower r when r is plotted against CAPE, nCAPE, omega, LS and DS compared to values associated with the non-extreme cases (Figure 14 a-e). However, extreme lightning is more present for high CAPE, nCAPE, LS and DS (Figure 14a, b, d and e). CAPE versus r shows the strongest relationship when considering extreme lightning only, but the pickup and density at high

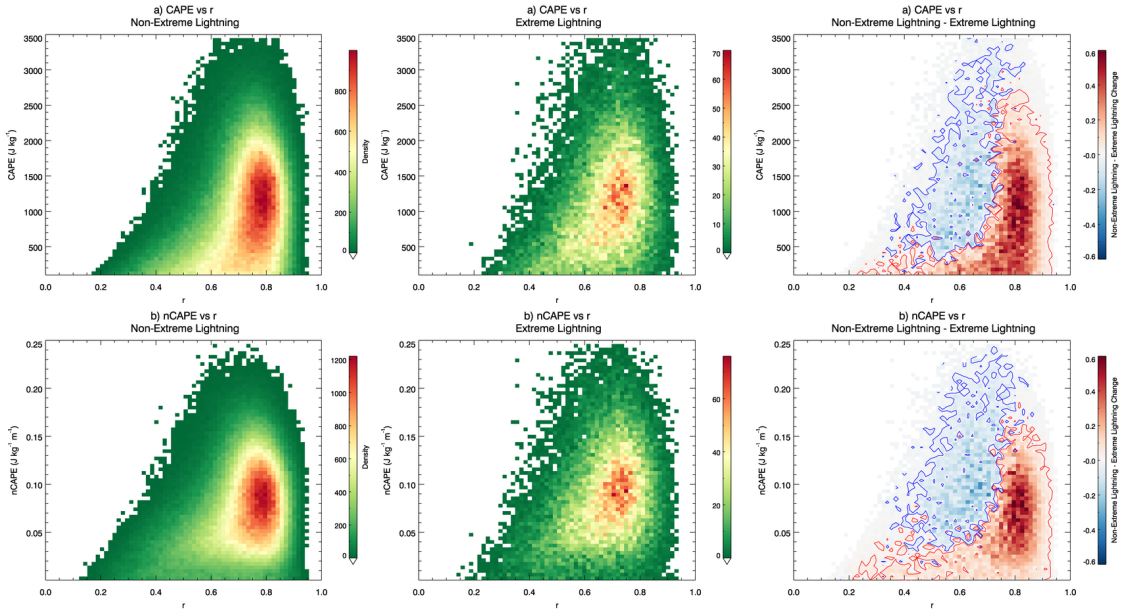

Figure 14. Two-variable extreme lightning density plots (1998-2013). The first column shows non-extreme lightning, the second column shows extreme lightning (values above the 95th percentile) and third column is the normalized difference between the two. Values of -0.03 and 0.03 are shown in the blue and red contours.

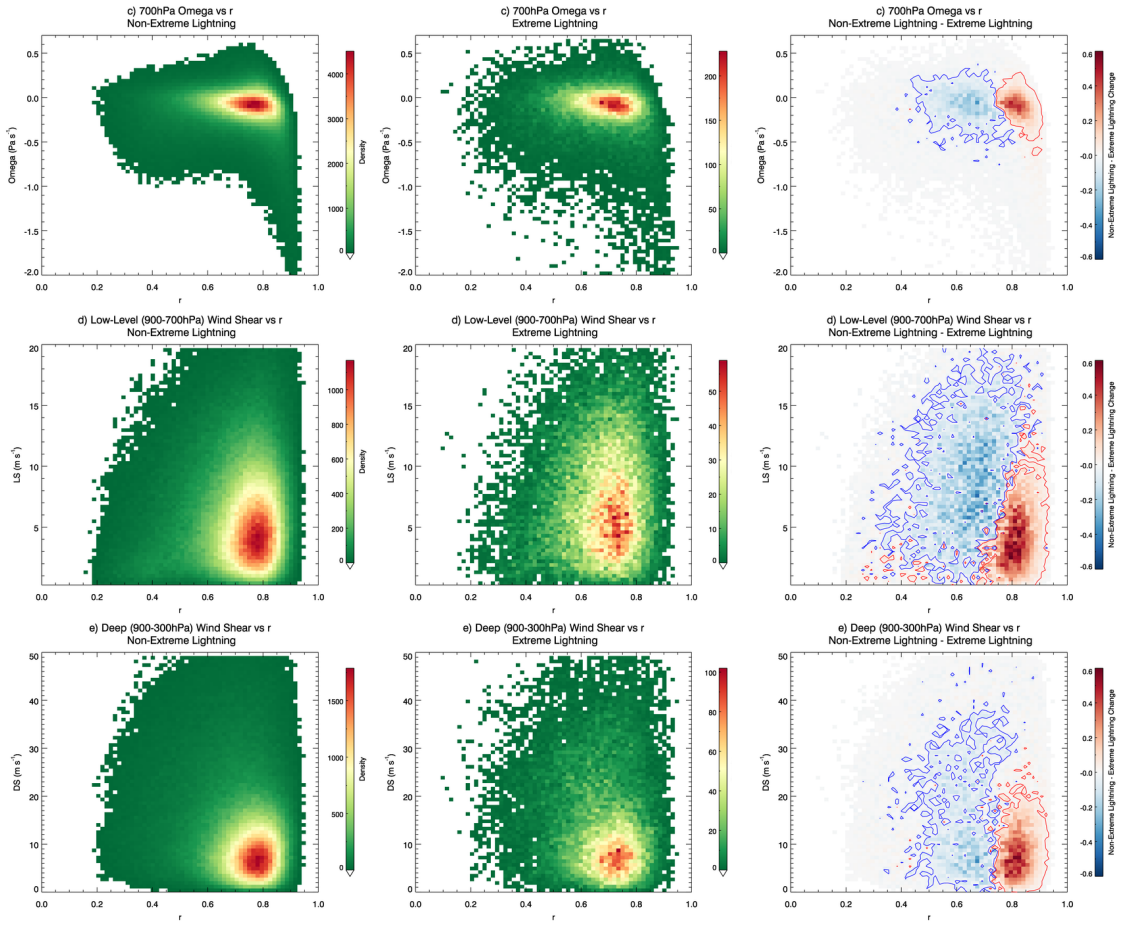

Figure 14. Continued

r/moderate CAPE isn't as strong as compared to non-extreme cases (Figure 14a). The omega/r relationship discussed in section 3.1.2 holds for extreme lightning but isn't stronger (Figure 14c). Overall, the environments for which extreme lightning occur appear to be relatively similar to non-extreme lightning conditions.

3.2. Statistical Models and Lightning Parameterizations

3.2.1. Generalized Linear Models Results

Following the methods presented in section 2.2, the results presented previously are quantified using a GLM. Table 1 shows the results obtained using the GLM given by equation (6) for the MERRA-2 data after the p_c threshold has been applied. The four cases discussed in section 2.2 are presented: (a) land and ocean (i.e., all), (b) land-only, (c) ocean-only, and (d) rare events (ReLogit). The GLM outputs "1" if lightning is predicted to occur in the specified bin and "0" if it is not. Furthermore, if TRMM LIS observes a lightning event within a bin, a "1" is recorded, and again, a "0" if a flash does not occur. Therefore, there are four possible combinations of output: (1) flash observed and predicted, (2) flash observed and not predicted, (3) flash not observed and not predicted, and (4) flash not observed and predicted. The proportion of the number of pixels that fall into each category is given in Table 1.

Table 1. Logisitic regression contingency table showing comparisons of observed and predicted lightning occurrence for TRMM LIS.

Model	p_c	Prediction	obs=0	obs=1
(a) All	4.0E-3	pred=0	0.819	0.001
		pred=1	0.176	0.004
(b) Land-Only	1.0E-2	pred=0	0.638	0.001
		pred=1	0.357	0.004
(c) Ocean-Only	1.7E-3	pred=0	0.758	0.002
		pred=1	0.237	0.003
(d) ReLogit	4.0E-3	pred=0	0.819	0.001
		pred=1	0.176	0.004

In general, the model performs well across experiments (a) to (d) when predicting which regions will experience no lightning. The best performers were (a) and (d), where the absence of lightning was predicted correctly 82% of the time. The worst performer was the land-only experiment (b), where the absence of lightning was only predicted correctly 64% of the time. Models (a), (b) and (d) predicted lightning correctly 80% of the time, and model (c) was the worst predictor of lightning occurrence, only predicting correctly 60% of the time. By these findings, models (a) and (d) are the best predictors of both the absence of lightning and lightning occurrences. However, the Relogit model does not improve on model (a), and therefore is removed for further analysis.

Table 2. Logistic regression estimates for (a) all events, (b) land-only events and (c) ocean-only events. P-values: " " = 0.1, "." = 0.05, * = 0.01, ** = 0.001, *** = 0.

	(a) All	(b) Land-Only	(c) Ocean-Only
CAPE	-1.350***	-1.707***	-1.180***
nCAPE	1.945***	2.183***	2.029***
Omega	-1.350***	-0.147***	-0.108***
r	0.630***	0.573***	0.925***
LS	0.060***	0.027***	0.172***
DS	-0.083***	-0.153***	0.008
LI	2.320***	NA	NA

In order to evaluate the significance and relative importance of each large-scale environmental variable for the lightning predictions above, the coefficients (β_i) in equation (6) for each variable are presented in Table 2, where positive (negative) coefficients show positive (negative) correlations. In addition to evaluating the large-scale

variables, a land indicator (LI) variable was also added to test the significance of the land-ocean contrasts shown in section 3.1. Experiment (a) shows positive correlations of nCAPE, r, LS and LI to lightning occurrence, and negative correlations for CAPE, omega and DS. A positive correlation means that as the variable increases, flash occurrence also increases. This is certainly reflected for r in Figure 6d, as flashes increase for high r. A negative correlation means that as the variable increases, flash occurrence decreases. CAPE and DS show strong evidence of this in Figure 6a and f, where lightning occurs for low values of CAPE and DS. A negative correlation is given for omega, but this is expected because it shows that as omega decreases, indicating rising motion, flash occurrences increase. We will therefore also consider this a positive correlation.

The land-only and ocean-only models show some contrasting results, particularly for DS where it is negatively correlated for (b) but positively correlated for (c). This could be attributed to more lightning occurring for only primarily small DS over land, compared to the ocean where more spread is observed (Figure 7f). For (c) all variables are positively correlated apart from CAPE. This suggests similar findings of those found in section 3.1, where for lightning to occur over the ocean, certain thresholds of each variable must be obtained. Through all models, nCAPE, omega, r and LS are positively correlated with lightning and CAPE is always negatively correlated. Also, for experiments (a) through (c), most variables have regression estimates that are statistically significant, thus making these reasonable predictors of lightning. The only variable that was not statistically significant was DS in the ocean-only model, so it will not be used in further analysis.

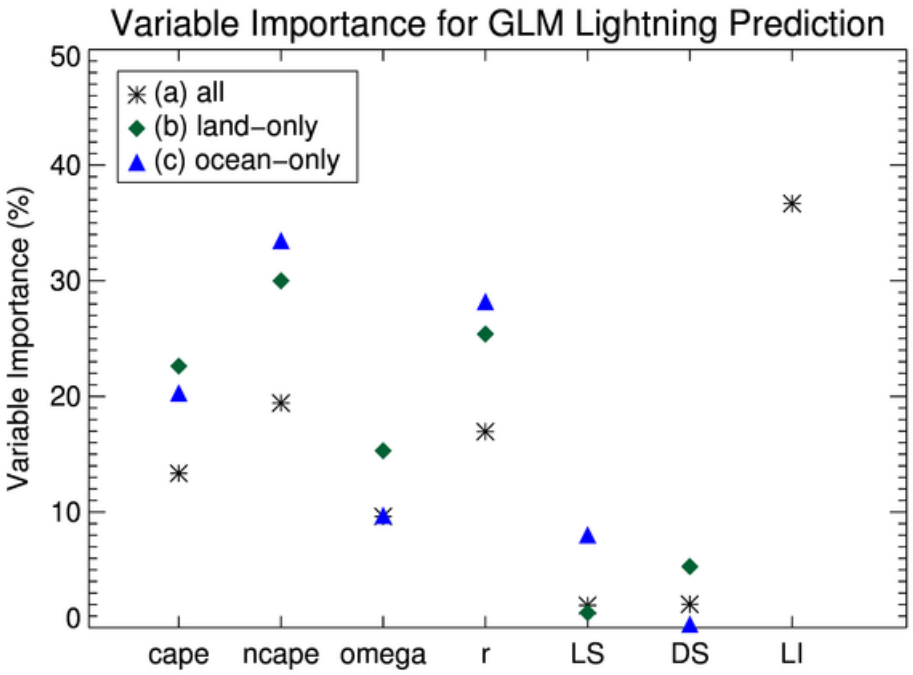

Figure 15. Relative importance of MERRA-2 variables in the GLM. Models (a) all (black stars), (b) land-only cases (green diamonds) and (c) ocean-only cases (blue triangles) are shown.

Figure 15 shows the relative importance of each variable for each model. For (a), the most important variables are nCAPE and r, with a heavy reliance on the land-ocean distinction variable. CAPE, nCAPE and r are the most important for the land-only model and similar results are shown for the ocean-only model, except that CAPE shows lesser importance compared to nCAPE and r. This is expected considering land-ocean contrasts and that CAPE does not distinguish between land and ocean (Williams et al. 2002). Omega, LS and DS all show little importance for each model, most likely due to lightning occurring at near-zero values of these variables.

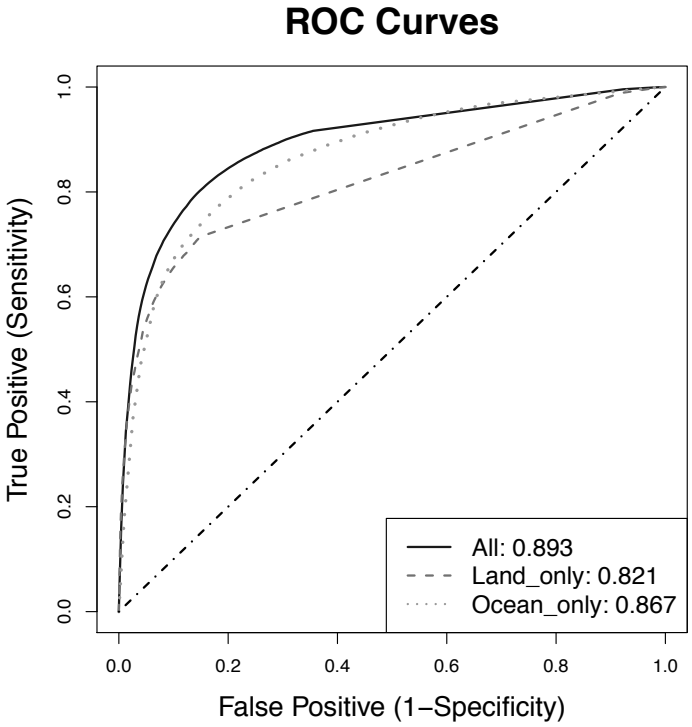

Figure 16. ROC curves for (a) all regression model, (b) land-only model and (c) ocean-only model. AUC denoted in legend.

To quantify the best performer of models (a)-(c), Receiver Operating Characteristic (ROC) curves are created and shown in Figure 16. From these plots, the Area Under the Curve (AUC) can be calculated, which is a measure of the performance of the binary classification (i.e., a result of lightning being predicted or not predicted). A larger AUC indicates better performance. For this study, the land-only AUC is the lowest, indicating the worst performer of the binary classification between the three cases, and the all case is the best performer of the three models because of its large AUC. As a result, in creating the lightning parameterization, model (a) will be used.

3.2.2. Log-Gaussian Cox Process Model Results

Next, a point-process model is used to not only be able to predict lightning occurrence, but lightning intensity as well. The LGCP model used here does not handle large regions well, so five regions of interest were chosen (Figure 17). These regions were chosen because of their varying lightning intensity and mean large-scale environments (Figure 5). For example, Argentina was chosen because of its intense lightning production for high nCAPE, rising motion, and LS but relatively low r and moderate CAPE (Figure 5 and Figure 9). Moreover, the Maritime Continent was chosen for its relatively high CAPE, high nCAPE, negative omega, and high r, but minimal lightning production (Figure 5 and Figure 9).

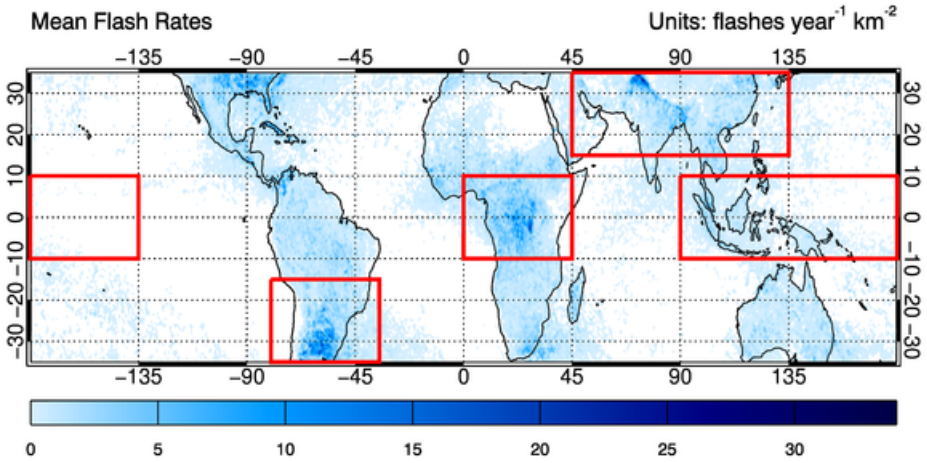

Figure 17. Mean flash rates for 2003 overlaid with LGCP regions of interest.

The computations for the spatial LGCP fittings are completed using R (www.r-project.org) package lgcp (Taylor et al. 2013). More specifically, for each region, the logarithm of the intensity process is assumed to follow the Gaussian process with mean

zero and the exponential covariance function (see section 2.2 for more detail). Parametric estimators are obtained by using the Metropolis-Hastings algorithm (Hastings 1970) with 50,000 repetitions. Table 3 reports the median of each estimator after 10% of the burn-in period.

Table 3. LGCP coefficient estimates for each region (displayed in Figure 17). Bolded values represent significant coefficients within a 95% credible interval.

	Argentina	Congo	Maritime Continent	Southeast Asia	Pacific Ocean
σ	1.853	1.971	1.671	2.092	1.333
ϕ (°)	5.841	6.496	5.658	6.546	1.989
$exp(\gamma_{CAPE})$	0.801	1.082	0.508	0.011	0.243
$exp(\gamma_{nCAPE})$	1.326	1.017	**2.100**	0.971	6.707
$exp(\gamma_{omega})$	0.971	1.112	**1.012**	1.028	1.198
$exp(\gamma_{r})$	1.134	0.965	1.205	1.007	0.974
$exp(\gamma_{LS})$	1.226	**0.821**	0.907	1.002	1.575
$exp(\gamma_{DS})$	1.295	**1.825**	0.734	0.914	**0.283**
$exp(\gamma_{LI})$	1.614	**2.002**	**2.025**	**1.924**	N/A

Table 3 presents the LGCP coefficients for each variable along with the standard deviation (σ) and the scale parameter (ϕ) for each of the five regions. A large scale parameter indicates more spatial dependence (i.e., more clustering) and a large standard deviation indicates large variability in the data. The largest standard deviation and scale parameter are in Southeast Asia, while the smallest are in the Pacific Ocean. This indicates more spatial dependence and variability in the data over Southeast Asia (i.e., where lightning flashes are present, they are clustered, but in varying degrees of intensity) and

more spread but more agreement between the data for the Pacific Ocean (i.e., where flashes are present, there may be spread but the intensities are similar through the region). Estimates in Table 3 greater than one show positive correlations, estimates less than one show negative correlations and bolded numbers represent significant coefficients with a 95% credible interval. In Argentina, all variables show positive correlations apart from CAPE. In the Congo region of Africa, omega, r and LS show negative correlations with significant coefficients for LS, DS and LI, while over the Maritime Continent, CAPE, omega, LS and DS all show negative correlations, with significant coefficients for nCAPE, omega and LI. Southeast Asia only shows positive correlations with r, LS and LI, but because the coefficients for r and LS are extremely close to one, they may be considered insignificant. These weak relationships may be caused due to the high elevation of the Himalayas driving convection rather than that large-scale environment. The Pacific Ocean only shows positive correlations with nCAPE and LS, but significant negative correlation is present for DS. Overall, only nCAPE and LI maintain positive correlations through most of the five regions and concludes that strong regional contrasts in relationships between lightning and the large-scale environment exist.

3.2.3. Creating and Evaluating a Lightning Parameterization

Next, a lightning parameterization is created using the best performing statistical model discussed in section 3.2.1, model (a). Figure 18 gives a comparison of observed lightning by TRMM LIS (top) with the predicted probabilities from the GLM (equation (6)) of lightning occurrence using MERRA-2 (middle) and CAM5 data (bottom); all

panels are averaged over the year 2003. The probabilities were scaled to match the flash rate values for panel (a). For complete data, areas in which a variable was missing in the MERRA-2 data (i.e., 700-hPa omega, LS and DS are not available at higher elevations) were set to zero and therefore in certain areas, like the Himalayas, lightning may be underpredicted. The CAM5 data handles extrapolates over these regions and therefore could result in incorrect predictions.

Predicted lightning occurrences using both MERRA-2 and CAM5 data generally agree well with the observed events, particularly over Africa, Australia and over the oceans. The GLM over predicts lightning occurrence in the Amazon and over the Maritime Continent, but under predicts in places like Argentina. Over predictions can be caused by a single variable; for example, where there are high omega values over Papua New Guinea (Figure 5), there are anomalously high predictions of lightning occurrence (Figure 18). The CAM5 results presented in Figure 18 are comparable with the MERRA-2 results. This is motivation for using lightning parameterizations in GCMs, and how using just a simple model, such as the one used here, can do a fairly good job for lightning prediction.

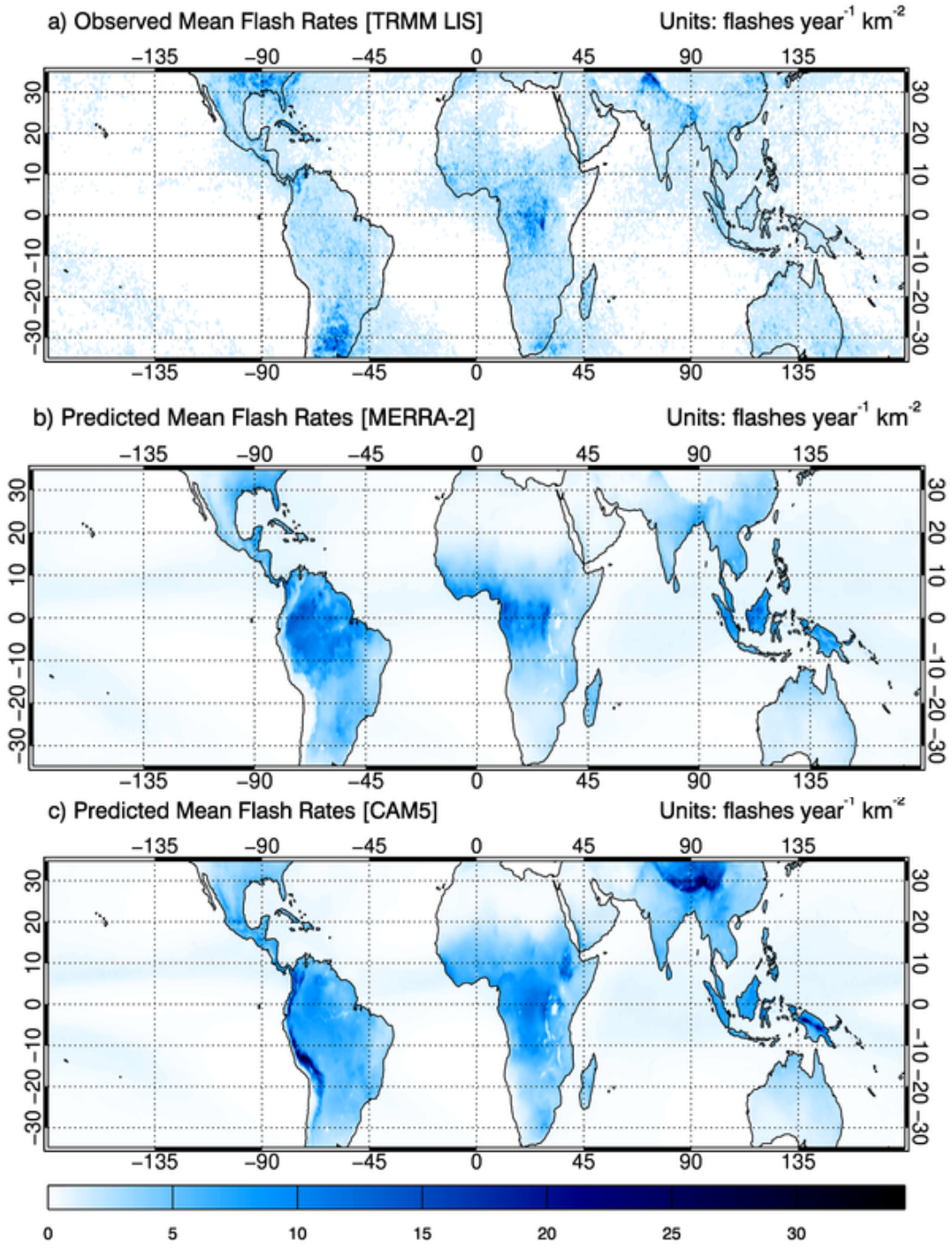

Figure 18. Mean predicted lightning from GLM parameterization. Mean flash rates for 2003 from TRMM LIS (top), predicted lightning from MERRA-2 variables using GLM (middle) and predicted lightning from CAM5 variables using GLM (bottom). The middle and bottom panel are scaled to match the values and color bar for the first panel.

4. CONCLUSIONS

Lightning data obtained from TRMM LIS from 1998 to 2013 is compared with six large-scale environmental variables derived from 3-hourly MERRA-2 reanalysis data to evaluate their relationships. These six variables include CAPE, nCAPE, r, 700-hPa omega, LS from 900-700 hPa and DS from 900-300 hPa. The data is binned at 0.5° x 0.5° resolution and matched if lightning occurred within 30 minutes of the MERRA-2 data.

Distributions of atmospheric variables for lightning and non-lightning environments show clear distinctions, particularly for CAPE, nCAPE and r (Figure 4). Density plots between each variable and observed lightning flash rates show that the greatest lightning density is associated with low-to-moderate CAPE, moderate nCAPE, slightly negative 700-hPa omega (i.e., rising motion), high r, low-to-moderate LS and low DS, although there is large spread for all variables (Figure 6). Clear geographical distinctions for flash occurrences exist between land and ocean for CAPE, nCAPE and r (Figure 7) and between tropical and sub-tropical areas for CAPE, nCAPE, r and DS (Figure 8). Figure 9 supports this result and shows that variable thresholds may differ from region to region in order for lightning to occur.

The relationship of r with other variables for lightning occurrences is evaluated and it is shown that CAPE and omega with r have the clearest relationships. As r increases, CAPE increases from low to moderate values and omega displays more rising motion (Figure 10). Figure 11 expands on the land-ocean contrasts observed in Figure 7, particularly over land where pickups occur at increasing r/low-to-moderate CAPE, increasing r/low-to-moderate nCAPE and decreasing r/low-to-moderate DS.

Relationships with the large-scale variables do not exhibit much change when examining extreme lightning cases (Figure 14).

A GLM is constructed using all data, land-only data and ocean-only data and nCAPE and r are found to be of utmost importance for lightning prediction. A LGCP model is also compiled and five regions are examined for lightning production. Overall, strong regional contrasts are observed and, with the exception of nCAPE and a land distinction term that should be used in all predictions, different variables and thresholds may need to be considered for differing regions. These results support the conclusion of Stolz et al. (2017) that the shear parameters are of secondary importance, particularly when compared to nCAPE, however this study refutes their conclusion that RH (i.e., r) is of secondary importance for lightning production. The GLM is used as the basis of a lightning parameterization to predict lightning from MERRA-2 and CAM5 variables. Predicted lightning from both datasets generally agrees with satellite observations, which is motivation for implementing a lightning parameterization in GCMs.

Evaluating the results found here with other lightning and environmental datasets may be useful, particularly the new additions of the International Space Station's LIS and the Geostationary Lightning Mapper (GLM). Strong latitudinal and elevation dependence is observed with the occurrence of lightning, so in future work, adding a latitude and elevation term may be warranted in statistical models. For comparison and evaluation, using the LGCP model in a parametrization may be useful, especially in showing lightning predictions on regional scales. It would also be useful to evaluate the parameterization in other models to see if the performance is consistent or worsens.

REFERENCES

Ahmed, F., and C. Schumacher, 2015: Convective and stratiform components of the precipitation-moisture relationship: PRECIPITATION AND COLUMN MOISTURE. *Geophysical Research Letters*, **42**, 10,453-10,462, doi:10.1002/2015GL066957.

——, and ——, 2017: Geographical differences in the tropical precipitation-moisture relationship and rain intensity onset: REGIONALITY AND INTENSITY IN P-*R* CURVES. *Geophysical Research Letters*, **44**, 1114–1122, doi:10.1002/2016GL071980.

Albrecht, R. I., S. J. Goodman, D. E. Buechler, R. J. Blakeslee, and H. J. Christian, 2016: Where Are the Lightning Hotspots on Earth? *Bulletin of the American Meteorological Society*, **97**, 2051–2068, doi:10.1175/BAMS-D-14-00193.1.

Bony, S., J. Dufresne, H. Treut, J. Morcette, and C. Senior, 2004: On dynamic and thermodynamic components of cloud changes.

Bretherton, C. S., M. E. Peters, and L. E. Back, 2004: Relationships between Water Vapor Path and Precipitation over the Tropical Oceans. *Journal of Climate*, **17**, 1517–1528, doi:10.1175/1520-0442(2004)017<1517:RBWVPA>2.0.CO;2.

Christian, H. J., and Coauthors, 2003: Global frequency and distribution of lightning as observed from space by the Optical Transient Detector. *Journal of Geophysical Research*, **108**, doi:10.1029/2002JD002347. http://doi.wiley.com/10.1029/2002JD002347 (Accessed July 17, 2018).

Churchill, D. D., and R. A. Houze, 1984: Mesoscale Updraft Magnitude and Cloud-Ice Content Deduced from the Ice Budget of the Stratiform Region of a Tropical Cloud Cluster. *Journal of the Atmospheric Sciences*, **41**, 1717–1725, doi:10.1175/1520-0469(1984)041<1717:MUMACI>2.0.CO;2.

Cox, D. R., 1955: Some Statistical Methods Connected with Series of Events. *Journal of the Royal Statistical Society*, **17**, 129–164.

Curran, E. B., R. L. Holle, and R. E. López, 2000: Lightning Casualties and Damages in the United States from 1959 to 1994. *Journal of Climate*, **13**, 3448–3464, doi:10.1175/1520-0442(2000)013<3448:LCADIT>2.0.CO;2.

Flannigan, M., B. Stocks, M. Turetsky, and M. Wotton, 2009: Impacts of climate change on fire activity and fire management in the circumboreal forest. *Global Change Biology*, **15**, 549–560, doi:10.1111/j.1365-2486.2008.01660.x.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, **30**, 5419–5454, doi:10.1175/JCLI-D-16-0758.1.

Hastings, W. K., 1970: Monte Carlo sampling methods using Markov chains and their applications. *Biometrika Trust*, **57**, 97–109.

King, G., and L. Zeng, 2003: Logistic Regression in Rare Events Data. *Journal of Statistical Software*, **8**, doi:10.18637/jss.v008.i02. http://www.jstatsoft.org/v08/i02/ (Accessed August 23, 2018).

Kummerow, C., C. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The Tropical Rainfall Measuring Mission (TRMM) Sensor Package.

Li, W., and C. Schumacher, 2011: Thick Anvils as Viewed by the TRMM Precipitation Radar. *Journal of Climate*, **24**, 1718–1735, doi:10.1175/2010JCLI3793.1.

MacKay, D. J. C., 2003: *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 612 pp.

Magi, B. I., 2015: Global Lightning Parameterization from CMIP5 Climate Model Output. *Journal of Atmospheric and Oceanic Technology*, **32**, 434–452, doi:10.1175/JTECH-D-13-00261.1.

Matérn, B., 1986: *Spatial Variation*. Springer New York, New York, NY, http://link.springer.com/10.1007/978-1-4615-7892-5 (Accessed August 29, 2018).

McCullagh, P., and J. A. Nelder, 1989: *Generalized Linear Models*. Second. Chapman and Hall,.

Molinari, J., P. K. Moore, V. P. Idone, R. W. Henderson, and A. B. Saljoughy, 1994: Cloud-to-ground lightning in Hurricane Andrew. *Journal of Geophysical Research*, **99**, 16665, doi:10.1029/94JD00722.

Moller, J., A. R. Syversveen, and R. P. Waagepetersen, 1998: Log Gaussian Cox Processes. *Scandinavian Journal of Statistics*, **25**, 451–482, doi:10.1111/1467-9469.00115.

Orville, R. E., and D. W. Spencer, 1979: Global Lightning Flash Frequency. *Monthly Weather Review*, **107**, 934–943, doi:10.1175/1520-0493(1979)107<0934:GLFF>2.0.CO;2.

——, and R. W. Henderson, 1986: Global Distribution of Midnight Lightning: September 1977 to August 1978. *Monthly Weather Review*, **114**, 2640–2653, doi:10.1175/1520-0493(1986)114<2640:GDOMLS>2.0.CO;2.

Reynolds, S. E., M. Brook, and M. F. Gourley, 1957: THUNDERSTORM CHARGE SEPARATION. *Journal of Meteorology*, **14**, 426–436, doi:10.1175/1520-0469(1957)014<0426:TCS>2.0.CO;2.

Rodger, C. J., J. B. Brundell, and R. L. Dowden, 2005: Location accuracy of VLF World-Wide Lightning Location (WWLL) network: Post-algorithm upgrade. *Annales Geophysicae*, **23**, 277–290, doi:10.5194/angeo-23-277-2005.

Romps, D. M., J. T. Seeley, D. Vollaro, and J. Molinari, 2014: Projected increase in lightning strikes in the United States due to global warming. *Science*, **346**, 851–854, doi:10.1126/science.1259100.

Rotunno, R., J. B. Klemp, and M. L. Weisman, 1988: A Theory for Strong, Long-Lived Squall Lines. *Journal of the Atmospheric Sciences*, **45**, 463–485, doi:10.1175/1520-0469(1988)045<0463:ATFSLL>2.0.CO;2.

Rutledge, S. A., E. R. Williams, and T. D. Keenan, 1992: The Down Upper Doppler and Electricity Experiment (DUNDEE): Overview and Preliminary Results. *Bulletin of the American Meteorological Society*, **73**, 3–16, doi:10.1175/1520-0477(1992)073<0003:TDUDAE>2.0.CO;2.

Samsury, C. E., and R. E. Orville, 1994: Cloud-to-Ground Lightning in Tropical Cyclones: A Study of Hurricanes Hugo (1989) and Jerry (1989). *Monthly*

Weather Review, **122**, 1887–1896, doi:10.1175/1520-0493(1994)122<1887:CTGLIT>2.0.CO;2.

Saunders, C. P. R., W. D. Keith, and R. P. Mitzeva, 1991: The effect of liquid water on thunderstorm charging. *Journal of Geophysical Research*, **96**, 11007, doi:10.1029/91JD00970.

Schumann, U., and H. Huntrieser, 2007: The global lightning-induced nitrogen oxides source.

Stensrud, D. J., 1996: Importance of Low-Level Jets to Climate: A Review. *Journal of Climate*, **9**, 1698–1711, doi:10.1175/1520-0442(1996)009<1698:IOLLJT>2.0.CO;2.

Stolz, D. C., S. A. Rutledge, and J. R. Pierce, 2015: Simultaneous influences of thermodynamics and aerosols on deep convection and lightning in the tropics: THERMODYNAMICS, AEROSOLS, AND CONVECTION. *Journal of Geophysical Research: Atmospheres*, **120**, 6207–6231, doi:10.1002/2014JD023033.

——, ——, ——, and S. C. van den Heever, 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, doi:10.1002/2016JD026220.

Takahashi, T., 1978: Riming Electrification as a Charge Generation Mechanism in Thunderstorms. *Journal of the Atmospheric Sciences*, **35**, 1536–1548, doi:10.1175/1520-0469(1978)035<1536:REAACG>2.0.CO;2.

Taylor, B. M., T. M. Davies, B. S. Rowlingson, and P. J. Diggle, 2013: lgcp: An R Package for Inference with Spatial and Spatio-Temporal Log-Gaussian Cox Processes. *Journal of Statistical Software*, **52**, doi:10.18637/jss.v052.i04. http://www.jstatsoft.org/v52/i04/ (Accessed August 30, 2018).

Virts, K. S., J. M. Wallace, M. L. Hutchins, and R. H. Holzworth, 2013: Highlights of a New Ground-Based, Hourly Global Lightning Climatology. *Bulletin of the American Meteorological Society*, **94**, 1381–1391, doi:10.1175/BAMS-D-12-00082.1.

Williams, E., and S. Stanfill, 2002: The physical origin of the land–ocean contrast in lightning activity. *Comptes Rendus Physique*, **3**, 1277–1292, doi:10.1016/S1631-0705(02)01407-X.

Williams, E., and Coauthors, 2002: Contrasting convective regimes over the Amazon: Implications for cloud electrification. *Journal of Geophysical Research*, **107**, doi:10.1029/2001JD000380. http://doi.wiley.com/10.1029/2001JD000380 (Accessed November 21, 2017).

Williams, E. R., S. G. Geotis, N. Renno, S. A. Rutledge, E. Rasmussen, and T. Rickenbach, 1992: A Radar and Electrical Study of Tropical "Hot Towers." *Journal of the Atmospheric Sciences*, **49**, 1386–1395, doi:10.1175/1520-0469(1992)049<1386:ARAESO>2.0.CO;2.

Zipser, E. J., and K. R. Lutz, 1994: The Vertical Profile of Radar Reflectivity of Convective Cells: A Strong Indicator of Storm Intensity and Lightning Probability? *Monthly Weather Review*, **122**, 1751–1759, doi:10.1175/1520-0493(1994)122<1751:TVPORR>2.0.CO;2.

Zipser, E. J., C. Liu, D. J. Cecil, S. W. Nesbitt, and D. P. Yorty, 2006: WHERE ARE THE MOST INTENSE THUNDERSTORMS ON EARTH? *Bulletin of the American Meteorological Society*, **87**, 1057–1071, doi:10.1175/BAMS-87-8-1057.