

Convection during the North American Monsoon across Central and Southern Arizona: Applications to Operational Meteorology

JARET W. ROGERS AND ARIEL E. COHEN

NOAA/NWS/NCEP/Storm Prediction Center, Norman, Oklahoma

LEE B. CARLAW

National Weather Service Forecast Office, Tucson, Arizona

(Manuscript received 5 August 2015, in final form 15 July 2016)

ABSTRACT

This comprehensive analysis of convective environments associated with thunderstorms affecting portions of central and southern Arizona during the North American monsoon focuses on both observed soundings and mesoanalysis parameters relative to lightning flash counts and severe-thunderstorm reports. Analysis of observed sounding data from Phoenix and Tucson, Arizona, highlights several moisture and instability parameters exhibiting moderate correlations with 24-h, domain-total lightning and severe thunderstorm counts, with accompanying plots of the precipitable water, surface-based lifted index, and 0–3-km layer mixing ratio highlighting the relationship to the domain-total lightning count. Statistical techniques, including stepwise, multiple linear regression and logistic regression, are applied to sounding and gridded mesoanalysis data to predict the domain-total lightning count and individual gridbox 3-h-long lightning probability, respectively. Applications of these forecast models to an independent dataset from 2013 suggest some utility in probabilistic lightning forecasts from the regression analyses. Implementation of this technique into an operational forecast setting to supplement short-term lightning forecast guidance is discussed and demonstrated. Severe-thunderstorm-report predictive models are found to be less skillful, which may partially be due to substantial population biases noted in storm reports over central and southern Arizona.

1. Introduction

This study of North American monsoon convective environments over central and southern Arizona builds on the information from [Carlaw et al. \(2017, hereafter CCR\)](#), which provides an overview of environments favorable for lightning and severe weather over the region. Observed soundings and gridded mesoanalysis data are examined to determine which commonly used variables exhibit the most predictive skill for lightning and severe thunderstorm reports across the domain (geographically and temporally illustrated by [CCR](#)).

Operational meteorologists often face substantial challenges in forecasting convective potential during the monsoon. The analysis of observed soundings remains an important component in assessing the

local environment and the potential for severe convective storms ([Craven and Brooks 2004](#); [Cohen et al. 2007](#)). Previous studies have established the relationship between low-level moisture from the Gulf of California and increased severe convective potential across Arizona ([Hales 1974](#); [Douglas et al. 2007](#); [CCR](#)). [Maddox et al. \(1995\)](#) offer an expanded overview of low-level moisture observations via soundings from Tucson, Arizona, and found that the 850-mb (1 mb = 1 hPa) dewpoint temperature could be used as a proxy for forecasting severe convective potential over central Arizona. However, a more comprehensive analysis describing the relationship of specific variables to convective potential still remains largely unexplored. [Section 2](#) of the present study will incorporate numerous sounding-based variables and examine their relationship with total lightning count and severe thunderstorm reports, using soundings from Phoenix (KPSR; hereafter, PHX) and Tucson (KTWC; hereafter, TUS), Arizona. Application of a

Corresponding author e-mail: Ariel Cohen, ariel.cohen@noaa.gov

TABLE 1. Correlation coefficients of variables observed in the 1200 UTC sounding to domain-total lightning count during the 1200–1159 UTC period. Values are sorted from highest to lowest in the Tucson lightning column.

Variable	Tucson lightning	Phoenix lightning	Tucson severe	Phoenix severe
PW	0.521	0.500	0.341	0.322
925-mb mixing ratio	0.494	0.432	0.311	0.291
0–3-km mixing ratio	0.492	0.469	0.328	0.317
0–1-km mixing ratio	0.479	0.444	0.325	0.298
700-mb mixing ratio	0.479	0.429	0.309	0.254
Surface–850-mb mixing ratio	0.478	0.450	0.324	0.307
SBCAPE	0.472	0.344	0.326	0.226
Surface dewpoint	0.470	0.443	0.308	0.292
925-mb dewpoint	0.467	0.423	0.291	0.286
850-mb mixing ratio	0.467	0.444	0.312	0.310
Wet-bulb zero height	0.460	0.458	0.298	0.280
MUCAPE	0.457	0.433	0.330	0.298
MLCAPE	0.456	0.404	0.338	0.269
850-mb dewpoint	0.433	0.435	0.286	0.300
500-mb mixing ratio	0.422	0.419	0.254	0.247
700-mb dewpoint	0.421	0.393	0.270	0.238
500-mb dewpoint	0.400	0.403	0.243	0.242
Surface relative humidity	0.375	0.285	0.244	0.184
250-mb geopotential height	0.358	0.322	0.228	0.206
300-mb geopotential height	0.357	0.320	0.227	0.203
500-mb geopotential height	0.316	0.285	0.210	0.189
700-mb geopotential height	0.300	0.265	0.189	0.166
300-mb temp	0.299	0.264	0.197	0.165
Surface temp	0.267	0.313	0.177	0.207
0–3 km AGL lapse rate	0.255	0.242	0.157	0.136
850-mb geopotential height	0.241	0.187	0.142	0.106
250-mb temp	0.238	0.192	0.154	0.129
925-mb temp	0.196	0.054	0.131	0.023
700-mb temp	0.147	0.193	0.115	0.154
500-mb temp	0.122	0.073	0.081	0.027
Melting level	0.100	0.121	0.077	0.091
Surface wind (<i>v</i> component)	0.090	0.067	0.054	0.041
0–3 km AGL MLCAPE	0.086	–0.005	0.041	–0.015
700–500-mb lapse rate	0.080	0.153	0.069	0.138
850-mb temp	0.077	0.084	0.049	0.044
850–500-mb lapse rate	0.048	0.076	0.032	0.056
MLCIN	–0.008	–0.127	–0.006	–0.169
300-mb wind (<i>v</i> component)	–0.009	0.013	–0.001	–0.009
250-mb wind (<i>v</i> component)	–0.015	0.034	–0.021	0.004
850-mb wind (<i>v</i> component)	–0.022	–0.023	–0.001	–0.014
Surface wind (<i>u</i> component)	–0.037	0.083	–0.010	0.061
3–6 km AGL lapse rate	–0.049	0.112	0.007	0.122
Downdraft CAPE	–0.061	–0.096	–0.013	–0.015
850-mb wind (<i>u</i> component)	–0.076	–0.072	–0.072	–0.057
0–1 km AGL storm relative helicity	–0.079	–0.071	–0.022	–0.028
SBCIN	–0.091	–0.174	–0.078	–0.178
MUCIN	–0.091	–0.110	–0.088	–0.100
0–3 km AGL storm relative helicity	–0.103	–0.094	–0.015	0
500-mb wind (<i>v</i>)	–0.108	–0.051	–0.053	–0.023
MLLFC	–0.137	–0.054	–0.070	–0.054
SBLFC	–0.157	–0.047	–0.074	–0.043
0–1 km AGL bulk shear	–0.170	–0.099	–0.083	–0.064
700-mb wind speed (<i>v</i> component)	–0.193	–0.246	–0.106	–0.147
0–3 km AGL bulk shear	–0.206	–0.230	–0.089	–0.110
MULFC	–0.214	–0.141	–0.150	–0.114
0–6 km AGL bulk shear	–0.252	–0.266	–0.126	–0.096
Convective temp	–0.255	–0.184	–0.151	–0.108

TABLE 1. (Continued)

Variable	Tucson lightning	Phoenix lightning	Tucson severe	Phoenix severe
700-mb wind speed (u component)	−0.314	−0.221	−0.224	−0.164
500-mb wind speed (u component)	−0.349	−0.312	−0.244	−0.240
250-mb wind speed (u component)	−0.357	−0.323	−0.250	−0.248
300-mb wind speed (u component)	−0.359	−0.331	−0.246	−0.258
MULCL	−0.360	−0.321	−0.243	−0.201
SBLCL	−0.377	−0.300	−0.248	−0.197
MLLCL	−0.380	−0.350	−0.256	−0.242
MULI	−0.463	−0.462	−0.318	−0.319
MLLI	−0.486	−0.469	−0.325	−0.323
SBLI	−0.496	−0.473	−0.327	−0.316

stepwise, multiple linear regression technique also is explored to determine the predictability of 24-h lightning counts and severe reports based on 1200 UTC soundings.

Gridded RAP-based objective analysis environmental data [Storm Prediction Center (SPC) Surface Objective Analysis (SFCOA); Bothwell et al. (2002)], including several thermodynamic and kinematic variables relevant to convective forecasting, are examined in section 3 to determine relationships between the hourly magnitudes of numerous variables with 3-h-long lightning counts. Localized terrain effects can have a significant impact on convective initiation and eventual convective evolution across surrounding areas, as established by CCR. Usage of 3-h-long gridded SFCOA data will allow for the development of unique probabilistic predictive equations for lightning, which will inherently incorporate terrain effects on convection using a sufficiently high spatial resolution, without explicitly including topography. A stepwise logistic regression technique will be applied and include multiple potential predictors of lightning, with examples of 3-h-long probabilistic output shown. An independent dataset from 2013 will also be tested to determine the forecast skill demonstrated by the predictive equations, as well as an accompanying discussion of potential applications to operational short-term forecasting.

2. Observed sounding analysis

a. Data and methodology

The domain for this study includes most of central and southern Arizona, as described in CCR, and covers the entire metropolitan areas of both Tucson and Phoenix and the higher-terrain areas of southeast and east-central Arizona. The domain includes the entire County Warning Areas (CWAs) of the National Weather Service

(NWS) Tucson Weather Forecast Office and the Arizona portion of the NWS Phoenix Weather Forecast Office. The period of this study includes the years from 2003 through 2013, as a result of the availability of archived SFCOA data. The dates of 1 June–30 September of each year are used to capture the peak period of precipitation associated with the monsoon, as shown in seasonal trends in CCR.

Hourly cloud-to-ground (CG) lightning counts from National Lightning Detection Network data, along with hail, wind, and tornado reports from the National Centers for Environmental Information's *Storm Data* publication, are assigned to each 40-km grid. Upper-air soundings are released at 1200 and 0000 UTC at TUS year-round, while those at PHX are released from June through September. Historical observed soundings are collected from the SPC's internal sounding archive. Aside from typical mandatory level variables, numerous derived meteorological parameters are included in the sounding record, using schemes developed by Hart and Korotky (1991) for the Skew- T and Hodograph Analysis Research Program (SHARP) sounding analysis software employed at SPC. In the present study, sounding data are compared with total daily lightning and severe counts across the domain for the 24-h period following the release time of the sounding.

TABLE 2. Percentile ranks and corresponding 1200–1159 and 0000–1159 UTC CG domain-total lightning count ranges.

Percentile range	1200–1159 UTC lightning count	0000–1159 UTC lightning count
<10	<86	<35
10–25	86–821	35–333
25–50	822–3285	334–1402
50–75	3286–7793	1403–3628
75–90	7794–13 086	3629–6881
>90	>13 086	>6881

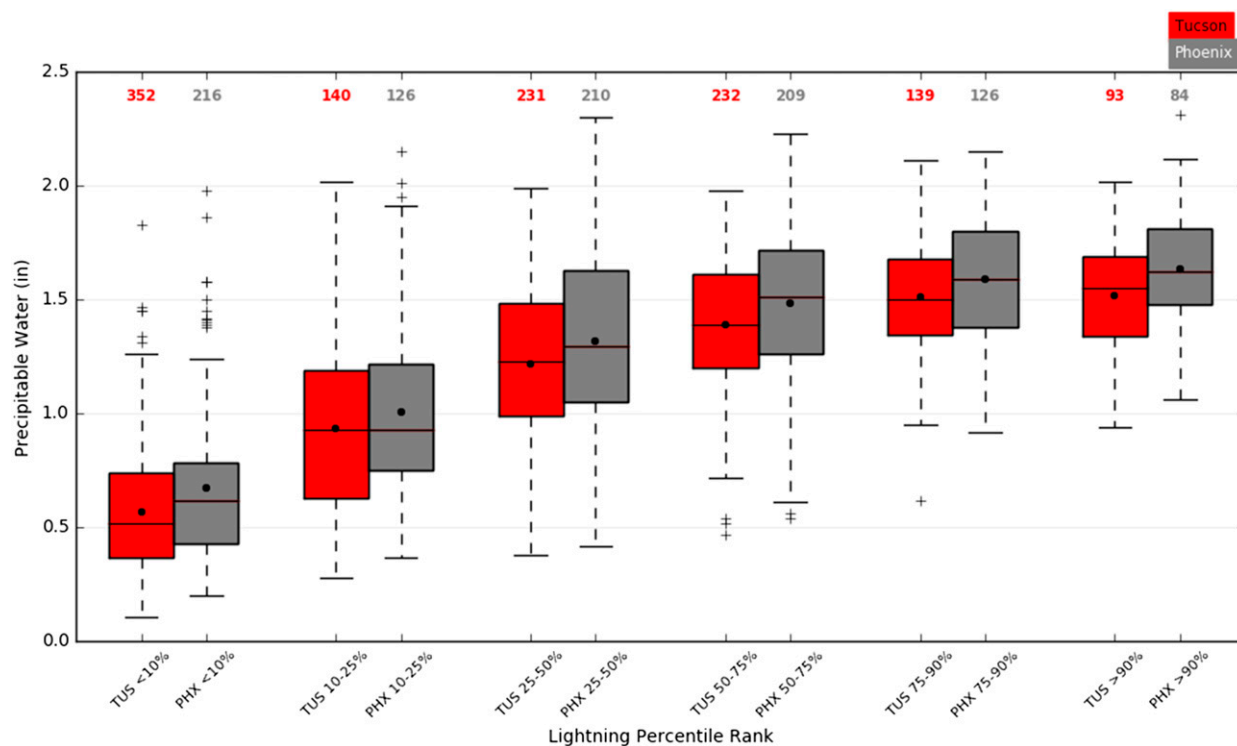


FIG. 1. Box-and-whisker plot of PW values (in.) at 1200 UTC for Tucson (red) and Phoenix (gray) compared to the domain-total lightning count for the 1200–1159 UTC period. Each distribution represents a specified range of lightning percentile ranks, as specified in Table 2. The black horizontal line segments represent the median value for each box. The black dot represents the mean value for each box. The numbers near the top of the plot represent the sample size for each box.

b. Domain-total lightning count correlation to sounding variables

Forecasts for lightning potential across central and southern Arizona often focus on moisture variables, such as precipitable water (PW) and low-level mixing ratios (e.g., Wallace et al. 1999). Examination of individual sounding-based variables compared with domain-total lightning count for the following 1200–1159 UTC period confirms that moisture-related variables have the highest correlation with lightning (Table 1) at both TUS and PHX. Most of the variables associated with low-level moisture exhibit correlation coefficients above 0.4, including layer-averaged and single-level mixing ratios and dewpoint variables. Instability variables, including most unstable (MU) and mixed layer (ML) convective available potential energy (CAPE), as well as wet-bulb zero height, are also shown to be moderately correlated with total lightning count. Lifted index (LI) values computed from surface-based (SB), ML, and MU parcels all exhibited negative correlations less than -0.4 for both TUS and PHX. The u components of the wind speed at 500, 300, and 250 mb exhibit weak negative correlations to total

lightning count, suggesting that a weaker westerly component or stronger easterly component of flow aloft may be more favorable for higher lightning counts. Temperatures at various levels of the troposphere generally exhibited little correlation to total lightning counts.

In general, sounding variables exhibit higher correlations to domain-total lightning count at TUS than PHX. This is likely due to the geographic location of the TUS sounding site within the domain, which is located in proximity to areas with higher climatological frequencies of convection, including the higher terrain east of Tucson. The eastern portion of the domain often experiences more thunderstorm activity than the western and central portions of the domain, as shown by CCR.

Observed distributions of selected variables exhibiting the highest correlations to domain-total lightning count are shown by the lightning count percentile rank (Table 2). The PW shows a nearly uniform increase in magnitude from the <10th percentile through the 75th percentile of domain-total lightning counts (Fig. 1). The rate of increase slows in the upper ranks of the domain-total lightning counts, suggesting that PW may show

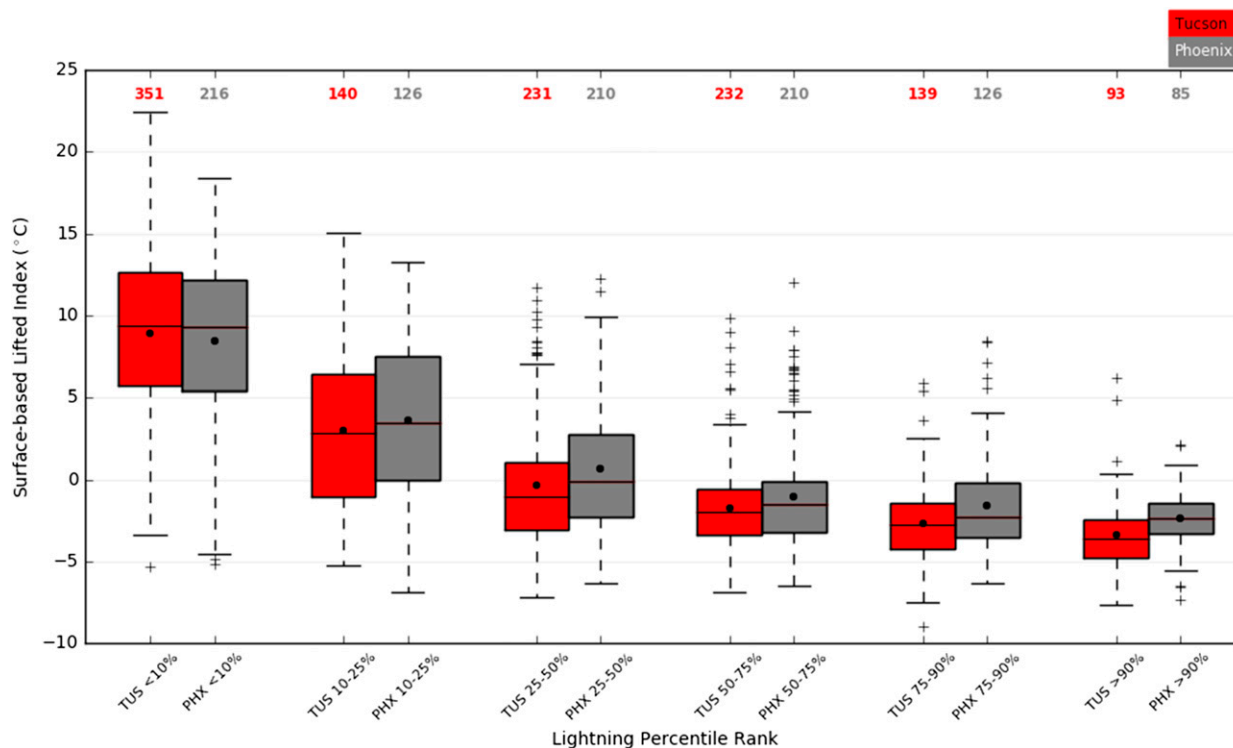


FIG. 2. As in Fig. 1, but for SBLI values ($^{\circ}\text{C}$) at 1200 UTC.

better discriminatory value in days with low-to-moderate lightning counts. Lightning days featuring lightning counts in the upper 50th percentile often feature an observed 1200 UTC PW value of 31.8 mm (1.25 in.) or greater at TUS, or 34.3 mm (1.35 in.) or greater at PHX. More prolific lightning days, above the 75th percentile, exhibit median and mean PW values at or above 38.1 mm (1.50 in.) and are almost exclusively above 25.4 mm (1 in.) in a majority of cases. It should be noted that high-PW values are occasionally associated with low lightning totals across the domain, as noted by the outlier values occurring within the less than 10th percentile of lightning counts in Fig. 1.

The SBLI, a measure of the instability present in the atmosphere (Galway 1956; DeRubertis 2006), exhibits a moderate negative correlation to the domain-total lightning count. Figure 2 shows a substantial decrease in SBLI values observed in the 1200 UTC soundings moving from the 10th percentile through the 50th percentile of the domain-total lightning counts. In most cases, SBLI values below 0°C are observed at 1200 UTC with domain-total lightning counts in the upper 50th percentile.

Low-level mixing ratios are often referenced for diagnosing the magnitude of low-level moisture present across central and southern Arizona (Maddox

et al. 1995). Figure 3 shows a similar trend as other moisture parameters, with days with fewer domain-total lightning counts (<10th percentile) exhibiting a median mixing ratio near 4 g kg^{-1} , while days with higher domain-total lightning counts (>50th percentile) are typically associated with median mixing ratios of 10 g kg^{-1} or greater.

Thunderstorm activity often continues well into the late evening and overnight across southern and central Arizona, with a peak across the greater Phoenix area observed after 0000 UTC (cf. Fig. 4 in CCR). Observed soundings from PHX and TUS at 0000 UTC may provide useful information to operational meteorologists in assessing the potential for nocturnal thunderstorms across the domain. The PW and SBLI parameter distributions observed in 0000 UTC soundings, relative to domain-total lightning counts during the 0000–1159 UTC period (Figs. 4 and 5), exhibit similar trends with increasing lightning counts, as shown in Figs. 1 and 2, albeit with slightly lower magnitudes at 0000 UTC. PW values less than 25.4 mm (1 in.) are rarely associated with prolific domain-total lightning counts, and the top 25 percentiles of domain-total lightning counts often exhibit PW values of 31.8 mm (1.25 in.) or higher, and SBLI values of -2°C or lower.

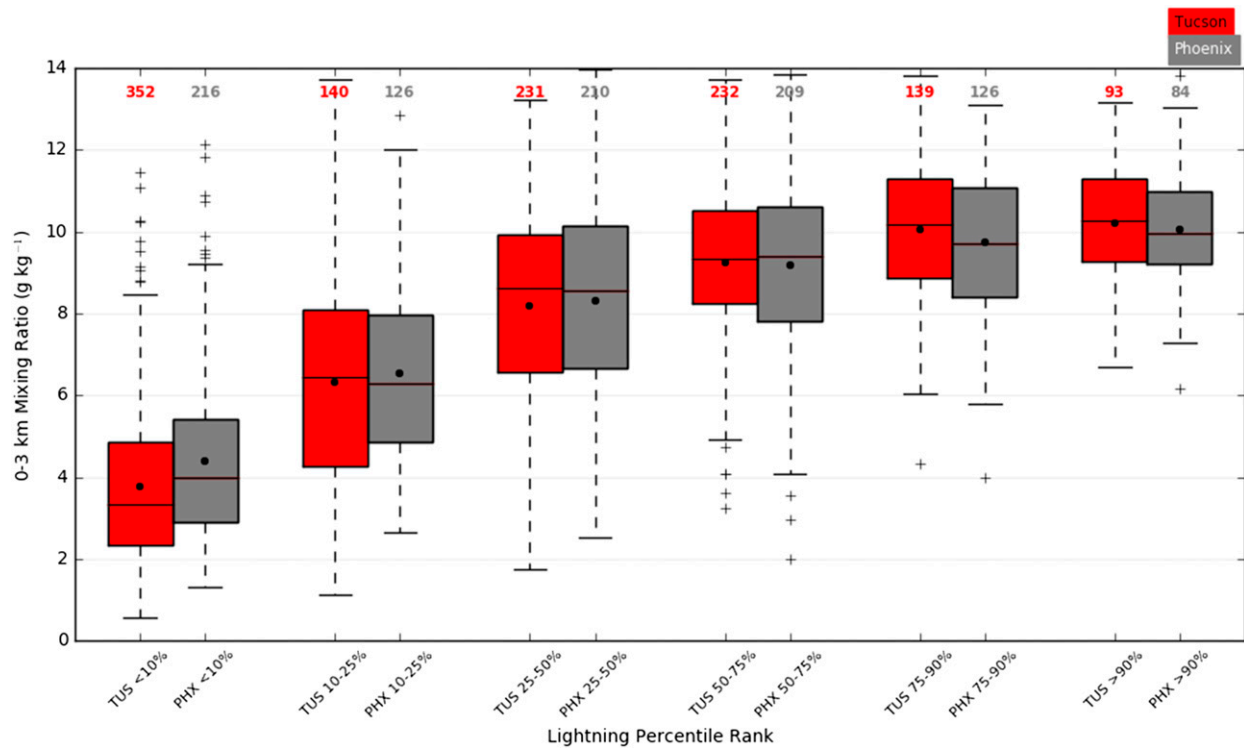


FIG. 3. As in Fig. 1, but for 0–3-km layer-mean mixing ratio values (g kg^{-1}) at 1200 UTC.

c. Domain-total severe reports count

Individual parameters from observed soundings generally exhibit a weaker correlation to the total severe report counts than to the lightning counts across the domain (Table 1). This may be partially explained by the high likelihood of an underrepresentation of actual severe events by the severe reports database (e.g., Shoemaker and Davis 2008; CCR), but also is attributed to the complex storm-scale processes that are not adequately captured by a single upper-air sounding (Rasmussen and Blanchard 1998; Thompson et al. 2003). The correlation of individual variables to domain-total severe reports is weaker than to total lightning counts, with PW the best-correlated variable at 0.341 at TUS and 0.322 at PHX at 1200 UTC. The rankings of individual variables are generally consistent with the total lightning count correlations, with low-level moisture and LI values showing relatively higher correlations than other types of variables for severe-thunderstorm reports.

Despite the generally low correlation coefficients that are observed, a notable difference in PW values occurs between days with no severe reports and one or greater severe reports observed (Fig. 6). Median values of PW associated with no severe reports but at least one lightning strike within the domain are near 30.5 mm (1.2 in.),

while the median PW value for one or more severe reports ranges from 35.6 to 40.6 mm (1.4–1.6 in.). Moisture parameters appear to be a better discriminator of the occurrence of severe reports, rather than a predictor of the number of total reports.

d. Multiple linear regression prediction model

Previous studies of atmospheric phenomena have used regression techniques to predict the magnitudes of specific variables drawn from meteorological datasets (e.g., Billet et al. 1997). This same technique can be applied to develop a predictive equation from the observed sounding dataset for the number of expected lightning strikes for the succeeding 24-h period. A multiple linear regression function assumes the form

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n, \quad (1)$$

where Y is the predictand, β_0 is a fitted constant, and β_1 through β_n are fitted coefficients of independent predictors x_1 through x_n (Montgomery et al. 2012). Potential predictors are assigned from only the 1200 UTC TUS sounding, owing to the consistently higher correlation coefficients observed with the total lightning count (Table 3), and to avoid the redundant

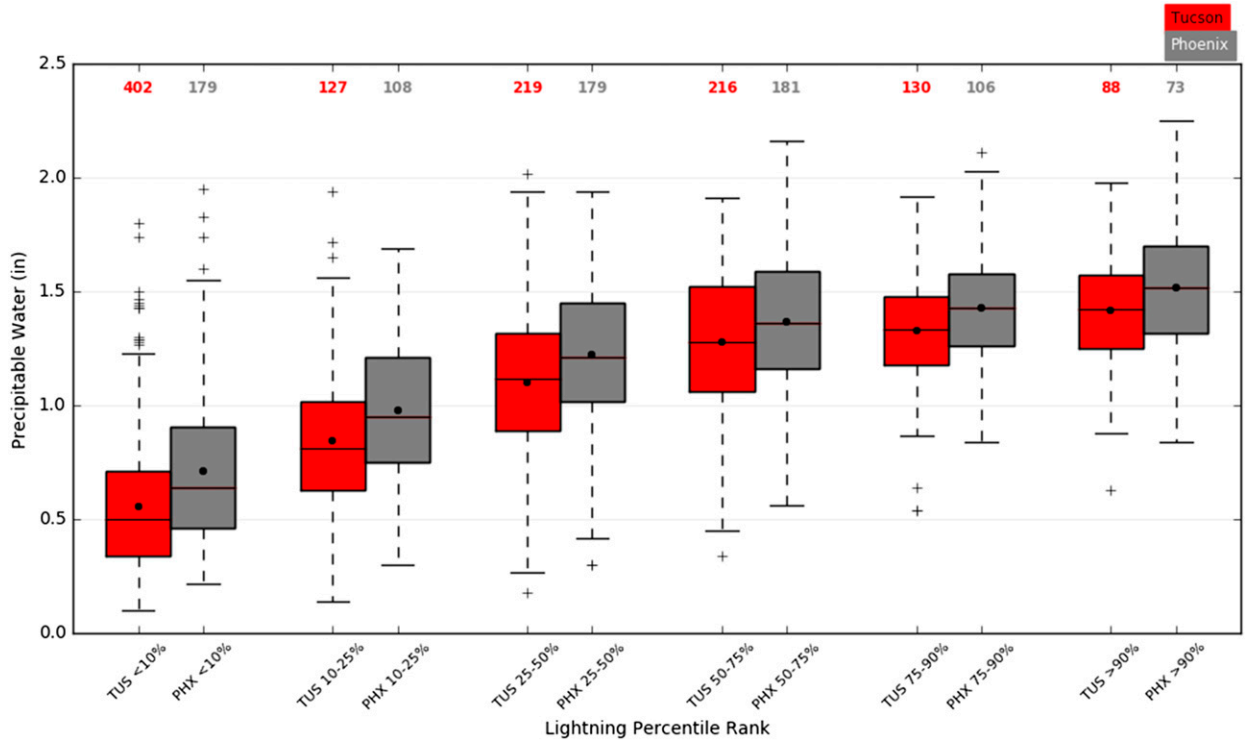


FIG. 4. PW values (in.) at 0000 UTC grouped by lightning percentile ranks of domain-total lightning counts for the 0000–1159 UTC period. Plot attributes are the same as described in Fig. 1.

inclusion of variables (e.g., PW from both PHX and TUS) in a single prediction model. It is difficult to select potential predictors that are entirely independent given the interdependency between many computed parameters and observed variables (e.g., moisture and CAPE), but predictors are selected based on the highest correlation coefficients from Table 1 that represent unique layers or physical processes (e.g., low-level moisture, total tropospheric moisture, computed instability parameters, etc.).

After including an initial set of predictors shown in Table 3, a stepwise regression technique, using both forward selection and backward elimination, is applied. Potential predictors that improve the fitted model performance are added during forward selection, and a new regression equation is computed. Next, potential predictors from the updated regression equation that do not improve model performance are removed during backward elimination (Miller 2002). The measure of relative model performance applied in the present study is the Akaike information criterion (AIC). The AIC measures the quality of a statistical model relative to earlier versions during the stepwise process, based on the maximum value of the likelihood function and the number of degrees of freedom within the model (Miller 2002).

The dependent dataset used to develop the equation incorporates 1200 UTC TUS sounding data from the

years 2003–2012. Applying the stepwise technique to the potential predictors reveals a best-fit model for predicting domain-total lightning count as

$$Y = -2442.493 + 3644.032x_1 + 2.635x_2 + 1.223x_3 - 158.629x_4 - 20.925x_5 - 64.217x_6, \quad (2)$$

where Y is the predicted domain-total lightning count for the 1200–1159 UTC period, x_1 is PW (in.), x_2 is SBCAPE (J kg^{-1}), x_3 is the SB lifted condensation level (LCL; m), x_4 is SBLI ($^{\circ}\text{C}$), x_5 is the u component of the 300-mb wind [in knots (kt), where $1 \text{ kt} = 0.51 \text{ m s}^{-1}$], and x_6 is the u component of the 700-mb wind (kt). SBCAPE, PW, and SBLCL are found to be the top three predictors using the AIC ranking method, and each of these predictors possesses a positive coefficient (Table 3). SBLI and the 700- and 300-mb u components of the wind all possess a negative coefficient. For the wind variables, a negative coefficient suggests that a weaker westerly component or stronger easterly component of midlevel and upper-level flow is favorable for increased lightning counts across the domain.

A summary of the regression model statistics, as provided in Table 3, shows that all predictors are found to be significant at the 95th percentile confidence interval, with the exception of the u component of the 300-mb wind.

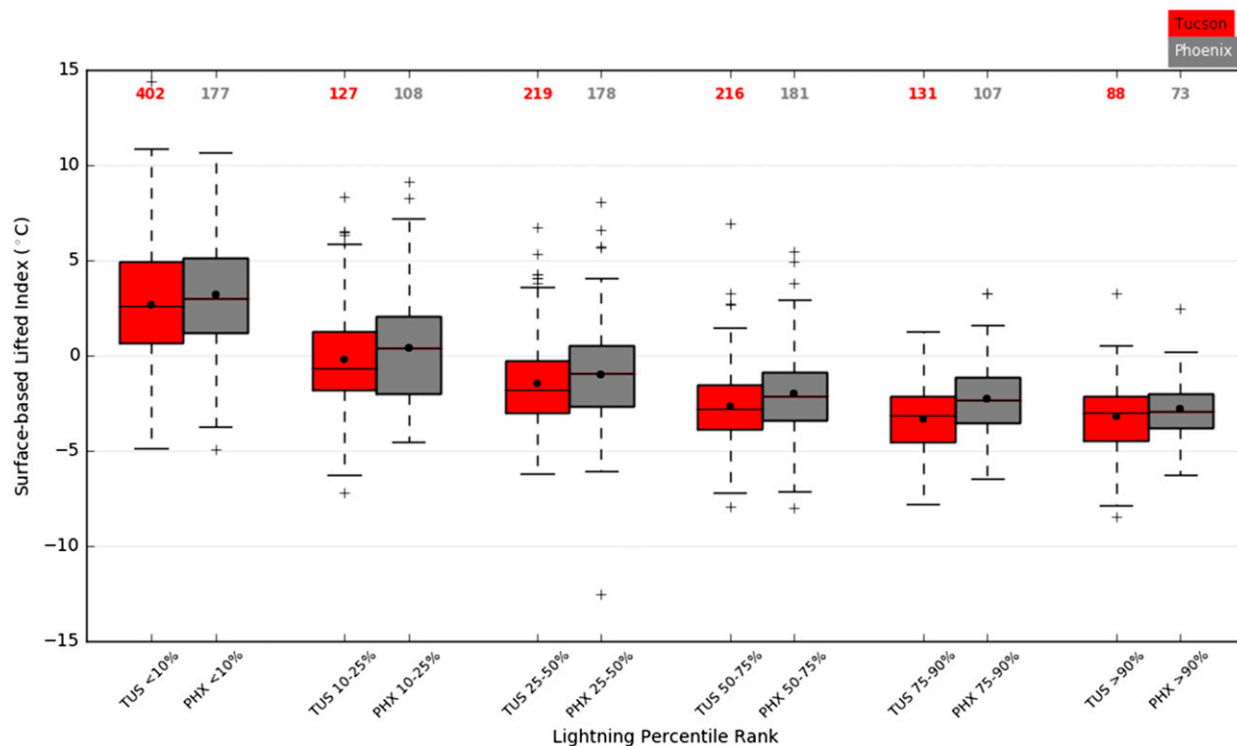


FIG. 5. As in Fig. 4, but for SBLI values (°C).

PW, SBCAPE, and SBLCL are all significant at the 99.9th percentile confidence interval, suggesting a high likelihood that these variables are significantly different from zero. The coefficient of determination r^2 is calculated as 0.3064, implying that the model explains 30.64% of the variance around the mean. However, given that most of the included predictors are statistically significant at the 95th percentile confidence interval, there does appear to be some predictive skill. The mean absolute error comparing the predicted values with the dependent dataset is 3126 strikes, and the bias for the prediction model is positive at 121.3 strikes.

The regression model is tested against an independent dataset from the year 2013. Variables obtained from the 1200 UTC TUS soundings are input into the prediction model [Eq. (2)] to determine a predicted domain-total lightning count for the subsequent 24-h period (Fig. 7). The mean absolute error for the predictive model using the 2013 dataset is 2058.1 strikes, which is lower than the training dataset from 2004 to 2012 used to develop the regression equation. The bias is still positive at 1152.3 strikes, indicating a tendency to overforecast lightning. No lightning occurred within the domain on days when zero lightning strikes were predicted.

Multiple linear regression is also applied to potential predictors listed in Table 3 in an attempt to estimate the

expected number of severe reports for the succeeding 24-h period. However, the model exhibits little predictive skill, which is likely due to the inconsistencies and biases within the reports database as described in CCR and, thus, results are not shown.

3. Gridded objective analysis

a. Logistic regression

Logistic regression analysis can be applied to produce a probabilistic function of a binary event occurring. A logit transformation is applied with the form

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}, \quad (3)$$

where P is the probability of a binary event occurring, β_0 is a fitted constant, and β_1 through β_n are fitted coefficients of independent predictors x_1 through x_n . A fitted generalized linear model (GLM) is applied to a given dataset with manually specified predictors (Everitt and Hothorn 2009). A stepwise algorithm, using both forward selection and backward elimination, is then applied to the initial GLM.

The resultant prediction model from the logistic regression technique allows for an estimated probability of occurrence of a binary event when applied to

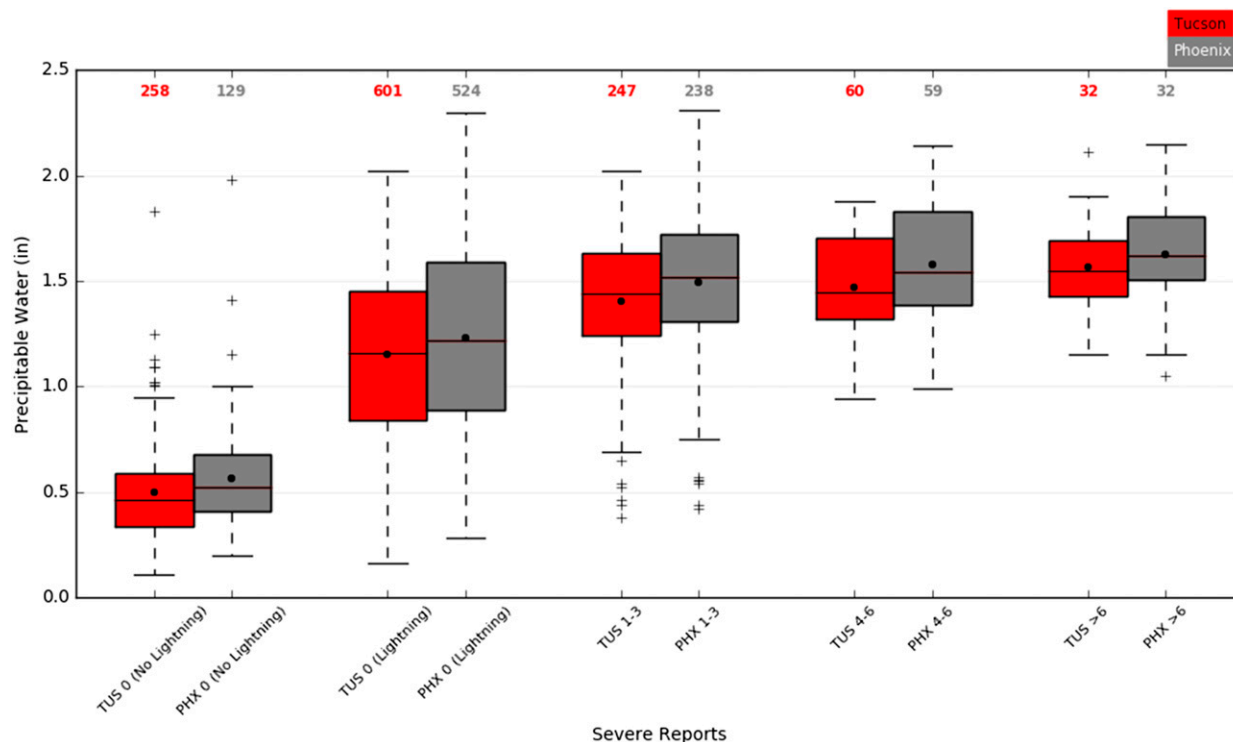


FIG. 6. As in Fig. 1, but for PW values (in.) compared to domain-total severe reports.

an independent dataset. The two binary events for the purpose of this study are the occurrence of CG lightning and severe reports. Rather than apply the probabilistic function to an observed sounding for the likelihood of lightning or a severe report across the domain, a more useful application may be in developing a logistic regression function for gridded objective analysis data. As described by CCR, 40-km gridded SFCOA data are collected across the domain for 2003–13 and include the variables listed in Table 4. For the purpose of this study, only the years 2004–13 are considered, as SFCOA lapse rate calculations are unavailable for 2003 and would otherwise preclude a

continuous dataset of all possible inputs for the development of predictive equations. Hourly severe report counts and CG lightning flashes are also assigned to each grid box. However, given the paucity of severe reports within any particular grid box during a 3-h period, logistic regression models are not shown to be skillful and would be of little use to operational meteorologists.

b. 3-h-long probabilistic lightning prediction models

Unique 3-h-long probabilistic prediction functions are generated for each grid box to predict the probability of lightning occurrence within the subsequent 3-h period.

TABLE 3. Statistics from the multiple regression equation for predicting the total lightning count across the domain. All initial predictors are included, with those in boldface remaining in the final stepwise regression equation. Data are from the years 2004–12.

Predictor	Coef	Standard error	<i>t</i> value	Pr > <i>t</i>
Y intercept	−2442.49	1298.02	−1.88	0.06
PW (in.)	3644.03	800.60	4.55	6×10^{-6}
SBCAPE (J kg^{−1})	2.63	0.40	6.57	8.4×10^{-11}
SBLCL (m)	1.23	0.35	3.53	4.4×10^{-4}
SBLI (°C)	−158.63	67.50	−2.35	0.02
300-mb <i>u</i> (kt)	−20.93	11.02	1.90	0.06
700-mb <i>u</i> (kt)	−64.21	23.03	−2.79	0.01
WBZ height (m)	—	—	—	—
500-mb <i>u</i> (kt)	—	—	—	—
0–3 km <i>w</i> (g kg ^{−1})	—	—	—	—

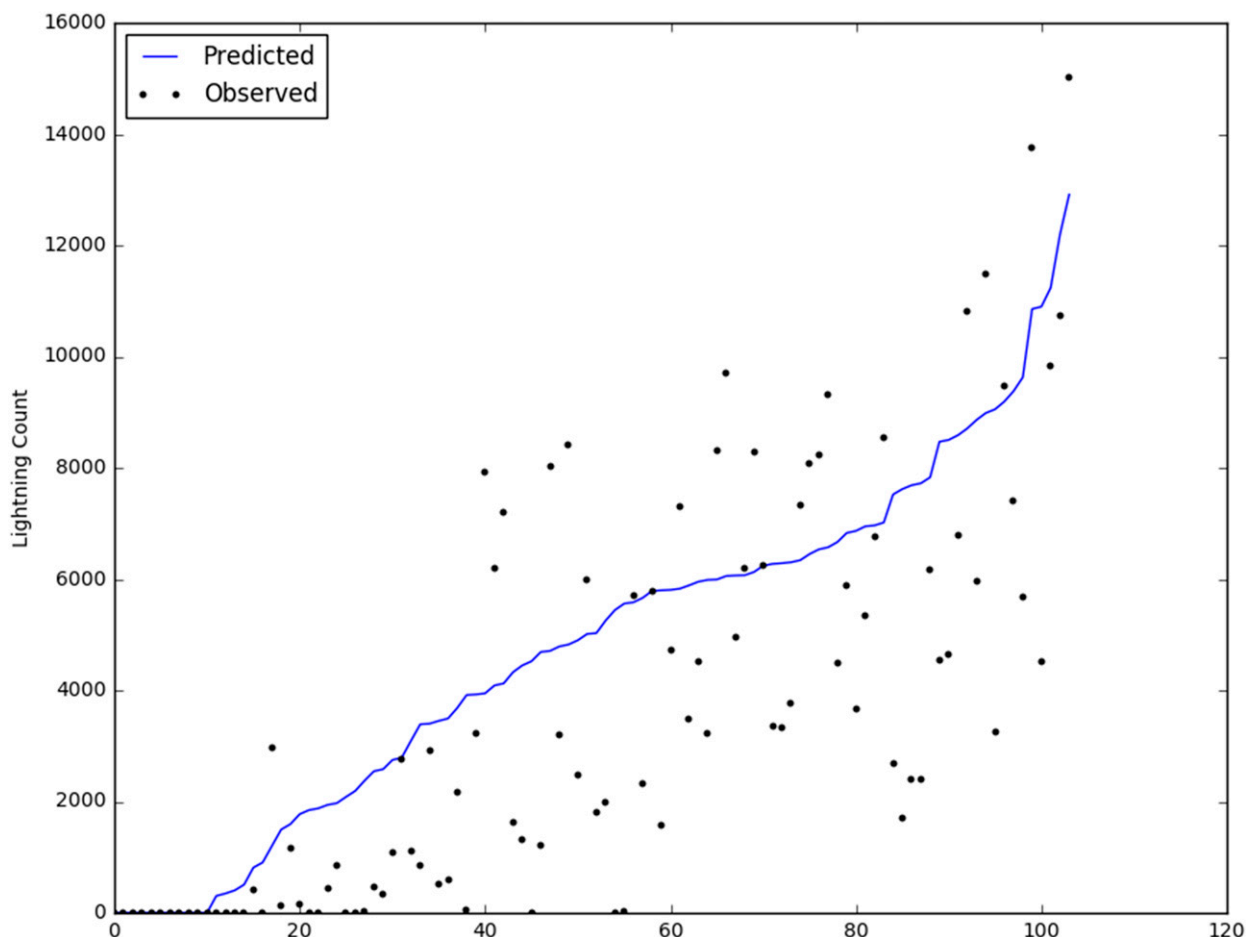


FIG. 7. Predicted and observed domain-total lightning count from 2013 (independent), based on 1200 UTC TUS sounding data. Plot is sorted from the smallest to largest predicted values along the x axis. Predicted values < 0 are manually set to 0.

Terrain variations across the domain can have significant effects on the potential for lightning, as shown by CCR. Determining a single probability of lightning for the domain total does not adequately represent terrain effects. Favored areas for thunderstorm development (i.e., higher-terrain locations) are likely to be better represented using a higher spatial resolution with the use of historical lightning data for each grid box. Additionally, producing higher spatial resolution lightning probability forecasts will be more useful for operational forecasting by showing spatial variations in lightning potential.

The dependent dataset used in the development of the logistic regression equations includes the years 2004–12, and includes a subset of input variables from the SFCOA database. The predictand is 3-h-long lightning occurrence per grid hour, where 1 is assigned for one or more lightning strikes and 0 is assigned for no lightning strikes within the following 3-h window. Potential predictors identified in Table 4 are selected from a single stepwise regression applied to the entire

dataset across the whole domain, in order to maintain the consistency of the predictors included in the unique equations assigned to each grid box. Most of the predictors in the final model, including moisture and instability parameters, exhibit a positive correlation with lightning, with the exception of 500-mb temperatures and the individual components of the 0–6-km bulk wind difference vector.

The performance of the prediction models is tested using an independent dataset from 2013. Rather than using goodness-of-fit statistics for each individual model, a general overview of the model performance is shown using a reliability diagram of all grid hours compared to the observed lightning occurrences (Fig. 8). Predicted probabilities of 0%–50% displayed an excellent fit with observed lightning occurrences. Probabilities above 50% tend to consistently overforecast lightning potential, with observed frequencies ranging between 40% and 55%. It is important to note that the sample size decreases substantially with higher predicted probabilities.

TABLE 4. List of predictors for individual lightning gridded logistic regression functions. Boldface predictors remained in the equation following application of the stepwise process. The predominant sign of the coefficient for each predictor is shown in the right-hand column.

Predictor	Sign
PW	+
500-mb temp	-
Surface temp	+
Surface dewpoint	+
MUCAPE	+
0–6-km bulk shear magnitude	+
850–500-mb lapse rate	+
0–1-km shear vector (<i>u</i> component)	+
0–1-km shear vector (<i>v</i> component)	+
0–6-km shear vector (<i>u</i> component)	-
0–6-km shear vector (<i>v</i> component)	-
MLCAPE	
MLCIN	
MUCIN	
Effective bulk shear magnitude	
700–500-mb lapse rate	
Mixing ratio (lowest 100-mb layer)	
Surface wind (<i>u</i> component)	
Surface wind (<i>v</i> component)	
0–3-km shear vector (<i>u</i> component)	
0–3-km shear vector (<i>v</i> component)	

Example spatial plots of the probabilistic lightning forecasts are shown in Figs. 9 and 10. A relatively active period that occurred on 11 July 2013 between 0000 and 0259 UTC over southeast Arizona is shown

in Fig. 9. The plot depicts the general lightning pattern quite well, with probabilities exceeding 70% over several grid boxes east and northeast of Tucson. Figure 10 shows an example plot from between 2200 and 0059 UTC 27 and 28 July 2013 and is representative of a moderately active period for lightning. An area of predicted probabilities of 40%–70% is well aligned with observed lightning strikes over southeast Arizona. Scattered occurrences of observed lightning also are noted east of Phoenix but are collocated with relatively low (i.e., less than 30%) predicted probabilities.

4. Conclusions

This work concludes a two-part investigation of convection occurring over Arizona during the North American monsoon. The primary focus of this particular portion of the study is on examining a large number of potential predictors for convective and severe weather using observational soundings and gridded objectively analyzed fields. Various moisture and instability parameters observed in 1200 UTC soundings possess the highest correlations to both lightning and severe-thunderstorm-report frequencies over central and southern Arizona. These parameters, including SBLI and PW, demonstrate some ability to discriminate between more active and less active convective days, which further supports the usage of 1200 UTC soundings in anticipating convective activity across the region.

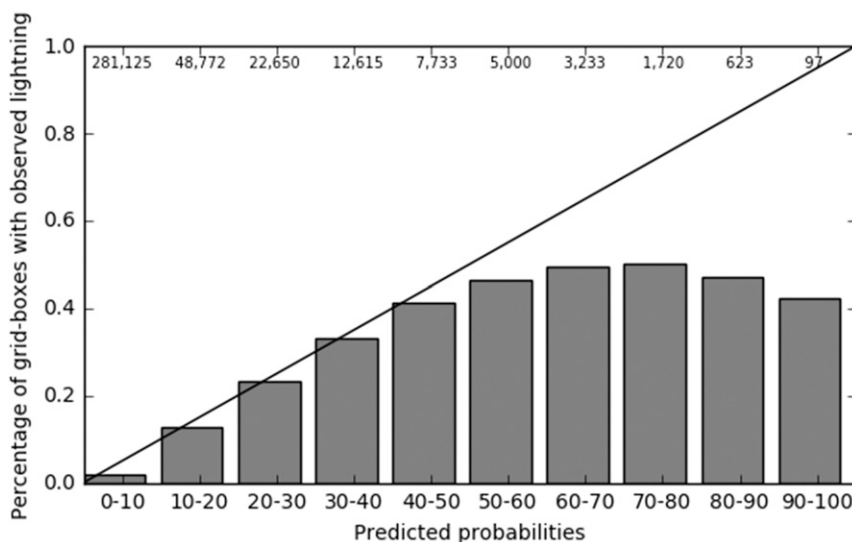


FIG. 8. Reliability diagram comparing observed lightning from 2013 with output from 3-h probabilistic forecasts for CG lightning strikes developed hourly for each grid box across the domain for all forecasts. Shaded gray bars represent the observed frequency of CG lightning strikes. Solid black line represents perfect reliability. Sample size of each bin is displayed at the top of the plot.

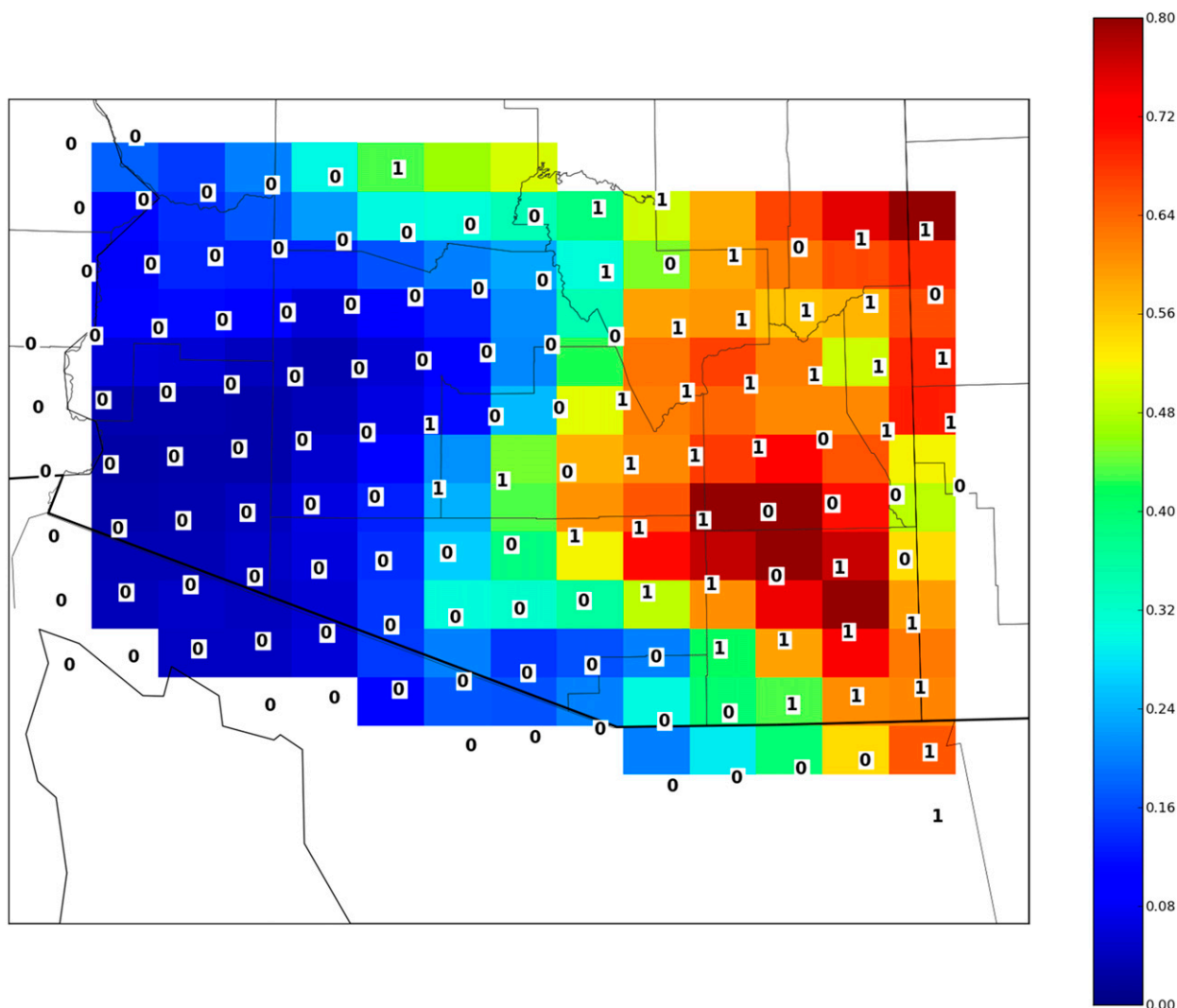


FIG. 9. Example plot of a probabilistic forecast of lightning occurrence between 0000 and 0259 UTC 11 Jul 2013. Color shaded grid boxes represent the predicted probability with the legend on the right-hand side of the plot. Boxes labeled 1 and 0 represent observed occurrences of lightning and no lightning, respectively.

Weaker correlations are observed between sounding-based parameters and the total number of severe thunderstorm reports across the domain, which may partially be affected by the inconsistency and population biases of the reports database (Shoemaker and Davis 2008), as reaffirmed by CCR.

Three-hour probabilistic forecasts of lightning, generated via a set of stepwise logistic regression equations, are derived from historical SFCOA data from 2004 through 2012. The equations can be applied operationally onto a 40-km grid to generate real-time plots used as short-term guidance, as shown in Figs. 9 and 10. Ultimately, a web-based portal for accessing the 3-h probabilistic lightning output could be developed for operational use to apply the findings of this study.

Additional development remains possible on the existing framework of the probabilistic models. While only 3-h forecasts are tested in this work, other time frames may be considered in future work to provide probabilistic forecasts of lightning. Additionally, 20-km spatial resolution RAP-based SFCOA data may become available in the near future and could potentially lead to higher-resolution and more skillful probabilistic lightning forecasts in the complex terrain of this region.

Acknowledgments. The authors thank Andy Dean of the Storm Prediction Center for providing the SFCOA data used in this study, and Dr. Israel Jirak also of the Storm Prediction Center for his assistance in improving this work. We additionally thank the two anonymous

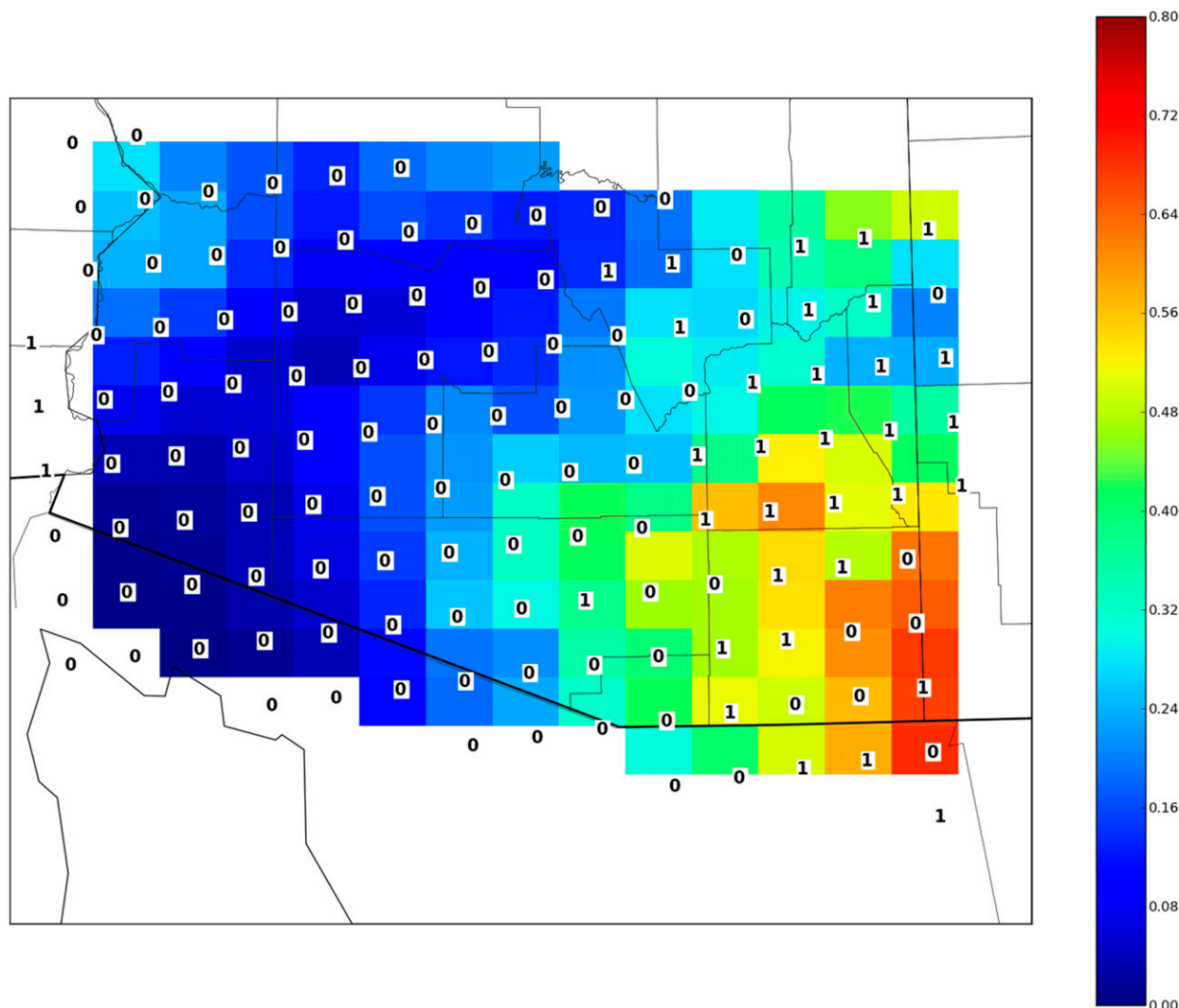


FIG. 10. As in Fig. 9, but between 2200 and 0059 UTC 27 and 28 Jul 2013.

reviewers who have greatly improved the readability and the content of this manuscript. Finally, great appreciation is given to the staff at the Tucson, Arizona, NWS Weather Forecast Office and the Storm Prediction Center for their assistance and input throughout the research process.

REFERENCES

- Billet, J., M. DeLisi, B. G. Smith, and C. Gates, 1997: Use of regression techniques to predict hail size and the probability of large hail. *Wea. Forecasting*, **12**, 154–164, doi:10.1175/1520-0434(1997)012<0154:UORTTP>2.0.CO;2.
- Bothwell, P. D., J. A. Hart, and R. L. Thompson, 2002: An integrated three-dimensional objective analysis scheme in use at the Storm Prediction Center. Preprints, *21st Conf. on Severe Local Storms/19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., JP3.1. [Available online at <https://ams.confex.com/ams/pdfpapers/47482.pdf>.]
- Carlaw, L. B., A. E. Cohen, and J. W. Rogers, 2017: Synoptic and mesoscale environment of convection during the North American monsoon across central and southern Arizona. *Wea. Forecasting*, **32**, 361–375, doi:10.1175/WAF-D-15-0098.1.
- Cohen, A. E., M. C. Coniglio, S. F. Corfidi, and S. J. Corfidi, 2007: Discrimination of mesoscale convective system environments using sounding observations. *Wea. Forecasting*, **22**, 1045–1062, doi:10.1175/WAF1040.1.
- Craven, J. P., and H. E. Brooks, 2004: Baseline climatology of sounding derived parameters associated with deep, moist convection. *Natl. Wea. Dig.*, **28**, 13–24.
- DeRubertis, D., 2006: Recent trends in four common stability indices derived from U.S. radiosonde observations. *J. Climate*, **19**, 309–323, doi:10.1175/JCLI3626.1.
- Douglas, M. W., R. A. Maddox, K. Howard, and S. Reyes, 2007: The Mexican monsoon. *J. Climate*, **20**, 1947–1954, doi:10.1175/JCLI4095.1.
- Everitt, B. S., and T. Hothorn, 2009: *A Handbook of Statistical Analyses Using R*. 2nd ed. Chapman and Hall/CRC, 376 pp.

- Galway, J. G., 1956: The lifted index as a predictor of latent instability. *Bull. Amer. Meteor. Soc.*, **37**, 528–529.
- Hales, J. E., Jr., 1974: Southwestern United States summer monsoon source—Gulf of Mexico or Pacific Ocean? *J. Appl. Meteor.*, **13**, 331–342, doi:[10.1175/1520-0450\(1974\)013<0331:SUSMS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1974)013<0331:SUSMS>2.0.CO;2).
- Hart, J. A., and W. Korotky, 1991: The SHARP workstation v1.50 users guide. National Weather Service, 30 pp. [Available from NWS Eastern Region Headquarters, 630 Johnson Ave., Bohemia, NY 11716.]
- Maddox, R. A., D. McCollum, and K. Howard, 1995: Large-scale patterns associated with severe summertime thunderstorms over central Arizona. *Wea. Forecasting*, **10**, 763–778, doi:[10.1175/1520-0434\(1995\)010<0763:LSPAWS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1995)010<0763:LSPAWS>2.0.CO;2).
- Miller, A., 2002: *Subset Selection in Regression*. 2nd ed. John Wiley and Sons, 256 pp.
- Montgomery, D. C., E. A. Peck, and G. G. Vining, 2012: *Introduction to Linear Regression Analysis*. 5th ed. John Wiley and Sons, 672 pp.
- Rasmussen, E. N., and D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148–1164, doi:[10.1175/1520-0434\(1998\)013<1148:ABCOSD>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<1148:ABCOSD>2.0.CO;2).
- Shoemaker, C., and J. T. Davis, 2008: Hazardous weather climatology for Arizona. NOAA Tech. Memo. NWS-WR-282, 47 pp. [Available online at https://docs.lib.noaa.gov/noaa_documents/NWS/NWS_WR/TM/NWS_WR_282.pdf.]
- Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261, doi:[10.1175/1520-0434\(2003\)018<1243:CPSWSE>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2).
- Wallace, C. E., R. A. Maddox, and K. W. Howard, 1999: Summertime convective environments in central Arizona: Local observations. *Wea. Forecasting*, **14**, 994–1006, doi:[10.1175/1520-0434\(1999\)014<0994:SCSEIC>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0994:SCSEIC>2.0.CO;2).