

## Statistical Prediction of Waterspout Probability for the Florida Keys

ANDREW DEVANAS

*NOAA/National Weather Service, Key West, Florida*

LYDIA STEFANOVA

*Center for Ocean–Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida*

(Manuscript received 17 July 2017, in final form 14 January 2018)

### ABSTRACT

A statistical model of waterspout probability was developed for wet-season (June–September) days over the Florida Keys. An analysis was performed on over 200 separate variables derived from Key West 1200 UTC daily wet-season soundings during the period 2006–14. These variables were separated into two subsets: days on which a waterspout was reported anywhere in the Florida Keys coastal waters and days on which no waterspouts were reported. Days on which waterspouts were reported were determined from the National Weather Service (NWS) Key West local storm reports. The sounding at Key West was used for this analysis since it was assumed to be representative of the atmospheric environment over the area evaluated in this study. The probability of a waterspout report day was modeled using multiple logistic regression with selected predictors obtained from the sounding variables. The final model containing eight separate variables was validated using repeated fivefold cross validation, and its performance was compared to that of an existing waterspout index used as a benchmark. The performance of the model was further validated in forecast mode using an independent verification wet-season dataset from 2015–16 that was not used to define or train the model. The eight-predictor model was found to produce a probability forecast with robust skill relative to climatology and superior to the benchmark waterspout index in both the cross validation and in the independent verification.

### 1. Introduction

The frequency of nonsupercell waterspouts in the Florida Keys is higher than any other location in the United States and may be one of the highest in the world, with hundreds of waterspouts occurring each year in the waters surrounding the Florida Keys (Golden 2003). Waterspouts are a marine hazard and do occasionally come on shore, resulting in damage to coastal structures. Objective waterspout guidance for the Key West area is limited, and forecasters rely mainly on subjective techniques and empirical rules of thumb to forecast the possibility of waterspout development. A reliable objective guidance that can be used in combination with available subjective tools would likely aid the forecast process.

Only a fraction of waterspouts are observed and recorded. During the Lower Keys Waterspout Project in

1969, at least 400 waterspouts were documented (Golden 1974). Golden (1977) estimated that between 50 and 500 waterspouts occur each year over the Florida Keys coastal waters. Given that approximately 25 waterspouts, on average, are reported during the wet season (estimated from 2006–14 reports discussed in section 2), waterspouts may be underreported by up to an order of magnitude, a conjecture also made by Golden and Bluestein (1994). Florida Keys waterspouts have been reported in every month of the year, although less frequently from October to May. Waterspout formation is noticeably more frequent during the wet season, June–September, with a primary maximum in June (Fig. 1), consistent with Golden (2003). The months of May and October are considered transition months (dry-to-wet and wet-to-dry seasons, respectively), during which midlatitude influences can vary greatly from year to year. Therefore, the transition months were excluded from this study. During the wet-season months in the present study waterspouts were reported on approximately 19% of all days.

---

*Corresponding author:* Andrew Devanas, [andrew.devanas@noaa.gov](mailto:andrew.devanas@noaa.gov)

DOI: 10.1175/WAF-D-17-0100.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

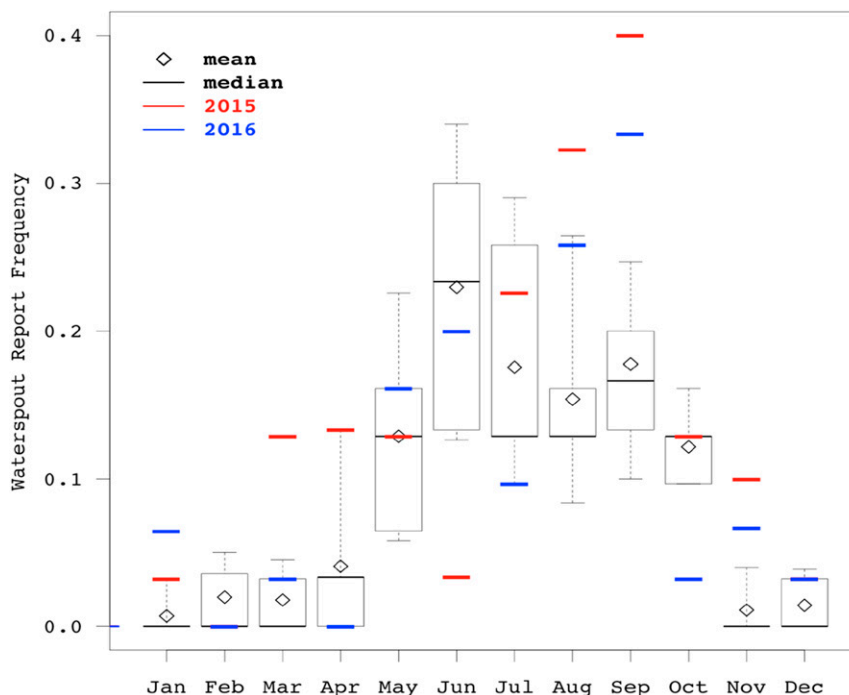


FIG. 1. Waterspout day frequency by month for the period 2006–14, with the median (thick black line) and mean (diamond) shown. Whiskers show the 10th and 90th percentiles. The values for 2015 (red lines) and 2016 (blue lines) are added to show the monthly frequency in the independent verification period. Data are extracted from NWS local storm reports from the field office in Key West.

Rossow (1970) noted that it is not possible to distinguish a likely waterspout day from one that is not, because the weather is nearly the same during the wet season. The atmospheric environment over the Florida Keys exhibits mainly incremental day-to-day variations in the temperature and moisture profile. Because of this, it is difficult to readily differentiate between days that are more favorable for waterspout development and days that are less favorable simply by examining soundings, sounding climatologies, or individual parameters or stability indices derived from the soundings. Nonetheless, the atmosphere over the Florida Keys does experience a degree of variability. For example, tropical features such as easterly waves and tropospheric upper-level trough cells traverse the area surrounding the Keys throughout the wet season. Changes in the position and strength of the Bermuda high, as well as the flow off the mainland and Cuba, may result in changes of wind patterns and temperature and moisture profiles.

Schwiesow (1981), as well as Sioutas and Keul (2007), noted that traditional stability indices in isolation did not perform well as indicators of a favorable waterspout development environment. Several previous efforts to quantify favorable conditions for waterspout development have been performed using sounding parameters

as well as synoptic-scale analysis. Golden and Bluestein (1994) noted that the thermodynamic soundings originating from the NWS Weather Forecast Office in Key West on waterspout days revealed relatively light low-level winds ( $5\text{--}8\text{ m s}^{-1}$ ), light shear, and abundant moisture in the boundary layer with drier air above. Golden (2003) expressed concern that a mean sounding climatology, or the NWS Key West sounding, could not adequately capture the superadiabatic lapse rates observed in the subcloud layer of waterspout-producing cumulus observed in his studies.

Spratt and Choy (1994) examined data from the period 1979–92, identifying days with at least one waterspout report in conjunction with soundings from Tampa, West Palm Beach, and Cape Canaveral. The results determined that light winds at discrete levels through the lower troposphere (up to 500 mb; 1 mb = 1 hPa) and total precipitable water vapor levels in excess of 4.1 cm (1.6 in.) produced favorable conditions for waterspout development. They combined these sounding-derived parameters with WSR-88D radar information to build a five-step waterspout forecast strategy.

Brown and Rothfuss (1998) developed a basic waterspout index for the Florida Keys and south Florida that included mean wind direction and speed below

600 mb, as well as a qualifier for whether a waterspout occurred on the previous day. The values obtained were assigned a value of 0.5 or 1 according to a table, and a sum equaling 2 or greater was considered favorable for waterspout development.

An index developed by [Kuiper and Van der Haven \(2007\)](#) used the product of four weighted parameters (0–3-km wind shear, 0–500-m lapse rate, 0–1-km average humidity, and 10-m wind speed) to indicate the risk of waterspout development over the shallow waters of the North Sea off the coast of the Netherlands. The Kuiper–Van der Haven spout index has been implemented in the locally run High Resolution Limited Area Model (HIRLAM) at the Royal Netherlands Meteorological Institute.

A study of waterspouts over the Adriatic, Ionian, and Aegean Seas by [Sioutas and Keul \(2007\)](#) examined several thermodynamic indices derived from proximity sounding data in relation to waterspout development. They found that “fair weather” waterspouts were mainly a result of a preexisting unstable air mass and low wind shear. Of the thermodynamic indices examined in the study, they stated that the K index and the total totals index were good indicators of an unstable environment and useful in identifying favorable tornadic and fair weather waterspout environments. Other thermodynamic indices they examined did not perform as well from a predictive standpoint.

A follow-up study ([Keul et al. 2009](#)) examined the performance of the Szilagyi waterspout forecasting technique ([Szilagyi 2009](#)) as a predictor for Mediterranean waterspout occurrence. The Szilagyi technique uses a nomogram to compare the difference between sea surface temperature and the 850-mb temperature against the convective cloud depth and is also available as an index (Szilagyi waterspout index). Values of the Szilagyi waterspout index range from –10 to 10, where values greater than zero indicate a favorable waterspout environment.

The NASA Applied Meteorology Unit (AMU) at the Kennedy Space Center in Florida used a combination of four factors (monthly climatology, vertical wind profile, precipitable water, and persistence) to determine the weak waterspout (fair weather waterspout) potential for the day ([Wheeler 2009](#)). The value produced by the methodology is compared to a table to ascertain whether the threat level for the day is low, medium, or high. An AMU follow-up study ([Watson 2011](#)) used a logistic regression technique similar to that in the present study to evaluate the probability of severe weather occurrence at the Kennedy Space Center. However, it did not include a reevaluation of the weak waterspout methodology outlined by [Wheeler \(2009\)](#). [Watson \(2011\)](#)

found that using logistic regression did not show an improvement over their existing severe weather likelihood methodology.

A common thread found through many studies ([Spratt and Choy 1994](#); [Brown and Rothfuss 1998](#); [Golden 2003](#); [Kuiper and Van der Haven 2007](#); [Sioutas and Keul 2007](#); [Wheeler 2009](#)) is the presence of weak lower-tropospheric winds and/or weak lower-tropospheric shear as a favorable environment for waterspout development. Additionally, some of these studies noted the importance of lapse rates at various levels in the lower troposphere ([Golden 2003](#); [Kuiper and van der Haven 2007](#)). These conditions were also found to be significant in the present study.

The commercial raob software created by Environmental Research Services, LLC ([Shewchuk 2017](#)), provides a waterspout index developed by P. Mohlin and F. Alsheimer of the NWS Forecast Office in Charleston, South Carolina. This index assigns risk points based on sounding-derived variables meeting certain thresholds ([Table 1](#); F. Alsheimer 2016, personal communication). The total number of risk points forms the Charleston waterspout index. The value of the index is used to qualitatively classify the risk for waterspout development as high, moderate, low, or none.

The goal of the present work was to develop a quantitative representation of the probability of waterspouts as a function of sounding-derived parameters for the Florida Keys wet season, June–September, and to compare the results with existing benchmarks. For this comparison, a logistic regression model (referred to hereafter as CWI) was constructed to convert the raob-provided Charleston waterspout index described in [Table 1](#) to a waterspout probability value. The CWI regression model was constructed using training period data (see [section 2](#)) with waterspout report days as a predictand and the Charleston waterspout index as a predictor. Similarly, logistic regression modeling during the training period was used to convert the Brown–Rothfuss index, the Kuiper–Van der Haven spout index, and the Szilagyi waterspout index to waterspout probabilities (hereafter referred to as BRI, KHI, and SWI, respectively). Of these four existing indices, the CWI produced the best-performing probabilistic model, and as such, it was chosen as a primary comparison benchmark.

Details of the waterspout and sounding observations are discussed in [section 2](#). The statistical methodology for model development and the metrics for assessing model performance are described in [section 3](#). The evaluation of performance for the proposed forecast model and a comparison with that of the benchmark CWI model is presented in [section 4](#) and further

TABLE 1. Parameters used in the Charleston waterspout index with thresholds and associated risk points. The EHI is defined by [Davies \(1993\)](#), and the Wind Index (WINDEX) is defined by [McCann \(1994\)](#).

Variable	Threshold 1 (risk points)	Threshold 2 (risk points)	Threshold 3 (risk points)	Threshold 4 (risk points)
Surface wind speed (kt)	>15 (0)	10–15 (2)	5–9 (6)	<5 (9)
0–3-km wind speed (kt)	>15 (0)	10–15 (2)	5–9 (6)	<5 (9)
Precipitable water (in.)	<1.00 (0)	1.00–1.49 (1)	1.50–1.99 (2)	>1.99 (3)
First dry layer (dewpoint depression > 5°C)	Below 950 mb (0)	950–751 mb (2)	750–600 mb (4)	Above 600 mb or none (6)
850-mb dewpoint depression (°C)	>7 (0)	5–7 (1)	2–4 (2)	0–1 (3)
850-mb mixing ratio ( $\text{g kg}^{-1}$ )	<11 (0)	11–13 (1)	14–16 (2)	>16 (3)
Surface–900-mb wet-bulb theta (K)	<291 (0)	291–293 (1)	294–295 (2)	>295 (3)
K index	<15 (0)	15–25 or >35 (1)	26–30 (2)	31–35 (3)
Lifted index	>0 (0)	0 to –2 (1)	–3 to –7 (2)	<–7 (3)
Surface-based CAPE (SBCAPE; $\text{J kg}^{-1}$ )	<500 (0)	500–999 (1)	1000–1499 or >1999 (2)	1500–1999 (3)
0–3-km CAPE ( $\text{J kg}^{-1}$ )	<40 (0)	40–74 (2)	75–124 or >174 (3)	125–174 (4)
Level of free convection (m)	>2200 (0)	1700–2200 (2)	1200–1699 (4)	<1200 (6)
EHI	>0.5 (0)	0.3–0.5 (2)	0.2 (3)	0.0–0.1 (4)
0–3-km storm-relative helicity ( $\text{m}^2 \text{s}^{-2}$ )	>45 (0)	30–45 (2)	10–29 (4)	<10 (6)
0–3-km lapse rate ( $\text{K km}^{-1}$ )	>–5.1 (0)	–5.1 to –5.5 (1)	–5.6 to –6.0 (2)	<–6.0 (3)
900–600-mb lapse rate ( $\text{K km}^{-1}$ )	>–5.1 (0)	–6.1 or less (1)	–5.1 to –5.5 (2)	–5.6 to –6.0 (3)
Downdraft CAPE ( $\text{J kg}^{-1}$ )	<240 or >1249 (0)	1000–1249 (1)	240–499 (2)	500–999 (3)
WINDEX (kt)	>49 (0)	<25 (1)	35–49 (2)	25–34 (3)

discussed in [section 5](#). Finally, some directions for future work are proposed in [section 6](#).

## 2. Data

The official NWS local storm reports (obtained from the National Centers for Environmental Information database) from the field office in Key West, which covers the Florida Keys nearshore coastal waters, were used to identify days on which waterspouts were reported. Days with one or more waterspout reports anywhere along the Florida Keys coastal waters were classified as waterspout report days. Days with no reports were classified as no-waterspout report days. [Figure 2](#) shows the geospatial distribution of waterspout reports along the Florida Keys from 2006 to 2016. These locations should be considered approximate, as they are often visually estimated at the time of observation.

Inherently, reports based on visual sightings have limitations. It is likely that a larger proportion of waterspouts are reported near the densely populated cities of Key West and Marathon. Conversely, there may be underreporting along the more sparsely populated islands. Many waterspouts are reported by mariners; therefore, waterspout reports would be more common in regions with large marinas and popular boating routes

(both recreational and commercial) and underreported elsewhere. Waterspouts sighted in areas outside of cellular phone or marine radio reception are more likely to go unreported, at least in real time. Also, few waterspouts have been reported in the hours between dusk and dawn. This is due to two main reasons. Physically, waterspouts are less likely to occur during evening hours because the generation of cumulus cloud lines from which waterspouts can spawn is less likely as a result of the loss of differential heating along the Keys islands. From an observation perspective, it is difficult to spot and identify waterspouts in the dark. Waterspout reports received via social media may at some point lead to an increase in the overall number of reports, but as of yet no such increase has been observed. This may in part be a result of the short data record since the onset of social media as a tool for reporting weather phenomena.

The 1200 UTC soundings were obtained at Key West for the wet seasons (June–September) of 2006–16. The first 9 years (2006–14) of the data were used as a training dataset to build a multiple logistic regression model of waterspout report probability. Once the model was finalized, data for the two withheld years (2015–16) were used as a testing dataset for independent forecast verification. Years prior to 2006 were not included in this study because of inconsistent reporting and waterspout

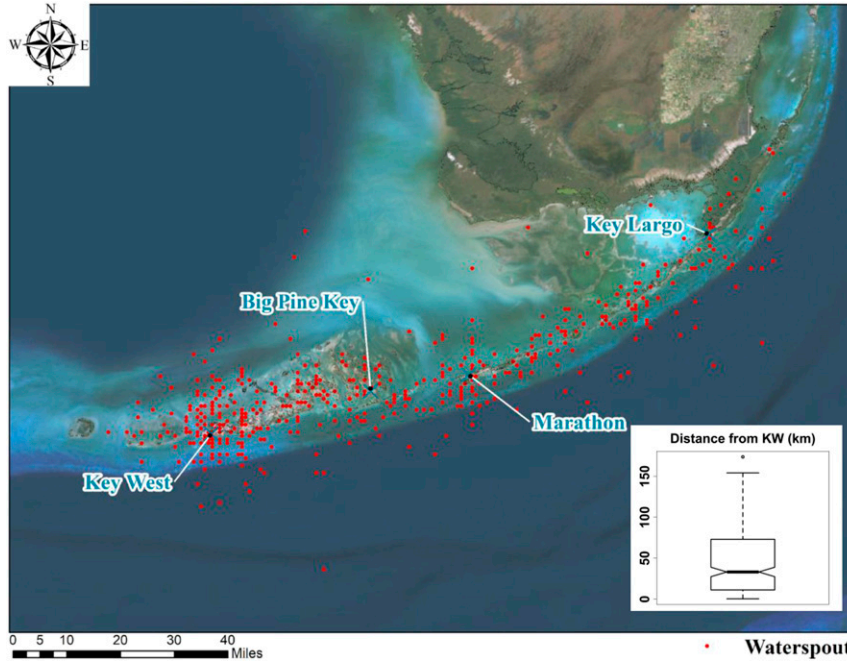


FIG. 2. Locations of all reported waterspouts during June–September 2006–16, extracted from NWS local storm reports from the field office in Key West. The distribution of waterspout distances (km) from the Key West sounding location is shown on the bottom right. The median, mean, and standard deviation of these distances are, respectively, 33.2, 47.3, and 42.1 km.

identification practices. The 1200 UTC soundings were selected because of their predictive potential for the daylight hours, which is when most waterspouts are reported (Fig. 3). The 1200 UTC Key West sounding was considered proximate to waterspout occurrences because of the generally spatially uniform structure of the atmosphere over the Florida Keys on a given wet-season day. The Key West location is representative of a tropical marine boundary layer, without landmass influences such as near-surface radiation inversions in the morning sounding or strong land–sea-breeze interactions. The sounding at Key West is within 73 km (about 46 mi) of 75% of the waterspouts in the dataset (see Fig. 2), and within 115 km (about 72 mi) of 90% of reported waterspouts.

Some of the limitations of using sounding data include the inability to capture localized mesogamma phenomena such as local sources of vorticity generated by cumulus cloud lines along the Keys, subcloud superadiabatic lapse rates, and convergence resulting from differential heating along the islands, to mention a few. These likely all play a role in waterspout development, but the data are not available for proper evaluation of these elements. Soundings are used in the hope of finding parameters

that may be precursors, or proxies, to an overall favorable waterspout development environment.

Dewpoint temperatures from 400 mb and above were omitted from this study. This is because of discrepancies in moisture measurements aloft from different

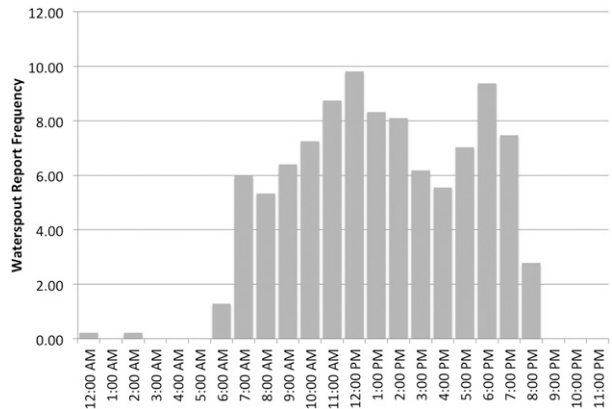


FIG. 3. Waterspout report frequency (%) by local hour of the day for the years 2006–14 during June–September. Data are extracted from NWS local storm reports from the field office in Key West. The 1200 UTC soundings are released at about 0700 eastern daylight time (EDT).



radiosondes used from 2006 through 2016. From the beginning of the study in 2006 to 22 March 2010, the Sippican VIZ-B2 radiosonde was used at Key West, which was replaced by the Lockheed Martin Mark IIA radiosonde, which in turn was replaced by the Vaisala RS92-NGP GPS radiosonde on 27 June 2012 and is in use to date. As is evident in the [appendix](#) (see [Fig. A1](#)), the upper-tropospheric moisture measured by the various radiosondes begins to diverge above 400 mb. The [appendix](#) also illustrates the monthly mean wet-season 100-mb dewpoint temperature by year ([Fig. A2](#)), showing a clear discontinuity in the data before 2010 and after 2012, as different radiosondes were introduced. Since the dewpoint temperature differences result from different radiosondes and not atmospheric changes, inclusion of the data could lead to statistically anomalous results. To include the data would have required a statistical scheme to remove the instrument bias, which is beyond the scope of this study.

Supercell waterspouts not associated with tropical cyclones during the wet-season months are relatively rare. One waterspout, which occurred on 1 June 2007 in association with Tropical Storm Barry moving past the Florida Keys, was removed from the dataset since this study is designed to investigate nonsupercell waterspouts, or fair weather waterspouts. This was the only waterspout in the dataset that was associated with a tropical cyclone.

The Key West radiosonde dataset included wind speed and direction at various levels or layers. This information was used to calculate the signs of the  $U$  and  $V$  components of the winds as additional derived variables that take values of +1, 0, or -1. The final dataset at that point comprised 238 sounding-derived parameters (listed in [Table A1](#)) that were considered as potential predictors for developing a waterspout probability model, as described in the following section.

### 3. Methodology

#### a. Modeling

The presence of a waterspout report can be considered a reliable indication of waterspout occurrence. However, as previously discussed, the absence of a waterspout report cannot be considered a reliable indication of waterspout nonoccurrence. This discrepancy between occurrence and report is an unavoidable property of the dataset. Given the known limitations of waterspout observations and of single-point radiosonde soundings as representations of the atmospheric environment for the Florida Keys and surrounding waters, it is not a priori obvious that it would be possible to

establish a statistical relationship between sounding variables and (reported) waterspout days. The present work, however, assumes that such a relationship can be established. The validity of this assumption is eventually evaluated and confirmed by demonstrating that the resulting model does indeed have skill above that of a climatological prediction.

A logistic regression approach was employed to model the probability of a waterspout report as a function of environmental variables derived from morning (1200 UTC) soundings. This approach required finding an appropriate subset of sounding variables that would serve as predictors in the final model. The large number of candidate predictors (in our case,  $n = 238$  separate parameters derived from the sounding) posed a problem for selecting an optimal subset of predictors. Since there are  $2^n$  possible subsets that can be obtained from a set of  $n$  elements, an exhaustive search for an optimal subset of predictors would require the calculation and comparison of  $2^{238} = 4.4 \times 10^{71}$  separate models. If the selection of subsets is restricted to require exactly  $k$  elements (out of  $n$ ), the number of possible subsets is given by “ $n$  choose  $k$ ,” or

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Building a robust logistic regression model for binary outcomes requires a minimum of 10–15 events per predictor variable ([Peduzzi et al. 1996](#)), although for achieving predictive accuracy some 20–40 events per variable may be required ([Austin and Steyerberg 2014](#)). In the present case, there are 202 events (waterspout observations), which could accommodate up to approximately 13–20 predictor variables under the first rule, or 5–10 by the second. If the optimal predictor set is thus restricted to contain no more than  $k = 10$  predictors, the number of possible subsets (NPSs) is given by

$$\text{NPS} = \sum_{k=0}^{10} \binom{n}{k}.$$

A graphical representation of the relationship between NPS and the number of candidate predictors  $n$  is shown in [Fig. 4](#). If all 238 variables are considered as candidate predictors,  $\text{NPS} = 1.4 \times 10^{17}$  ([Fig. 4](#), circle symbol), which is still a very large number. An exhaustive search through that many combinations would be computationally prohibitive, which makes finding a definitive “best model” not computationally feasible.

It is therefore necessary to shorten the list of candidate predictors to a more manageable number. As a first

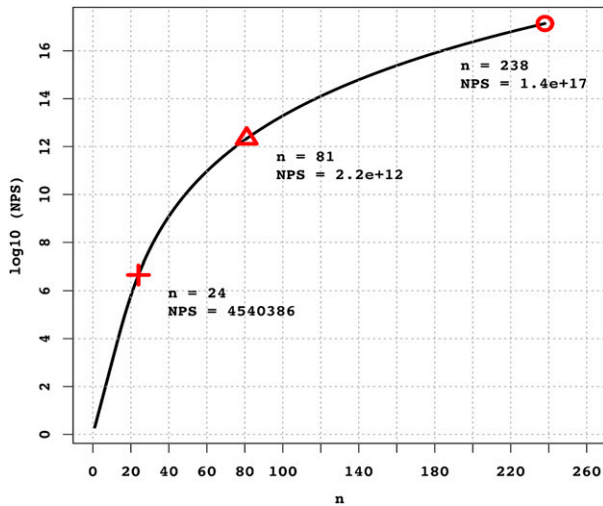


FIG. 4.  $\log_{10}$  of the NPS (vertical axis) of up to  $k = 10$  predictors that can be selected from a set of  $n$  (horizontal axis) candidate predictors. Graph obtained by calculating  $NPS = \sum_{k=0}^{10} \binom{n}{k}$  as a function of  $n$ . The red circle, triangle, and cross indicate the number of such subsets that can be selected, respectively, from a set of  $n = 238$ ,  $n = 81$ , and  $n = 24$  candidate predictors.

step in this process, the statistical significance of the difference of means (using a  $t$  test) for each variable between waterspout report days versus no-report days was calculated (see Table A1, where the candidate predictors are arranged in order of increasing  $p$  value). If the means for a particular variable were significantly different at the 95% confidence level ( $p$  value of 0.05) or higher, that variable was added to a shortened candidate predictors list. This testing excluded 157 variables and retained  $n = 81$  variables (Table A1, column  $N_0$ ), reducing the NPS to  $2.2 \times 10^{12}$  (Fig. 4, triangle symbol). To further reduce the number of candidate predictors, and to reduce the risk of multicollinearity in the final model, a requirement was added that no two candidate predictors be highly correlated. A commonly used rule of thumb for the level of correlation between regression model predictors that must not be exceeded is 0.6 (MacInnes 2017). This requirement was implemented as follows. Sequentially moving through the shortened list of candidate predictors, a variable was retained if its absolute correlation with any of the previously retained variables was less than the defined threshold of 0.6. To illustrate this point, the first variable on the list (0–3 kft AGL wind speed) was retained by default, since it is on the top of the list and there are no preceding entries. The next variable (1000-mb wind speed) was not retained because of its high correlation to the previously retained variable (0–3 kft AGL wind speed). Similarly, the

3rd–14th variables [925-mb wind speed, transport mean speed, surface wind speed, 0–10 kft AGL wind speed, fire danger index, shear 1, 850-mb wind speed, 925–70-mb average wind speed, 3–10 kft AGL wind speed, energy helicity index (EHI), transport peak speed, fog, and fog stability index (FSI)] were not retained for the same reason. The 15th variable (0–10 kft AGL  $V$ -wind direction) was retained because its correlation with the only previously retained variable (0–3 kft AGL wind speed) did not reach the specified threshold. The 16th variable (1000–500-mb lapse rate) did not reach the specified correlation threshold with either of the two previously retained variables (0–3 kft AGL wind speed and 0–10 kft AGL longitudinal wind direction) and was thus also retained. This process is continued until the 81st variable (the last on the significant candidate predictors list) is reached. At the end of the process,  $n = 24$  variables are retained, forming the further shortened candidate predictors list (Table A1, column  $N_1$  and boldface). The NPS for this set is a much more manageable  $4.5 \times 10^6$ .

These 24 variables were then supplied as candidate predictors to an exhaustive search for a subset of up to 10 predictors that minimized the Akaike information criterion (AIC; Akaike 1974). The AIC was chosen as a selection criterion because it rewards good fit while penalizing the number of predictors, thus reducing the risk of overfitting. The search was performed using the “bestglm” package in R (<https://www.r-project.org/>) and resulted in a final eight-variable model (Table A1, column  $N_2$  and boldface and italic).

During the predictor selection phase described above, days with missing data for any of the initial 238 variables were temporarily excluded from consideration. Once the predictors were finalized, the model was constructed using all days for which none of the final predictors contained missing data.

The contribution of individual predictors to the informational content (measured by AIC and residual deviance) of the final eight-variable logistic regression model (LRM) is illustrated in Table 2. To construct this table, as a first step, separate logistic regressions ( $k = 1$  in Table 2) were carried out for each individual predictor, and the residual deviance and AIC were documented. The predictor with the lowest value of AIC (representing the biggest improvement over the null model, i.e., the one with intercept only) was selected. For the next step ( $k = 2$ ), the best single predictor, the average wind speed at 0–3 kft AGL, was combined in a logistic regression with each of the other predictors one at a time. Next, the resulting two-predictor model with the lowest value of AIC was retained, and the remaining

TABLE 2. Residual deviance (RES) and AIC for models with  $k$  predictors built from subsets of the final eight predictors using a sequential selection approach based on AIC. Individual predictors are abbreviated as in the text. The lowest values of RES and AIC for each  $k$  are marked in boldface. The null deviance is 1032.5.

$k$	0-3kSpd	0-10kV+/-	Craven	1000-700LR	100U+/-	TransU+/-	100T	TT	RES	AIC
1	X								<b>984.0</b>	<b>988.0</b>
1		X							1009.9	1013.9
1			X						1025.4	1029.4
1				X					1022.9	1026.9
1					X				1021.3	1025.3
1						X			1021.6	1025.6
1							X		1026.5	1030.5
1								X	1024.8	1028.8
2	X	X							<b>970.5</b>	<b>976.5</b>
2			X						974.1	980.1
2				X					977.2	983.2
2					X				975.6	981.6
2						X			979.8	985.8
2							X		979.5	985.5
2								X	977.0	983.0
3	X	X	X						<b>958.1</b>	<b>966.1</b>
3				X					961.1	966.1
3					X				963.5	971.5
3						X			966.3	974.3
3							X		966.2	974.2
3								X	962.2	970.2
4	X	X	X	X					<b>950.6</b>	<b>960.6</b>
4					X				952.8	962.8
4						X			953.9	963.9
4							X		953.5	963.5
4								X	954.4	964.4
5	X	X	X	X	X				<b>945.4</b>	<b>957.4</b>
5						X			946.4	958.4
5							X		948.5	960.5
5								X	947.7	959.7
6	X	X	X	X	X	X			<b>941.5</b>	<b>955.5</b>
6							X		942.6	956.6
6								X	942.4	956.4
7	X	X	X	X	X	X	X		<b>937.6</b>	<b>953.6</b>
7								X	938.2	954.21
8	X	X	X	X	X	X	X	X	<b>934.9</b>	<b>952.9</b>

predictors were added, again one at a time, to determine the best three-predictor ( $k = 3$ ) model, and so forth, until the full  $k = 8$  model is reached. The first step of the procedure illustrates that out of the set of eight, when considered separately, the best predictors are, in order 1) average wind speed at 0–3 kft AGL (0–3kSpd), 2) sign of the  $V$ -wind component at 0–10 kft AGL (0–10kV+/-), 3) Craven significant severe parameter (Craven; Craven and Brooks 2004), 4) 1000–700-mb lapse rate (1000–700LR), 5) sign of the  $U$ -wind component of the 100-mb wind (100U+/-), 6) sign of the  $U$ -wind component of the transport mean wind (TransU+/-), 7) temperature at 100 mb (100T), and 8) total totals index (TT; Miller 1972). The transport mean wind as defined here is the average wind in a layer from the surface to the mixing height, with the mixing height found using the

Heffter method (Heffter 1980). An estimate of the relative contribution of each predictor in the full model can be seen in the standardized coefficients for the LRM (Fig. 5). According to these standardized coefficients, waterspout probability is largest for weak 0–3 kft AGL average wind speeds, stronger negative 1000–700-mb lapse rates, and larger values of the Craven parameter, total totals, and 100-mb temperature; the most favorable flows have positive  $U$  components of the transport mean winds and the 100-mb winds, as well as a negative  $V$  component in the lowest 10 kft of the atmosphere.

The nature of these relationships holds when the predictors are considered separately (Fig. 6). For each individual quantitative variable, the frequency distribution over the training dataset, the conditional frequency of waterspout report days in each of 15 equally



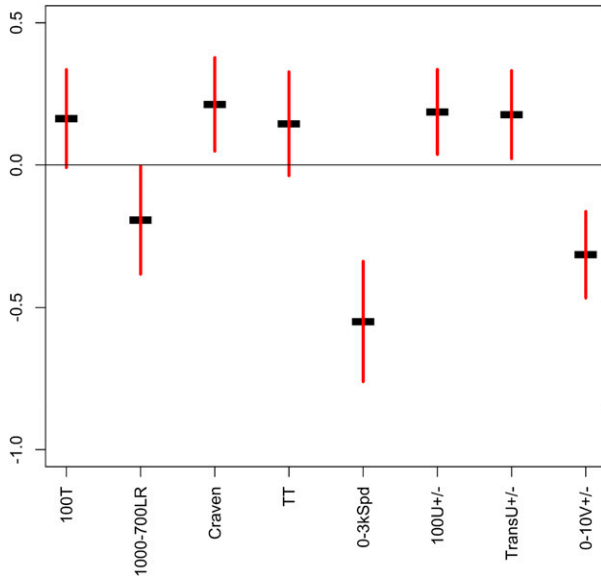


FIG. 5. Standardized coefficients (black) for the LRM with 95% confidence intervals (red lines).

populated bins, and the single-variable logistic regression fit are shown for each predictor of the LRM.

#### b. Performance metrics

Three complementary sets of metrics were used to assess the performance of the LRM and compare the results with those of the benchmark model (CWI) and climatology, specifically (i) the distribution of predictions on days with and without waterspout reports; (ii) the reliability diagram and associated reliability, resolution, and Brier skill score of the predictions; and (iii) their receiver operating characteristics (ROCs) and the areas under the curve (AUCs). The first set of metrics addresses the question of how different the forecast distributions are for days with and without waterspout reports and whether the differences between these distributions are statistically significant. The reliability diagram and the associated Brier skill decomposition and score (Wilks 2006) address the question of how reliable the forecasts are (i.e., an event with a predicted probability of  $X\%$  occurs in  $X\%$  of cases), how well resolving they are (i.e., the degree to which the model produces predictions away from climatology), and whether they are, overall, more skillful than a climatological forecast. Finally, the ROC illustrates how the hit (true positive) rates and false alarm (false positive) rates compare at different classification thresholds. In other words, the ROC diagram shows, for any given value of probability taken as a threshold to differentiate between a categorical “yes” and a “no”

forecast, the hit and false alarm rates for such forecast. AUC is the area under the ROC curve. For a random (no skill) forecast, the hit and false alarm rates are equal at any forecast probability threshold. For such forecasts, the ROC follows the diagonal line of the diagram, connecting the bottom-left and top-right corners, and its AUC value is 0.5. For a perfect categorical forecast, the ROC is made of two straight lines connecting the bottom-left, top-left, and top-right corners of the diagram, with an AUC value of 1. A value of AUC of 0.7 is often accepted as a threshold for moderately discriminating forecasts (Swets 1988). Unlike the reliability diagram analysis, the ROC/AUC does not penalize miscalibration, reflecting the fact that even miscalibrated (biased) forecasts can, in practice, still provide useful guidance and can be improved by proper calibration.

The performance of the LRM and the benchmark model was evaluated both in a repeated cross-validation setting using data from 2006 to 2014 and in an independent verification mode using the withheld data for 2015–16. The repeated cross validation (20 times five folds) consists of randomly splitting the training dataset into five “folds” of similar size; one fold is withheld for validation, while the remaining four are used for calculating model coefficients; this process is repeated for each of the five folds; this random splitting procedure is repeated 20 times. A similar repeated fivefold cross validation was performed for the logistic regression built using the CWI. The reason fivefold cross validation was selected is that the length of each fold (approximately 213 days) is similar in length to the independent verification dataset (235 days), allowing for a comparison of several independent verification measures with those based on training-period results for a dataset of similar length.

## 4. Results

### a. Comparison of model distributions

The distribution of modeled probabilities of waterspouts on report versus no-report days is shown in Fig. 7 as box-and-whisker diagrams. Here, and in subsequent diagrams, colored boxes represent the interquartile range (25th–75th percentile), whiskers span from the 10th to the 90th percentile, and the notch around the median represents the 95% confidence interval for the median’s true value. Only the CWI results are shown here as a benchmark for the LRM performance, since the BRI, KHI, and SWI demonstrated no appreciable skill in differentiating between report and no-report days.

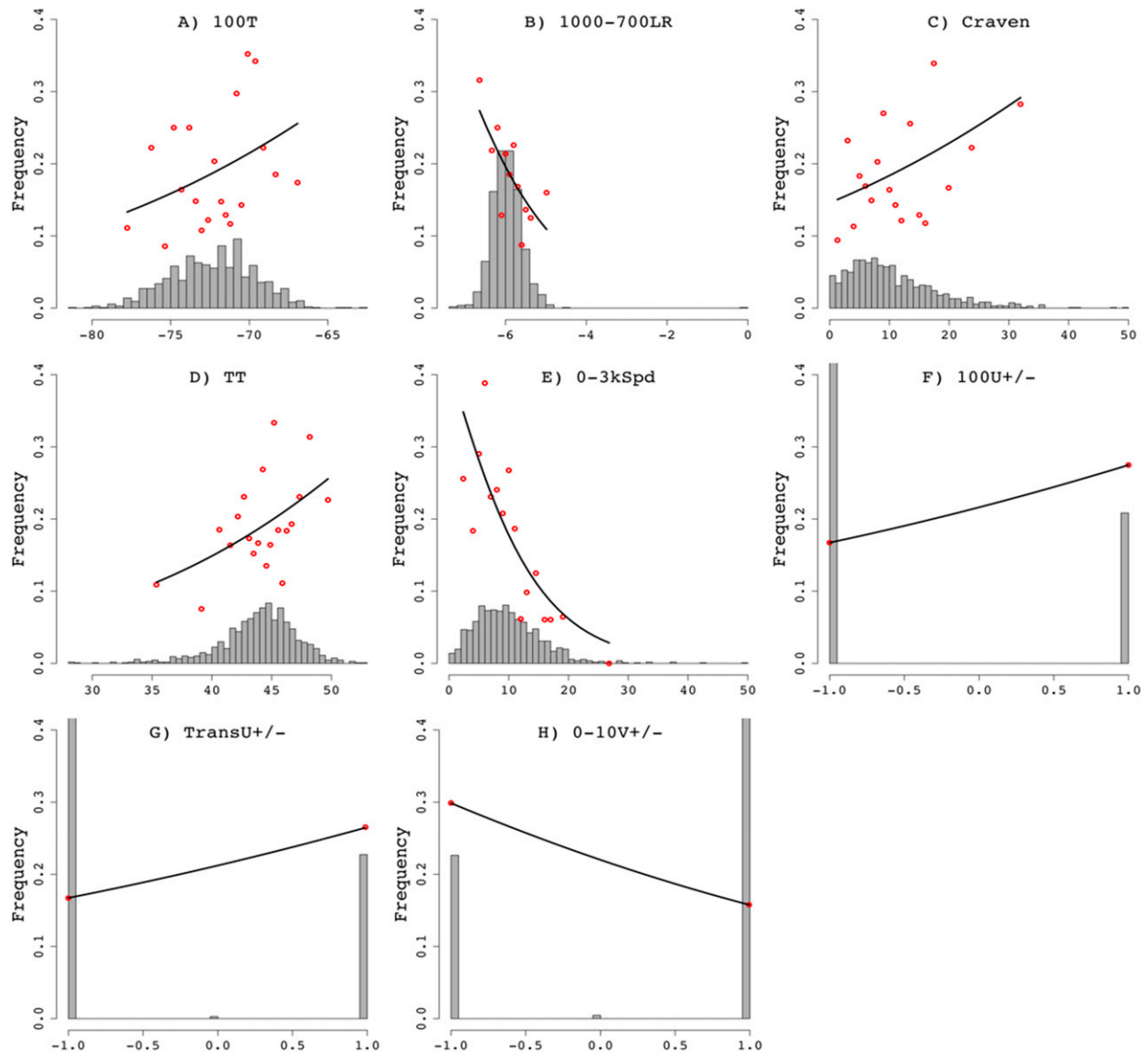


FIG. 6. Frequency distribution (gray bars), conditional frequency of waterspout reports (red dots), and single-variable logistic regression fit (black line) for (a) 100-mb temperature (K), (b) 1000–700-mb lapse rates ( $\text{K km}^{-1}$ ), (c) the Craven significant severe parameter ( $\times 10^3 \text{ m}^3 \text{ s}^{-3}$ ), (d) the TT index (unitless), and (e) 0–3-kft wind speed (kt, where  $1 \text{ kt} = 0.51 \text{ m s}^{-1}$ ), (f) 100-mb  $U$ -wind sign (unitless), (g) transport mean  $U$ -wind sign (unitless), and (h) 0–10-kft  $V$ -wind sign (unitless).

The climatological probability of waterspout reports for the wet season, based on the training period, 2006–14, is 19.0%. On no-report days, the cross-validated LRM predicted a median probability of 14.3% (mean 17.4%), while on waterspout report days the median predicted probability was 22.8% (mean 26.0%) (Fig. 7a, red). The difference in means across the two groups is statistically significant ( $p$  value  $< 2.2 \times 10^{-16}$ ). The 95% confidence intervals of the medians of the forecasts on no-report and report days do not overlap, indicating statistically significant differences in the medians as well.

The cross-validated CWI model median probability on no-report days was 17.8% (mean 18.5%), and on report days it was 20.6% (mean 20.8%) (Fig. 7a, blue), and the difference in means was also statistically significant ( $p$  value  $< 2.2 \times 10^{-16}$ ). Here, too, the 95% confidence intervals of the medians of the forecasts on no-report and report days do not overlap, indicating statistically significant differences in the medians as well.

Also statistically significant are the differences between the mean and median forecasts produced by the two models (LRM and CWI) on no-report days

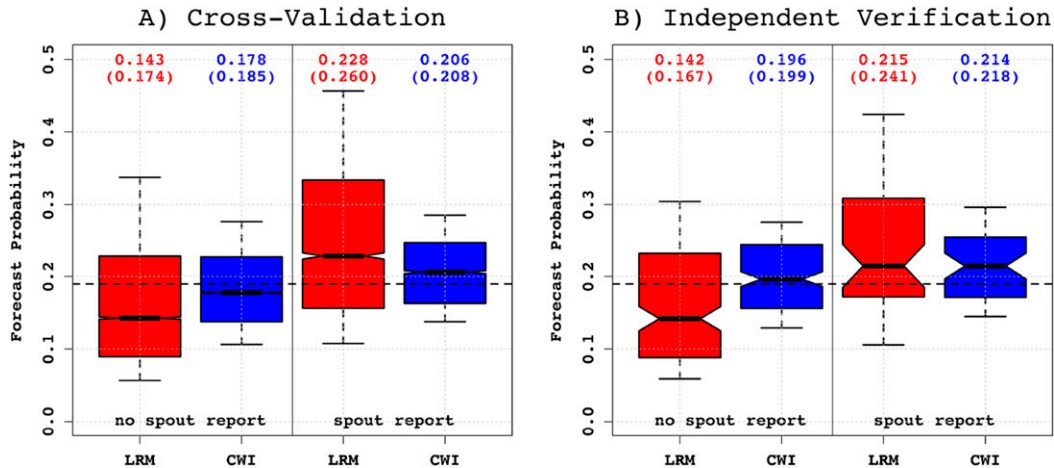


FIG. 7. Boxplot of the LRM and CWI modeled waterspout probability on report vs no-report days from (a) cross validation and (b) independent verification. Whiskers show the 10th and 90th percentiles. The notch around the median (thick black line) indicates the 95% confidence interval for the median’s true value. The horizontal dashed line indicates the climatological frequency of waterspouts. The median value for each box-and-whisker plot is shown along the top of the plot, and the mean value is shown in parentheses beneath it.

( $p$  value  $< 2.2 \times 10^{-16}$  for the difference of means), as well as on report days ( $p$  value  $< 2.2 \times 10^{-16}$  for the difference of means). The comparison in Fig. 7 shows that there is a wider range of probabilities produced by the LRM relative to the CWI model on both no-report and report days, and that the two CWI distributions are less distinguishable from each other than are the two LRM distributions. The 10th–90th percentile range for the LRM spans from 5.7% to 33.7% on no-report days and from 10.8% to 45.7% on report days. For the CWI model, the 10th–90th percentile spans from 10.6% to 27.6% on no-report days and from 13.8% to 28.5% on report days.

The average probability of waterspout reports for the wet season, based on the 2015–16 dataset, was roughly 22.5%, somewhat higher than that of the training period (see Fig. 1 for the monthly mean frequencies during the verification versus the training period). On no-report days, the independent verification with the LRM predicted a median probability of 14.2% (mean 16.7%), while on waterspout report days the median predicted probability was 21.5% (mean 24.1%) (Fig. 7b, red). The difference in means across the two groups is statistically significant ( $p$  value =  $9.1 \times 10^{-5}$ ), and the 95% confidence intervals for the median forecast values on report versus no-report days are clearly separated.

The independent verification with the CWI model for no-report days had a median probability of 19.6% (mean 19.9%), while on report days it was 21.4% (21.8%) (Fig. 7b, blue), and the difference in means had a weaker statistical significance ( $p$  value = 0.04). The 95% confidence intervals for the median forecast

values on report versus no-report days are marginally separated.

Also statistically significant are the differences between the mean and median forecasts produced by the two models (LRM and CWI) on no-report days ( $p$  value =  $3.0 \times 10^{-4}$  for the difference of means). The difference between the two models on report days is not clearly statistically significant, with a  $p$  value = 0.22 for the difference of means test and some overlap between the 95% confidence intervals for the medians. Similarly to the cross-validation results, the two CWI distributions are less distinguishable from each other than are the two LRM distributions. The 10th–90th percentile range for the LRM spans from 5.9% to 30.4% on no-report days and from 10.6% to 42.4% on report days. For the CWI model, the 10th–90th percentiles span from 12.9% to 27.5% on no-report days and from 14.5% to 29.6% on report days.

The above comparison of model distributions demonstrates that both the LRM and CWI outperform climatology, and that the LRM outperforms the CWI in terms of differentiating between days with no waterspout reports and days with waterspout reports. The average predicted frequency on days during which no waterspouts are reported is lower in the LRM model than in the CWI model. Conversely, on days for which waterspouts are reported, the average predicted frequency is higher in the LRM than in the CWI. The CWI forecast frequency distribution has a similar and highly overlapping range on report versus no-report days and is more clustered around climatology. The LRM distributions demonstrate a greater difference in ranges and a

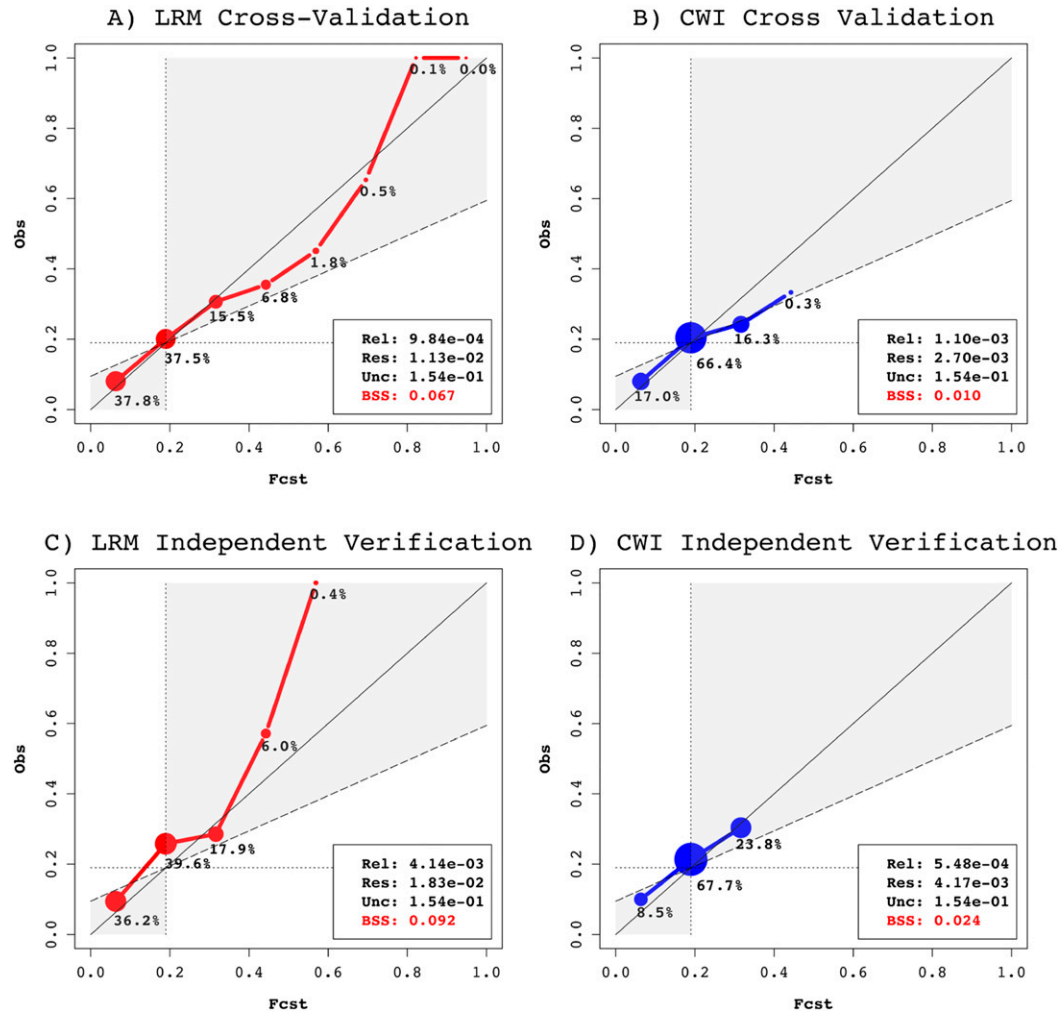


FIG. 8. Reliability diagram for modeled waterspout probability for the (a) cross-validated LRM, (b) cross-validated CWI, (c) independent verification LRM, and (d) independent verification CWI. The point size reflects the forecast bin frequency, also denoted as a percentage. Values for the Brier score decomposition into reliability (Rel), resolution (Res), and uncertainty (Unc), as well as the BSS, are printed in the bottom-right corner of each panel.

smaller overlap between report and no-report days and are less clustered around climatology. The above results are seen both in cross-validation and independent verification mode.

#### b. Reliability diagram and Brier skill score

For a complementary look of the two models' performances, reliability diagrams were constructed (Fig. 8). Here, the predicted probabilities were separated into evenly spaced bins between 0 and 1 (horizontal axis) with width equal to  $\frac{2}{3}$  of the climatological probability for 2006–14, and the forecast probability in each bin was assigned its midpoint value. The observed probability (vertical axis) for each bin was calculated as the mean of the corresponding validating observations.

The Brier score decomposition (Wilks 2006) into reliability, resolution, and uncertainty terms, and the Brier skill score (BSS) relative to climatology, are indicated in each panel in Fig. 8. The uncertainty term is determined purely by the observations' distribution and is not influenced by the forecast. Note that a more skillful forecast will have a smaller reliability term (indicating well-calibrated forecasts) and a larger resolution term (indicating the presence of forecasts farther from the climatological mean). The Brier skill score, on the other hand, is larger for more skillful forecasts and indicates skill improvement relative to a forecast of climatology. The BSS for a forecast as skillful as climatology is 0, it is negative for forecast less skillful than climatology, and it is 1 for a perfect deterministic forecast. This theoretical

upper limit of the BSS is, in practice, not a realistic target for probabilistic forecasts (Palmer et al. 2000), and even small positive values can be indicative of forecast skill (Jolliffe and Stephenson 2008).

For comparison, we show that for the repeated cross validation the LRM (Fig. 8a) is an improvement over the CWI model (Fig. 8b). The LRM reliability term is slightly smaller than that of the CWI, indicating a slightly better-calibrated model. On the graph this is evidenced by the closer proximity of the LRM forecast/observation data points to the diagonal (perfect reliability) line. The resolution term of the LRM model is larger than that of the CWI model, indicating that the LRM produces more frequent forecasts that are farther away from the climatological mean. On the graph this is evidenced by the size of the dots representing the forecast/observation data points (and specified as a percentage of all forecasts), which reiterates that the CWI model is more conservative, with most forecasts lying near the climatological probability of waterspout occurrence. In contrast, the LRM more frequently, and reliably, predicts probabilities above or below the climatological value. Only 37.5% of LRM forecasts fall into the “climatology” bin, compared to some 66.4% of CWI forecasts. The Brier skill score of the LRM model (0.067) is several times larger than that of the CWI (0.010).

Box-and-whisker diagrams (Fig. 9) of the Brier skill scores were obtained from the individual folds of the 20-times repeated fivefold cross validation for the LRM (red) and CWI (blue). Also shown for reference are the Brier skill scores of similarly obtained results for the three other models (BRI, KHI, and SWI). There is a clear separation between the Brier skill score distributions for the two models, as illustrated by the lack of overlap in the interquartile ranges. The difference in means is statistically significant ( $p$  value  $< 2.2 \times 10^{-16}$ ), and the 95% confidence intervals for the medians are well separated. The 10th–90th percentile range of Brier skill scores for single folds of the LRM cross validation spans from 0.014 to 0.121 with a median value of 0.065, while for the CWI model this span is from  $-0.014$  to 0.039 with a median value of 0.010. The range for BRI spans from 0 to  $8.5 \times 10^{-5}$  with a median of  $4 \times 10^{-11}$ , for KHI it is from  $-6 \times 10^{-10}$  to  $9 \times 10^{-5}$  with a median of  $4 \times 10^{-11}$ , and for SWI it is from  $-9 \times 10^{-3}$  to  $4 \times 10^{-3}$  with a median of  $9 \times 10^{-4}$ .

The results from independent verification of the LRM and CWI in terms of reliability diagrams and Brier skill scores (Figs. 8c,d) are generally similar to those from the cross validation. Although the reliability term for the LRM is larger than that for the CWI, indicating a better calibration of the latter model, this is counterbalanced

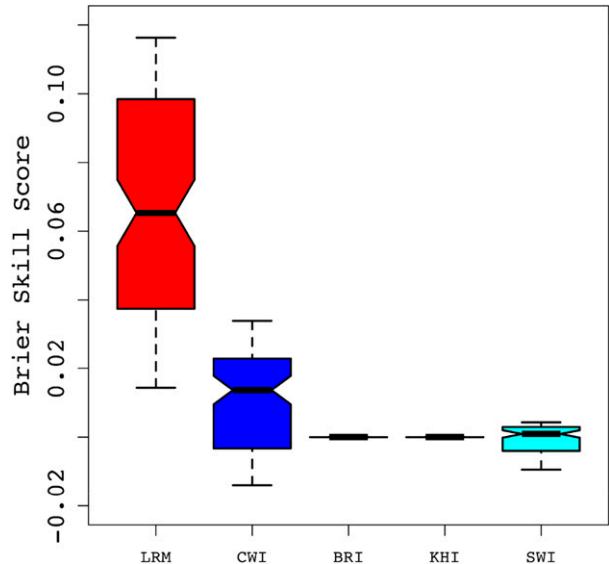


FIG. 9. BSSs for individual folds of the 20-times fivefold cross validation for the LRM (red) and CWI (blue), as well as the BRI (green), KHI (orange), and SWI (cyan).

by the improved (larger) resolution term in the LRM compared to the CWI. The overall Brier skill score for the LRM is 0.092, larger than that of the CWI (0.024). These values are within the range of values for individual folds of the cross validation (recall that the length of data in an individual fold is similar to the length of data in the independent verification period) seen in Fig. 9, confirming that, when equitably compared, the model performance in independent verification mode is within the range of performance seen in cross validation.

The above comparison of the reliability diagrams demonstrates that both the LRM and CWI are well calibrated, but the resolution of the LRM forecasts is an improvement over that of the CWI forecasts, that is, that the LRM more frequently (and statistically correctly) produces forecasts that are farther away from climatology. This is reflected in the improvement of the Brier skill score of the LRM relative to the CWI. These results are seen both in cross-validation and independent verification mode. The performance of the remaining three models evaluated here (BRI, KHI, and SWI) is significantly below that of both the LRM and CWI and only slightly above climatology.

### c. Receiver operating characteristic and area under the curve

Examination of the ROC curves (Fig. 10) provides a complementary comparison of the performance of the LRM and CWI models, in terms of their ability to discriminate between the two outcomes (report versus



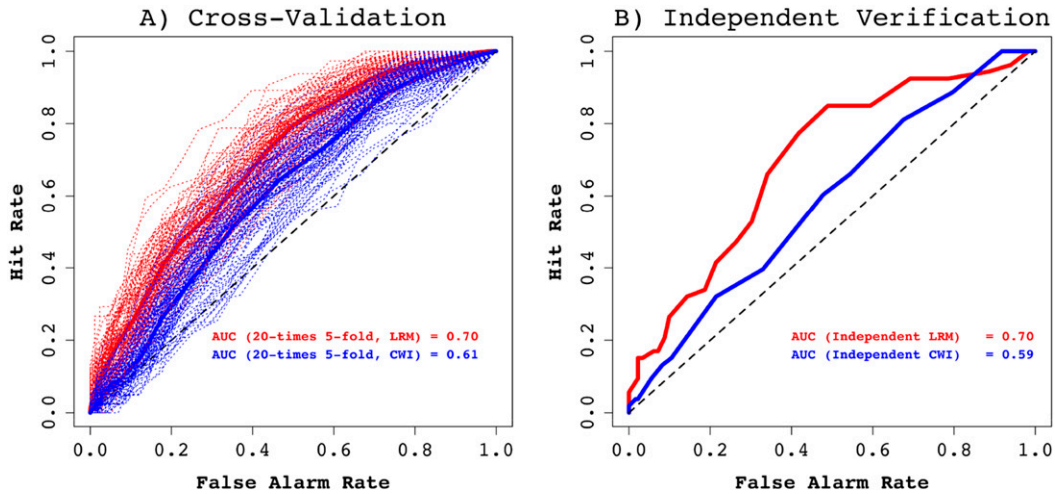


FIG. 10. ROC curves for the (a) cross validation and (b) independent verification of waterspout probability for the LRM (red) and for the CWI model (blue). AUC values are indicated in the bottom-right corner of each panel. In (a), the dashed lines show the ROC curves for individual folds, and solid lines are for the entire cross validation.

no-report days). The ROC/AUC for repeated cross validation demonstrates that the discrimination of the LRM (Fig. 10a, solid red line) is an improvement over CWI (Fig. 10a, solid blue line), since the LRM ROC is above and to the left of the CWI ROC. The AUC for the LRM is 0.70, greater than the CWI AUC value of 0.61. The ROC values for individual cross-validation folds are depicted in Fig. 10a with dashed red lines for the LRM and dashed blue lines for the CWI, thus illustrating the variability within subsets of the data and providing a reference range for the independent verification period.

The box-and-whisker plot of AUC values for individual folds is depicted in Fig. 11. There is a clear separation between the AUC distributions for the two models, as illustrated by the lack of overlap in the interquartile ranges. The difference in means is statistically significant ( $p$  value  $< 2.2 \times 10^{-16}$ ), and the 95% confidence intervals for the medians are well separated. The 10th–90th percentile range of AUC for single folds of the LRM cross validation spans from 0.66 to 0.74, with a median value of 0.70, while for the CWI model this span is from 0.55 to 0.66, with a median value of 0.62. The BRI spans from 0.47 to 0.59, with a median value of 0.54; the KHI spans from 0.49 to 0.58, with a median value of 0.54; and the SWI spans from 0.48 to 0.57, with a median value of 0.52.

The results from independent verification of the LRM and CWI in terms of ROC/AUC (Fig. 10b) are generally similar to those from the cross validation. The LRM ROC is almost entirely above and to the left of the CWI ROC. The AUC for the LRM is 0.70, greater than the value for the CWI of 0.59. These values are within the range of values for individual folds of the cross

validation (recall that the length of data in an individual fold is similar to the length of data in the independent verification period) seen in Fig. 11, illustrating that the model performance in independent verification mode is within the range of performance obtained during cross validation.

The ROC/AUC results of both the LRM and CWI models demonstrate that both models are an improvement over a random forecast, and the performance of

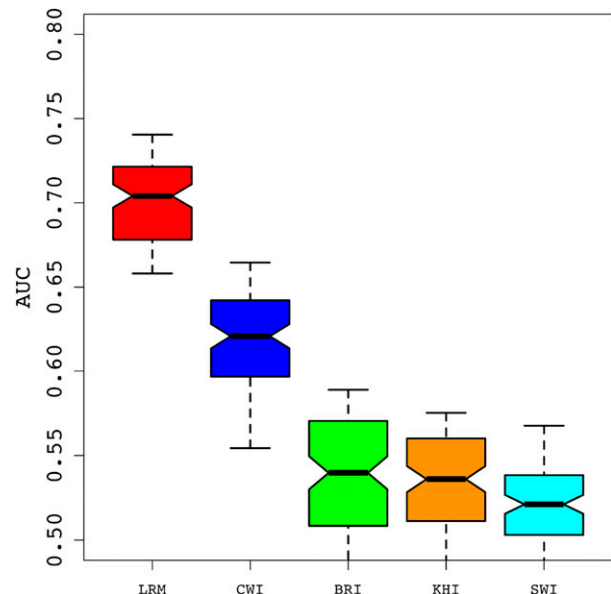


FIG. 11. AUC for individual folds of the 20-times fivefold cross validation for the LRM (red) and CWI (blue), as well as the BRI (green), KHI (orange), and SWI (cyan).

the LRM is superior to that of the CWI in both the cross validation and independent verification. The LRM AUC is at the threshold of what is considered a moderately well discriminating model, where the CWI fails to meet that threshold. The performance of the remaining three models evaluated here (BRI, KHI, and SWI) is significantly below that of both the LRM and CWI and only slightly above climatology.

## 5. Summary and discussion

More than 200 separate variables were extracted from 1200 UTC Key West soundings for the period 2006–14 for the wet-season months from June through September. The variables were separated into two subsets: days when at least one waterspout was reported and days with no waterspout reports. The number of variables considered needed to be reduced in order to obtain a workable number of candidate predictors for logistic regression analysis. Each variable was examined for a statistically significant difference in the means between the two subsets. Variables for which the difference in means between the two subsets met or exceeded the 95% confidence level were retained for further analysis, and the rest were excluded from further consideration. The number of candidate predictors was further reduced by requiring that no two variables had an absolute correlation exceeding 0.6. AIC was used as a criterion to perform an exhaustive search to select a best model based on this reduced set of predictors. This resulted in an eight-predictor model, the LRM. The LRM obtained in this manner is but one model solution that resulted from the particular choice of the model selection procedure, and other model configurations, obtained by alternative means, may perform equally well or better.

The impact of each predictor in the LRM was examined by comparing against the null model and ensuring that the inclusion of each successive predictor, in order of importance, reduced the model AIC and residual deviance. The LRM was then compared against the benchmark model (CWI) using repeated cross validation, and its potential for operational forecasting was further evaluated by building the logistic regression model with the 2006–14 wet-season data and applying the resulting model to an independent, unused set constructed with data from the 2015–16 wet seasons. These evaluations demonstrated that the multivariable logistic regression, LRM, built using a small set of predictor variables derived from morning sounding data at Key West, produced a quantitative probability forecast with robust skill relative to climatology, and yields results that exceeded those of an existing operational index, CWI, validated using the same methodology. The

models' performances were compared in terms of several complementary measures: discrimination, reliability, resolution, Brier skill score, receiver operating characteristic (ROC) curves, and area under the curve (AUC). It was found that the LRM outperformed the benchmark model in terms of all evaluated metrics for both cross validation and independent verification, thus providing a more skillful probabilistic forecast with greater potential utility.

Even though the LRM performance exceeds that of the CWI model, it is important to note the results make no statement of the performance of the CWI for the Charleston coastal environment, or elsewhere. The CWI was used here only as a benchmark because the two models have similar origins, being built from sounding-derived parameters, and the CWI demonstrates an improvement over climatology. Three other additional indices described in the literature were considered as potential benchmarks, but most of these results are not shown since their performance was only marginally better than climatology.

The LRM models the chance of a waterspout being reported, rather than directly the chance of a waterspout occurring. However, it is assumed that the probability of a waterspout being reported is proportional to waterspout existence. This leaves the potential for bias since waterspouts occurring farther offshore may be witnessed and/or reported at a lesser rate. If indeed waterspouts farther offshore are underrepresented in the reports relative to nearshore waterspouts, the model may favor flow regimes that are preferentially associated with nearshore waterspouts.

The LRM is trained for waterspout development over the waters surrounding the Florida Keys and likely will not perform as well over other coastal waters since, for example, the shape of the coastline is reflected in favorable surface wind flow, and the upper troposphere may not carry the same importance in other locations as it does over the Florida Keys. Additionally, the climatological occurrence of waterspouts will differ by location, which will affect the training of the model. One possible impediment to evaluating the LRM methodology in other locations would be a low occurrence rate of waterspouts. Given a low climatological occurrence in other locations, the resulting model may not be statistically stable ([World Meteorological Organization 1999](#), section 3.3.2). However, the methodology used to produce the LRM is expected to produce favorable results in other areas and will be examined in further work.

The LRM may prove useful in daily operational forecasting of waterspout development, either alone or in conjunction with other guidance tools. We found that when the LRM forecast was above climatology, the

TABLE 3. Relative forecast frequency and corresponding waterspout report frequency for the following forecast outcomes: the LRM and CWI individually or jointly above/below climatology (0.19) and individually or jointly above 0.30. Cross-validation results and independent verification results (in parentheses).

Forecast outcome	Relative forecast frequency (%)	Waterspout report frequency (%)
LRM > 0.19	40.4 (39.1)	29.1 (34.8)
LRM < 0.19	59.6 (60.9)	12.1 (14.7)
	CWI > 0.19	24.9 (26.4)
	CWI < 0.19	13.9 (17.9)
LRM > 0.19	CWI > 0.19	30.0 (36.8)
LRM > 0.19	CWI < 0.19	27.2 (29.2)
LRM < 0.19	CWI > 0.19	17.6 (14.8)
LRM < 0.19	CWI < 0.19	09.5 (14.6)
LRM > 0.3		35.5 (42.9)
	CWI > 0.3	23.5 (33.3)
LRM > 0.3	CWI > 0.3	25.7 (66.7)

probability of a waterspout being observed was about 29% (35%) in cross-validation (independent verification) mode (Table 3), about 1.5 times higher than the climatological probability. When the LRM forecast was below climatology, the corresponding probability of waterspouts was about 12% (15%). For CWI forecasts these numbers were about 25% (26%) and 14% (18%), respectively. When both LRM and CWI forecasts were above climatology, waterspouts were observed in 30% (37%) of cases for cross validation (independent verification). When both forecasts were below climatology, these numbers were about 10% (15%). When only forecasts exceeding approximately 1.5 times the climatological probability were considered, however, there was no benefit to considering the CWI in addition to the LRM. This suggests that in some circumstances, the specifics of which merit further investigation, there may be a marginal utility in considering both models in combination. Such guidance would likely be most beneficial when used in conjunction with long-established forecaster rules of thumb.

## 6. Future work

One of the surprises to come from this research was the apparent influence of the upper troposphere, which was uncovered statistically in the development of the LRM. Waterspouts were previously thought to be a shallow phenomenon, with cloud tops generally around 3 km (Rossow 1970), and Golden (2003) alluded to the importance of weak lower-tropospheric troughs on days when large waterspouts or multiple waterspouts were reported. A statistical evaluation of favorable synoptic conditions through the depth of the troposphere should be examined, with an emphasis on transient upper-tropospheric conditions as well as the potential effects of tropical waves and/or perturbations in the easterly

flow. Additionally, examination of waterspout breakout days (e.g., subjectively defined as days with three or more reports) may be helpful in defining an overall favorable synoptic pattern or relevant intraseasonal and interseasonal changes.

It may be worth expanding the dataset to years prior to 2006. Care will have to be taken to ensure following this course does not degrade the quality of the dataset. One of the waterspout reporting practices employed prior to 2006 was to identify waterspouts as funnel clouds if a spray ring was not spotted, or if the funnel did not descend to the horizon. Therefore, reports of funnel clouds over water, in “fair weather” conditions, would likely need to be considered waterspout reports. Even then, occasionally a storm report would not be issued for a brief funnel cloud.

Even though the radiosonde sounding at the Miami, Florida, NWS office is well inland (approximately 15 km), and may not be representative of the marine boundary layer, it should be examined for utility in developing future iterations of the LRM. Along these lines, the use of modeled soundings (with an emphasis on high-resolution mesoscale models) may also be useful in identifying upcoming favorable waterspout days. This would be helpful in extended outlook products, such as the NWS’s hazardous weather outlooks.

*Acknowledgments.* The authors are grateful for the help of Kennard Kasper, Sean Daida, and Melody Lovin at the National Weather Service in Key West for their work on graphics used in this paper as well as edits and comments. The authors also acknowledge the work of Hollings Scholar student Taylor Adams of Valparaiso University on background material and additional research. This work was funded by the United States Government, Department of Commerce, National Weather Service.

TABLE A1. Comprehensive list of all candidate variables obtained from raob (numbered in column  $N$ ), candidate variables retained after the  $t$ -test screening (numbered in column  $N_0$ ), candidate variables retained after subsequent correlation screening (numbered in column  $N_1$ ), and final list of variables after subsequent best model selection based on AIC criterion (numbered in column  $N_2$ ). The abbreviated names for candidate variables retained after the correlation screening are shown in boldface, and the names for the final variables are shown in boldface italics. Also listed are the mean values of each variable on no-report vs report days and the  $p$  values for the  $t$  test of the difference of those means.

$N$	$N_0$	$N_1$	$N_2$	Abbreviated variable name	Mean value on no-report days	Mean value on report days	$p$ value for $t$ test
1	1	1	1	<b><i>0–3 kft AGL wind speed</i></b>	$1.08 \times 10^1$	$8.00 \times 10^0$	$2.15 \times 10^{-16}$
2	2			1000-mb wind speed	$9.14 \times 10^0$	$6.85 \times 10^0$	$2.86 \times 10^{-15}$
3	3			925-mb wind speed	$1.19 \times 10^1$	$8.91 \times 10^0$	$9.61 \times 10^{-15}$
4	4			Transport mean speed	$9.24 \times 10^0$	$6.90 \times 10^0$	$2.59 \times 10^{-13}$
5	5			Surface wind speed	$5.53 \times 10^0$	$4.40 \times 10^0$	$7.80 \times 10^{-11}$
6	6			0–10-kft wind speed	$1.02 \times 10^1$	$7.93 \times 10^0$	$8.46 \times 10^{-11}$
7	7			Fire danger index	$9.52 \times 10^0$	$7.56 \times 10^0$	$1.17 \times 10^{-10}$
8	8			Wind shear 0–1 km	$3.89 \times 10^0$	$2.99 \times 10^0$	$1.32 \times 10^{-9}$
9	9			850-mb wind speed	$1.10 \times 10^1$	$8.80 \times 10^0$	$1.11 \times 10^{-8}$
10	10			925–700-mb wind speed	$1.04 \times 10^1$	$8.32 \times 10^0$	$1.98 \times 10^{-8}$
11	11			3–10 kft AGL wind speed	$1.04 \times 10^1$	$8.40 \times 10^0$	$7.27 \times 10^{-8}$
12	12			EHI	$2.81 \times 10^{-1}$	$1.88 \times 10^{-1}$	$6.90 \times 10^{-7}$
13	13			Transport peak speed	$9.95 \times 10^0$	$8.09 \times 10^0$	$2.55 \times 10^{-6}$
14	14			Fog stability index	$4.07 \times 10^1$	$3.87 \times 10^1$	$2.00 \times 10^{-5}$
15	15	2	2	<b><i>0–10 kft AGL V+/-</i></b>	$5.99 \times 10^{-1}$	$2.96 \times 10^{-1}$	$3.33 \times 10^{-5}$
16	16	3		<b><i>1000–500-mb lapse rate</i></b>	$-5.86 \times 10^0$	$-5.93 \times 10^0$	$1.99 \times 10^{-4}$
17	17			3–10 kft AGL V+/-	$5.41 \times 10^{-1}$	$2.77 \times 10^{-1}$	$3.38 \times 10^{-4}$
18	18	4		<b><i>Significant tornado parameter (STP)</i></b>	$4.42 \times 10^{-2}$	$2.18 \times 10^{-2}$	$3.42 \times 10^{-4}$
19	19	5	3	<b><i>1000–700-mb lapse rate</i></b>	$-5.89 \times 10^0$	$-5.99 \times 10^0$	$4.84 \times 10^{-4}$
20	20	6		<b><i>250-mb temp</i></b>	$-4.25 \times 10^1$	$-4.28 \times 10^1$	$6.28 \times 10^{-4}$
21	21			925–700-mb V+/-	$5.51 \times 10^{-1}$	$3.01 \times 10^{-1}$	$6.30 \times 10^{-4}$
22	22			400-mb temp	$-1.74 \times 10^1$	$-1.78 \times 10^1$	$6.83 \times 10^{-4}$
23	23	7		<b><i>700-mb U+/-</i></b>	$-2.83 \times 10^{-1}$	$-2.43 \times 10^{-2}$	$8.63 \times 10^{-4}$
24	24	8		<b><i>850–500-mb thickness</i></b>	$1.43 \times 10^4$	$1.42 \times 10^4$	$9.39 \times 10^{-4}$
25	25			600–400-mb thickness	$1.04 \times 10^4$	$1.03 \times 10^4$	$1.08 \times 10^{-3}$
26	26			700–500-mb thickness	$8.89 \times 10^3$	$8.88 \times 10^3$	$1.30 \times 10^{-3}$
27	27			600-mb U+/-	$-2.29 \times 10^{-1}$	$1.94 \times 10^{-2}$	$1.43 \times 10^{-3}$
28	28	9	4	<b><i>100-mb U+/-</i></b>	$-6.15 \times 10^{-1}$	$-3.89 \times 10^{-1}$	$1.45 \times 10^{-3}$
29	29			500-mb temp	$-6.76 \times 10^0$	$-7.06 \times 10^0$	$1.80 \times 10^{-3}$
30	30	10		<b><i>700-mb wind speed</i></b>	$1.08 \times 10^1$	$9.46 \times 10^0$	$1.95 \times 10^{-3}$
31	31	11		<b><i>100-mb height</i></b>	$1.66 \times 10^4$	$1.66 \times 10^4$	$2.11 \times 10^{-3}$
32	32	12	5	<b><i>Transport mean U+/-</i></b>	$-5.81 \times 10^{-1}$	$-3.64 \times 10^{-1}$	$2.31 \times 10^{-3}$
33	33			150-mb height	$1.42 \times 10^4$	$1.42 \times 10^4$	$2.53 \times 10^{-3}$
34	34			700-mb temp	$9.05 \times 10^0$	$8.78 \times 10^0$	$2.61 \times 10^{-3}$
35	35	13	6	<b><i>TT index</i></b>	$4.38 \times 10^1$	$4.46 \times 10^1$	$2.70 \times 10^{-3}$
36	36	14		<b><i>Fog advection probability</i></b>	$4.12 \times 10^1$	$3.96 \times 10^1$	$2.74 \times 10^{-3}$
37	37			925-mb V+/-	$6.86 \times 10^{-1}$	$4.90 \times 10^{-1}$	$3.02 \times 10^{-3}$
38	38			200-mb temp	$-5.43 \times 10^1$	$-5.46 \times 10^1$	$3.03 \times 10^{-3}$
39	39			300-mb temp	$-3.25 \times 10^1$	$-3.28 \times 10^1$	$3.25 \times 10^{-3}$
40	40			850-mb U+/-	$-4.49 \times 10^{-1}$	$-2.33 \times 10^{-1}$	$3.95 \times 10^{-3}$
41	41	15		<b><i>700-mb V+/-</i></b>	$3.76 \times 10^{-1}$	$1.60 \times 10^{-1}$	$4.30 \times 10^{-3}$
42	42			1000–500-mb thick	$1.89 \times 10^4$	$1.89 \times 10^4$	$4.60 \times 10^{-3}$
43	43	16		<b><i>1000-mb V+/-</i></b>	$3.93 \times 10^{-1}$	$1.80 \times 10^{-1}$	$4.82 \times 10^{-3}$
44	44			Boyden index	$9.65 \times 10^1$	$9.67 \times 10^1$	$4.98 \times 10^{-3}$
45	45			Surface U+/-	$-6.27 \times 10^{-1}$	$-4.42 \times 10^{-1}$	$5.49 \times 10^{-3}$
46	46			600-mb temp	$1.65 \times 10^0$	$1.39 \times 10^0$	$6.23 \times 10^{-3}$
47	47			Transport mean V+/-	$4.78 \times 10^{-1}$	$2.77 \times 10^{-1}$	$6.43 \times 10^{-3}$
48	48			Transport peak U+/-	$-5.42 \times 10^{-1}$	$-3.54 \times 10^{-1}$	$8.76 \times 10^{-3}$
49	49	17	7	<b><i>Craven parameter</i></b>	$1.06 \times 10^1$	$1.23 \times 10^1$	$9.28 \times 10^{-3}$
50	50			200-mb height	$1.24 \times 10^4$	$1.24 \times 10^4$	$9.41 \times 10^{-3}$
51	51			850–700-mb thickness	$5.37 \times 10^3$	$5.37 \times 10^3$	$9.43 \times 10^{-3}$
52	52	18		<b><i>Corfidi downshear speed</i></b>	$1.61 \times 10^1$	$1.81 \times 10^1$	$9.55 \times 10^{-3}$

TABLE A1. (Continued)

<i>N</i>	<i>N</i> <sub>0</sub>	<i>N</i> <sub>1</sub>	<i>N</i> <sub>2</sub>	Abbreviated variable name	Mean value on no-report days	Mean value on report days	<i>p</i> value for <i>t</i> test
53	53			850-mb <i>V</i> +/-	$5.85 \times 10^{-1}$	$4.08 \times 10^{-1}$	$1.02 \times 10^{-2}$
54	54			Significant hail parameter (SHIP)	$2.10 \times 10^{-1}$	$2.51 \times 10^{-1}$	$1.03 \times 10^{-2}$
55	55			1000-mb <i>U</i> +/-	$-6.30 \times 10^{-1}$	$-4.56 \times 10^{-1}$	$1.08 \times 10^{-2}$
56	56			Freezing level height	$1.55 \times 10^4$	$1.54 \times 10^4$	$1.08 \times 10^{-2}$
57	57	19		<b>500-mb</b> <i>V</i> +/-	$4.58 \times 10^{-3}$	$-1.89 \times 10^{-1}$	$1.08 \times 10^{-2}$
58	58			Bulk wind shear 6 km	$5.31 \times 10^0$	$5.95 \times 10^0$	$1.09 \times 10^{-2}$
59	59	20	8	<b>100-mb temp</b>	$-7.24 \times 10^1$	$-7.18 \times 10^1$	$1.11 \times 10^{-2}$
60	60			Wind shear speed 0-1 km	$1.03 \times 10^1$	$1.16 \times 10^1$	$1.18 \times 10^{-2}$
61	61			Freezing level pressure	$5.79 \times 10^2$	$5.82 \times 10^2$	$1.19 \times 10^{-2}$
62	62			0-3 kft AGL <i>U</i> +/-	$-5.83 \times 10^{-1}$	$-4.08 \times 10^{-1}$	$1.21 \times 10^{-2}$
63	63	21		<b>Cloud depth ratio</b>	$4.41 \times 10^{-1}$	$4.35 \times 10^{-1}$	$1.56 \times 10^{-2}$
64	64	22		<b>400-mb</b> <i>U</i> +/-	$-6.07 \times 10^{-2}$	$1.26 \times 10^{-1}$	$1.59 \times 10^{-2}$
65	65			250-mb height	$1.09 \times 10^4$	$1.09 \times 10^4$	$1.60 \times 10^{-2}$
66	66			TQ index	$1.49 \times 10^1$	$1.55 \times 10^1$	$1.76 \times 10^{-2}$
67	67			0-3-kft <i>V</i> +/-	$5.78 \times 10^{-1}$	$4.17 \times 10^{-1}$	$2.07 \times 10^{-2}$
68	68			Cross totals	$1.91 \times 10^1$	$1.96 \times 10^1$	$2.11 \times 10^{-2}$
69	69			850-70-mb lapse rate	$-5.45 \times 10^0$	$-5.56 \times 10^0$	$2.11 \times 10^{-2}$
70	70			300-mb height	$9.68 \times 10^3$	$9.68 \times 10^3$	$2.48 \times 10^{-2}$
71	71			S index	$3.38 \times 10^1$	$3.51 \times 10^1$	$2.48 \times 10^{-2}$
72	72			Transport peak <i>V</i> +/-	$4.00 \times 10^{-1}$	$2.33 \times 10^{-1}$	$2.55 \times 10^{-2}$
73	73			Showalter	$1.15 \times 10^0$	$7.51 \times 10^{-1}$	$2.93 \times 10^{-2}$
74	74			850-500-mb lapse rate	$-5.69 \times 10^0$	$-5.75 \times 10^0$	$2.95 \times 10^{-2}$
75	75	23		<b>Vorticity generation parameter</b>	$2.36 \times 10^{-1}$	$2.22 \times 10^{-1}$	$3.66 \times 10^{-2}$
76	76			Surface <i>V</i> +/-	$3.24 \times 10^{-1}$	$1.70 \times 10^{-1}$	$3.75 \times 10^{-2}$
77	77			925-700-mb <i>U</i> +/-	$-3.88 \times 10^{-1}$	$-2.33 \times 10^{-1}$	$3.85 \times 10^{-2}$
78	78			3-10-kft <i>U</i> +/-	$-3.83 \times 10^{-1}$	$-2.28 \times 10^{-1}$	$3.95 \times 10^{-2}$
79	79	24		<b>500-mb</b> <i>U</i> +/-	$-9.97 \times 10^{-2}$	$5.83 \times 10^{-2}$	$4.23 \times 10^{-2}$
80	80			0-10-kft <i>U</i> +/-	$-4.60 \times 10^{-1}$	$-3.16 \times 10^{-1}$	$4.67 \times 10^{-2}$
81	81			400-mb height	$7.59 \times 10^3$	$7.59 \times 10^3$	$4.75 \times 10^{-2}$
82				Vertical totals	$2.47 \times 10^1$	$2.49 \times 10^1$	$5.68 \times 10^{-2}$
83				1000-850-mb lapse rate	$-6.40 \times 10^0$	$-6.49 \times 10^0$	$5.88 \times 10^{-2}$
84				Supercell composite parameter (SCP)	$5.32 \times 10^{-2}$	$2.77 \times 10^{-2}$	$6.25 \times 10^{-2}$
85				Mixing height <i>q</i> method	$1.80 \times 10^4$	$1.68 \times 10^4$	$6.88 \times 10^{-2}$
86				Mixing height <i>q</i> pressure	$5.49 \times 10^2$	$5.76 \times 10^2$	$7.15 \times 10^{-2}$
87				Fog threat	$-9.86 \times 10^{-1}$	$-6.63 \times 10^{-1}$	$7.35 \times 10^{-2}$
88				Lifted index	$-4.81 \times 10^0$	$-5.03 \times 10^0$	$7.70 \times 10^{-2}$
89				LFC-LCL	$6.20 \times 10^2$	$5.53 \times 10^2$	$8.03 \times 10^{-2}$
90				600-mb <i>V</i> +/-	$1.40 \times 10^{-1}$	$4.85 \times 10^{-3}$	$8.19 \times 10^{-2}$
91				Thompson index	$3.32 \times 10^1$	$3.42 \times 10^1$	$8.79 \times 10^{-2}$
92				1000-700-mb thick	$1.00 \times 10^4$	$1.00 \times 10^4$	$8.87 \times 10^{-2}$
93				0-3 kft AGL RH	$8.33 \times 10^1$	$8.26 \times 10^1$	$8.88 \times 10^{-2}$
94				925-700-mb mixed-layer CAPE (MLCAPE)	$7.59 \times 10^1$	$8.54 \times 10^1$	$8.90 \times 10^{-2}$
95				0-3 kft AGL dewpoint depression (Tdd)	$3.76 \times 10^1$	$3.79 \times 10^1$	$9.18 \times 10^{-2}$
96				Max wind <i>V</i> +/-	$1.15 \times 10^{-3}$	$-1.26 \times 10^{-1}$	$9.82 \times 10^{-2}$
97				Jefferson index	$2.72 \times 10^1$	$2.77 \times 10^1$	$9.97 \times 10^{-2}$
98				300-mb <i>V</i> +/-	$-2.35 \times 10^{-1}$	$-3.54 \times 10^{-1}$	$1.00 \times 10^{-1}$
99				400-mb <i>V</i> +/-	$-1.55 \times 10^{-1}$	$-2.77 \times 10^{-1}$	$1.02 \times 10^{-1}$
100				Vent index	$6.63 \times 10^3$	$5.58 \times 10^3$	$1.02 \times 10^{-1}$
101				Storm wind 0-3-km <i>U</i> +/-	$-3.07 \times 10^{-1}$	$-1.84 \times 10^{-1}$	$1.07 \times 10^{-1}$
103				Convective inhibition	$-2.61 \times 10^1$	$-2.20 \times 10^1$	$1.12 \times 10^{-1}$
104				Max vertical velocity	$6.23 \times 10^1$	$6.38 \times 10^1$	$1.27 \times 10^{-1}$
105				Hail CAPE	$6.73 \times 10^2$	$7.00 \times 10^2$	$1.41 \times 10^{-1}$
106				K index	$2.83 \times 10^1$	$2.92 \times 10^1$	$1.54 \times 10^{-1}$
107				850-mb temp	$1.80 \times 10^1$	$1.79 \times 10^1$	$1.64 \times 10^{-1}$
108				Storm wind 0-3-km <i>V</i> +/-	$-1.78 \times 10^{-1}$	$-2.82 \times 10^{-1}$	$1.65 \times 10^{-1}$
109				150-mb <i>U</i> +/-	$2.99 \times 10^{-2}$	$1.37 \times 10^{-1}$	$1.66 \times 10^{-1}$
110				925-mb <i>U</i> +/-	$-5.23 \times 10^{-1}$	$-4.27 \times 10^{-1}$	$1.66 \times 10^{-1}$



TABLE A1. (Continued)

<i>N</i>	<i>N</i> <sub>0</sub>	<i>N</i> <sub>1</sub>	<i>N</i> <sub>2</sub>	Abbreviated variable name	Mean value on no-report days	Mean value on report days	<i>p</i> value for <i>t</i> test
111				Fog point	$2.15 \times 10^1$	$2.12 \times 10^1$	$1.71 \times 10^{-1}$
112				Modified Richardson index	$-9.87 \times 10^0$	$-2.66 \times 10^{-1}$	$1.72 \times 10^{-1}$
113				MLCAPE	$2.01 \times 10^3$	$2.09 \times 10^3$	$1.82 \times 10^{-1}$
114				MLCAPE positive	$2.04 \times 10^3$	$2.11 \times 10^3$	$1.95 \times 10^{-1}$
115				Max parcel level	$9.38 \times 10^1$	$9.07 \times 10^1$	$1.99 \times 10^{-1}$
116				Corfidi downshear <i>U</i> +/-	$1.91 \times 10^{-1}$	$2.86 \times 10^{-1}$	$2.03 \times 10^{-1}$
117				400-mb wind speed	$1.19 \times 10^1$	$1.26 \times 10^1$	$2.08 \times 10^{-1}$
118				Wind shear 0–3 km	$1.93 \times 10^0$	$1.83 \times 10^0$	$2.13 \times 10^{-1}$
119				200-mb <i>V</i> +/-	$-3.65 \times 10^{-1}$	$-4.51 \times 10^{-1}$	$2.13 \times 10^{-1}$
120				925-mb height	$8.05 \times 10^2$	$8.07 \times 10^2$	$2.15 \times 10^{-1}$
121				Mean TdD	$5.15 \times 10^0$	$4.96 \times 10^0$	$2.20 \times 10^{-1}$
122				Max <i>U</i> +/-	$-7.57 \times 10^{-1}$	$-6.89 \times 10^{-1}$	$2.20 \times 10^{-1}$
123				MLCAPE 0–3-km layer	$7.95 \times 10^1$	$8.42 \times 10^1$	$2.21 \times 10^{-1}$
124				LFC height	$4.19 \times 10^3$	$4.01 \times 10^3$	$2.23 \times 10^{-1}$
125				Warm cloud depth	$2.16 \times 10^3$	$2.21 \times 10^3$	$2.25 \times 10^{-1}$
126				100-mb <i>V</i> +/-	$-4.94 \times 10^{-1}$	$-5.71 \times 10^{-1}$	$2.30 \times 10^{-1}$
127				KO index	$-1.55 \times 10^1$	$-1.59 \times 10^1$	$2.36 \times 10^{-1}$
128				LCL height	$2.16 \times 10^3$	$2.20 \times 10^3$	$2.37 \times 10^{-1}$
129				500-mb height	$5.88 \times 10^3$	$5.88 \times 10^3$	$2.40 \times 10^{-1}$
130				Lifted parcel level (MLCAPE)	$9.88 \times 10^2$	$9.89 \times 10^2$	$2.40 \times 10^{-1}$
131				Surface pressure	$1.01 \times 10^3$	$1.01 \times 10^3$	$2.40 \times 10^{-1}$
132				1000-mb height	$1.19 \times 10^2$	$1.21 \times 10^2$	$2.43 \times 10^{-1}$
133				Mean RH	$7.39 \times 10^1$	$7.46 \times 10^1$	$2.43 \times 10^{-1}$
134				600-mb wind speed	$1.05 \times 10^1$	$1.00 \times 10^1$	$2.46 \times 10^{-1}$
135				200-mb <i>U</i> +/-	$5.63 \times 10^{-2}$	$1.46 \times 10^{-1}$	$2.46 \times 10^{-1}$
136				925-mb Td	$1.88 \times 10^1$	$1.86 \times 10^1$	$2.50 \times 10^{-1}$
137				850-mb height	$1.54 \times 10^3$	$1.54 \times 10^3$	$2.61 \times 10^{-1}$
138				Corfidi up shear <i>U</i> +/-	$4.96 \times 10^{-1}$	$4.17 \times 10^{-1}$	$2.62 \times 10^{-1}$
139				LFC pressure	$8.78 \times 10^2$	$8.83 \times 10^2$	$2.82 \times 10^{-1}$
140				LFC-EL pressure	$1.66 \times 10^2$	$1.63 \times 10^2$	$2.97 \times 10^{-1}$
141				Downdraft CIN	$-1.28 \times 10^0$	$-8.98 \times 10^{-1}$	$3.05 \times 10^{-1}$
142				LCL pressure	$9.41 \times 10^2$	$9.40 \times 10^2$	$3.12 \times 10^{-1}$
143				CCL-EL pressure	$1.53 \times 10^2$	$1.52 \times 10^2$	$3.19 \times 10^{-1}$
144				250-mb <i>U</i> +/-	$5.50 \times 10^{-2}$	$1.31 \times 10^{-2}$	$3.23 \times 10^{-1}$
145				Storm relative 4–6-km speed	$1.57 \times 10^1$	$1.62 \times 10^1$	$3.28 \times 10^{-1}$
146				600–400-mb lapse rate	$-6.04 \times 10^0$	$-6.07 \times 10^0$	$3.38 \times 10^{-1}$
147				Sfc–900-mb wet-bulb theta	$2.97 \times 10^2$	$2.96 \times 10^2$	$3.42 \times 10^{-1}$
148				850-mb Td	$1.24 \times 10^1$	$1.26 \times 10^1$	$3.48 \times 10^{-1}$
149				LFC temp	$1.94 \times 10^1$	$1.96 \times 10^1$	$3.55 \times 10^{-1}$
150				CCL-EL temp	$-6.60 \times 10^1$	$-6.63 \times 10^1$	$3.61 \times 10^{-1}$
151				Corfidi up shear <i>V</i> +/-	$-5.78 \times 10^{-1}$	$-5.19 \times 10^{-1}$	$3.68 \times 10^{-1}$
152				600-mb Td	$-8.07 \times 10^0$	$-7.59 \times 10^0$	$3.71 \times 10^{-1}$
153				SWEAT index	$1.84 \times 10^2$	$1.81 \times 10^2$	$3.75 \times 10^{-1}$
154				Wet microburst severe index	$5.89 \times 10^1$	$6.07 \times 10^1$	$3.93 \times 10^{-1}$
155				GOES microburst index	$1.87 \times 10^0$	$2.28 \times 10^0$	$3.96 \times 10^{-1}$
156				LFC-EL temp	$-6.33 \times 10^1$	$-6.38 \times 10^1$	$3.96 \times 10^{-1}$
157				150-mb <i>V</i> +/-	$-4.94 \times 10^{-1}$	$-5.49 \times 10^{-1}$	$4.01 \times 10^{-1}$
158				Heat burst index	$4.99 \times 10^1$	$5.41 \times 10^1$	$4.19 \times 10^{-1}$
159				1000-mb Td	$2.31 \times 10^1$	$2.31 \times 10^1$	$4.24 \times 10^{-1}$
160				Haines index	$3.56 \times 10^0$	$3.52 \times 10^0$	$4.29 \times 10^{-1}$
161				LCL temp	$2.19 \times 10^1$	$2.19 \times 10^1$	$4.33 \times 10^{-1}$
162				Mean mixing ratio	$1.80 \times 10^1$	$1.79 \times 10^1$	$4.43 \times 10^{-1}$
163				Humidity index	$2.56 \times 10^1$	$2.48 \times 10^1$	$4.43 \times 10^{-1}$
164				500-mb speed	$1.07 \times 10^1$	$1.11 \times 10^1$	$4.48 \times 10^{-1}$
165				250-mb <i>V</i> +/-	$-3.02 \times 10^{-1}$	$-3.54 \times 10^{-1}$	$4.70 \times 10^{-1}$
166				Bulk Richardson shear	$6.51 \times 10^0$	$6.95 \times 10^0$	$4.72 \times 10^{-1}$
167				Shear 0–1-km <i>U</i> +/-	$1.79 \times 10^{-1}$	$2.33 \times 10^{-1}$	$4.73 \times 10^{-1}$

TABLE A1. (Continued)

<i>N</i>	<i>N</i> <sub>0</sub>	<i>N</i> <sub>1</sub>	<i>N</i> <sub>2</sub>	Abbreviated variable name	Mean value on no-report	Mean value on report	<i>p</i> value for <i>t</i>
					days	days	test
168				WINDEX	$4.11 \times 10^1$	$4.15 \times 10^1$	$4.80 \times 10^{-1}$
169				300-mb <i>U</i> +/-	$1.49 \times 10^{-2}$	$6.80 \times 10^{-2}$	$4.94 \times 10^{-1}$
170				Storm relative 9–11-km speed	$2.07 \times 10^1$	$2.12 \times 10^1$	$5.00 \times 10^{-1}$
171				Total precipitable water	$1.91 \times 10^0$	$1.92 \times 10^0$	$5.02 \times 10^{-1}$
172				Storm relative 0–2-km <i>V</i> +/-	$4.40 \times 10^{-1}$	$3.93 \times 10^{-1}$	$5.11 \times 10^{-1}$
173				Normalized CAPE	$1.58 \times 10^{-1}$	$1.50 \times 10^{-1}$	$5.13 \times 10^{-1}$
174				Storm relative helicity (SRH) 0–1 km	$5.43 \times 10^0$	$4.43 \times 10^0$	$5.47 \times 10^{-1}$
175				Max parcel level height	$5.67 \times 10^4$	$5.69 \times 10^4$	$5.51 \times 10^{-1}$
176				Dry microburst potential index	$1.82 \times 10^0$	$2.14 \times 10^0$	$5.57 \times 10^{-1}$
177				Storm relative 0–2-km <i>U</i> +/-	$2.41 \times 10^{-2}$	$-1.94 \times 10^{-2}$	$5.76 \times 10^{-1}$
178				Raob parameter metric	$2.36 \times 10^1$	$2.39 \times 10^1$	$5.82 \times 10^{-1}$
179				Shear 0–1-km <i>V</i> +/-	$-2.21 \times 10^{-1}$	$-2.62 \times 10^{-1}$	$5.84 \times 10^{-1}$
180				CCL temp	$2.15 \times 10^1$	$2.14 \times 10^1$	$5.89 \times 10^{-1}$
181				Storm relative 4–6-km <i>U</i> +/-	$2.81 \times 10^{-1}$	$3.20 \times 10^{-1}$	$5.90 \times 10^{-1}$
182				PBL mixing height	$4.62 \times 10^3$	$4.89 \times 10^3$	$5.96 \times 10^{-1}$
183				PBL parcel height	$4.62 \times 10^3$	$4.89 \times 10^3$	$5.96 \times 10^{-1}$
184				Liquid water path	$5.44 \times 10^{-2}$	$5.10 \times 10^{-2}$	$6.03 \times 10^{-1}$
185				LFC-EL height	$4.49 \times 10^4$	$4.51 \times 10^4$	$6.14 \times 10^{-1}$
186				700–500-mb lapse rate	$-5.84 \times 10^0$	$-5.85 \times 10^0$	$6.17 \times 10^{-1}$
187				PBL mix pressure	$8.79 \times 10^2$	$8.72 \times 10^2$	$6.23 \times 10^{-1}$
188				PBL parcel pressure	$8.79 \times 10^2$	$8.72 \times 10^2$	$6.23 \times 10^{-1}$
189				250-mb speed	$2.06 \times 10^1$	$2.01 \times 10^1$	$6.24 \times 10^{-1}$
190				Corfidi downshear <i>V</i> +/-	$-3.73 \times 10^{-1}$	$-4.08 \times 10^{-1}$	$6.28 \times 10^{-1}$
191				150-mb temp	$-6.65 \times 10^1$	$-6.65 \times 10^1$	$6.28 \times 10^{-1}$
192				Storm speed (Bunkers)	$1.63 \times 10^1$	$1.61 \times 10^1$	$6.31 \times 10^{-1}$
193				925-mb temp	$2.20 \times 10^1$	$2.20 \times 10^1$	$6.35 \times 10^{-1}$
194				100-mb speed	$1.59 \times 10^1$	$1.62 \times 10^1$	$6.48 \times 10^{-1}$
195				Max parcel level temp	$-6.92 \times 10^1$	$-6.94 \times 10^1$	$6.64 \times 10^{-1}$
196				Hail size	$3.64 \times 10^1$	$3.73 \times 10^{-1}$	$6.65 \times 10^{-1}$
197				Downdraft CAPE	$9.36 \times 10^2$	$9.27 \times 10^2$	$6.85 \times 10^{-1}$
198				Theta-e 4000 ft	$6.89 \times 10^1$	$6.91 \times 10^1$	$6.91 \times 10^{-1}$
199				PBL theta height	$1.91 \times 10^4$	$1.89 \times 10^4$	$6.97 \times 10^{-1}$
200				Fog radiation probability	$4.19 \times 10^1$	$4.17 \times 10^1$	$6.99 \times 10^{-1}$
201				Max wind speed	$5.26 \times 10^1$	$5.22 \times 10^1$	$7.04 \times 10^{-1}$
202				700-mb height	$3.17 \times 10^3$	$3.18 \times 10^3$	$7.04 \times 10^{-1}$
203				CCL-EL height	$4.64 \times 10^4$	$4.64 \times 10^4$	$7.30 \times 10^{-1}$
204				1000-mb temp	$2.71 \times 10^1$	$2.71 \times 10^1$	$7.37 \times 10^{-1}$
205				500-mb Td	$-1.80 \times 10^1$	$-1.82 \times 10^1$	$7.39 \times 10^{-1}$
206				600-mb height	$4.43 \times 10^3$	$4.43 \times 10^3$	$7.61 \times 10^{-1}$
207				700-mb Td	$3.10 \times 10^{-1}$	$4.37 \times 10^{-1}$	$7.62 \times 10^{-1}$
208				CCL height	$2.97 \times 10^3$	$2.99 \times 10^3$	$7.67 \times 10^{-1}$
209				Max wind pressure	$4.84 \times 10^1$	$4.70 \times 10^1$	$7.98 \times 10^{-1}$
210				Shear 0–2 km	$1.15 \times 10^0$	$1.16 \times 10^0$	$8.05 \times 10^{-1}$
211				200-mb speed	$2.47 \times 10^1$	$2.45 \times 10^1$	$8.12 \times 10^{-1}$
212				CCL pressure	$9.15 \times 10^2$	$9.14 \times 10^2$	$8.16 \times 10^{-1}$
213				300-mb speed	$1.64 \times 10^1$	$1.66 \times 10^1$	$8.24 \times 10^{-1}$
214				SRH 0–2 km	$1.94 \times 10^1$	$1.97 \times 10^1$	$8.35 \times 10^{-1}$
215				SRH 0–3 km	$3.09 \times 10^1$	$3.04 \times 10^1$	$8.38 \times 10^{-1}$
216				PBL theta pressure	$5.22 \times 10^2$	$5.25 \times 10^2$	$8.46 \times 10^{-1}$
217				1000–850-mb thickness	$4.65 \times 10^3$	$4.65 \times 10^3$	$8.60 \times 10^{-1}$
218				Chandler burn index	$3.37 \times 10^0$	$3.34 \times 10^0$	$8.63 \times 10^{-1}$
219				150-mb speed	$2.31 \times 10^1$	$2.30 \times 10^1$	$8.72 \times 10^{-1}$
220				Corfidi upshear speed	$1.15 \times 10^1$	$1.15 \times 10^1$	$8.72 \times 10^{-1}$
221				Surface Td	$2.29 \times 10^1$	$2.29 \times 10^1$	$8.95 \times 10^{-1}$
222				Convective temp	$3.03 \times 10^1$	$3.03 \times 10^1$	$8.97 \times 10^{-1}$
223				Storm relative 9–11-km <i>U</i> +/-	$3.08 \times 10^{-1}$	$3.01 \times 10^{-1}$	$9.23 \times 10^{-1}$
224				Integrated water vapor	$4.40 \times 10^0$	$4.40 \times 10^0$	$9.27 \times 10^{-1}$

TABLE A1. (Continued)

$N$	$N_0$	$N_1$	$N_2$	Abbreviated variable name	Mean value on no-report days	Mean value on report days	$p$ value for $t$ test
225				Wet-bulb zero height	$1.31 \times 10^4$	$1.31 \times 10^4$	$9.32 \times 10^{-1}$
226				Wet-bulb zero pressure	$6.34 \times 10^2$	$6.34 \times 10^2$	$9.34 \times 10^{-1}$
227				SRH 0–0.5 km	$-1.68 \times 10^0$	$-1.78 \times 10^0$	$9.45 \times 10^{-1}$
228				Surface temp	$2.78 \times 10^1$	$2.78 \times 10^1$	$9.49 \times 10^{-1}$
229				Wind chill index	$2.78 \times 10^1$	$2.78 \times 10^1$	$9.49 \times 10^{-1}$
230				Temp humidity index	$3.10 \times 10^1$	$3.10 \times 10^1$	$9.54 \times 10^{-1}$
231				Microburst day potential index	$1.02 \times 10^0$	$1.02 \times 10^0$	$9.57 \times 10^{-1}$
232				Downrush temp	$1.90 \times 10^4$	$1.90 \times 10^1$	$9.63 \times 10^{-1}$
233				Max wind height	$8.85 \times 10^4$	$8.87 \times 10^4$	$9.63 \times 10^{-1}$
234				Storm relative 4–6-km $V_{+/-}$	$2.50 \times 10^{-1}$	$2.52 \times 10^{-1}$	$9.71 \times 10^{-1}$
235				Storm relative 9–11-km $V_{+/-}$	$-8.93 \times 10^{-2}$	$-8.74 \times 10^{-2}$	$9.80 \times 10^{-1}$
236				Storm relative 0–2-km speed	$1.51 \times 10^1$	$1.51 \times 10^1$	$9.83 \times 10^{-1}$
237				Delta theta-e	$2.79 \times 10^1$	$2.79 \times 10^1$	$9.89 \times 10^{-1}$
238				CAP (lid strength index)	$1.04 \times 10^0$	$1.04 \times 10^0$	$9.89 \times 10^{-1}$

APPENDIX

Details of the Variables Considered in the Construction of the LRM

Shown in Table A1 is a list of the candidate predictors (parameters obtained from raob) that were

used as a starting point for building the LRM model. Also indicated in Table A1 are the candidate predictors that were retained at each step of the selection procedure described in section 3a. Figures A1 and A2 are provided to support the motivation for excluding raob-derived values of dewpoint temperatures above 400 mb from consideration, as described in section 2.

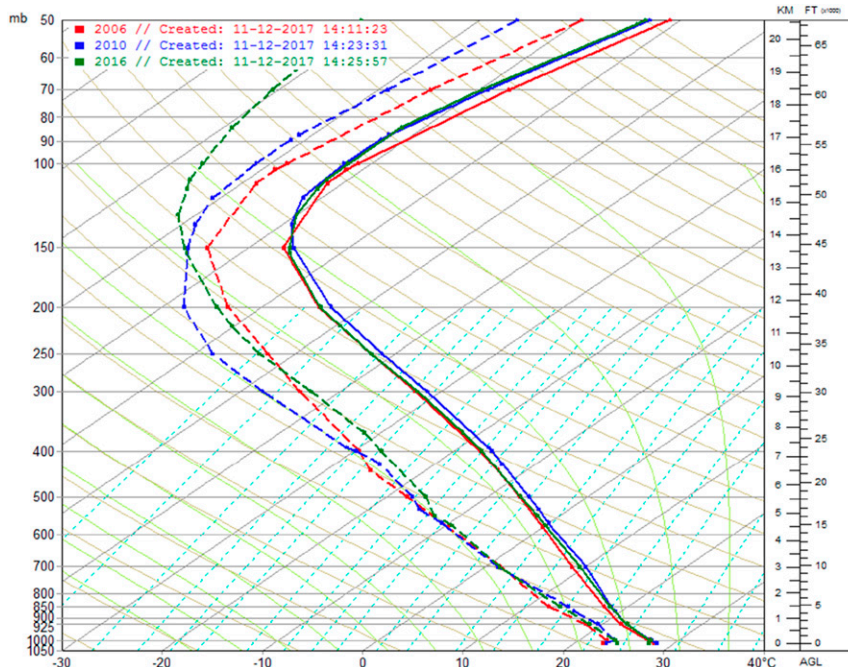


FIG. A1. Wet season composite soundings for the years 2006 (red), 2010 (blue), and 2016 (green). The dashed line is the dewpoint temperature, and the solid line is the temperature. Note the divergence in dewpoint temperatures above 400 mb, reflecting the change in radio-sonde instruments in 2010 and 2012.

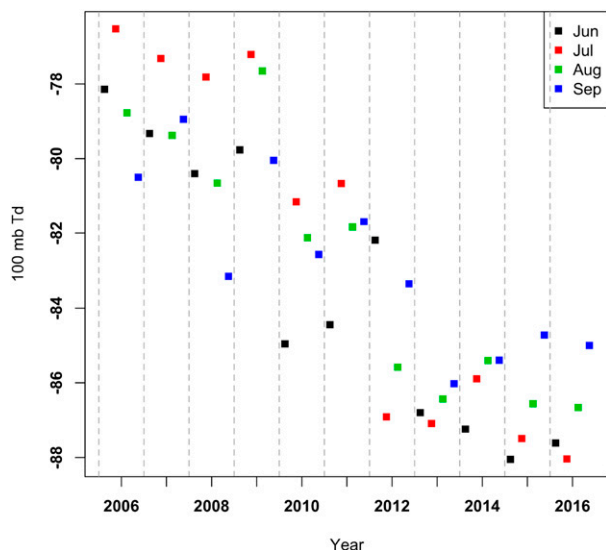


FIG. A2. Monthly mean 100-mb dewpoint temperature ( $^{\circ}\text{C}$ ) from the rawinsonde observations at Key West. Note the discontinuity between 2006–10 and 2012–16 reflecting instrumentation changes.

#### REFERENCES

- Akaike, H., 1974: A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, **19**, 716–723, <https://doi.org/10.1109/TAC.1974.1100705>.
- Austin, P. C., and E. W. Steyerberg, 2014: Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat. Methods Med. Res.*, **26**, 796–808, <https://doi.org/10.1177/0962280214558972>.
- Brown, D., and J. Rothfuss, 1998: An approach to waterspout forecasting for south Florida and the Keys. NOAA/NWS Internal Rep., Miami, FL, [https://www.weather.gov/mfl/waterspout\\_fcsting](https://www.weather.gov/mfl/waterspout_fcsting).
- 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.
- Davies, J. M., 1993: Hourly helicity, instability, and EHI in forecasting supercell tornadoes. Preprints, *17th Conf. on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc., 107–111.
- Golden, J., 1974: The life cycle of Florida Keys' waterspouts. I. *J. Appl. Meteor.*, **13**, 676–692, [https://doi.org/10.1175/1520-0450\(1974\)013<0676:TLCOFK>2.0.CO;2](https://doi.org/10.1175/1520-0450(1974)013<0676:TLCOFK>2.0.CO;2).
- , 1977: An assessment of waterspout frequencies along the U.S. East and Gulf Coasts. *J. Appl. Meteor.*, **16**, 231–236, [https://doi.org/10.1175/1520-0450\(1977\)016<0231:AAOWFA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1977)016<0231:AAOWFA>2.0.CO;2).
- , 2003: Waterspouts. *Encyclopedia of Atmospheric Sciences*, G. North, J. Pyle, and F. Zhang, Eds., Elsevier, 2510–2525.
- , and H. B. Bluestein, 1994: The NOAA–National Geographic Society Waterspout Expedition (1993). *Bull. Amer. Meteor. Soc.*, **75**, 2281–2288, [https://doi.org/10.1175/1520-0477\(1994\)075<2281:TNGSW>2.0.CO;2](https://doi.org/10.1175/1520-0477(1994)075<2281:TNGSW>2.0.CO;2).
- Heffter, J. L., 1980: Air Resources Laboratories Atmospheric Transport and Dispersion Model (ARL-ATAD). NOAA Tech. Memo. ERL ARL 81, 29 pp., <https://www.arl.noaa.gov/documents/reports/arl-81.pdf>.
- Jolliffe, I. T., and D. B. Stephenson, 2008: Proper scores for probability forecasts can never be equitable. *Mon. Wea. Rev.*, **136**, 1505–1510, <https://doi.org/10.1175/2007MWR2194.1>.
- Keul, A. G., M. V. Sioutas, and W. Szilagyi, 2009: Prognosis of central-eastern Mediterranean waterspouts. *Atmos. Res.*, **93**, 426–436, <https://doi.org/10.1016/j.atmosres.2008.10.028>.
- Kuiper, J., and M. Van der Haven, 2007: A new index to calculate risk of waterspout development. *Fourth European Conf. on Severe Storms*, Trieste, Italy, European Severe Storms Laboratory, <https://www.esrl.org/ECSS/2007/abs/06-Forecasts/1179250265.kuiper.pdf>.
- MacInnes, J., 2017: *An Introduction to Secondary Data Analysis with IBM SPSS Statistics*. Sage Publications, 336 pp.
- McCann, D. W., 1994: WINDEX—A new index for forecasting microburst potential. *Wea. Forecasting*, **9**, 532–541, [https://doi.org/10.1175/1520-0434\(1994\)009<0532:WNIFFM>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0532:WNIFFM>2.0.CO;2).
- Miller, R. C., 1972: Notes on analysis and severe storm forecasting procedures of the Air Force Global Weather Central. Air Weather Service Tech. Rep. 200, 184 pp., <http://www.dtic.mil/dtic/tr/fulltext/u2/744042.pdf>.
- Palmer, T. N., C. Brankovic, and D. S. Richardson, 2000: A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Quart. J. Roy. Meteor. Soc.*, **126**, 2012–2033, <https://doi.org/doi:10.1002/qj.49712656703>.
- Peduzzi, P., J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein, 1996: A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.*, **49**, 1373–1379, [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3).
- Rossov, V., 1970: Observations of water spouts and their parent clouds. NASA Tech. Note D-5854, 63 pp., <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19700020540.pdf>.
- Schwiesow, R. L., 1981: Horizontal velocity structure in waterspouts. *J. Appl. Meteor.*, **20**, 349–360, [https://doi.org/10.1175/1520-0450\(1981\)020<0349:HVSIV>2.0.CO;2](https://doi.org/10.1175/1520-0450(1981)020<0349:HVSIV>2.0.CO;2).
- Shewchuk, J., 2017: The Universal Rawinsonde Observation Program (version 6.8). Eosonde Research Services, <http://www.raob.com>.
- Sioutas, M. V., and A. G. Keul, 2007: Waterspouts of the Adriatic, Ionian and Aegean Sea and their meteorological environment. *Atmos. Res.*, **83**, 542–557, <https://doi.org/10.1016/j.atmosres.2005.08.009>.
- Spratt, S. M., and B. K. Choy, 1994: Employing the WSR-88D for waterspout forecasting. Postprints, *NEXRAD Users Conf.*, Norman, OK, NWS, <https://www.weather.gov/media/mlb/research/EmployingtheWSR88DforWaterspoutForecasting.pdf>.
- Swets, J. A., 1988: Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293, <https://doi.org/10.1126/science.3287615>.
- Szilagyi, W., 2009: A waterspout forecasting technique. *Fifth European Conf. on Severe Storms*, Landshut, Germany, European Severe Storms Laboratory, <https://www.esrl.org/ECSS/2009/preprints/O05-14-sziladgyi.pdf>.
- Watson, L. W., 2011: Upgrade summer severe weather tool. NASA Contractor Rep. CR-2011-216299, 32 pp.
- Wheeler, M. M., 2009: Severe weather and weak waterspout checklist in MIDDs. NASA Contractor Rep. CR-2009-214760, 15 pp., <https://science.ksc.nasa.gov/amu/final-reports/svr-wx-wksht-midds.pdf>.
- Wilks, D. S., 2006: *Statistical Methods in Atmospheric Sciences*. 2nd ed. Elsevier, 627 pp.
- World Meteorological Organization, 1999: Methods of interpreting numerical weather prediction output for aeronautical meteorology. WMO Tech. Note 195, 124 pp., [https://library.wmo.int/opac/doc\\_num.php?explnum\\_id=3447](https://library.wmo.int/opac/doc_num.php?explnum_id=3447).