Probabilistic 0–1-h Convective Initiation Nowcasts that Combine Geostationary Satellite Observations and Numerical Weather Prediction Model Data

JOHN R. MECIKALSKI

Atmospheric Science Department, University of Alabama in Huntsville, Huntsville, Alabama

JOHN K. WILLIAMS

National Center for Atmospheric Research, Boulder, Colorado*

CHRISTOPHER P. JEWETT

Atmospheric Science Department, University of Alabama in Huntsville, Huntsville, Alabama

DAVID AHIJEVYCH

National Center for Atmospheric Research, Boulder, Colorado

ANITA LEROY AND JOHN R. WALKER

Atmospheric Science Department, University of Alabama in Huntsville, Huntsville, Alabama

(Manuscript received 31 May 2014, in final form 30 January 2015)

ABSTRACT

The Geostationary Operational Environmental Satellite (GOES)-R convective initiation (CI) algorithm predicts CI in real time over the next 0–60 min. While GOES-R CI has been very successful in tracking nascent clouds and obtaining cloud-top growth and height characteristics relevant to CI in an object-tracking framework, its performance has been hindered by elevated false-alarm rates, and it has not optimally combined satellite observations with other valuable data sources. Presented here are two statistical learning approaches that incorporate numerical weather prediction (NWP) input within the established GOES-R CI framework to produce probabilistic forecasts: logistic regression (LR) and an artificial-intelligence approach known as random forest (RF). Both of these techniques are used to build models that are based on an extensive database of CI events and nonevents and are evaluated via cross validation and on independent case studies. With the proper choice of probability thresholds, both the LR and RF techniques incorporating NWP data produce substantially fewer false alarms than when only GOES data are used. The NWP information identifies environmental conditions (as favorable or unfavorable) for the development of convective storms and improves the skill of the CI nowcasts that operate on GOES-based cloud objects, as compared with when the satellite IR fields are used alone. The LR procedure performs slightly better overall when 14 skill measures are used to quantify the results and notably better on independent case study days.

1. Introduction

Considerable effort has been spent on obtaining highly accurate convective initiation (CI) nowcasts, or 0–1-h forecasts, in light of the impacts convective storms have on infrastructure, travel, and society (Fritsch et al. 1998; Curran et al. 2000; Weckwerth et al. 2004; Weckwerth and Parsons 2006; Brooks and Dotzek 2007). Given the high cost of convective weather to various forms of travel,

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: John R. Mecikalski, Atmospheric Science Department, University of Alabama in Huntsville, National Space Science and Technology Center, 320 Sparkman Drive, Huntsville, AL 35805-1912.
E-mail: johnm@nsstc.uah.edu

DOI: 10.1175/JAMC-D-14-0129.1

© 2015 American Meteorological Society

particularly aviation-related disruptions (Evans and Ducot 2006; Wolfson and Clark 2006; Iskenderian et al. 2010, 2012), and the cost of storm-related hazards (hail, high winds, flooding, tornadoes; Brooks et al. 2003; Brooks and Dotzek 2007; Dixon et al. 2011), more precise and timely thunderstorm forecasts are needed. It is also important to further our understanding of the physical processes that drive CI (Wilson et al. 1998; Ziegler et al. 2007; Lima and Wilson 2008; Wakimoto and Murphey 2009) given the correspondence between correct CI nowcasts and subsequent forecast skill (e.g., Brooks et al. 1992), which then allows us to use radar, satellite, and other observations in more intelligent ways within nowcasting systems.

A number of 0–2-h, mostly extrapolation-based systems that emphasize the use of radar observations have been developed to predict where CI is likely to occur and where existing convective storms are likely to propagate. These systems include the AutoNowcaster (Mueller et al. 1993, 2003; Wilson et al. 2010; Roberts et al. 2012) for CI and storm propagation; the Thunderstorm Identification, Tracking, Analysis and Nowcasting (TITAN; Dixon and Wiener 1993); the Warning Decision Support System–Integrated Information (Lakshmanan et al. 2007); the Canadian Radar Detection System (CARDS; Joe et al. 2003); the Thunderstorm Strike Probability Nowcasting Algorithm (THESPA; Dance et al. 2010); and the Collaborative Adaptive Sensing of the Atmosphere (CASA) Distributed Collaborative Adaptive Sensing network (Ruzanski et al. 2011) for storm cell evolution and tracking. Other nowcast systems like the Global/Regional Assimilation and Prediction System–Severe Weather Forecast Tool (GRAPES–SWIFT; Hu et al. 2007), with an overview provided in Wilson et al. (2010), and the Corridor Integrated Weather System (CIWS; Wolfson and Clark 2006), also include fields from numerical weather prediction (NWP) models as a means of forming high-quality analyses and modeled initial conditions (Feng et al. 2007) for storm-tracking purposes.

Returning to CI, an ideal platform for providing an early detection capability for the potential initiation of thunderstorms is geostationary satellite data (Purdom 1976, 1982; Roberts and Rutledge 2003; Mecikalski and Bedka 2006; Rosenfeld et al. 2008; Mecikalski et al. 2010a,b). Geostationary satellites like the Geostationary Operational Environmental Satellite (GOES) over North and South America, Meteosat over Europe, the Fengyun series over China, the Multifunctional Transport Satellite (MTSAT), and the *Himawari-8/-9* series over Japan and surrounding oceanic regions, offer 500-m- to 4-km-resolution views in visible, near-infrared (NIR), and infrared (IR) channels at 5–15-min temporal resolution (with 2.5-min rapid scan modes). Methods developed by Roberts and Rutledge (2003), Mecikalski and Bedka (2006), Lensky and Rosenfeld (2006), Rosenfeld et al. (2008), Harris et al. (2010), Mecikalski et al. (2010a,b), Sieglaff et al. (2011), Walker et al. (2012), Merk and Zinner (2013), and Nisi et al. (2014) demonstrate an ability to nowcast CI using geostationary satellites, and help identify locally strong, newly forming convective storms. For this present study, a legacy radar-based CI definition is used, namely the first occurrence of a ≥35-dBZ echo at the −10°C level within a cumulus cloud (Browning and Atlas 1965; Wilson and Schreiber 1986; Wilson et al. 1992; Wilson and Mueller 1993; Mueller et al. 2003).

Until recently, most satellite-based approaches to CI nowcasting have emphasized use of critical thresholds in one (Roberts and Rutledge 2003; Sieglaff et al. 2011) or more (Mecikalski and Bedka 2006; Rosenfeld et al. 2008; Mecikalski et al. 2008, 2010a,b) IR "interest fields" derived from the satellite imagery to predict which cumulus clouds will develop into cumulonimbus clouds over an ~60-min time frame. Table 1 shows interest fields for GOES (the top nine fields) that have been used successfully to nowcast CI occurrence using IR data, while other studies focus on the use of visible data (Setvák et al. 2003; Mecikalski et al. 2010b; Merino et al. 2014). Static thresholds are useful yet scoring a convective object (i.e., an individual cumulus cloud) typically only provides a "yes" or "no" CI nowcast of a pending new storm. Such binary, deterministic nowcasts are somewhat less useful to forecasters of imminent convective weather (Siewert and Kuhlman 2011), whereas probabilistic forecasts (i.e., 0%–100%) are found to be easier to interpret (Terborg and Gravelle 2012). Probabilistic forecasts also provide useful information on the uncertainty of the event (Dance et al. 2010; Steiner et al. 2010). The AMS Council (2008) stated, "surveys have consistently indicated that users desire information about uncertainty or confidence of weather forecasts. [And that information] is likely to yield substantial economic and social benefits, because users can make decisions that explicitly account for this uncertainty." This statement supports the need for an evolution toward combining data from many sources to address challenging forecast problems (in this case CI), while presenting the data in a meaningful way for end users (e.g., probabilistically).

Although satellites "see" cumulus clouds growing well in advance of a radar echo, incorporation of nonsatellite fields becomes valuable when forecasting CI for several reasons, including being able to constrain the forecast problem, and by providing information that otherwise is not contained in satellite fields. Examples of the former include predicting CI only where convective instability is positive [e.g., where the convective

TABLE 1. The satellite and NWP model predictors used for 0–1-h CI nowcasts. Each variable was collected for cloud objects representing potential CI events, and the dataset was used to train and evaluate the statistical learning approaches used within this study, logistic regression and random forest. Here, the dominant cloud type is obtained from the Berendes et al. (2008) cloud classification method. The object size is obtained from the Walker et al. (2012) CI algorithm, as GOES satellite pixels are grouped into cloud objects and tracked from an initial time 1 to a time 2, which is nominally 15 min later than time 1. All NWP fields were retrieved from the RAP model as described in the text. Here and in later tables, "temp" indicates temperature.

Satellite and NWP predictors
1) 10.7-μm T_B (cloud-top temp) (K)
2) 13.3–10.7-μm T_B diff (K)
3) 6.5–10.7-μm T_B diff (K)
4) 15-min 10.7-μm cloud-top cooling rate (K)
5) 15-min time change of 13.3–10.7-μm T_B diff (K)
6) Dominant cloud type at time 2
7) Dominant cloud type at time 1
8) Object size at time 1 (composed of 1-km pixels)
9) Object size at time 2 (composed of 1-km pixels)
10) Surface-based CIN (J kg^{-1})
11) LCL height (m)
12) Most-unstable CIN (J kg^{-1})
13) LFC height (m)
14) Most-unstable CAPE (MUCAPE) (J kg^{-1})
15) Nearness to convective temp (K)
16) Surface-based CAPE (J kg^{-1})
17) Convective temp (K)
18) Surface lifted index (K)
19) CCL height (m)
20) Freezing-level height (m)
21) Surface–500-hPa wind shear (s^{-1})
22) Surface–700-hPa wind shear (s^{-1})
23) Best lifted index (K)
24) 700–500-hPa wind shear (s^{-1})
25) Land/sea flag (0 or 1)

available potential energy (CAPE) is positive], indicating the potential for CI. For the latter, NWP soundings of moisture and wind shear complement cloud-top temperature and inferred in-cloud dynamics offered by geostationary satellite fields by identifying conducive environmental conditions.

The CI nowcast problem is well suited for statistical learning methods that permit fusion of a variety of data. This is true for two reasons: 1) CI is very much regulated by local conditions, such that a combination of satellite-derived predictors that produces a forecast with high skill in one situation may fail in another, despite conditions appearing similar in the predictor variables (IR fields in this case); and 2) it is difficult to know where most of the "importance" lies within a set of predictors, which again can be a function of relatively local phenomena (e.g., meso-γ scale, 2–25 km, with respect to developing thunderstorms). The motivation for the present study is to demonstrate procedures that make use of both satellite and nonsatellite predictors to provide probabilistic CI nowcasts. This study diverges from previous efforts in that nonsatellite NWP datasets are considered for use in combination with IR datasets as a means of increasing the value that geostationary satellite brightness temperature (T_B) fields bring to the problem. The outcomes from this study demonstrate the value of combining NWP data and satellite observations in a manner that meets the needs of short-term weather forecasters. The paper proceeds as follows: Section 2 provides a more in-depth background on the algorithms related to this study, while section 3 presents the methodology and datasets as well as the analysis techniques employed herein. Section 4 shows the results, and section 5 discusses the main findings and concludes the paper.

2. Background

The GOES-R CI [also known as Satellite Convection Analysis and Tracking (SATCAST)] algorithm as presented in Walker et al. (2012) was developed as a geostationary satellite-based, deterministic style CI nowcast methodology, with an emphasis on "cloud object tracking." The Walker et al. (2012) approach used the same geostationary satellite IR interest fields as prior binary yes/no algorithms (Mecikalski and Bedka 2006; Mecikalski et al. 2008), yet possessed a more robust cloud object-tracking framework to accumulate information on cloud-top heights, growth rates, glaciation, and updraft widths when making a determination of CI; the GOES IR field values are then used to score a cloud object as either a yes or no CI forecast.

While a great deal of information about growing clouds can be provided by satellite data alone (e.g., Rosenfeld et al. 2008; Zinner et al. 2008; Bedka et al. 2010; Mecikalski et al. 2010a; Wang et al. 2010; Setvák et al. 2013), there is no directly available knowledge provided about the atmospheric environment in which clouds will grow. Numerous studies have shown the fundamental importance of vertical (0–6 km) wind shear, CAPE, convective inhibition (CIN), and low-level moisture (to name a few) to the occurrence and organization of convective storms (e.g., Brooks et al. 1994). For example, if a newly developed cumulus cloud shows very strong signals of vertical development as observed in satellite observations, the GOES-R CI algorithm is designed to output a yes that CI will occur. However, this output could be very misleading if the rapid vertical cloud growth is occurring beneath a strong midlevel capping inversion, where in reality the chances of continued development and CI are very low. With the

inclusion of NWP data for environmental characterization used in tandem with satellite-derived information about individual clouds, a more accurate set of nowcasts can be produced. This additional NWP information would act to appropriately reduce the CI nowcast likelihood in situations like the one mentioned above, despite strong satellite-retrieved cloud growth signals, and to increase the CI nowcast likelihood when both the satellite-based cloud growth signals and the environment are more conducive to mature convective development.

When moving from deterministic methods toward development of probabilistic approaches, a valuable exercise to perform is to test more than one algorithm. This study provides a comparison between logistic regression (LR) and random forest (RF) using the same training database, determining if there is a tendency for one method to perform better overall, or in certain convective environments or situations. A necessary component when developing a statistically based forecast method is a sizeable database, in this case of IR satellite and NWP field indicators for CI and non-CI events, that establishes truth data that is used to both train and test predictive models. Such a database was assembled as described in the following section.

As of 2013–14, the GOES-R CI nowcasting methodology is used by forecasters in the Federal Aviation Administration (Iskenderian et al. 2012), National Weather Service (NWS) Forecast Offices, the Aviation Weather Center Testbed (AWT; Terborg and Gravelle 2012), and the National Oceanic and Atmospheric Administration Hazardous Weather Testbed (HWT; Terborg et al. 2013). An outcome of this study will be an improved algorithm incorporating NWP data that can be transitioned to these entities. The evolution to a probabilistic approach is motivated by forecaster feedback when using the GOES-R CI algorithm (see Terborg and Gravelle 2012; Terborg et al. 2013), and also for reasons given in Dance et al. (2010) and Steiner et al. (2010) on the value of probabilistic versus deterministic forecasts.

3. Methodology

a. Enhanced CI nowcasting framework

The present study utilizes in part the Walker et al. (2012; see their Fig. 2) 0–1-h CI object-tracking framework, with per-cloud object assessments of CI potential. The Walker et al. (2012) method is expanded by incorporating NWP fields and enhanced statistical methods to form probabilistic 0–1-h CI predictions. The cloud-object-tracking methods involve use of the Berendes et al. (2008) convective cloud classification scheme and so-called mesoscale atmospheric motion vectors (Bedka and Mecikalski 2005) to identify and advect convective clouds, respectively. All GOES IR and NWP fields are mapped to the cloud objects.

Two well-established statistical, probabilistic methods are well suited to the CI nowcasting challenge: LR (Hosmer and Lemeshow 1989) and RF (Breiman 2001). LR measures the relationship between a categorical dependent variable and one or more independent variables, which are usually continuous, applying a sigmoid (logistic) function to the result to constrain the output to between 0 and 1, as the predicted values of the dependent variable. LR has been successfully used to forecast maintenance of mesoscale convective systems (Coniglio et al. 2007) and to identify conditions conducive to contrail formation (Duda and Minnis 2009), as well as many other environmental science problems, and has the advantage of providing a relatively simple mathematical model that can be executed quickly. However, the training procedure can fail to converge if the ratio of predictor variables to training instances is high, or if predictor variables are highly correlated. RFs are collections of decision trees, each of which is formed using a random subset of the training dataset and with a random subset of the predictor variables considered for constructing each decision node. This randomized training procedure ensures that the decision trees are distinct from one another; therefore, they can serve as an ensemble of experts that ''vote'' on the classification of a new data instance. The resulting RF vote counts can be calibrated into reliable probability forecasts using an independent calibration dataset. RFs have been widely used in the biomedical field (e.g., Díaz-Uriarte and de Andrés 2006), in satellite remote sensing (Pal 2005), in a number of atmospheric science applications including convective nowcasting (e.g., Williams et al. 2008), and in diagnosing atmospheric turbulence for aviation users (Williams 2014). RF models are capable of representing quite complex predictive functions; however, they are generally more complex and slower to build and execute than LR models. For this study, more than one LR and RF models were developed, using portions of the training dataset (i.e., when cross validation was performed), toward identifying the most accurate statistical model of each type. The LR and RF methods are described in greater detail below.

b. Logistic regression

LR is a regression technique used for modeling dichotomous dependent variables from a set of several independent, or predictor, variables. More simply, it is a

means of estimating a probability of an event in which only two outcomes are possible (a yes/no event, e.g., CI or no CI, represented by the values 1 or 0), based on the values of several other "predictor" variables. The logistic model is given by the formula

$$E(Y) = \frac{1}{1 + \exp\left[-\left(\beta_o + \sum_{j=1}^{k} \beta_j X_j\right)\right]}, \qquad (1)$$

in which the expectation value E of the dependent variable Y lies in the interval [0, 1] and the values of the predictor variables are represented by $\{X_1, \ldots, X_k\}$. The parameters $\{\beta_0, \ldots, \beta_k\}$ are linear coefficients or "weights" for the predictor variables and k is the number of predictor variables. Because the output of the LR model lies between 0 and 1, the output can be thought of as a probabilistic prediction of whether the event will occur given the values of the predictor variables.

For this study, the Excel statistical analysis package (XLSTAT) was used to build LR models. XLSTAT is an add-in capability of the widely used Microsoft Excel program. Linear and LR methods belong to the same family of models, called generalized linear models, where an event (in this case CI) is linked to a linear combination of explanatory or predictor variables (e.g., those listed in Table 1). For LR, the dependent variable or response variable follows a Bernoulli distribution for parameter P, where P is the mean probability that an event will occur (if an experiment is done once), or a binomial (n, P) distribution (if the experiment is repeated n times). Here P is the logistic function of a linear combination of the predictor variables. Specifically, the so-called Logit model within XLSTAT was used to form $E(Y)$ as described by Eq. (1). The Logit model of the XLSTAT LR is a transformation; basically, instead of assessing the likelihood, or "odds" that an event will occur, the algorithm assesses mathematically the natural log odds that an event will occur. XLSTAT applies maximum likelihood estimation (using the Newton–Raphson method), and the predictors can potentially be both categorical and continuous (e.g., a land/sea flag vs CAPE). The choice was made to use the Logit model with the maximum likelihood estimation to account for both types of predictor variables.

To achieve a probabilistic CI nowcast, the LR equation for the probability of CI was developed from the sample containing 9015 observations (CI and non-CI events) of tracked objects. Five GOES satellite, Lagrangian predictor variables were taken directly from Walker et al. (2012) as based originally on Mecikalski and Bedka (2006): 10.7-μm T_B, $\partial T(10.7\,\mu\text{m})/\partial t$, instantaneous 6.5–10.7 μm, instantaneous 13.3–10.7 μm, and $\partial T(13.3\text{–}10.7\,\mu\text{m})/\partial t$ (first five fields listed in Table 1). A sixth GOES-R CI interest field, $\partial T(6.5\text{–}10.7\,\mu\text{m})/\partial t$, was observed to possess multiple colinearity with other fields, and hence this field was removed from the sample dataset before a final regression was performed.

The coefficients in Eq. (1) were determined from the training database (see section 3d) using XLSTAT, and to display a probability the $E(Y)$ was scaled from 0 to 100 and associated with a standard color bar in which warmer colors represent higher CI probabilities, that is, a greater likelihood that the specific tracked object will become a mature convective storm. The tracked cloud objects for which the CI nowcasts were made were overlaid with contemporaneous GOES visible satellite imagery to assist interpretation by the end user. This model is labeled LR-Sat in the results and case study sections to indicate that *only GOES data* were used as inputs. A second model with an expanded set of predictor variables (Table 1), including 14 variables derived from the 13-km Rapid Update (RAP; Benjamin et al. 2009) NWP model plus a land/sea flag, was created similarly; this second model is referred to as LR-SatNWP. Similarly, as described below, a satellite-only RF was formed, called RF-Sat, with a second model using satellite+NWP model predictor data referred to as RF-SatNWP.

c. *Random forest*

In contrast to LR, the RF method (Breiman 2001) is a nonparametric procedure based on the consensus of a collection of decision trees (Dattatreya 2009). Each tree is formed with a degree of randomness in both the selection of instances used for training and the choice of candidate variables for splitting at each decision node. The training set for each tree is drawn randomly *with replacement* from the original set of N training instances, so that some training instances are used more than once and some are not used at all. For each node, a random subset of predictor variables are selected, and a threshold for the one that best discriminates positive and negative instances is selected to split the dataset into two subsets; this process is repeated recursively until the final subsets (leaves) satisfy a homogeneity or minimum size criterion. Once trained, the decision tree sorts any new instance into a "leaf" based on the sequence of splits that it satisfies, and the consensus value of the training set instances in that leaf determine the "tree's" output. The rules for building the tree generally ensure that the set of predictand values in a leaf are relatively homogeneous. Geometrically, one can think of a decision tree as splitting the predictor variable space $X_1 \times X_2 \times \cdots \times X_k$ into hypercubes, each of whose boundaries are perpendicular to one of the coordinate axes. For

a dichotomous discrimination problem, the decision tree assigns each of these hypercubes a value of either 0 (no CI) or 1 (CI). By aggregating a collection of decision trees, a finer set of hypercubes is defined, each of which is associated with a vote count between 0 and the number of trees in the collection, and the fraction of trees voting 1 may be interpreted as a probability. A large enough collection of decision trees can represent any probability function arbitrarily closely, similar to a Riemann sum approximating an integral, although there is no guarantee that the RF training procedure will necessarily generate an optimal approximation, particularly for small training datasets. In practice, RFs have the advantage of being straightforward to train and use, are relatively insensitive to parameter settings, and tend not to "overfit" as some other statistical methods do (Breiman 2001). Thus, the RF method is an ideal candidate to compare with LR for the purpose of fusing multiple satellite and NWP variables together to nowcast CI.

To create a probabilistic model for CI, an RF was trained from the same dataset as LR, comprising 9015 labeled instances (section 3d). As for LR, the RF was applied to five satellite predictors to create the RF-Sat model. The RF was also applied to the full set of 25 GOES and RAP model predictor variables in Table 1, including object-based fields, satellite radiance trends, cloud types, observations, and surface type (land/sea flag) to produce the RF-SatNWP model. For both LR and RF models, all predictor variables were the same, whether all from the GOES satellite, or from GOES and the RAP model.

Although the RF has several adjustable parameters that affect the training of the model, the default parameters generally provide very good performance. However, to optimize the equitable threat score, several different parameter values were tested. Breiman (2001) noted the primary parameter to which the RF is sensitive is the number of candidate predictor variables chosen at each node. This was varied between 1 and 4 for the 5-predictor RF-Sat model, and between 1 and 13 for the 25-predictor RF-SatNWP model; the RF default is to use the square root of the total number of predictors. Along with this parameter, three forest sizes were tested (200, 400, and 800 trees), and the number of votes required for a positive forecast was also varied. The best forest size was found to be 400 trees for group 1 and 800 trees for group 2, suggesting that performance asymptotes by 400 trees. It was found that allowing selection of three candidate predictor variables per node was best for RF-Sat/group 1, and only allowing one predictor was best for RF-Sat/group 2. One predictor was also best for RF-SatNWP/group 1 and two predictors were best for RF-SatNWP/group 2. Other parameters were left unchanged from the default settings. The event and nonevent classifications were weighted equally, and any node that could be split at all was analyzed for splitting.

For the three independent case study days, the RF vote count was mapped to a probability that ensured the final distribution of probabilities from the RF matched the final distribution of probabilities from the LR model to facilitate comparison of the two models. This was done by converting the RF vote counts and LR probabilities to percentiles and then mapping each vote count to the LR probability with the same percentile. An alternative calibration would use multiple cross-validation experiments to build a relationship between votes and observed event frequencies, as described in Williams (2014).

d. Training datasets and model evaluation

Given the need for a training database for both LR and RF models, incipient cloud objects identified in spring and summer 2010 and 2011 GOES satellite imagery over the continental United States and nearshore locations along the Gulf of Mexico coast were tracked to determine whether radar reflectivity at the −10°C isotherm eventually exceeded 35 dBZ, representing a case of CI, or whether the cloud dissipated (a nonevent). The multiradar/multisensor (MRMS; Lakshmanan et al. 2006, 2007) products were used for assessing radar reflectivity at the −10°C altitude. A total of 9015 cloud objects were tracked to form this database. Of these, manual analysis found that 3270, or 36.3%, eventually met the CI criteria of the GOES-R algorithm as shown in Table 1 from Walker et al. (2012). The remaining tracked objects did not. Antecedent predictor data associated with all 9015 objects were collected, including RAP model fields (fields 10–24 in Table 1). Instead of using a strictly routine nearest-neighbor technique in mapping the RAP data to the objects, a 3 × 3 box of RAP pixels was used. For example, an *average* of the three highest CAPE values from the 3 × 3 box would be used, instead of the CAPE in the RAP pixel closest to a cloud object. This approach helped reduce possible errors associated with misplacement of features by the NWP model and reduced sensitivity to pixel-sized noise. Table 1 shows the GOES satellite fields that were recorded into the database (for the LR-Sat and RF-Sat models). For both LR-SatNWP and RF-SatNWP models, four additional satellite fields were added pertaining to the dominant cloud type and size of an object at the two times used to measure cloud evolution (Walker et al. 2012).

For evaluation purposes, the 9015-object database was divided into separate sets of training and testing cases. Cases from a particular date were either placed in

TABLE 2. List of cases used to generate the training database. Groups 1 and 2 refer to the ~50% split in the main training dataset, as a means of performing cross-validation training and testing to evaluate the logistic regression and random forest methods' skills. See the text for further description.

Group 1 (10 days):	2010: 11 Jul, 15 Jul, 2 Aug, 4 Aug, 6 Aug, 8 Aug, 30 Aug 2011: 7 Aug, 20 Aug, 27 Oct
Group 2 (11 days):	2010: 10 Jul, 14 Jul, 16 Jul, 18 Jul, 20 Jul, 1 Aug, 3 Aug, 5 Aug, 7 Aug 2011: 24 May, 21 Aug

group 1 (10 days) or group 2 (11 days) as shown in Table 2. To help ensure independence of the training and testing sets, cases from the same day were not split between groups. An effort was made to objectively split the data into two groups with adequate diversity in each group to ensure similar results from each group. The number of cases from each day varied: two days accounted for almost half (46%) of the CI cases and were assigned to different groups. Both the LR and RF methods were trained on group 1 and tested on group 2, and then the reverse was done, to obtain estimates of each method's performance and to allow a comparison of the two methods' skills. Therefore, two cross-validation performance scores are presented in the results section for each of a variety of forecast skill metrics. This cross-validation procedure provides a basis for comparing the LR and RF methods. For example, the resolved LR Eq. (1) with coefficients obtained from group 1 was run against the group 2 dataset, and each object whose estimated probability of CI exceeded 0.5 [$P(0.5)$] was considered to be a predicted CI event. Because the training dataset was selected to have roughly half CI and half non-CI cases per group, these skill scores provide estimates of the generalization performance expected when the CI nowcast methods are used in practice, and form a solid basis for comparing the various methods and sets of predictor variables, though some skill scores dependent on the relative frequencies of positive and negative instances will vary in other datasets. The RF skill scores were obtained similarly, though with the RF vote threshold chosen to optimize the equitable threat score. Last, the performance of LR and RF nowcast models trained on the entire dataset was evaluated on three independent case study days.

4. Results

The following sections evaluate the predictive performance of both the LR and RF 0–1-h CI nowcast models. As a means of assessing the overall algorithm performance, and to demonstrate an evolution of the GOES-R CI algorithm through this study's inclusion of NWP predictor variables, Table 3 (from Walker et al. 2012) lists performance statistics as obtained over four regions in the United States from the previous version of the GOES-R CI algorithm that used five GOES predictors in a deterministic (yes/no) CI nowcast model. Although the Walker et al. approach did not use LR, the same input satellite fields were used to determine the likelihood for CI for a given cloud object, namely fields 1–5 in Table 1. The Table 3 results serve then as a benchmark for measuring improvements to the GOES-R CI method, with the aim being to decrease false CI detections through use of convective environment information contained in the NWP fields, and to show an evolution from a deterministic approach toward probabilistic LR and RF models.

Accuracy and performance statistics are presented in Table 4 for cross evaluation using groups 1 and 2 for the four combinations of data and methods described above:

TABLE 3. From Walker et al. (2012). Validation statistics and averaged forecast lead times of CI from the GOES-R CI algorithm over four regions of the United States that used only five GOES interest fields, and no numerical weather prediction fields, within a deterministic yes/no 0–1-h CI forecasting methodology. These results are broken into the three classes of statistical evaluation for each of the four study regions. Note that only the POD and accuracy statistics change for the different classes. Class 1 relates to all CI forecasts and corresponding CI events that were associated only with tracked cloud objects, and class 2 represents all CI events documented in the validation study, whether they were associated with algorithm output forecasts or not, with the exception of those masked or affected by cirrus contamination, which were omitted. The definitions of POD, POFD, FAR, and Accuracy are provided, with H, M, F, and C indicative of hits, misses, false alarms, and correct negatives, respectively.

	Melbourne, FL		Memphis, TN		Central United States/Great Plains		Northeastern United States	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
POD [$H/(H + M)$]	54%	43%	74%	45%	85%	78%	75%	67%
POFD [$F/(F + C)$]	4%	4%	6%	6%	3%	3%	2%	2%
FAR [$F/(F + H)$]	60%	60%	48%	48%	55%	55%	54%	54%
Accuracy [$(H + C)/(H + C + M + F)$]	94%	92%	92%	87%	97%	97%	97%	97%
Lead time (minutes)	23.8	23.8	32.7	32.7	32.9	32.9	27.4	27.4

TABLE 4. Performance metrics for LR and RF CI nowcasts. All LR statistics are formed using a CI probability threshold of 0.5, while the RF threshold was chosen to optimize the ETS. Two performance metrics are listed per column, with the first being for group 1 data and the second for group 2 data, as described in the main text. The mathematical formula used to generate each skill score is shown in the far right column. Definitions for H, M, F, and C are as in Table 3, and r = (total forecasts of the event) × (total observations of the event)/(sample size), with sample size being $H+F+M+C$.

Method	LR-Sat		LR-SatNWP		RF-Sat		RF-SatNWP		Formula
POD	0.54	0.80	0.68	0.85	0.49	0.74	0.72	0.87	$H/(H + M)$
POFD	0.20	0.51	0.20	0.36	0.21	0.40	0.27	0.40	$F/(F + C)$
FAR	0.32	0.30	0.28	0.22	0.36	0.27	0.33	0.24	$F/(F + H)$
Accuracy/total performance	0.69	0.67	0.74	0.76	0.66	0.68	0.72	0.76	$(H + C)/(H + F + M + C)$
% correct CI nowcasts	53.7%	79.8%	67.6%	84.6%	49.5%	73.5%	71.9%	87.2%	$H/(H + M)$
% correct non-CI nowcasts	80.2%	48.6%	79.6%	64.3%	78.8%	59.5%	72.5%	60.4%	$C/(F + C)$
Positive predictive value	67.8%	69.5%	72.1%	77.7%	64.5%	72.8%	67.0%	76.4%	$H/(H + F)$
Negative predictive value	69.0%	62.1%	75.9%	73.9%	66.7%	60.5%	76.8%	76.3%	$C/(C + M)$
Bias	1.98	0.98	1.81	1.18	1.96	1.13	1.54	1.12	$(H + F)/(H + M)$
AUC	0.73	0.71	0.81	0.83	0.69	0.73	0.80	0.82	
CSI	0.43	0.59	0.54	0.68	0.39	0.58	0.53	0.69	$H/(H + F + M)$
ETS	0.21	0.17	0.31	0.33	0.17	0.20	0.28	0.33	$(H - r)/(H + F + M - r)$
TSS	0.34	0.28	0.47	0.49	0.28	0.33	0.44	0.48	$[(H)(C) - (F)(M)]/(H + M)(F + C)$
Votes/total trees in RF:					318/400	301/800	252/400	278/800	
Candidate predictors on which to split data at each RF node:					3	1	1	2	

1) LR trained and applied to only GOES CI interest fields (LR-Sat), 2) LR trained and applied to GOES CI and NWP model fields (LR-SatNWP), 3) RF trained and applied to only GOES CI interest fields (RF-Sat), and 4) RF trained and applied to GOES CI and NWP-model fields (RF-SatNWP). Last, three case days showing the relative performances of each approach are presented. Receiver operating characteristic (ROC) curves for the case study days are shown in Fig. 1, and the LR-SatNWP and RF-SatNWP CI nowcasts for the case studies are shown along with radar reflectivity (Figs. 2–4).

a. General performance

Table 4 lists the group 1 and 2 cross-validation performance metrics for each method (LR and RF) and predictor variable set (Sat or SatNWP). One outcome evident in Table 4 is that the performance skill scores for the group 1 and group 2 cross evaluations are often significantly different. This is likely due in part to differences in the meteorological scenarios and types of convection represented in the two groups, which particularly affect skill scores that depend on the ratio of positive to negative instances. Another reason for these differences is that our training dataset is small, and therefore does not capture the variability across a wide variety of convective storm environments. One of the most robust statistics, the area under the ROC curve (AUC; Mason 1982), is an exception. ROC curves plot probability of detection (POD) versus probability of false detection (POFD) as the forecast decision threshold changes, providing a useful summary of the discrimination capability of the predictive model. An AUC near 1 represents a nearly perfect forecast. In Table 4, the AUC statistic shows relatively little difference between group 1 and group 2. In general, a comparison of the group 1 and group 2 results is not expected to be particularly informative, while the more useful comparisons are between the results for the two different methods and the two-predictor variable sets.

As developed in Mecikalski et al. (2008) and Walker et al. (2012), the GOES IR fields provide considerable information about the status of cumulus development over time in terms of physical processes involved in CI (i.e., in-cloud updraft magnitude and depth, cloud-top glaciation, cloud growth), but as is often the case, nearly identical GOES observations can be observed for a CI event in an unstable environment versus for a growing cumulus in a suppressed/stable environment that never produces a ≥35-dBZ echo. The use of NWP fields, therefore, should act to constrain the GOES observations, and should provide increased skill for cases where the atmosphere surrounding the growing cumulus clouds is favorable/unfavorable for sustained growth toward a fully developed convective storm or organized convective system. The results in Table 4 suggest that this is indeed the case. For both the LR and RF methods, nearly every skill score statistic for the models using the satellite-plus-NWP variables shows a significant improvement over the same method using only the satellite

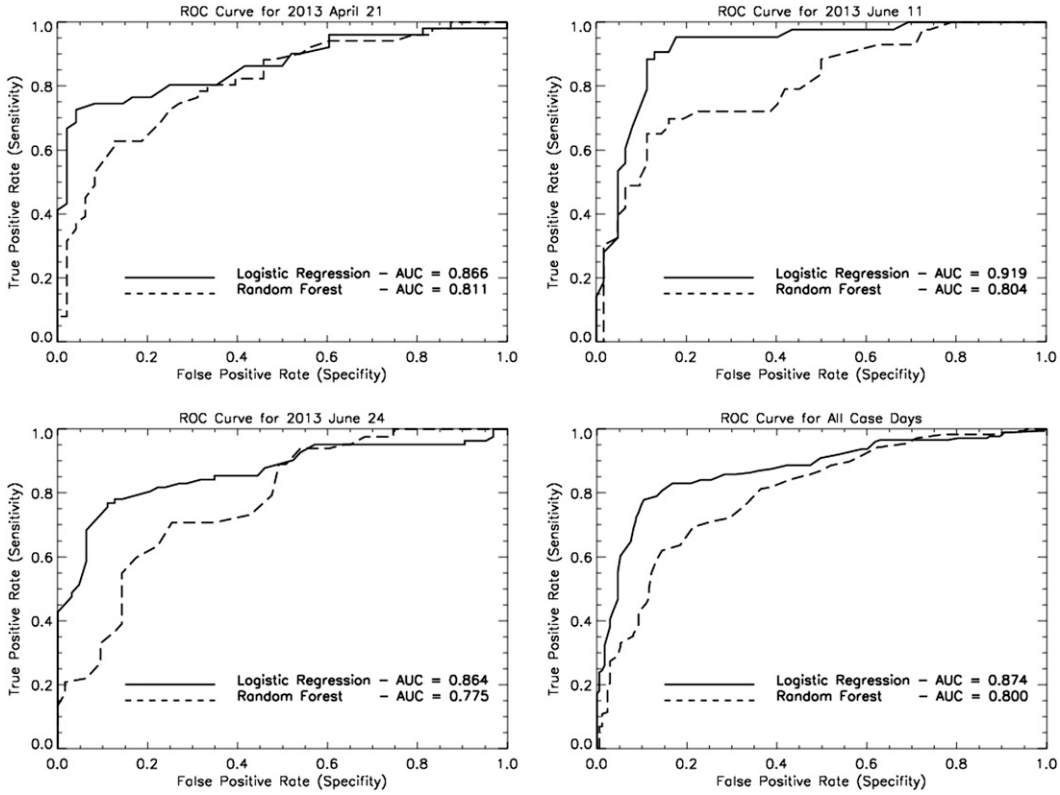

FIG. 1. ROC curves for the three case days analyzed—(top left) 21 Apr 2013, (top right) 11 Jun 2013, and (bottom left) 24 Jun 2013—and (bottom right) a summary of all three days. For these plots, the LR-SatNWP and RF-SatNWP are used, as shown in Table 4. These ROC diagrams provide a comparative summary of skills of the convective initiation nowcasts using logistic regression (solid curve) and random forest (dashed curve). The AUC values are listed per procedure, LR and RF.

variables. For example, the AUC improves by an average of 0.10 (~15%) for both LR-SatNWP over LR-Sat and for RF-SatNWP over RF-Sat. Similarly, the equitable threat score (ETS) improved by an average of 0.13, a relative increase of 68%, between LR-Sat and LR-SatNWP, and by 0.12 (65%) between RF-Sat and RF-SatNWP. The same trends are seen for the critical success index (CSI) and true skill statistic (TSS) measures, with the skills being higher when NWP data are included. It is also interesting to note that the difference in performance between group 1 and group 2 is significantly less for the SatNWP predictor variable set than for the Sat variables alone. For example, the difference in ''percent correct CI nowcasts'' is 26% for LR-Sat and 25% for LR-Sat, but only 17% for LR-SatNWP and 15% for RF-SatNWP. Thus, the GOES satellite+NWP predictor variables appear to provide both LR and RF methods a better capability for generalizing than do the satellite fields alone. In summary, it is clear that the addition of the NWP data provides a notable improvement in the CI nowcasting capability of both statistical learning methods. Note that the best combination of forest size, number of candidate predictors at each node, and vote threshold, computed separately for the RF-Sat and RF-SatNWP models and for group 1 data or group 2 data, is listed in Table 4 beneath the RF-Sat and RF-SatNWP columns.

With respect to the Walker et al. (2012) study, the Table 4 LR-Sat and RF-Sat statistics can be compared with those in Table 3 toward assessing the improvement to the CI nowcasting algorithm given that the same GOES-only variables were used, although the methodologies differ (simple per-CI object interest field scoring in Walker et al. vs LR or RF). The satellite-only POD results provide similar or higher skill when LR and RF are used, as seen when comparing classes 1 and 2 in Table 3, and between groups 1 and 2 in Table 4. POD scores when using satellite and NWP fields reached 87% for the RF method. When the group 1 and 2 false-alarm ratio (FAR) scores are compared (Table 3 vs Table 4), they range from 22% to 36% for all LR and RF categories (vs 48%–60% from the Walker et al. study),

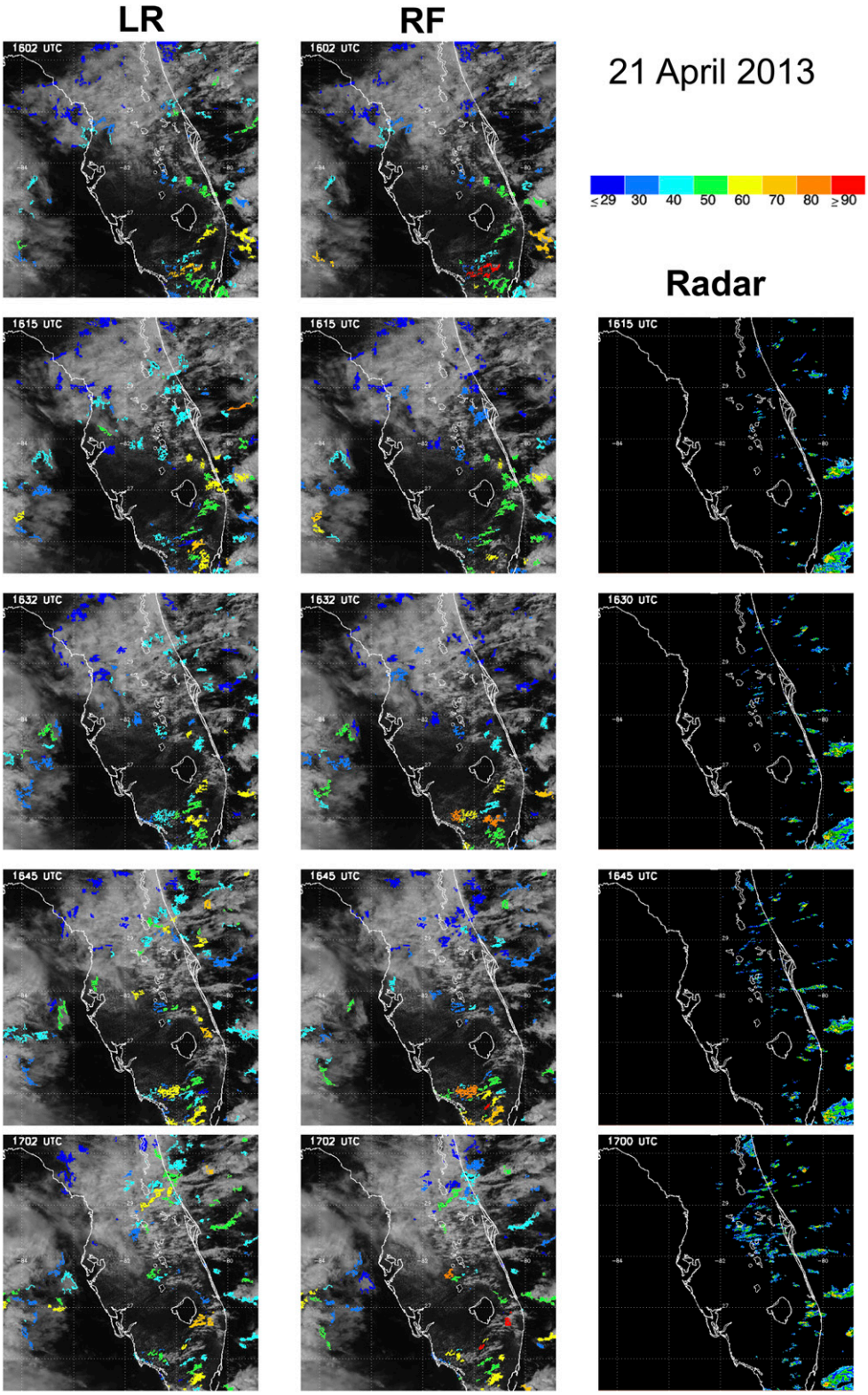

FIG. 2. Example of GOES-R CI nowcasts performed using (left) LR, (center) RF, and (right) WSR-88D radar for 21 Apr 2013. See text for discussion of these images. Times for the GOES satellite/GOES-R CI and WSR-88D are shown as UTC.

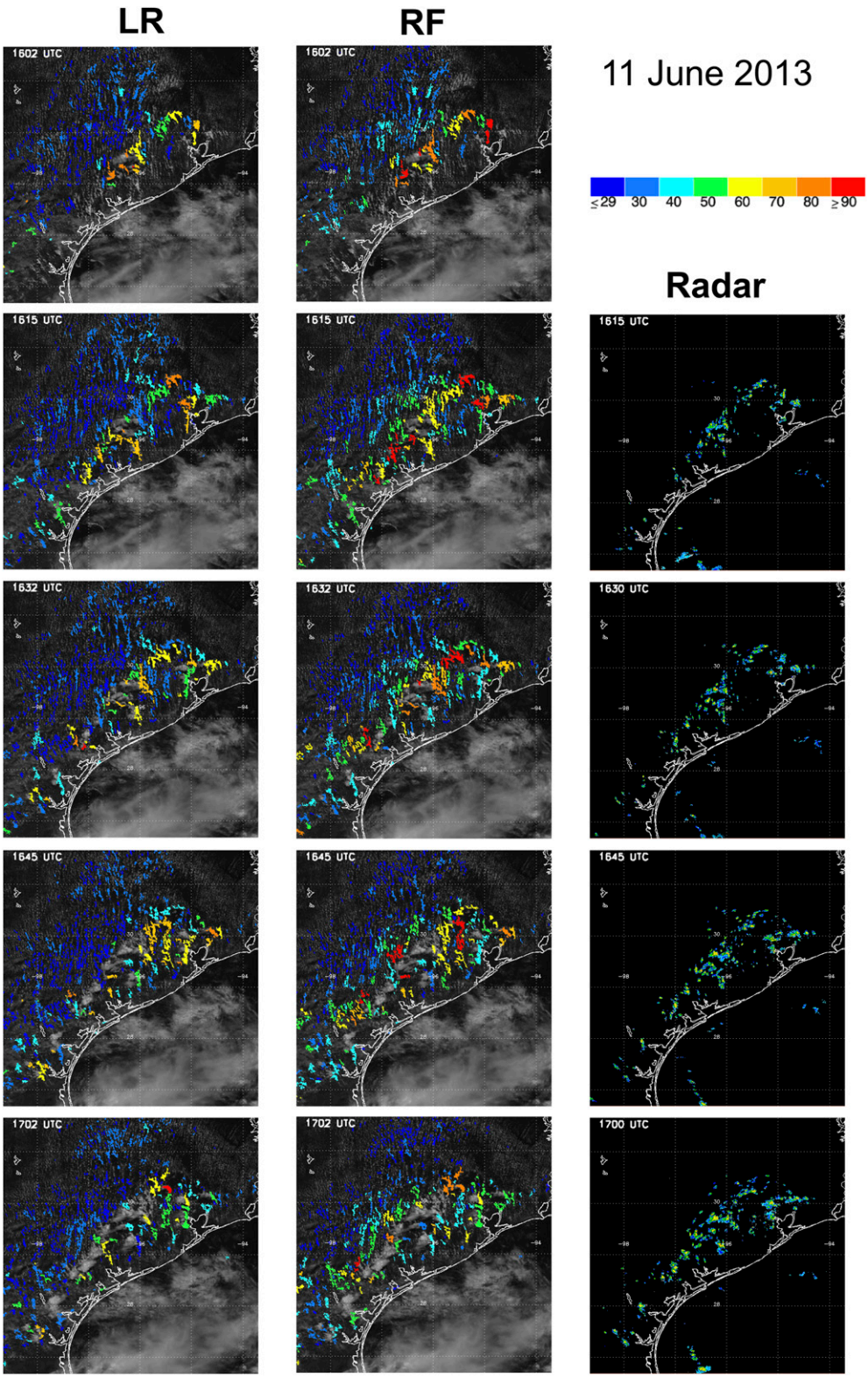

FIG. 3. As in Fig. 2, but for 11 Jun 2013.

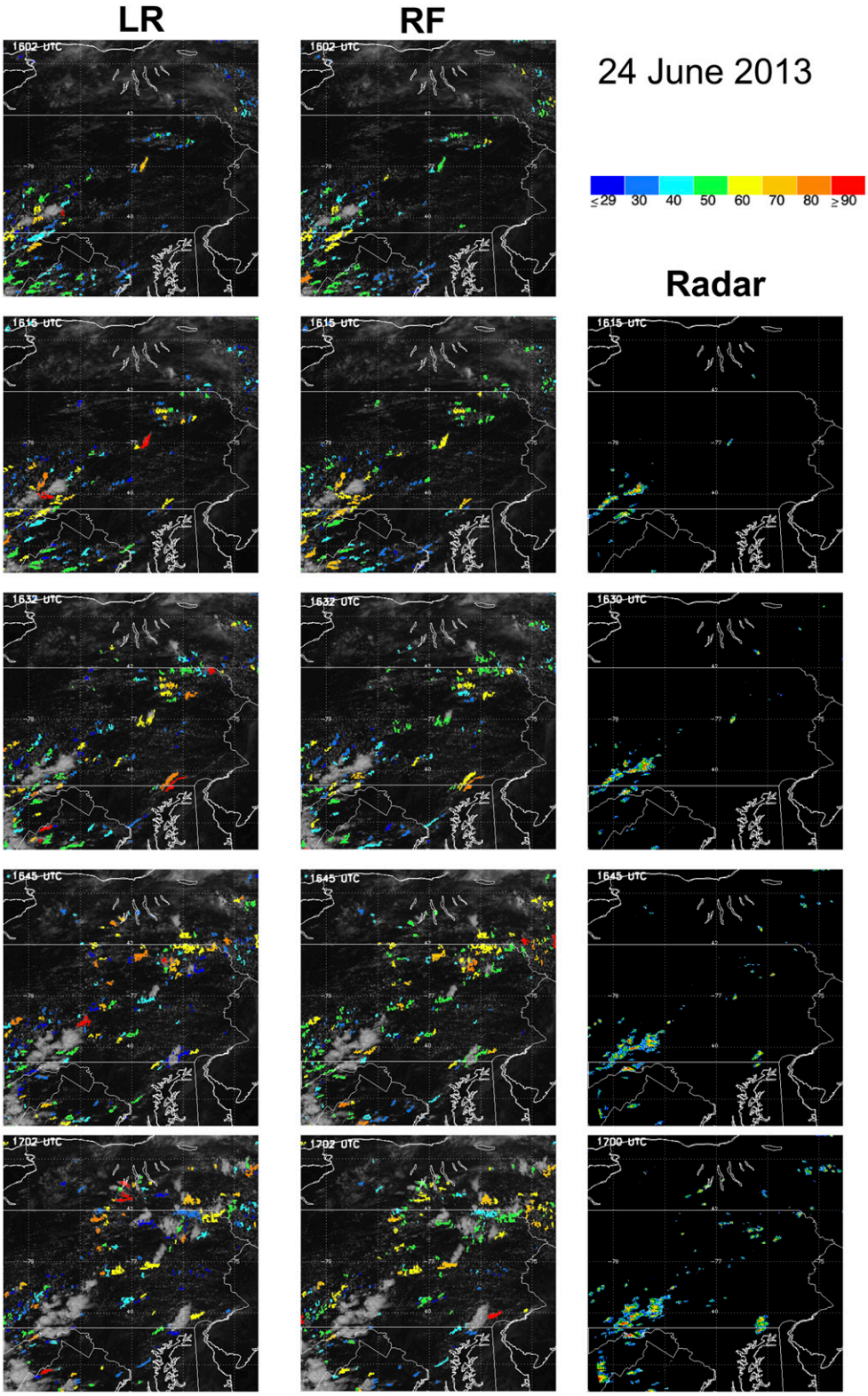

FIG. 4. As in Fig. 2, but for 24 Jun 2013.

suggesting again that the use of more advanced statistical methods overall leads to *significantly* better 0–1-h CI nowcasts. Noteworthy is that the FAR scores were several percentage points lower when NWP data were included, for both the LF and RF models. Note that the POFD and accuracy scores in the Walker et al. (2012) study are significantly better than for the statistical models developed in this study; this is because of the much larger fraction of non-CI events in the 2012 study, and hence the larger numbers of correct non-CI nowcasts (i.e., correct negative forecasts), as compared to the roughly 50/50 ratio used here (depending on group and dataset partitioning).

Comparing the two statistical methods themselves based on these results is not as straightforward. From Table 4, it is seen that the percent correct CI nowcasts for the LR-Sat and RF-Sat compare well, with the LR-Sat showing slightly higher skill, but the reverse is true in the comparison of LR-SatNWP and RF-SatNWP. The skill scores of the percentage correct non-CI nowcasts also compare well between LR-Sat and RF-Sat, yet the difference between groups for RF is much lower. The remaining skill measures in Table 4 give a slight advantage to the LR approach when only the GOES fields are used. One of the most robust measures of skill, the AUC (Fig. 1), shows that the RF and LR approaches are nearly identical when GOES fields are used alone, providing AUCs of ~0.72. For LR and RF with NWP fields included, the percent correct CI and non-CI nowcasts again vary considerably between groups 1 and 2, yet not as much as when NWP data are not included. The other performance measures for LR-SatNWP and RF-SatNWP similarly compare well. Overall, the LR results again show a slight (~2%–4%) advantage over the corresponding RF scores. For example, the AUC reaches 0.83 for LR-SatNWP, versus 0.82 for RF-SatNWP, although it is not clear that this margin is statistically significant. On the other hand, LR models are much simpler and can execute more quickly in making predictions, so in the face of this ambiguity, the LR model likely has the advantage.

b. *Case examples*

Figure 1 shows the ROC curves for three example case days in 2013 (21 April, 11 June, and 24 June, which are not in the training database), and then for the average of all three days, with the AUC values provided for both the LR-SatNWP and RF-SatNWP models, trained now using the entire dataset. As seen in Table 4, the AUC for LR is greater than for RF, for all case days and for the average. In all cases, the ROC curve increases more rapidly for the LR model than the corresponding RF model. The AUC values in Fig. 1 can be interpreted as accepting an ~0.15–0.20 false-positive rate to achieve a true positive (CI nowcast) rate between 0.75 and 0.90. The best performances were seen on 21 April and 11 June 2013. For the ROC curves obtained from all three case study days, the AUC is 0.87 for LR-SatNWP and 0.80 for RF-SatNWP, and is comparable to the results shown in Table 4. In all cases, the AUC is at least 0.05 larger using LR-SatNWP than when using the RF-SatNWP methodology, reflecting the somewhat better performance of the LR-SatNWP method.

The three case days are shown visually for select periods when CI was particularly active, in Figs. 2–4, respectively, with accompanying National Weather Service Weather Surveillance Radar-1988 Doppler (WSR-88D) data (Zhang et al. 2011). These figures provide examples of how both LR-SatNWP (left column) and RF-SatNWP (center column) CI nowcasts would appear to a user, with the colors of a cumulus object representing an estimated percentage probability (30%–100%) that CI will occur within the coming 15–60 min. For 21 April 2013 (Fig. 2), an example is shown over Florida, while on 11 June 2013 (Fig. 3) and 24 June 2013 (Fig. 4), examples are given over southeastern Texas and centered on Pennsylvania, respectively. The accompanying radar (right column) needs to be checked 15–60 min later for the validating radar echo corresponding to a given highlighted CI object. Radar echoes in yellow and red colors correspond to echo intensities ≥35 dBZ.

Table 5 shows the LR-SatNWP and RF-SatNWP performance statistics as a function of CI probability bin for the three case study days (Figs. 2–4), from 30% to 39% (LR_{30} and RF_{30}) to 90% to 99% (LR_{90} and RF_{90}), toward identifying when peak deterministic performance can be expected. From Table 5, using accuracy/total performance, CSI, ETS, and TSS, it is seen that both the LR and RF methods peak in performance when the CI probabilities are between 50% and 60%. As another measure of skill, the Brier score for each case day was computed. The Brier score (Brier 1950) is a measure of probabilistic prediction accuracy (Wilks 2011, 332–333). A Brier score of 0 is considered a perfect forecast (i.e., the better the forecast verification), while a score of 1 indicates a useless forecast. For 21 April for LR and RF (LR/RF) the Brier score values are 0.180/0.183, for 11 June, they are 0.155/0.198, and for 24 June they are 0.161/0.202. These results suggest that both methods provide robust skill beyond a reference forecast, but that the LR performance is superior to RF when both are evaluated probabilistic predictions.

c. *Variable importance*

For RFs, an ''importance'' measure is provided for each predictor variable in the course of model training.

TABLE 5. Performance metrics for the three case study days as shown in Figs. 2–4 for LR and RF CI nowcasts. Results are shown for LR-SatNWP and RF-SatNWP as in Table 4 yet are binned by CI probability (as predicted by LR and RF) from 30% to 39% (i.e., LR_{30} and RF_{30}), 40% to 49% (i.e., LR_{40} and RF_{40}), and so forth to 90% to 99%. The mathematical formula used to generate each skill score are as in Table 3. In bold are maximum values of accuracy, total performance, CSI, ETS, and TSS, showing that both the LR and RF methods peak in performance when the CI probabilities are between 50% and 60%.

Method	LR_{30}	LR_{40}	LR_{50}	LR_{60}	LR_{70}	LR_{80}	LR_{90}	RF_{30}	RF_{40}	RF_{50}	RF_{60}	RF_{70}	RF_{80}	RF_{90}
POD	0.966	0.909	0.841	0.716	0.542	0.216	0.074	0.989	0.949	0.841	0.551	0.324	0.222	0.108
POFD	0.775	0.497	0.231	0.087	0.023	0.006	0.000	0.908	0.659	0.428	0.121	0.052	0.029	0.012
FAR	0.441	0.350	0.213	0.106	0.059	0.026	0.000	0.474	0.406	0.333	0.178	0.136	0.114	0.095
Accuracy/total performance	0.599	0.708	0.805	**0.814**	0.664	0.675	0.533	0.544	0.648	0.708	**0.713**	0.633	0.593	0.544
% correct CI nowcasts	96.6%	90.9%	84.1%	71.6%	54.2%	21.6%	7.4%	98.9%	94.9%	84.1%	55.1%	32.4%	22.2%	10.8%
% correct non-CI nowcasts	22.5%	50.3%	76.9%	91.3%	97.7%	99.4%	100.0%	9.2%	34.1%	57.2%	87.9%	94.8%	97.1%	98.8%
Positive predictive value	55.9%	65.0%	78.7%	89.4%	94.1%	97.4%	100.0%	52.6%	59.4%	66.7%	82.2%	86.4%	88.6%	90.5%
Negative predictive value	86.7%	84.5%	82.6%	76.0%	59.7%	55.5%	51.5%	88.9%	86.8%	78.0%	65.8%	58.0%	55.1%	52.1%
Bias	1.727	1.398	1.068	0.801	0.382	0.222	0.074	1.881	1.597	1.261	0.670	0.375	0.250	0.119
CSI	0.548	0.611	**0.685**	0.660	0.352	0.215	0.074	0.523	0.576	**0.592**	0.492	0.308	0.215	0.107
ETS	0.107	0.261	**0.439**	0.287	0.200	0.117	0.038	0.043	0.171	0.261	**0.273**	0.156	0.106	0.050
TSS	0.191	0.412	0.610	**0.629**	0.336	0.210	0.074	0.081	0.290	0.413	**0.430**	0.272	0.193	0.096

Although importance is calculated deep in the RF algorithm (Topić and Šmuc 2014), it is conceptually fairly simple. In essence, the method replaces each variable in turn with a randomized resampling, and then measures the impact of that replacement on the RF trees' prediction accuracy. The predictor variables with the highest importance are those whose randomization degrades accuracy the most; they are the variables that are frequently chosen by the random forest for splitting and that most meaningfully separate the data. Ideally, the expected importance of a random predictor will be near zero; a higher importance value suggests the predictor is more useful than a random variable, and the highest values indicate the greatest importance.

The RF importance values for the 25 satellite+NWP fields in group 1 and group 2 are listed in Table 6. This ranked list shows two measures of CIN and two measures of CAPE to be the most important, followed by the 10.7-μm T_B (longwave IR, representing cloud-top temperature) and LFC height. Nearness to convective temperature, which is very similar to CIN, is next. Six out of the first seven predictors are derived from the thermodynamic stability profile provided by the NWP model.

One downside of the RF importance calculation is that it does not reflect correlations between variables. When several highly correlated variables are utilized, they are all assigned approximately the same importance, although only one may be needed for a skillful model. Thus, one should not read too much into the presence of two versions of CIN at the top of the list; likely only one of them is truly needed. Surface-based and most-unstable parcel CIN are very similar and are in fact highly correlated (correlation coefficient ≈ 0.9 for this dataset). At the same time, this list does show that some predictors are more useful than others. For example, LCL height is more important than convective condensation level (CCL) height, and dominant cloud type (time 1) and land/sea flag were the least useful of their satellite+NWP peers. Variables that show low RF importance are likely not useful for the RF model and may have the potential to harm the generalization error of the learned model if they are included.

While RF variable importance is a good indicator of the value of individual predictor variables, it does not determine a minimal set of variables that can be used collectively by a model to provide the best predictive performance. To achieve this, a forward–backward selection method was used in which two forward steps (adding variables that most improved the cross-validation performance of the model) were followed by a backward step (removing the variable that harmed the cross-validation performance of the model the least). Thus, at any stage of forward–backward selection, this method provides an approximation to the best performance that can be achieved with that given number of predictor variables. This method can be applied to any statistical learning method, including RF and LR.

Performing forward–backward selection with group 1 as the training set and group 2 as the cross-validation set,

TABLE 6. Random forest variable importance ranks and scores obtained from groups 1 and 2. See text for the interpretation of these results.

Rank	Importance	Predictor
1	3.61	Surface-based CIN
2	3.31	Most-unstable CIN
3	3.29	Surface-based CAPE
4	3.22	MUCAPE
5	2.98	10.7-μm T_B (cloud-top temp)
6	2.65	LFC height
7	2.37	Nearness to convective temp
8	2.36	LCL height
9	2.32	Object size at time 2
10	2.29	Convective temp
11	2.22	13.3–10.7-μm T_B diff
12	2.09	Freezing-level height
13	2.04	Surface–700-hPa wind shear
14	2.03	Surface lifted index
15	2.00	Surface–500-hPa wind shear
16	1.95	CCL height
17	1.75	Best lifted index
18	1.65	6.5–10.7-μm T_B diff
19	1.56	10.7-μm T_B cooling rate
20	1.31	Object size at time 1
21	1.04	700–500-hPa wind shear
22	1.00	13.3–10.7-μm T_B temporal diff
23	0.93	Dominant cloud type at time 2
24	0.43	Land/sea flag
25	0.39	Dominant cloud type at time 1

and then vice versa, and aggregating the results showed that for LR, the cross-validation performance peaked at 7 variables and declined after 11 variables, whereas for RF the performance peaked at 11 variables and declined only after about 20. This result suggests that using too many predictor variables in either method causes overfitting, but more quickly and significantly for LR. The variable rankings based on their number of occurrences in the forward–backward selection (Occ) are shown in Table 7 for LR and RF. The results for RF are notably different than the variable importance rankings: similar variables (e.g., most-unstable CIN and surface-based CIN, or surface-based CAPE and most-unstable CAPE) are now separated. This is because once the more useful of the pair has been selected for inclusion, the other has little marginal value. The land/sea flag, which had an RF importance rank near the bottom (24), was selected as the 10th variable. It is also notable that the variable rankings are significantly different between LR and RF. For instance, 10.7-μm T_B has rank 3 for RF but only 17 for LR; similarly, surface-based CAPE has rank 7 for RF and 15 for LR. These results underscore that the inherently nonlinear RF method uses the information in the variables differently from the linear LR method.

For LR, an additional way to evaluate variable importance is to observe the *absolute value* of their coefficients in a trained LR model, assuming that the predictor variables have been normalized to have mean zero and variance one prior to calculating the LR fit. A higher magnitude weight for a predictor variable indicates that it will contribute more to the linear combination, and therefore have a bigger impact on the LR model output. Removing variables with small coefficients can reduce overfitting and therefore improve the LR model's generalization performance. Table 8 lists the weights in order of absolute value magnitudes for the LR equation for all input variables in Table 1. The weights in Table 8 are for the LR-SatNWP model that produced the highest AUC of 0.83 in Table 4. The largest weight is for the most-unstable CAPE of 0.666, with the second highest being for the surface-based CAPE of −0.626, followed by the cumulus cloud object size at time 2 (0.520), the 13.3–10.7-μm channel difference (0.513), and the most-unstable CIN (0.496). The variable with the lowest weight was LFC height (−0.005). Interpretation of these weights is that CI is most sensitive to and associated with high instability (most-unstable CAPE, surface-based CAPE), low CIN, rapidly developing cumulus clouds (object size at time 2), and preexisting nearby convection (the 13.3–10.7-μm channel difference). Specifically, the correspondence between nearby more mature convection (as identified as small 13.3–10.7-μm values) and new CI occurrence is interesting with the association being that CI is more likely to occur in unstable environments (high CAPE, low CIN) when convective storms are ongoing in the near vicinity (within a distance of several GOES 4-km-resolution pixels). It is important to note that because the coefficients of most-unstable CAPE and surface-based CAPE have opposite signs but similar absolute values (Table 8), the two variables may be highly correlated. A goal of future LR analysis will be to remove highly correlated predictor variables and repeat the analysis (which would likely lead to one type of CAPE with a positive coefficient being overall less important in the GOES-R CI algorithm, and perhaps less important than other satellite fields).

5. Conclusions

The study presented here demonstrates how GOES IR satellite observations of cumulus cloud objects can be combined with fields from NWP models (Table 1) to predict short-term (0–1 h) CI nowcasts using two statistical learning approaches. In a previous study by Walker et al. (2012) the GOES-R CI algorithm methodology was developed, and therefore this study extends that previous analysis by incorporating NWP fields, and developing probabilistic LR and RF statistical modeling techniques. As a means of quantifying the benefits of both LR and RF, and of comparing these two approaches, 13 measures

TABLE 7. Forward–backward variable selection ranks and occurrences (Occ) for LR and RF, respectively, aggregated from groups 1 and 2, as described in the text.

LR			RF		
Rank	Occ	Predictor variable name	Rank	Occ	Predictor variable name
1	64	Most-unstable CIN	1	70	Most-unstable CIN
2	63	13.3–10.7-μm T_B diff	2	68	Object size at time 2
3	55	13.3–10.7-μm T_B temporal diff	3	59	10.7-μm T_B
4	52	10.7-μm T_B cooling rate	4	59	Dominant cloud type at time 2
5	50	Dominant cloud type at time 2	5	48	6.5–10.7-μm T_B diff
6	47	Object size at time 2	6	47	10.7-μm T_B cooling rate
7	45	6.5–10.7-μm T_B diff	7	47	Surface-based CAPE
8	45	Dominant cloud type at time 1	8	47	Dominant cloud type at time 1
9	44	Surface-based CIN	9	46	13.3–10.7-μm T_B diff
10	44	700–500-hPa wind shear	10	40	Land/sea flag
11	38	Nearness to convective temp	11	39	LCL height
12	36	Object size at time 1	12	35	Surface-based CIN
13	36	MUCAPE	13	34	Object size at time 1
14	36	Convective temp	14	34	Surface lifted index
15	35	Surface-based CAPE	15	33	CCL height
16	35	Freezing-level height	16	33	13.3–10.7-μm T_B temporal diff
17	30	10.7-μm T_B	17	32	Freezing-level height
18	29	CCL height	18	28	Surface–700-hPa wind shear
19	25	Land/sea flag	19	27	MUCAPE
20	24	Surface–700-hPa wind shear	20	26	Nearness to convective temp
21	24	Surface–500-hPa wind shear	21	25	Surface–500-hPa wind shear
22	23	Surface lifted index	22	19	Convective temp
23	17	LCL height	23	17	LFC height
24	17	Best lifted index	24	13	Best lifted index
25	14	LFC height	25	4	700–500-hPa wind shear

of skill were used to evaluate cross-validation experiments using two disjoint subgroups of the training dataset (Table 4). As noted, the main motivation of this research was not to show which statistical learning approach is superior but rather to highlight two points: 1) the value of combining GOES satellite indicators with information on the convective environment (as obtained in this case from NWP) toward generating short-term CI predictions with higher quality than can be obtained when using GOES data alone (e.g., as in Mecikalski et al. 2008; Sieglaff et al. 2011) and 2) how statistical-learning-based methods can be used in constructing applications that benefit NWS and other forecast systems by providing improved forecasts. The outcome of this study is to provide improved GOES-R CI nowcasts to NWS forecasters, which is the broader impact of this work. The CI nowcasts will be evaluated objectively in test beds like HWT and AWT.

The training dataset used here consists of 16 convectively active storm days in 2010 and 5 similar days in 2011. This main dataset was divided into two roughly equal groups (Table 2) that were used for cross-validation evaluations (i.e., one for training and the other for testing, then vice versa). To help ensure independence of the training and testing sets, cases from the same day were not split between groups. The LR and RF models were then trained on the entire dataset, and the predictive models were run on three new days from the spring and summer of 2013 (as shown in Figs. 1–4). The main conclusions are summarized as follows: 1) With the proper choice of probability thresholds, both the LR and RF techniques using NWP data produce fewer false alarms and show better skill for a variety of evaluation metrics than the GOES-only method. Use of NWP information helps identify environmental conditions (as favorable or unfavorable) for the development of convective storms, information that is not provided by the satellite observations alone. Thus, the NWP data are valuable for improving the skill of CI nowcasts that operate on a GOES-satellite-based cloud object, as compared to when only IR fields are used. 2) The LR procedure as outlined here performed slightly better than the RF on the training set cross validations when a variety of skill measures were used to quantify the results, but it was not clear whether this advantage was significant. 3) The LR method performed better than the RF on the three case studies from 2013 based on several skill scores, and the Brier scores for the LR method were better than for RF, which is designed to evaluate probabilistic predictions. Given that the

TABLE 8. Per-field weights as produced by XLSTATS for the LR model, listed in order of the highest to lowest absolute value of the weights. The weights shown are for the model that provided the largest area under the ROC curve (0.83 for LR-SatNWP in Table 4). See Table 1 for definitions. For those not in Table 1: t1 is the first time of cloud object tracking in the GOES-R CI algorithm, and t2 is the second time of cloud object tracking in the GOES-R CI algorithm. The largest weight magnitude was found for the MUCAPE (0.666). See text for description of the physical interpretation of these results.

Source—all parameters	Weights
MUCAPE	0.666
Surface-based CAPE	−0.626
Object size at t2	0.520
13.3–10.7-μm diff	0.513
Most-unstable CIN	0.496
Surface lifted index	−0.379
6.5–10.7-μm diff	−0.275
Convective temp	0.242
Surface-based CIN	0.239
Best lifted index	0.229
10.7-μm cooling rate	−0.219
Freezing-level height	−0.177
Surface–700-hPa wind shear	−0.176
700–500-hPa wind shear	−0.173
13.3–10.7-μm temporal diff	−0.166
Nearness to convective temp	−0.141
Surface–500-hPa wind shear	0.137
Object size at t1	0.115
10.7-μm temp	−0.097
Land/sea flag (0—land, 1—sea)	0.086
CCL height	0.075
Dominant cloud type at t2	−0.063
LCL height	0.033
Dominant cloud type at tl	0.009
LFC height	−0.005

implementation of the trained LR model to make CI nowcasts is significantly simpler than the RF, these results suggest that the LR method is more appropriate for this application. 4) The performance of both the LR and RF methods when using satellite and NWP fields together peaks for several deterministic skill scores when the predicted CI probability threshold is near 60%. And 5), in terms of variable importance between both statistical methods, CAPE and CIN fields that measure the amount of stability and lack of an inhibiting mechanism for the release of the instability (low CIN) are most important, along with GOES information on the depth of the clouds being analyzed (10.7-μm T_B), the presence of nearby convection (low 13.3–10.7-μm channel differences), and the size of the convective cloud object at the second 15-min GOES time of analysis (t2). Last, a forward–backward variable selection procedure showed that performance peaked at 7 variables and declined after 11 variables for LR, while for RF the performance peaked at 11 variables and declined only after about 20.

The superior performance of the LR method for the CI nowcast problem was surprising to the authors, given previous comparisons of these two approaches for other complex meteorological forecast problems (e.g., Williams 2014). We speculate that two aspects of the CI nowcast problem account for this result: 1) The satellite and NWP predictor variables selected for this study were generally monotonically related to the likelihood of CI, and thus were ideally suited for being combined using a linear method. The RF method may be better suited to domains in which the predictive relationships of the variables are more complicated. 2) As a nonlinear model capable of capturing complex predictive relationships, the trained RF model may have keyed on some idiosyncratic aspects of the 2010 and 2011 data and thus did not generalize as well to predicting CI for the 2013 cases. This failure to generalize well could be addressed by performing more careful variable selection, using only those predictor variables that were most useful and thereby simplifying the trained model. Related to this unexpected performance is the impact of the relatively small training database size used in this study, which was collected mainly near the Gulf of Mexico (as guided by the project's funding—see the acknowledgments). Present work is focused toward automating the validation procedure, forming a training database >20 times in size toward capturing CI in a wide variety of regimes and cloud conditions (e.g., from completely clear to cumulus clouds partially obscured by higher clouds), which will be reported in subsequent science papers. Another goal of future work will be to remove correlated predictor fields and arrive at a refined set of only the most important CI predictors.

An important future consideration is how this research will be improved once the GOES-R generation of geostationary satellites becomes operational. These improvements fall into four themes: 1) Increased visible channel resolution to 500 m, IR spatial resolution to 2 km, and increased time resolution to 5 min will help automated algorithms discern cumulus clouds better than is possible today with the current GOES satellites. This increased resolution will afford enhanced nowcast lead times for CI. Furthermore, the use of 5-min-resolution data will improve the tracking of small-scale cumulus clouds, especially in a GOES-R-like system where cloud object overlap can be relied upon. Presently, with 15-min GOES data, many smaller clouds are missed at early stages of growth given significant cloud evolution between image scenes, which is especially the case for more rapidly moving clouds (Mecikalski et al.

2008). The higher temporal resolution will also facilitate improved measures of the rates of change of cloud properties, which should further increase CI nowcast accuracy. 2) The additional channels (16 vs the present-day 4) will help improve our ability to determine cloud-top glaciation, and likely cloud-top heights. The current GOES lacks channels that can be used to infer glaciation, whereas having the 8.5- and 12.3-μm channels, as well as three water vapor channels (6.19, 6.95, and 7.34 μm, instead of just the 6.5-μm channel), will greatly improve our ability to detect ice versus water particles (see, e.g., Strabala et al. 1994). The added 10.35-μm channel on GOES-R (in addition to 11.2 μm) will also improve estimates of cloud height. All of these will be critical satellite indicators of rapidly growing, tall clouds. 3) Presently, efforts are being made toward evaluating how properties that describe cloud-top microphysics can be exploited in the GOES-R CI system to determine updraft strength, updraft width, and future storm intensity, along the lines of work done by Rosenfeld et al. (2008). Use of statistical learning and other statistical methods (beyond the more simple scoring methods initially employed in satellite-based CI nowcasting) can reasonably be trained for use in producing short-term predictions of storm intensity. Also, products that estimate cloud optical depth have shown value in the detection of growing cumulus beneath thin cirrus (Minnis et al. 2011a,b; Mecikalski et al. 2013). 4) Finally, this research on improved methods to nowcast CI will be synergistic with the forthcoming Geostationary Lightning Mapper (GLM) instrument also on GOES-R, which will provide 8-km spatial and ~20 s temporal resolution lightning observations over the GOES-R field of view (Goodman et al. 2013). Thus, an eventual plan would be to nowcast CI and use the GLM to help monitor storm evolution for high-impact severe weather events. Statistical learning methods like LR and RF may provide an ideal test bed for determining which of this rich set of future predictors will be most valuable for CI nowcasting, and for combining them to make skillful forecasts.

Last, the improved satellite-based CI nowcasts that utilize statistical learning can be employed within existing systems that monitor and track convection for a variety of users. Systems of this kind include the CIWS (Wolfson and Clark 2006), CbTRAM (Zinner et al. 2008), and the Rapidly Developing Thunderstorm (RDT; Autones 2012).

Acknowledgments. This project was supported by the National Aeronautics and Space Administration (NASA) Gulf of Mexico Research Grant NNX10AO07G. The authors thank Dr. Dan Lindsey (NOAA Center for Satellite Applications and Research) and two other anonymous reviewers for comments and suggestions that significantly improved the quality of this paper.

REFERENCES

AMS Council, 2008: Enhancing weather information with probability forecasts. *Bull. Amer. Meteor. Soc.,* **89,** 1049–1053. [Available online at http://www.ametsoc.org/policy/2008enhancingweatherinformation_amsstatement.pdf.]

Autones, F., 2012: Product user manual for "Rapid Development Thunderstorms" (RDT-PGE11 v3.0d). EUMETSAT Network of Satellite Application Facilities, Météo France, 27 pp. [Available online at http://www.nwcsaf.org/scidocs/Documentation/SAF-NWC-CDOP2-MFT-SCI-PUM-11_v3.0d.pdf.]

Bedka, K. M., and J. R. Mecikalski, 2005: Application of satellite-derived atmospheric motion vectors for estimating mesoscale flows. *J. Appl. Meteor.,* **44,** 1761–1772, doi:10.1175/JAM2264.1.

——, J. Brunner, R. Dworak, W. Feltz, J. Otkin, and T. Greenwald, 2010: Objective satellite-based detection of overshooting tops using infrared window channel brightness temperature gradients. *J. Appl. Meteor. Climatol.,* **49,** 181–202, doi:10.1175/2009JAMC2286.1.

Benjamin, S. G., and Coauthors, 2009: Rapid Refresh/Rapid Update Cycle (RR/RUC) technical review. NOAA/ESRL/GSD Internal Review, 168 pp. [Available online at http://ruc.noaa.gov/pdf/RR-RUC-TR_11_3_2009.pdf.]

Berendes, T. A., J. R. Mecikalski, W. M. Mackenzie, K. M. Bedka, and U. S. Nair, 2008: Convective cloud detection in satellite imagery using standard deviation limited adaptive clustering. *J. Geophys. Res.,* **113,** D20207, doi:10.1029/2008JD010287.

Breiman, L., 2001: Random forests. *Mach. Learn.,* **45,** 5–32, doi:10.1023/A:1010933404324.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probabilities. *Mon. Wea. Rev.,* **78,** 1–3, doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

Brooks, H. E., and N. Dotzek, 2007: The spatial distribution of severe convective storms and an analysis of their secular changes. *Climate Extremes and Society,* H. F. Diaz and R. Murnane, Eds., Cambridge University Press, 35–53.

——, C. A. Doswell III, and R. A. Maddox, 1992: On the use of mesoscale and cloud-scale models in operational forecasting. *Wea. Forecasting,* **7,** 120–132, doi:10.1175/1520-0434(1992)007<0120:OTUOMA>2.0.CO;2.

——, ——, and J. Cooper, 1994: On the environments of tornadic and nontornadic mesocyclones. *Wea. Forecasting,* **9,** 606–618, doi:10.1175/1520-0434(1994)009<0606:OTEOTA>2.0.CO;2.

——, ——, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting,* **18,** 626–640, doi:10.1175/1520-0434(2003)018<0626:CEOLDT>2.0.CO;2.

Browning, K. A., and D. Atlas, 1965: Initiation of precipitation in vigorous convective clouds. *J. Atmos. Sci.,* **22,** 678–683, doi:10.1175/1520-0469(1965)022<0678:IOPIVC>2.0.CO;2.

Coniglio, M. C., H. E. Brooks, S. J. Weiss, and S. F. Corfidi, 2007: Forecasting the maintenance of quasi-linear mesoscale convective systems. *Wea. Forecasting,* **22,** 556–570, doi:10.1175/WAF1006.1.

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

Dance, S., E. Ebert, and D. Scurrah, 2010: Thunderstorm strike probability nowcasting. *J. Atmos. Oceanic Technol.,* **27,** 79–93, doi:10.1175/2009JTECHA1279.1.

Dattatreya, G. R., 2009: Decision trees. *Artificial Intelligence Methods in the Environmental Sciences,* S. E. Haupt, C. Marzban, and A. Pasini, Eds., Springer, 424 pp.

Díaz-Uriarte, R., and S. A. de Andrés, 2006: Gene selection and classification of microarray data using random forest. *BMC Bioinf.,* **7,** 3, doi:10.1186/1471-2105-7-3.

Dixon, M., and G. Wiener, 1993: TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A radar-based methodology. *J. Atmos. Oceanic Technol.,* **10,** 785–797, doi:10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2.

Dixon, P., A. Mercer, J. Choi, and J. Allen, 2011: Tornado risk analysis: Is Dixie Alley an extension of Tornado Alley? *Bull. Amer. Meteor. Soc.,* **92,** 433–441, doi:10.1175/2010BAMS3102.1.

Duda, D. P., and P. Minnis, 2009: Basic diagnosis and prediction of persistent contrail occurrence using high-resolution numerical weather analyses/forecasts and logistic regression. Part II: Evaluation of sample models. *J. Appl. Meteor. Climatol.,* **48,** 1790–1802, doi:10.1175/2009JAMC2057.1.

Evans, J. E., and E. R. Ducot, 2006: Corridor Integrated Weather System. *Lincoln Lab. J.,* **16,** 59–80. [Available online at https://www.ll.mit.edu/publications/journal/pdf/vol16_no1/16_1_4EvansDucot.pdf.]

Feng, Y., Y. Wang, T. Peng, and J. Yan, 2007: An algorithm on convective weather potential in the early rainy season over the Pearl River Delta in China. *Adv. Atmos. Sci.,* **24,** 101–110, doi:10.1007/s00376-007-0101-2.

Fritsch, J. M., and Coauthors, 1998: Quantitative precipitation forecasting: Report of the Eighth Prospectus Development Team, U.S. Weather Research Program. *Bull. Amer. Meteor. Soc.,* **79,** 285–299, doi:10.1175/1520-0477(1998)079<0285:QPFROT>2.0.CO;2.

Goodman, S. J., and Coauthors, 2013: The GOES-R Geostationary Lightning Mapper (GLM). *Atmos. Res.,* **125–126,** 34–49, doi:10.1016/j.atmosres.2013.01.006.

Harris, R. J., J. R. Mecikalski, W. M. MacKenzie, P. A. Durkee, and K. E. Nielsen, 2010: The definition of GOES infrared lightning initiation interest fields. *J. Appl. Meteor. Climatol.,* **49,** 2527–2543, doi:10.1175/2010JAMC2575.1.

Hosmer, D. W., and S. Lemeshow, 1989: *Applied Logistic Regression.* John Wiley & Sons, 307 pp.

Hu, S., S. Gu, X. Zhuang, and H. Luo, 2007: Automatic identification of storm cells using Doppler radars. *Acta Meteor. Sin.,* **21,** 353–365.

Iskenderian, H., and Coauthors, 2010: Satellite data applications for nowcasting of convective initiation. *14th Conf. on Aviation, Range, and Aerospace Meteorology,* Atlanta, GA, Amer. Meteor. Soc., 5.2. [Available online at https://ams.confex.com/ams/90annual/recordingredirect.cgi/id/11827.]

——, L. Bickmeier, J. Mecikalski, and C. P. Jewett, 2012: Satellite data applications for nowcasting of cloud-to-ground lightning initiation. *18th Conf. on Satellite Meteorology, Oceanography and Climatology/First Joint AMS-Asia Satellite Meteorology Conf.,* New Orleans, LA, Amer. Meteor. Soc., 13C.4. [Available online at https://ams.confex.com/ams/92Annual/recordingredirect.cgi/id/20310.]

Joe, P., M. Falla, P. V. Rijn, L. Stamadianos, T. Falla, D. Magosse, L. Ing, and J. Dobson, 2003: Radar data processing for severe weather in the national radar project of Canada. Preprints, *21st Conf. on Severe Local Storms,* San Antonio, TX, Amer. Meteor. Soc., P4.13. [Available online at http://ams.confex.com/ams/pdfpapers/47421.pdf.]

Lakshmanan, V., T. Smith, K. Hondl, G. J. Stumpf, and A. Witt, 2006: A real-time, three-dimensional, rapid updating, heterogeneous technique radar merger technique for reflectivity, velocity, and derived products. *Wea. Forecasting,* **21,** 802–823, doi:10.1175/WAF942.1.

——, ——, G. J. Stumpf, and K. Hondl, 2007: The Warning Decision Support System–Integrated Information (WDSS-II). *Wea. Forecasting,* **22,** 596–608, doi:10.1175/WAF1009.1.

Lensky, I. M., and D. Rosenfeld, 2006: The time-space exchangeability of satellite retrieved relations between cloud top temperature and particle effective radius. *Atmos. Chem. Phys.,* **6,** 2887–2894, doi:10.5194/acp-6-2887-2006.

Lima, M. A., and J. W. Wilson, 2008: Convective storm initiation in a moist tropical environment. *Mon. Wea. Rev.,* **136,** 1847–1864, doi:10.1175/2007MWR2279.1.

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.,* **30,** 291–303.

Mecikalski, J. R., and K. M. Bedka, 2006: Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime GOES imagery. *Mon. Wea. Rev.,* **134,** 49–78, doi:10.1175/MWR3062.1.

——, ——, S. J. Paech, and L. A. Litten, 2008: A statistical evaluation of GOES cloud-top properties for predicting convective initiation. *Mon. Wea. Rev.,* **136,** 4899–4914, doi:10.1175/2008MWR2352.1.

——, W. M. Mackenzie, M. Koenig, and S. Muller, 2010a: Cloud-top properties of growing cumulus prior to convective initiation as measured by Meteosat Second Generation. Part I: Infrared fields. *J. Appl. Meteor. Climatol.,* **49,** 521–534, doi:10.1175/2009JAMC2344.1.

——, ——, ——, and ——, 2010b: Cloud-top properties of growing cumulus prior to convective initiation as measured by Meteosat Second Generation. Part II: Use of visible reflectance. *J. Appl. Meteor. Climatol.,* **49,** 2544–2558, doi:10.1175/2010JAMC2480.1.

——, P. Minnis, and R. Palikonda, 2013: Use of satellite derived cloud properties to quantify growing cumulus beneath cirrus clouds. *Atmos. Res.,* **120–121,** 192–201, doi:10.1016/j.atmosres.2012.08.017.

Merino, A., L. López, J. L. Sánchez, E. García-Ortega, E. Cattani, and V. Leivizzani, 2014: Daytime identification of summer hailstorm cells from MSG data. *Nat. Hazards Earth Syst. Sci.,* **14,** 1017–1033, doi:10.5194/nhess-14-1017-2014.

Merk, D., and T. Zinner, 2013: Detection of convective initiation using Meteosat SEVIRI: Implementation in and verification with the tracking and nowcasting algorithm Cb-TRAM. *Atmos. Meas. Tech.,* **6,** 1903–1918, doi:10.5194/amt-6-1903-2013.

Minnis, P., and Coauthors, 2011a: CERES Edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data—Part I: Algorithms. *IEEE Trans. Geosci. Remote Sens.,* **49,** 4374–4400, doi:10.1109/TGRS.2011.2144601.

——, and Coauthors, 2011b: CERES Edition-2 cloud property retrievals using TRMM VIRS and *Terra* and *Aqua* MODIS data—Part II: Examples of average results and comparisons with other data. *IEEE Trans. Geosci. Remote Sens.,* **49,** 4401–4430, doi:10.1109/TGRS.2011.2144602.

Mueller, C. K., J. W. Wilson, and N. A. Crook, 1993: The utility of sounding and mesonet data to nowcast thunderstorm initiation. *Wea. Forecasting,* **8,** 132–146, doi:10.1175/1520-0434(1993)008<0132:TUOSAM>2.0.CO;2.

——, T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, and J. Yee, 2003: NCAR Auto-Nowcast System. *Wea. Forecasting,* **18,** 545–561, doi:10.1175/1520-0434(2003)018<0545:NAS>2.0.CO;2.

Nisi, L., P. Ambrosetti, and L. Clementi, 2014: Nowcasting severe convection in the Alpine region: The COALITION approach. *Quart. J. Roy. Meteor. Soc.,* **140,** 1684–1699, doi:10.1002/qj.2249.

Pal, M., 2005: Random forest classifier for remote sensing classification. *Int. J. Remote Sens.,* **26,** 217–222, doi:10.1080/01431160412331269698.

Purdom, J. F. W., 1976: Some uses of high resolution GOES imagery in the mesoscale forecasting of convection and its behavior. *Mon. Wea. Rev.,* **104,** 1474–1483, doi:10.1175/1520-0493(1976)104<1474:SUOHRG>2.0.CO;2.

——, 1982: Subjective interpretations of geostationary satellite data for nowcasting. *Nowcasting,* K. Browning, Ed., Academic Press, 149–166.

Roberts, R. D., and S. Rutledge, 2003: Nowcasting storm initiation and growth using GOES-8 and WSR-88D data. *Wea. Forecasting,* **18,** 562–584, doi:10.1175/1520-0434(2003)018<0562:NSIAGU>2.0.CO;2.

——, A. R. S. Anderson, E. Nelson, B. G. Brown, J. W. Wilson, M. Pocernich, and T. Saxen, 2012: Impacts of forecaster involvement on convective storm initiation and evolution nowcasting. *Wea. Forecasting,* **27,** 1061–1089, doi:10.1175/WAF-D-11-00087.1.

Rosenfeld, D., W. L. Woodley, A. Lerner, G. Kelman, and D. T. Lindsey, 2008: Satellite detection of severe convective storms by their retrieved vertical profiles of cloud particle effective radius and thermodynamic phase. *J. Geophys. Res.,* **113,** D04208, doi:10.1029/2007JD008600.

Ruzanski, E., V. Chandrasekar, and Y. Wang, 2011: The CASA nowcasting system. *J. Atmos. Oceanic Technol.,* **28,** 640–655, doi:10.1175/2011JTECHA1496.1.

Setvák, M., R. M. Rabin, C. A. Doswell, and V. Levizzani, 2003: Satellite observations of convective storm tops in the 1.6, 3.7, and 3.9 μm spectral bands. *Atmos. Res.,* **67–68,** 607–627, doi:10.1016/S0169-8095(03)00076-0.

——, K. Bedka, D. T. Lindsey, A. Sokol, Z. Charvát, J. Šťástka, and P. K. Wang, 2013: A-Train observations of deep convective storm tops. *Atmos. Res.,* **123,** 229–248, doi:10.1016/j.atmosres.2012.06.020.

Sieglaff, J. M., L. M. Cronce, W. F. Feltz, K. M. Bedka, M. J. Pavolonis, and A. K. Heidinger, 2011: Nowcasting convective storm initiation using satellite-based box-averaged cloud-top cooling and cloud-type trends. *J. Appl. Meteor. Climatol.,* **50,** 110–126, doi:10.1175/2010JAMC2496.1.

Siewert, C., and K. Kuhlman, 2011: Hazardous Weather Testbed—Final evaluation. NOAA Rep., 21 pp. [Available online at http://www.goes-r.gov/users/docs/pg-activities/PGFR-HWT-2011-Final.pdf.]

Steiner, M., R. Bateman, D. Megenhardt, Y. Liu, M. Xu, M. Pocernich, and J. Krozel, 2010: Translation of ensemble weather forecasts into probabilistic air traffic capacity impact. *Air Traffic Quart.,* **18,** 229–254.

Strabala, K. I., S. A. Ackerman, and W. P. Menzel, 1994: Cloud properties inferred from 8–12-μm data. *J. Appl. Meteor.,* **33,** 212–229, doi:10.1175/1520-0450(1994)033<0212:CPIFD>2.0.CO;2.

Terborg, A., and C. Gravelle, 2012: GOES-R desk final evaluation. Aviation Weather Testbed 2012 Summer Experiment, 20 pp. [Available online at http://www.goes-r.gov/users/docs/pg-activities/PGFR-AWC-2012-Final.pdf.]

——, K. Calhoun, C. Gravelle, and W. Line, 2013: Hazardous Weather Testbed—GOES-R Proving Ground final evaluation. NOAA Rep., 27 pp. [Available online at http://www.goes-r.gov/users/docs/pg-activities/PGFR-HWT-2013-Final.pdf.]

Topić, G., and T. Šmuc, 2014: Parallel random forest algorithm usage. Accessed 26 June 2014. [Available online at http://code.google.com/p/parf/wiki/Usage.]

Wakimoto, R. M., and H. V. Murphey, 2009: Analysis of a dryline during IHOP: Implications for convection initiation. *Mon. Wea. Rev.,* **137,** 912–936, doi:10.1175/2008MWR2584.1.

Walker, J. R., W. M. MacKenzie, J. R. Mecikalski, and C. P. Jewett, 2012: An enhanced geostationary satellite–based convective initiation algorithm for 0–2-h nowcasting with object tracking. *J. Appl. Meteor. Climatol.,* **51,** 1931–1949, doi:10.1175/JAMC-D-11-0246.1.

Wang, P. K., S.-H. Su, M. Setvák, L.-H. Lin, and R. M. Rabin, 2010: Ship wave signature at the cloud top of deep convective storms. *Atmos. Res.,* **97,** 294–302, doi:10.1016/j.atmosres.2010.03.015.

Weckwerth, T. M., and D. B. Parsons, 2006: A review of convection initiation and motivation for IHOP_2002. *Mon. Wea. Rev.,* **134,** 5–22, doi:10.1175/MWR3067.1.

——, and Coauthors, 2004: An overview of the International H_2O Project (IHOP_2002) and some preliminary highlights. *Bull. Amer. Meteor. Soc.,* **85,** 253–277, doi:10.1175/BAMS-85-2-253.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.

Williams, J. K., 2014: Using random forests to diagnose aviation turbulence. *Mach. Learn.,* **95,** 51–70, doi:10.1007/s10994-013-5346-7.

——, D. Ahijevych, S. Dettling, and M. Steiner, 2008: Combining observations and model data for short-term storm forecasting. *Remote Sensing Applications for Aviation Weather Hazard Detection and Decision Support.,* W. Feltz and J. Murray, Eds., International Society for Optical Engineering (SPIE Proceedings, Vol. 7088), 708805, doi:10.1117/12.795737.

Wilson, J. W., and W. E. Schreiber, 1986: Initiation of convective storms by radar–observed boundary layer convergent lines. *Mon. Wea. Rev.,* **114,** 2516–2536, doi:10.1175/1520-0493(1986)114<2516:IOCSAR>2.0.CO;2.

——, and C. K. Mueller, 1993: Nowcasts of thunderstorm initiation and evolution. *Wea. Forecasting,* **8,** 113–131, doi:10.1175/1520-0434(1993)008<0113:NOTIAE>2.0.CO;2.

——, G. B. Foote, N. A. Crook, J. C. Fankhauser, C. G. Wade, J. D. Tuttle, C. K. Mueller, and S. K. Kruger, 1992: The role of boundary-layer convergence zones and horizontal rolls in the initiation of thunderstorms: A case study. *Mon. Wea. Rev.,* **120,** 1785–1815, doi:10.1175/1520-0493(1992)120<1785:TROBLC>2.0.CO;2.

——, N. A. Crook, C. K. Mueller, J. Sun, and M. Dixon, 1998: Nowcasting thunderstorms: A status report. *Bull. Amer. Meteor. Soc.,* **79,** 2079–2099, doi:10.1175/1520-0477(1998)079<2079:NTASR>2.0.CO;2.

——, Y. Feng, M. Chen, and R. D. Roberts, 2010: Nowcasting challenges during the Beijing Olympics: Successes, failures, and implications for future nowcasting systems. *Wea. Forecasting,* **25,** 1691–1714, doi:10.1175/2010WAF2222417.1.

Wolfson, M. M., and D. A. Clark, 2006: Advanced aviation weather forecasts. *Lincoln Lab. J.,* **16,** 31–58. [Available

online at http://www.ll.mit.edu/publications/journal/pdf/vol16_no1/16_1_3Wolfson.pdf.]

Zhang, J., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) System: Description, results, and future plans. *Bull. Amer. Meteor. Soc.,* **92,** 1321–1338, doi:10.1175/2011BAMS-D-11-00047.1.

Ziegler, C. L., E. N. Rasmussen, M. S. Buban, Y. P. Richardson, L. J. Miller, and R. M. Rabin, 2007: The ''Triple Point'' on 24 May 2002 during IHOP. Part II: Ground–radar and in situ boundary layer analysis of cumulus development and convection initiation. *Mon. Wea. Rev.,* **135,** 2443–2472, doi:10.1175/MWR3411.1.

Zinner, T., H. Mannstein, and A. Tafferner, 2008: Cb-TRAM: Tracking and monitoring severe convection from onset over rapid development to mature phase using multi-channel *Meteosat-8* SEVIRI data. *Meteor. Atmos. Phys.,* **101,** 191–210, doi:10.1007/s00703-008-0290-y.