An Empirical Model for Assessing the Severe Weather Potential of Developing Convection

JOHN L. CINTINEO

Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, Wisconsin

MICHAEL J. PAVOLONIS

NOAA/NESDIS/Center for Satellite Applications and Research/Advanced Satellite Products Team, Madison, Wisconsin

JUSTIN M. SIEGLAFF

Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, Wisconsin

DANIEL T. LINDSEY

NOAA/NESDIS/Center for Satellite Applications and Research/Regional and Mesoscale Meteorology Branch, Fort Collins, Colorado

(Manuscript received 1 October 2013, in final form 21 February 2014)

ABSTRACT

The formation and maintenance of thunderstorms that produce large hail, strong winds, and tornadoes are often difficult to forecast due to their rapid evolution and complex interactions with environmental features that are challenging to observe. Given inherent uncertainties in storm development, it is intuitive to predict severe storms in a probabilistic manner. This paper presents such an approach to forecasting severe thunderstorms and their associated hazards, fusing together data from several sources as input into a statistical model. Mesoscale numerical weather prediction (NWP) models have been developed in part to forecast environments favorable to severe storm development. Geostationary satellites, such as the Geostationary Operational Environmental Satellite (GOES) series, maintain a frequently updating view of growing cumulus clouds over the contiguous United States to provide temporal trends in developing convection to forecasters. The Next Generation Weather Radar (NEXRAD) network delivers repeated scans of hydrometeors inside storms, monitoring the intensification of hydrometeor size and extent, as well as hydrometeor motion. Forecasters utilize NWP models, and GOES and NEXRAD data, at different stages of the forecast of severe storms, and the model described in this paper exploits data from each in an attempt to predict severe hazards in a more accurate and timely manner while providing uncertainty information to the forecaster. A preliminary evaluation of the model demonstrates good skill in the forecast of storms, and also displays the potential to increase lead time on severe hazards, as measured relative to the issuance times of National Weather Service (NWS) severe thunderstorm and tornado warnings and occurrence times of severe events in local storm reports.

1. Introduction

Severe thunderstorms are a relatively common occurrence in the United States, yet determining which thunderstorms will be severe is sometimes complicated. These storms produce hazards such as tornadoes, large

hail, high winds, and often flash flooding, which are a serious threat to life and property (e.g., Lubber 2013). The literature is extensive on the topic of severe thunderstorms, with authors contributing both to our theoretical (e.g., Lemon and Doswell 1979; Bluestein 1993; Markowski and Richardson 2010b) and applied (e.g., Rasmussen 2003; Brooks et al. 2003; Doswell et al. 2005) knowledge of severe convective storms. Severe thunderstorms need several ingredients in order to form (Markowski and Richardson 2010a,b): 1) potential instability in the atmosphere, manifested by generally at

least 100 J kg^{-1} of convective available potential energy (CAPE) and often greater than 1000 J kg^{-1} of CAPE; 2) adequate atmospheric moisture, whereby lifted air becomes saturated and remains saturated when lifted to great heights; 3) a lifting mechanism for parcels of air to reach their level of free convection (LFC) (e.g., surface synoptic fronts, drylines, outflow boundaries, sea-breeze boundaries, or intense surface heating); and 4) vertical wind shear, to organize and sustain the storm's updraft (generally 0–6-km shear ≥18 m s^{-1}, or 35 kt, 1 kt = 0.5144 m s^{-1}). There are complicating factors for forecasters to consider when predicting severe thunderstorms, however. Convective initiation alone is a complex function of vertical motions from the microscale (e.g., thermals) to the synoptic scale, as well as mesoscale temperature and moisture gradients and nonuniformities (Markowski and Richardson 2010a). There are a myriad of other uncertainties, including if the lifting will be adequate to reach the LFC of a storm environment and persistent enough to support a storm; if dry air will entrain into the updraft of the storm, inhibiting vertical development; and if the environmental shear will be too little to organize the storm, or perhaps too great and destroy the updraft of the storm, or be oriented such that a storm may merge with other storms, possibly impacting the severity of the storm (e.g., Markowski and Richardson 2010b, and references therein).

Because of the danger such storms pose, and in response to a great deal of research, observational networks have been installed, in part, to monitor severe weather conditions in situ (e.g., rawinsonde launching sites, surface observation stations) and remotely [e.g., Next Generation Weather Radar (NEXRAD), Crum and Alberty (1993); Geostationary Operational Environmental Satellite (GOES), Menzel and Purdom (1994)]. Numerical weather prediction (NWP) models are also used to forecast severe weather, largely to identify favorable environments for such storms. Rapidly updating mesoscale numerical models, such as the Rapid Refresh (RAP; Benjamin et al. 2006) can assimilate observations and generate near-term (0–18 h) forecasts useful for storm prediction in a timely manner, over a continental domain.

A probabilistic approach to this forecasting problem is preferred because of complexities that can be difficult to observe, features whose automated identification is challenging, or phenomena that are not fully understood. For instance, outflow boundaries from other storms can sometimes aid or detract from storm development; warm-air intrusions aloft that are unobserved in soundings or NWP data can create significant convective inhibition (CIN); high spatial variability of surface moisture may be difficult to resolve, impacting a parcel's buoyancy; or orographic features can greatly enhance or suppress convection. Because of these and other complexities, a probabilistic statistical model was chosen for severe hazard prediction.

Despite these shortcomings, remote sensing of storms has proved fruitful for decades. Adler and Fenn (1979a, b), Adler et al. (1985), and Reynolds (1980) quantitatively used temporal trends in GOES infrared (IR) brightness temperature (BT) to infer the intensity of thunderstorms. Numerous satellite-derived convective initiation and cloud-top cooling methodologies have been developed in the last decade to forecast thunderstorm initiation and intensification (e.g., Roberts and Rutledge 2003; Mueller et al. 2003; Mecikalski and Bedka 2006; Vila et al. 2008; Zinner et al. 2008; Bedka et al. 2010; Sieglaff et al. 2011; Mecikalski et al. 2011). Cintineo et al. (2013, hereafter C13) recently employed temporal trends in GOES-derived cloud properties to infer storm severity from developing convection.

NEXRAD, which provides three-dimensional information of the hydrometeor and kinematic properties (via Doppler velocity) of convective clouds, has proven to be the most important tool for storm surveillance and nowcasting. Polger et al. (1994) quantifies the increase in forecast skill using NEXRAD over legacy weather radar; critical success index scores for severe thunderstorm and tornado warnings almost doubled relative to the pre-NEXRAD era, while tornado warning lead time increased from 4.4 min pre-NEXRAD to 8.8 min post-NEXRAD implementation. Algorithms derived from radar data have also shown some skill for diagnosing severe weather potential (e.g., Greene and Clark 1972; Witt et al. 1998a,b; Stumpf et al. 1998; Wilson et al. 2009).

High spatial resolution NWP models have steadily become more skillful, due to improved observations, data assimilation techniques, physical understanding, and computing power (e.g., Benjamin et al. 2004), such that they are now used frequently in the short term to identify regions with atmospheric environments conducive to severe weather (e.g., Bothwell et al. 2002). This paper synthesizes derived data from rapidly updating remotely sensed observation networks (GOES and NEXRAD) and NWP model output (RAP) to automatically predict the probability that a developing convective cloud or storm will produce severe weather within 120 min. The product described in this paper is meant to link together quantifiable information available to forecasters so they may be able to quickly discern the potential severity of storms of interest.

2. Data and methods

a. Probabilistic model

A flexible and simple statistical model was chosen for this forecast problem: the naïve Bayesian classifier (Kuncheva 2006; Domingos and Pazzani 1997), which assumes that predictors are independent. Despite the clear violation of this "naïve" assumption, the model has demonstrated good performance in meteorological applications (e.g., Kossin and Sitkowski 2009). This model computes a conditional probability of class membership dependent on a set of quantifiable features. For this problem, there are two classes: storms that will produce severe hazards [a tornado, a wind gust of $58\,\mathrm{mi\,h}^{-1}$ ($\sim 26\,\mathrm{m\,s}^{-1}$) or greater, or a hailstone with diameter 25.4 mm or greater] and storms that will not (C_{sev} and C_{ns}, respectively). Using Bayes's theorem, the probability of a storm producing severe hazards given a set of observed predictors $\mathbf{F}$ is defined by

$$P(C_{\mathrm{sev}}\,|\,\mathbf{F}) = \frac{P(C_{\mathrm{sev}})P(\mathbf{F}\,|\,C_{\mathrm{sev}})}{P(\mathbf{F})}, \qquad (1)$$

where the pipe (|) represents a conditional probability and $P\,(C_{\mathrm{sev}})$ is the probability that would be assigned if there were no predictors. We refer to $P\,(C_{\mathrm{sev}})$ as the prior probability, or a priori probability. In this study, the a priori has been transformed into a function of environmental data (see sections 2d and 2f), and serves as a "first guess" of the probability of severe, or $P(C_{\mathrm{sev}}\,|\,\mathbf{F})$. Naturally, $P(C_{\mathrm{ns}}\,|\,\mathbf{F}) = 1 - P(C_{\mathrm{sev}}\,|\,\mathbf{F})$. The assumption of predictor independence allows for practical implementation of Bayes's theorem through reduction of dimensionality (see Kossin and Sitkowski 2009) and Eq. (1) can be rewritten as

$$P(C_{\mathrm{sev}}\,|\,\mathbf{F}) = \frac{P(C_{\mathrm{sev}})\prod_{i=1}^{N} P(F_i\,|\,C_{\mathrm{sev}})}{P(\mathbf{F})}, \qquad (2)$$

with F_i denoting the value of the ith predictor, and N the number of predictors. The Π is the product operator, multiplying the probability of each ith predictor conditional on the storm being a member of the severe class. Thus, only the a priori and conditional probability distribution for each predictor is needed to compute the final probability conditional on the observed predictor set $\mathbf{F}$.

b. Satellite data and algorithms

GOES-12 and GOES-13 imager data (variable temporal resolution, from 5 to 60 min) are used as input into satellite algorithms that compute a cloud mask (Heidinger and Straka 2013), cloud-top phase (Pavolonis 2010a,b), and 11-μm top-of-the-troposphere cloud emissivity $\varepsilon_{\mathrm{tot}}$ (Pavolonis 2010b). The cloud-top phase and $\varepsilon_{\mathrm{tot}}$ fields are remapped using a nearest-neighbor method to a cylindrical equidistant projection with 0.04° latitude × 0.04° longitude spatial resolution, and are interpolated to the radar resolution (see section 2c) when the statistical model is executed. The remapping software package used is Mapx from the National Snow and Ice Data Center (NSIDC 2013). The $\varepsilon_{\mathrm{tot}}$ is the 11-μm emissivity of the cloud when the effective cloud height (Cox 1976) is taken to be the top of the troposphere [see Eq. (2) in Pavolonis (2010b)]. For clouds with a large infrared optical depth (e.g., growing cumulus clouds), $\varepsilon_{\mathrm{tot}}$ is a measure of how close the cloud radiative center is to the tropopause (on a scale from 0.0 to 1.0). Please refer to Pavolonis (2010b) for a complete physical description of $\varepsilon_{\mathrm{tot}}$. This derived product is used to compute vertical growth rates in clouds as opposed to 11-μm BT, since it is less sensitive to the thermodynamic state of the cloud-free atmosphere and maintains the spatial gradients observed in the 11-μm BT field (see discussion in C13 and their Fig. 1).

The GOES cloud-top phase algorithm also uses multispectral GOES imager observations as well as output from a multispectral GOES cloud mask (Heidinger and Straka 2013) to determine the phase of cloudy GOES pixels. The GOES cloud-top phase output is used to determine the rate at which the top of a developing convective cloud glaciates (i.e., converts from liquid water to ice). The temporal rates of change in $\varepsilon_{\mathrm{tot}}$ ($\Delta\varepsilon_{\mathrm{tot}}$) and the ice cloud fraction (Δice, or glaciation rate) are used as predictors in the probabilistic model, and both use the same units (min^{-1}). These two temporal trends depict vertical cloud growth, while only Δice captures microphysical changes at the cloud top. C13 demonstrated that both metrics discriminated between severe and nonsevere convection skillfully. Thus, both are used as predictors in the statistical model. The training for these two predictors is described in C13, and summarized in section 2f.

c. Radar data and algorithms

Level-II NEXRAD-derived products have been provided by the University of Oklahoma Cooperative Institute for Mesoscale Meteorological Studies (OU/CIMMS) and the National Oceanic and Atmospheric Administration National Severe Storms Laboratory (NOAA/ NSSL). NSSL quality controls (QCs) radar reflectivity from throughout the contiguous United States (CONUS), and creates a 3D composite reflectivity field (5-min temporal resolution) using the Warning Decision Support System-Integrated Information (WDSS-II;

Lakshmanan et al. 2006; Lakshmanan et al. 2007a; Lakshmanan et al. 2007b). Radar data are merged using an exponential distance relationship, such that data from closer radars are weighted more than data from more distant radars, at any given point (Lakshmanan et al. 2006). In this fashion, the radar fields are remapped (with WDSS-II) using a cylindrical equidistant projection with $0.01°$ latitude $\times 0.01°$ longitude spatial resolution. The remapped QC merged reflectivity (REF) is then input into the NSSL enhanced hail detection algorithm (along with RAP environmental data) to compute the maximum expected size of hail (MESH; Witt et al. 1998a). MESH is empirically derived from the severe hail index (SHI), which is a reflectivity-weighted vertical integration from the melting level ($0°C$) to storm top that neglects reflectivities less than $40\,dBZ$, in an effort to capture the ice content in a storm. It has been used to diagnose hail potential in severe storms and shown to have some skill (e.g., Witt et al. 1998b; Ortega et al. 2009; Wilson et al. 2009; Cintineo et al. 2012). The instantaneous maximum MESH in a storm (at any given time) is another predictor in the probabilistic model. Sources of error for the MESH product include errors in the heights of temperature levels of RAP analysis data, affecting the vertical integration of SHI, as well as single-radar surveillance of storms, especially at ranges far from the radar, when beam widening may be a problem, resulting in an underestimation of the maximum MESH (Cintineo et al. 2012, Bunkers and Smith 2013).

d. RAP data fields

The RAP was the NWP model of choice in this study due to its superior update frequency (60 min) and spatial resolution (13.5 km) over any operational North American model. The RAP assimilates data from rawinsondes, wind profilers, aircraft, surface stations, ship reports, NEXRAD, satellite radiances, GOES, GPS, and other sources using the community gridpoint statistical interpolation (GSI) system. Two NWP-derived fields that help discern severe storm environments are used to create the a priori probability of severe weather (described in section 2f). They are the effective bulk shear (EBS; Thompson et al. 2007) and the most-unstable CAPE (MUCAPE). Bulk shear (0–6 km) has shown good discrimination between environments that produce severe storms and those that produce ordinary storms (e.g., Thompson et al. 2003). The EBS normalizes the shear values for storms with shallow and deep inflow layers, enabling a better comparison between vertical wind profiles of storms in different environments. The use of MUCAPE as opposed to mixed layer or surface-based CAPE is better for identifying environments supportive of elevated convection that can become severe. When a storm overcomes the environmental CIN, the MUCAPE and EBS of the environment can help determine the intensity of the convective updraft and storm mode (e.g., supercell, multicell, or ordinary storm), to the first order. Because mesoscale NWP forecast output often contains errors in the timing, placement, and intensity of frontal features, the moisture field, and wind speeds (e.g., Coniglio 2012), a conservative approach is used to create blended environmental products. First, a WDSS-II algorithm (gribToNetcdf) is used to remap RAP fields from their native Lambert conformal projection onto a cylindrical equidistant projection. Next, MUCAPE and EBS are calculated for every point in the remapped RAP grid for a given forecast time. For both MUCAPE and EBS, the maximum value at every point in the forecast grid is taken over the span of five forecast hours. The forecast hours used for this operation are the current analysis (t_0), the previous hour analysis (t_{-1}), and the 1-, 2-, and 3-h forecasts (t_{F1}, t_{F2}, and t_{F3}, respectively). This "off centered" approach is used since the RAP output valid at t_0 is usually unavailable to a forecaster until approximately 1 h later. Thus, when new RAP data become available, the derived output valid at t_{F1} is meant to represent the current state of the atmosphere.

After this temporal compositing is complete, a 5×5 gridpoint ($\sim 67.5\,km \times 67.5\,km$) spatial filter is applied to each composite field (with a Gaussian kernel and a smoothing radius equal to three standard deviations), creating a smoothed estimate of the instability and shear in the environment. This process is performed in the hope that poor timing or location of features in the RAP output will not be detrimental to the output probability of severe weather determined by the statistical model. Finally, a second WDSS-II algorithm (w2cropconv, using the default Cressman interpolation) converts the grid to one with $0.04°$ latitude $\times 0.04°$ longitude spatial resolution (still cylindrical equidistant). These RAP-derived fields are interpolated to the resolution of the radar fields when the probabilistic model is run, in similar fashion to the GOES-derived fields. Thus, the GOES-derived, NEXRAD-derived, and RAP-derived fields are all on the same cylindrical equidistant grid ($0.01°$ latitude $\times 0.01°$ longitude) during the execution of the statistical model.

e. Storm identification and tracking

The satellite-object identification and tracking system developed by Sieglaff et al. (2013, hereafter S13) is employed to create satellite objects needed for the automated computation of temporal trends in the previously described metrics ($\Delta\varepsilon_{tot}$ and Δice). A satellite object is defined as a collection of spatially connected pixels that contain a local maximum of ε_{tot}. A unique

identification number (ID) is assigned to the satellite object, which is tracked through space and time in subsequent GOES imager observations. The S13 framework allows for the history of a variety of parameters (satellite, radar, and NWP) to be maintained through a satellite object's lifetime. The reader is encouraged to consult S13 for complete details. Hartung et al. (2013), Sieglaff et al. (2014), and C13 also utilized the satellite-object tracking described in S13. The tracking system utilizes the WDSS-II identification and tracking algorithm (Lakshmanan et al. 2003), applied to the ε_{tot} field. In addition, a postprocessing algorithm, which uses three different spatial scales of a satellite object to correct noncontinuous storm tracks (i.e., owing to misidentifications by the WDSS-II algorithm) is employed.

Radar object identification and tracking is performed on composite reflectivity (REF$_{comp}$) using the WDSS-II algorithm (Lakshmanan et al. 2003) without the postprocessing scheme described in S13. The postprocessing scheme is not needed given the finer spatial and temporal resolutions of NEXRAD data. WDSS-II uses an enhanced watershed algorithm to create objects. In this case, the algorithm searches for local maxima of REF$_{comp} \geq 35\,dBZ$. It then spatially grows the objects in increments of $10\,dBZ$ until a size of at least 20 pixels is reached (approximately $20\,km^2$). For example, if a maximum of $42\,dBZ$ is identified, the algorithm will search for pixels spatially connected to the maximum pixel greater than or equal to $32\,dBZ$. If this yields an object of at least 20 pixels, the object will stop growing. Please see Lakshmanan et al. (2003) for more details. The maximum MESH is extracted at each scan time during the life of a radar object.

The greatest source of error for both $\Delta\varepsilon_{tot}$ and Δice occurs when there is a noncontinuous storm track. The method of S13 works well to correct these tracks, but begins to break down when the amount of time elapsed between GOES scans becomes large (30 min or more). Thus, during times of GOES full-disk scans (30 min between scans, every 3 h), the accuracy in satellite-based tracking and the computed $\Delta\varepsilon_{tot}$ and Δice may decline, especially for spatially small storms (roughly five satellite pixels). Tracking radar objects in REF$_{comp}$ every 5 min does not suffer from this decline in accuracy. The storm ID of a radar object may change if the object spatially changes greatly from one scan to the next (e.g., during splits or mergers). Since only the latest instantaneous MESH is being used as a predictor, however, a change in storm ID is not cause for concern, since no temporal trend is computed using radar data.

It is necessary to identify and track convective cloud/storm objects in GOES and NEXRAD because of the differing sensitivities to hydrometeor properties. More specifically, rapid cloud growth (observable by GOES) often occurs prior to the development of precipitation-sized hydrometeors required for robust NEXRAD-based remote sensing of cloud properties. Information derived from GOES and NEXRAD is combined by associating each parallax-corrected GOES cloud object with one or more spatially overlapping NEXRAD objects (if a NEXRAD object or objects exists). In the case of large cloud shields, when no satellite objects are discernable, the model will only utilize NEXRAD and NWP information to compute a probability of severe weather for identified radar objects.

f. Training dataset

The identification of satellite and radar objects was also necessary to create datasets to train the statistical model. C13 describes the criteria for the training in detail, as well as the spatial and seasonal distributions of selected storms (see S13 for details about the tracking method of satellite objects). To summarize, the severe storm class comprises 130 storms, while the nonsevere storm class has 976 storms. Severe storms were selected if they produced at least one tornado or hail report, since those two hazards are more directly related to updraft strength than wind reports, and wind reports often carry larger ambiguity (Trapp et al. 2006). These training storms may have produced wind reports in addition to hail and tornado reports, however. Storms that only resulted in reports of straight-line wind damage (no severe hail or tornado reports) were completely excluded from the classifier training. Conditional probability distributions for the two satellite-derived predictors, $\Delta\varepsilon_{tot}$ and Δice, are shown in Figs. 1a and 1b, respectively for both severe and nonsevere classes. The lifetime maximum $\Delta\varepsilon_{tot}$ and Δice for each member storm compose the distributions, which have also been smoothed via kernel density estimation (KDE; see C13 for details) to approximate the true distribution of the population. The maximum MESH at every scan time during the lifetime of a storm object was also extracted, creating the conditional probability distributions seen in Fig. 1c (also smoothed via KDE). The conditional probability of severe for each predictor, $P(F_i | C_{sev})$, is drawn from these distributions.

The creation of the a priori probability table required a much larger dataset (since it is two-dimensional). All vertically growing satellite objects ($\Delta\varepsilon_{tot} > 0$) were initially used from 28 days in 2008–12, over the CONUS GOES-East scan sector (see C13 for their seasonal distribution). Of these growing satellite objects, any of which that achieved at least $35\,dBZ$ at $-10°C$ at some point in their lifetime were defined as thunderstorms. This criterion was selected based on simulations of
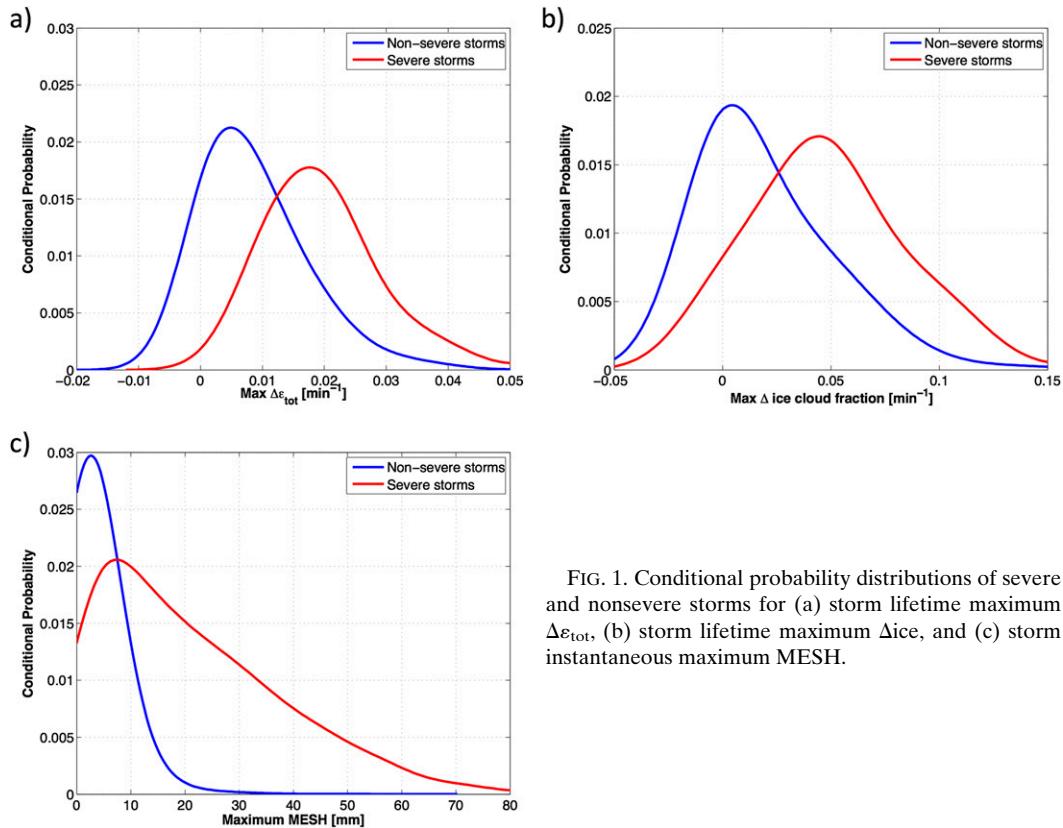
FIG. 1. Conditional probability distributions of severe and nonsevere storms for (a) storm lifetime maximum $\Delta\varepsilon_{tot}$, (b) storm lifetime maximum Δice, and (c) storm instantaneous maximum MESH.

convective initiation by Kain et al. (2013). These criteria resulted in 5635 thunderstorms. Of these thunderstorms, any storm that achieved a MESH of 25.4 mm (1.0 in.) or greater was characterized as severe, resulting in 324 storms. This threshold is arbitrary, but matches a criterion established by the National Weather Service (NWS) for severe thunderstorm warning verification (NWS 2013). All thunderstorms and severe thunderstorms were plotted on a 2D phase space, defined by their lifetime maximum, spatial median EBS, and MUCAPE. Each plot was smoothed with 2D KDE, with a Gaussian kernel and an optimally chosen bandwidth. The smoothed thunderstorm distribution was finally divided by the smoothed severe thunderstorm distribution to create a 2D conditional probability table (Fig. 2). Thus, this table yields the probability that a storm will produce severe weather, given only its environmental EBS and MUCAPE, that is, a storm's a priori probability.

g. Verification dataset

The model (object-tracking and statistical classifier) operated in real time at the University of Wisconsin's Cooperative Institute for Meteorological Satellite Studies

(UW-CIMSS) from April through November 2013, providing numerous days with severe weather over the CONUS. The model was scored against both preliminary severe local storm reports (LSRs) from NOAA's Storm Prediction Center (SPC), and NWS-issued severe
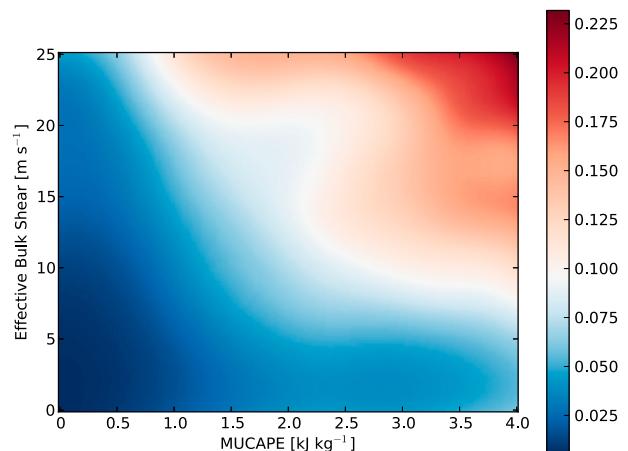

FIG. 2. Table of a priori probability of severe for identified storms, given their median effective bulk shear and median MUCAPE.

thunderstorm and tornado warnings. Since one goal of this model is to increase the lead time of NWS warnings, only the *first* severe thunderstorm or tornado warning issued on a storm was used to validate the model (as well as the first LSR). LSRs have clear biases and inconsistencies (e.g., Trapp et al. 2006, Doswell et al. 2005, Witt et al. 1998b, Kelly et al. 1985, Morgan and Summers 1982), including reduced reporting in sparsely populated regions and after sunset, and incomplete reporting (e.g., Ortega et al. 2009). Nevertheless, LSRs [a tornado, a wind gust of $58\,\mathrm{mi}\,\mathrm{h}^{-1}$ ($\sim 26\,\mathrm{m}\,\mathrm{s}^{-1}$) or greater, or a hailstone with diameter 25.4 mm or greater] are used for verification in addition to warnings since they are widely regarded as the ''ground truth'' (though LSRs certainly do not convey the *entire* truth). While the predictors in this model do not specifically relate to tornadogenesis, tornadoes can often occur in storms with strong rotating updrafts (e.g., supercell thunderstorms), which may be inferred from the statistical model. Thus, tornado LSRs are included as a verification metric; only 2 out of 80 first LSRs for storms were for tornadoes, however, making it difficult to measure meaningful skill of the model for tornadoes alone.

Scoring was performed subjectively because of numerous challenges in automatically validating storms as severe. These challenges include storms that form or travel into or out of the CONUS, storms that merge or split prior to being warned or producing severe weather (thus changing ID), denoting the ''first'' severe weather warning for a storm, and the presence of radar anomalies owing to nonmeteorological scatterers, which occur infrequently. Thus, because of the possibility of a scoring algorithm introducing nontrivial bias, a manual evaluation was performed at this time. Two independent days (i.e., days that were not used to train the classifier) were selected for the evaluation: 18 May and 17 June 2013. These days were selected due to the fairly large geographic coverage of storms (Fig. 3), and since storms initiated predominantly as discrete or small multicellular convection events before growing upscale. The SPC issued a moderate risk for severe storms on 18 May and several slight risk areas on 17 June. Despite only scoring on two days, over 800 storms were evaluated (warned and unwarned), providing an ample and representative *preliminary* validation for the model.

3. Results

a. Real-time operation overview

There are three streams of real-time data essential for the proper functioning of this model: 1) GOES-East imagery, 2) NEXRAD-derived products, and 3) RAP model data. GOES-derived cloud products and RAP-derived fields are processed accordingly as the data arrive (see sections 2b and 2d). Real-time tracking algorithms identify and track storms using both $\varepsilon_{\mathrm{tot}}$ and $\mathrm{REF}_{\mathrm{comp}}$ once those data are computed or available. The probabilistic model is run with the latest satellite and RAP fields once tracking is completed on a $\mathrm{REF}_{\mathrm{comp}}$ field, providing probability of severe output data every 5 min for each identified storm in the eastern two-thirds of the CONUS. The model computes or extracts the predictor information ($\Delta\varepsilon_{\mathrm{tot}}$, Δice, MESH, MUCAPE, and EBS) and associates satellite-tracked objects with radar objects, sharing information between the two. Each satellite object may share information with one or more overlapping radar objects (or none if the storm is not sufficiently mature to be identified on radar). Each radar object may only contain information from at most one overlapping satellite object. A parallax correction is also performed (using a constant cloud-top height of 7 km) to allow satellite objects to better match corresponding radar objects. The $P(C_{\mathrm{sev}}|\mathbf{F})$ is computed based on lookup tables derived from the distributions in Figs. 1 and 2. The predictor values and probability information for each storm are then stored to a file, which is read when a new radar scan is received and objects are identified and tracked.

b. An example case: 3 July 2013

Once the model completes a time step, the severe probability data can be visualized. Figure 4 demonstrates the output from the model for severe storms over western New York on 3 July 2013 between 1900 and 2130 UTC, annotated for clarity. The arrows are colored by the calculated $P(C_{\mathrm{sev}}|\mathbf{F})$ and point to storms identified by the model. On this day, the SPC only forecasted a 5% chance of high wind [within 25 mi ($\sim$40 km) of a given point] for western New York in their 2000 UTC day 1 outlook, and never issued a severe thunderstorm watch for the region. Thus, forecasted severe weather potential was quite marginal for western New York. Nevertheless, storm 1 is identified at 1904 UTC (Fig. 4a, depicted by the box labeled ''1'' and an attached larger arrow) with $P(C_{\mathrm{sev}}|\mathbf{F}) < 0.1$ and contains $\mathrm{REF}_{\mathrm{comp}}$ generally less than 35 dBZ. The parent satellite object resided in a favorable environment, with approximately $3000\,\mathrm{J}\,\mathrm{kg}^{-1}$ of MUCAPE and $20\,\mathrm{m}\,\mathrm{s}^{-1}$ of EBS, and exhibited moderate $\Delta\varepsilon_{\mathrm{tot}}$ and Δice (see Fig. 1). Fifty minutes later (Fig. 4b), the storm exhibited maximum $\mathrm{REF}_{\mathrm{comp}}$ near 50 dBZ, maximum MESH of 2.54 mm, and a probability of severe of 0.1. At 2014 UTC (Fig. 4c), the probability increases again to
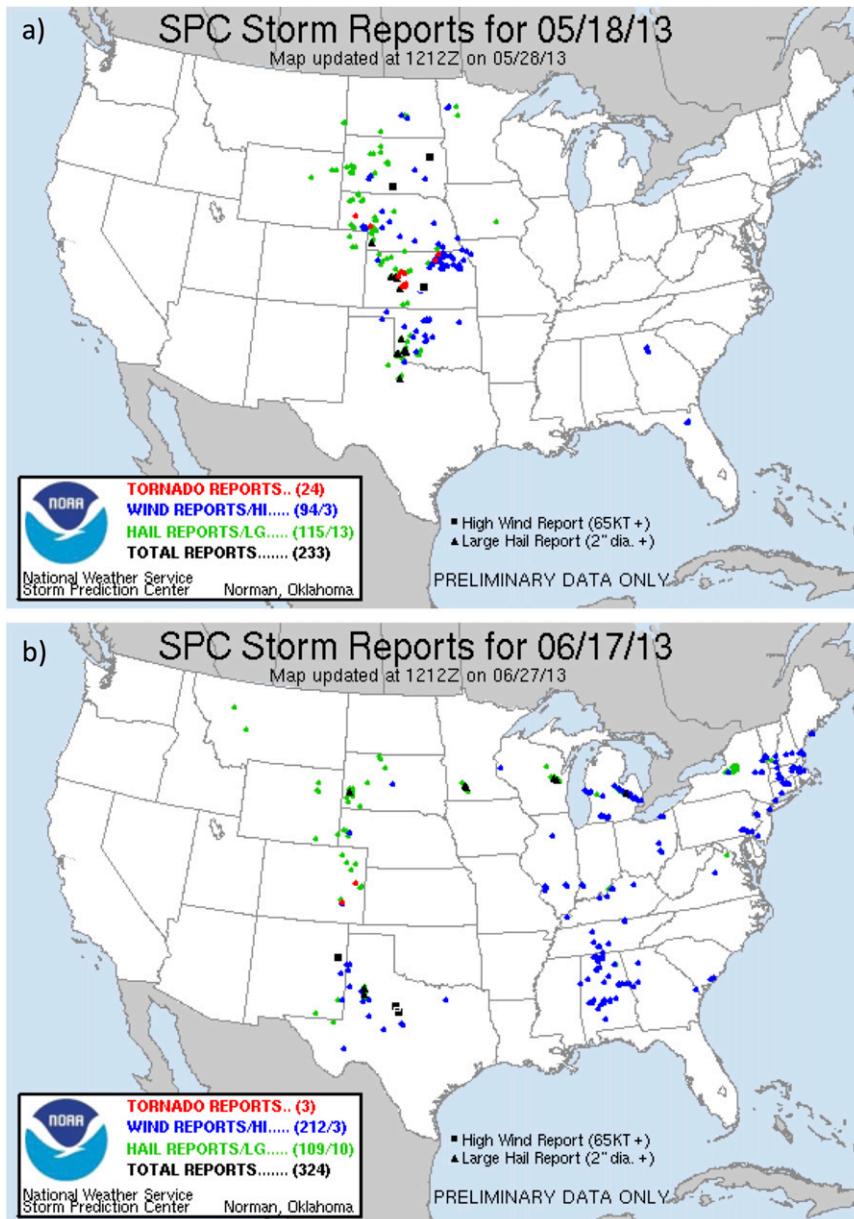
FIG. 3. SPC map of preliminary severe weather reports for (a) 18 May and (b) 17 Jun 2013.

0.46, and again at 2019 UTC (Fig. 4d) to 0.56, largely due to slowly increasing maximum MESH. This storm produced a measured wind gust of 58 mi h^{-1} ($\sim$26 m s^{-1}) at the Rochester International Airport at 2039 UTC [20 min after the first computed $P(C_{\text{sev}} | \mathbf{F}) > 0.5$], and would go on to produce multiple severe wind reports in the region. The NWS office in Buffalo, New York, issued a severe thunderstorm warning at 2052 UTC (approximately Fig. 4e), 33 min after the model first produced a severe probability greater than 0.5 for this storm, and 13 min after the first severe wind report was recorded.

This storm obtained a maximum probability of severe of 0.89 (Fig. 4f). The reader should note that even though the model was not trained with storms producing only severe wind, it may still be useful for certain storms that do produce severe wind, since severe wind is often associated with storms that may produce severe hail and tornadoes (e.g., storms with strong, persistent updrafts).

Storm 2 (denoted in Figs. 4c–f) measures $P(C_{\text{sev}} | \mathbf{F}) = 0.56$ at 2124 UTC. The storm was warned 6 min later at 2130 UTC (not shown) by the NWS, and would
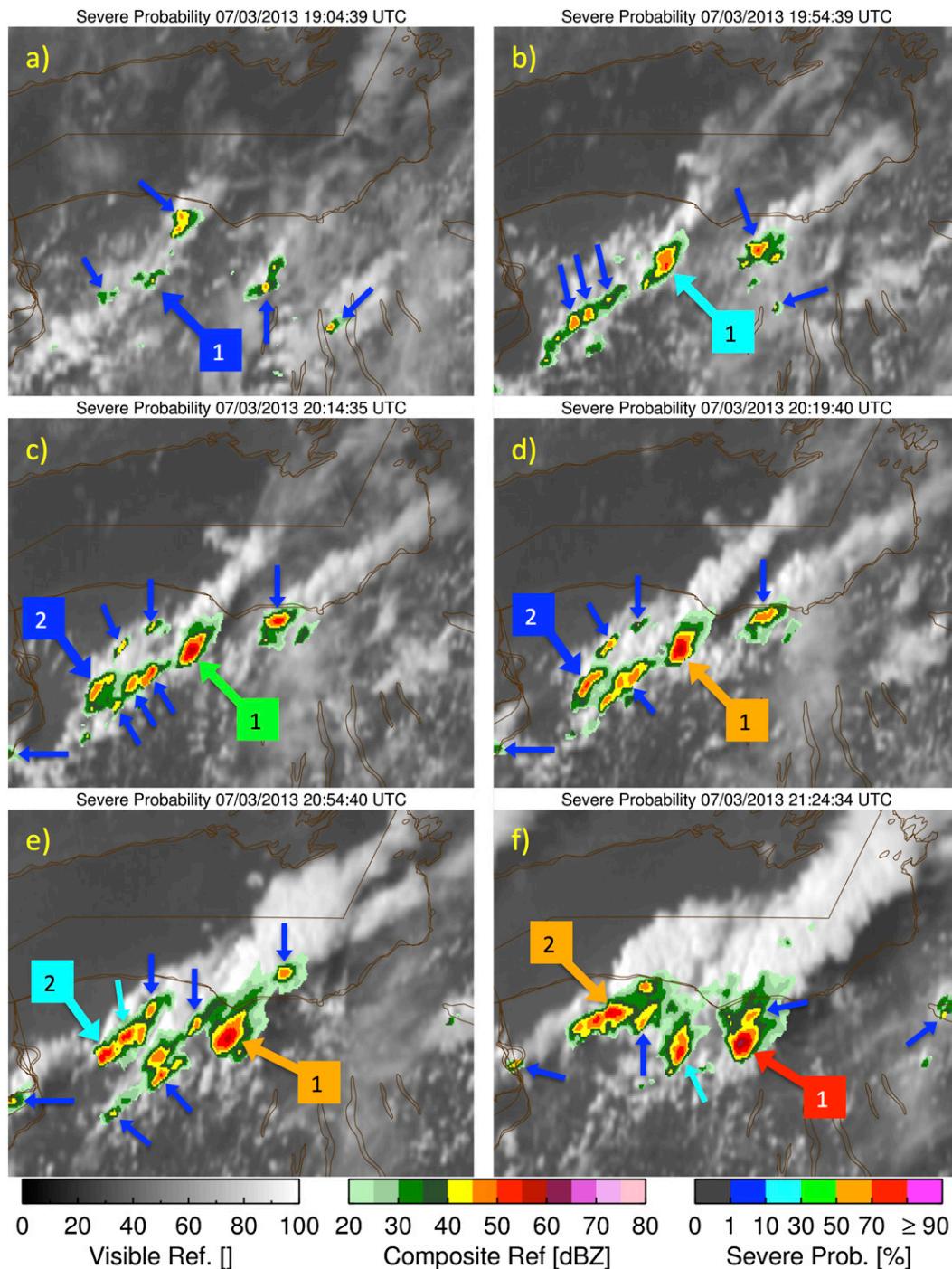
FIG. 4. An example from 3 Jul 2013 of GOES-East visible reflectance overlaid with shaded REF$_{comp}$. The arrows point to storms identified by the model. Arrows are colored by the computed probability of severe. If an arrow does not point to a REF$_{comp}$ maximum, then that REF$_{comp}$ maximum was not identified as a storm by the model at this time. Storms 1 and 2 are indicated by larger arrows with attached labeled boxes, and are referenced in the text.

go on to produce 25.4-mm hail at 2143 UTC. The maximum severe probability exhibited by this storm was 0.93 at 2134 UTC. The first severe thunderstorm warning issued for storm 2 did not measure as much lead time as the warning issued on storm 1, since the observed $\Delta\varepsilon_{tot}$ and Δice were smaller for storm 2, but storm 2 still exhibited $P(C_{sev} \,|\, \mathbf{F}) > 0.5$ before the warning was issued.

c. Model performance

The model was scored against initial LSRs (2 tornado, 52 hail, 26 wind) and initial severe thunderstorm and tornado warnings issued by the NWS. Additionally, the skill of the model was compared to the skill of NWS warnings. Probabilistic forecasts from the model were divided into six bins: 1) $0.01 \leq x < 0.1$, 2) $0.1 \leq x < 0.3$, 3) $0.3 \leq x < 0.5$, 4) $0.5 \leq x < 0.7$, 5) $0.7 \leq x < 0.9$, and 6) $x \geq 0.9$. The midpoint probability for each bin is used to represent the entire bin in several figures. To evaluate the model against LSRs and NWS warnings, a trained human analyst subjectively examined geospatial images of the output probability (at 5-min resolution), geospatial images of warning polygons (at 1-min resolution), the SPC's preliminary severe LSR log, and Iowa State University's (ISU) valid time extent code (VTEC) application, which aided in matching NWS warnings to storms and often provides verification LSRs as well (ISU 2013). VTEC is an application that allows the user to navigate archived NWS warnings, showing the full warning text and maps with the warning polygon and location of LSRs (if they occurred). The time a forecast probability first exceeded a threshold was recorded for each storm (using the six aforementioned bins), as well as the time when the first warning was issued on a storm (if it was ever warned), and the time of the first severe hail, wind, or tornado report associated with the storm (if a report ever occurred). Storms were evaluated in the eastern two-thirds of the CONUS, approximately east of the 105°W meridian. An attempt was made to evaluate every storm that occurred, but some storms with low lifetime maximum probability of severe (less than 0.1) may have been excluded, simply because of their very high frequency. Diligent effort was given to storms exceeding a probability of 0.1 at some point in their lifetime (and storms that produced LSRs or became warned).

Traditional metrics of scoring were computed, including the probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), and Peirce skill score (PSS; Peirce 1884). These metrics were plotted as a function of probability threshold (Fig. 5). A storm has a "hit" at a certain probability threshold if it attained or exceeded the probability prior to the occurrence of the truth metric (an LSR or NWS warning). A "miss" denotes occurrence of the truth metric prior to when a probability threshold was attained for a storm (if it ever reached the threshold). A "false alarm" occurs when the probability of a storm meets a certain threshold, but the truth metric is never observed. A "correct null" happens when no truth metric is observed, and the probability threshold for a storm is never met. Thus, a single storm may have a combination of hits and misses
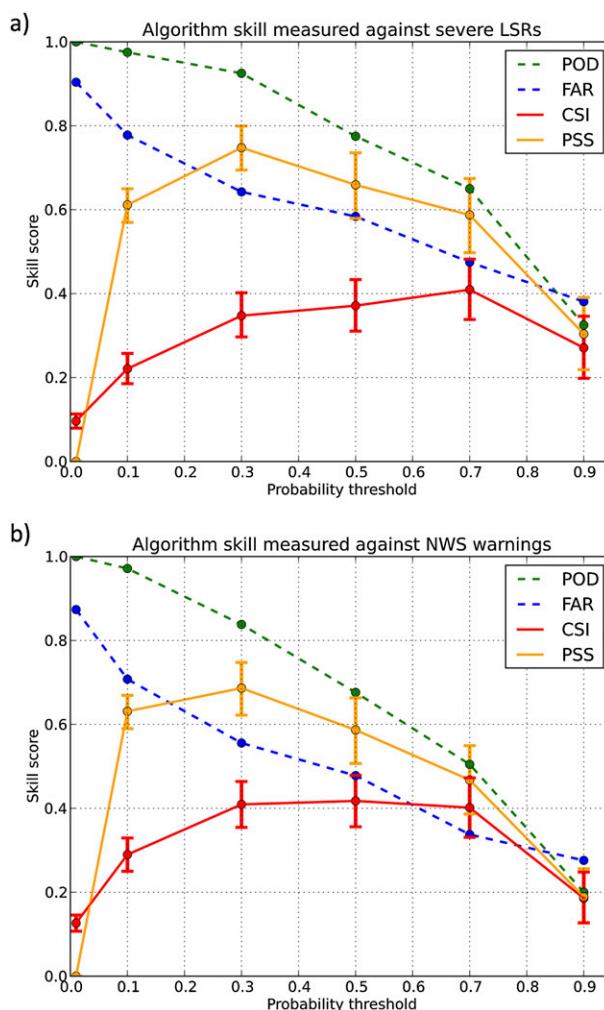

FIG. 5. POD, FAR, CSI, and PSS as functions of forecast probability threshold (from the statistical model) scored against (a) LSRs and (b) NWS warnings. Error bars represent 95% confidence intervals for the CSI and PSS.

or false alarms and correct nulls, corresponding to different probability thresholds. The skill metrics were formulated following Wilks (2006). The CSI is useful in this evaluation since the occurrence of severe weather is much less frequent than its nonoccurrence, while the PSS measures skill against the sample climatology of occurrence. Figures 5a and 5b demonstrate the model skill measured to severe LSRs and NWS warnings, respectively. The vertical bars represent the 95% confidence intervals for the CSI and PSS at each probability threshold, which were computed empirically, from 5000 samples via bootstrapping [i.e., resampling with replacement; see Wilks (2006)]. The model exhibited maximum CSI > 0.4 against both criteria (from 0.3 to 0.7 for NWS warnings, and at 0.7 for LSRs), while the POD and FAR were lower versus NWS warnings. LSRs

typically occur after an NWS warning has been issued, contributing to the higher POD for the LSR truth metric. The higher FAR for the LSR criterion may be partially due to some biases in reporting (e.g., a dearth of reports in sparsely populated regions).

The lead time for each storm hit was measured for each truth metric. Distributions of lead time are shown in Fig. 6. The notches in each boxplot denote the 95% confidence interval for the median lead time (indicated by the horizontal red line), computed empirically via bootstrapping the sample 5000 times. The top and bottom edges of the boxes encompass the middle 50% of lead times, or the interquartile range (IQR). Plus signs indicate outlier lead times beyond 1.5 times the IQR. Lead times are substantial when computed relative to the LSR time (Fig. 6a), as evidenced by median lead times of 25 and 15 min, respectively, for the two highest probability bins. Even when lead time was measured relative to NWS warning issuance (Fig. 6b), 50% of the storms had at least 12.5 min of lead time in the 0.6 bin. Median lead times for the 0.8 and 0.9 bins were approximately 10 min. Thus, this model may be able to help forecasters extend some severe storm warnings from one or two radar volume scans (approximately 5 min each) to perhaps 30 min or more, even with the model's current latency of 5 min (using modest hardware).

A reliability diagram (also called an attributes diagram) can also elucidate the skill of the model. It expresses the conditional probability of event occurrence, as a function of probability forecast. Unique probabilistic forecasts (not thresholded) were used for each probability bin. Figure 7 illustrates the reliability of the model to both LSR (Fig. 7a) and NWS warning occurrence (Fig. 7b). The inset in Fig. 7a represents the forecast frequency for each probability bin. A perfectly reliable model corresponds to the blue 1:1 line. Points that fall on or near this line signify that the probability of the event occurring, given a forecast probability, is approximately equal to the probability of the forecast. The horizontal no-resolution line represents the value of the sample climatology (~0.09 for LSRs and ~0.12 for warnings). A "no skill" line is drawn halfway between the no-resolution line and the perfect reliability line, and forms two regions: one area closer to the perfect reliability line and one area closer to the no-resolution or climatology line. Points in the region closer to the perfect reliability line than to the no-resolution line add to the model skill, which is equivalent to the Brier skill score (Wilks 2006). Thus, the forecasts in a given probability bin contribute positively to model skill where any red point resides in the green-shaded region. The skill is measured relative to the climatology of the sample (i.e., the average forecast probability). The model is well
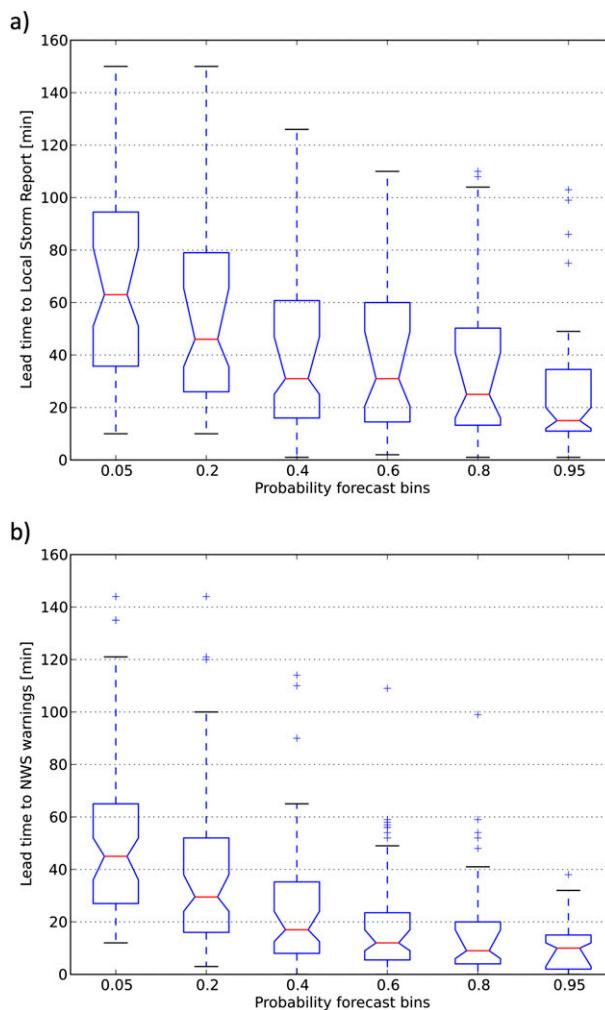

FIG. 6. Distributions of lead time for model hits measured to (a) LSRs and (b) NWS warnings, given the forecast probability. Horizontal red lines are median lead times, boxes represent the IQR of lead-time values, the bounds of notches are 95% confidence intervals for the median lead time, and plus signs represent lead-time outliers.

calibrated when measured against NWS warnings, since the red line corresponds closely to the 1:1 line, with each probability bin contributing to the forecast skill. The model is less reliable when the truth metric is LSR occurrence. In fact, an overforecasting bias is evident at probability bins 0.4 and greater. The sources of error discussed in sections 2c and 2e may contribute to this, as well as potential LSR shortcomings, similar to the discussion regarding Fig. 5. Despite this, each bin above 0.05 still contributes positively to the forecast skill.

Finally, model skill was compared against the skill of NWS warnings, measured to LSR occurrence. The most skillful (as measured to LSRs) probabilistic threshold is used for the model (0.7). Figure 8 shows the POD, FAR,
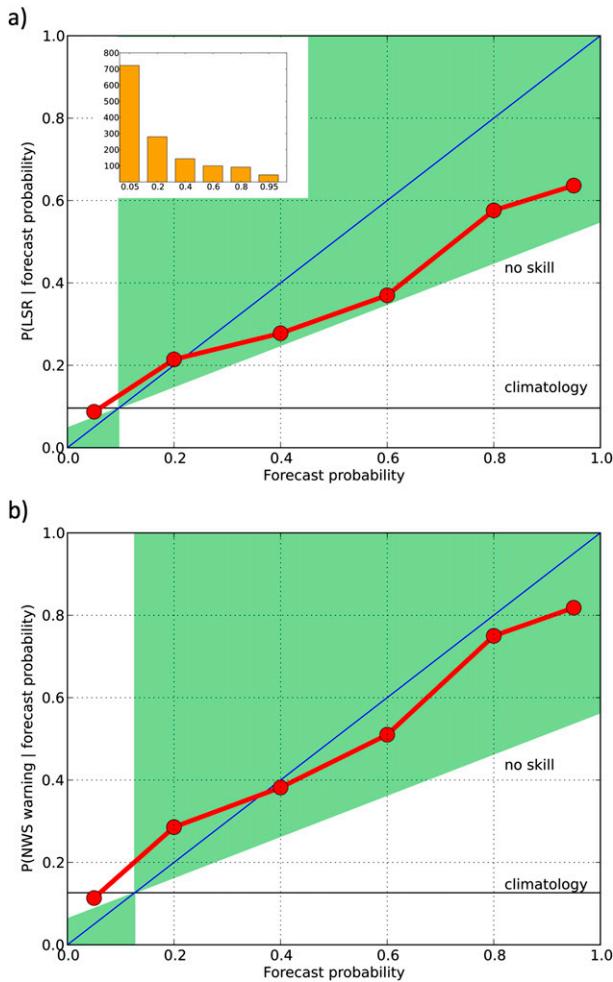
a)

b)

FIG. 7. Reliability diagrams for model skill measured against (a) LSRs and (b) NWS warnings. The inset figure in (a) shows counts of unique forecast probabilities. The blue line is the 1:1 line and represents perfect reliability. The no-skill line delineates regions where the model contributes positively to skill (green region) and negatively to skill (as measured against the sample "climatology"; the horizontal line). A red point that resides in the green-shaded region demonstrates a positive skill contribution for the associated forecast probability bin.

CSI, and median lead time for the statistical model and the NWS-issued severe weather warnings (the *initial* warnings for storms). The NWS has overall higher skill than the statistical model (0.51 versus 0.41 CSI), but shorter median lead time than the model (15 versus 25 min). This lead-time statistic is heavily weighted toward large hail and high wind reports. Regardless, this model may provide forecasters more confidence to issue some severe weather warnings sooner than they otherwise would have, which may be especially important for users who require more lead time (e.g., hospitals, stadiums, concert venues).
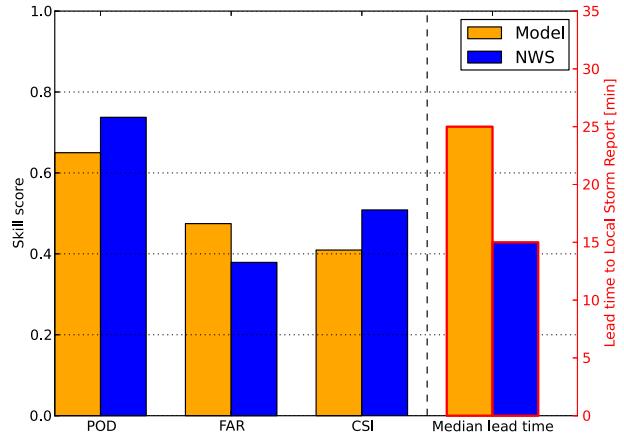

FIG. 8. The POD, FAR, and CSI for the model presented in this paper (orange) and the NWS warnings (blue), measured to severe LSRs. The most skillful probabilistic threshold was used for the statistical model. The bars on the right represent the median lead time to severe weather occurrence for both the statistical model and the NWS warnings (scale on right side).

4. Conclusions and future work

This paper describes the method and skill of the real-time synthesis of data from GOES, NEXRAD, and the RAP model to statistically forecast the probability that a developing storm will produce severe weather. This research leverages refined data and concepts from recent work regarding cloud properties (Pavolonis 2010a,b), convective cloud growth (C13, Sieglaff et al. 2011), multiradar multisensor products (e.g., Ortega et al. 2006; Lakshmanan et al. 2007a), rapidly updating numerical models (Benjamin et al. 2006), and object tracking (S13, Lakshmanan et al. 2003). Object identification and tracking from both satellite and radar imagery enables the sharing of storm trends and information for data derived from each sensor, and facilitates the association of features on an object level instead of a pixel level, thus reducing gigabytes of input to kilobytes of output.

The simple naïve Bayesian model exhibited good performance using traditional skill metrics during a preliminary evaluation of over 800 storms for two days, with maximum CSI > 0.4 and maximum PSS > 0.65 measured to both LSRs and NWS-issued severe weather warnings. The model demonstrated sizeable lead times relative to initial LSRs and NWS warnings for those storms that were correctly identified, evidenced by a median lead time of 12.5 min for NWS warnings and a median lead time of 31 min for LSRs (for the 0.6 probability bin). The two days evaluated contained a variety of storm types in geographically diverse regions; skill was measured relative to initial NWS warnings and LSRs when most storms were single-cell or smaller multicell entities, however. Thus, the model's

skill should be investigated further based on varying storm type and morphology (e.g., supercell, ordinary storm, multicell, convective line), which may indeed have an impact on forecast skill (Guillot et al. 2008). It is important to note that mesoscale convective systems (MCSs) that tend to produce predominantly high winds may not be handled well by this model, since there are often processes driving these systems (e.g., large, deep cold pools and rear-inflow jets) that are captured by few if any of the predictors of this model. MCSs generally form from upscale growth of single-cell or multicellular convection, or when the deep-layer wind shear has a large component parallel to the initiating boundary of the convection. The model in this paper has been mainly designed to predict the severity of storms where updraft strength can be inferred.

Several measures have been identified to decrease false alarms and improve the detection of severe weather. The introduction of several additional NWP-derived products could help sharpen the a priori probability, including products that help capture the melting of hailstones (e.g., height of the 0°C isotherm), and tornadogenesis (e.g., lifted condensation level, low-level shear). Percentile values of predictors (such as the 75th or 90th percentiles) may be less noisy than the maximum value of predictors, enhancing the discrimination between severe and nonsevere storms. GOES temporal resolution also varies from 5 to 30 min, greatly influencing the $\Delta\varepsilon_{tot}$ and Δice for storms. A more precise training method based on temporal resolution could help discern severe from nonsevere convection better.

One data source that is missing from this model is lightning data. Total lightning flash rates, which have been shown to be skillful when predicting severe weather (Schultz et al. 2011), will be available on the Geostationary Lightning Mapper instrument (GLM; Goodman et al. 2006) aboard the next-generation GOES (GOES-R; Schmit et al. 2005), which is anticipated to be operational in 2017. In the interim, cloud-to-ground (CG) lightning flash rates or GLM-proxy total lightning flash rates (from ground-based lightning mapping arrays) may be investigated as potential predictors, but actual GLM data will likely be utilized in this model once GOES-R becomes operational. This real-time model also readily lends itself to creating statistical models for specific severe hazards, by adding or subtracting certain predictors for each hazard. Models for severe hail, severe wind (from single-cell convection and MCSs), tornado, and flash-flooding probability may also be pursued.

The end goal of this research is to provide skillful guidance to forecasters to discriminate between severe and nonsevere convection and increase warning lead time for severe hazards. This fused approach may help forecasters make better warning decisions by distilling the pertinent information from an increasing amount of data at the forecaster's disposal into a focused probabilistic product. With the advent of more numerous and more capable satellites, radar networks, numerical models, and in situ measurements, products that incorporate some or all of these data sources may provide meteorologists a better answer to certain forecast problems than any single product alone.

REFERENCES

Adler, R. F., and D. D. Fenn, 1979a: Thunderstorm intensity as determined from satellite data. *J. Appl. Meteor.,* **18,** 502–517, doi:10.1175/1520-0450(1979)018<0502:TIADFS>2.0.CO;2.

——, and ——, 1979b: Thunderstorm vertical velocities estimated from satellite data. *J. Atmos. Sci.,* **36,** 1747–1754, doi:10.1175/1520-0469(1979)036<1747:TVVEFS>2.0.CO;2.

——, M. J. Markus, and D. D. Fenn, 1985: Detection of severe Midwest thunderstorms using geosynchronous satellite data. *Mon. Wea. Rev.,* **113,** 769–781, doi:10.1175/1520-0493(1985)113<0769:DOSMTU>2.0.CO;2.

Bedka, K., 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, 2004: An hourly assimilation–forecast cycle: The RUC. *Mon. Wea. Rev.,* **132,** 495–518, doi:10.1175/1520-0493(2004)132<0495:AHACTR>2.0.CO;2.

——, and Coauthors, 2006: From the 13-km RUC to the Rapid Refresh. Preprints, *12th Conf. on Aviation, Range, and Aerospace Meteorology,* Atlanta, GA, Amer. Meteor. Soc., 9.1. [Available online at https://ams.confex.com/ams/pdfpapers/104851.pdf.]

Bluestein, H. B., 1993: Precipitation systems in the midlatitudes. *Observations and Theory of Weather Systems,* H. Bluestein, Ed., Vol II, *Synoptic–Dynamic Meteorology in Midlatitudes,* Oxford University Press, 426–568.

Bothwell, P. D., J. A. Hart, and R. L. Thompson, 2002: An integrated three-dimensional objective analysis scheme in use at the Storm Prediction Center. Preprints, *21st Conf. on Severe Local Storms,* San Antonio, TX, Amer. Meteor. Soc., JP3.1. [Available online at https://ams.confex.com/ams/SLS_WAF_NWP/techprogram/paper_47482.htm.]

Brooks, H. E., C. A. Doswell III, 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.

Bunkers, M. J., and P. L. Smith, 2013: Comments on ''An objective high-resolution hail climatology of the contiguous United States.'' *Wea. Forecasting,* **28,** 915–917, doi:10.1175/WAF-D-13-00020.1.

Cintineo, J. L., T. M. Smith, V. Lakshmanan, H. E. Brooks, and K. L. Ortega, 2012: An objective high-resolution hail climatology of the contiguous United States. *Wea. Forecasting,* **27,** 1235–1248, doi:10.1175/WAF-D-11-00151.1.

——, M. J. Pavolonis, J. M. Sieglaff, and A. K. Heidinger, 2013: Evolution of severe and nonsevere convection inferred from GOES-derived cloud properties. *J. Appl. Meteor. Climatol.,* **52,** 2009–2023, doi:10.1175/JAMC-D-12-0330.1.

Coniglio, M. C., 2012: Verification of RUC 0–1-h forecasts and SPC mesoscale analyses using VORTEX2 soundings. *Wea. Forecasting,* **27,** 667–683, doi:10.1175/WAF-D-11-00096.1.

Cox, S. K., 1976: Observations of cloud infrared effective emissivity. *J. Atmos. Sci.,* **33,** 287–289, doi:10.1175/1520-0469(1976)033<0287:OOCIEE>2.0.CO;2.

Crum, T. D., and R. L. Alberty, 1993: The WSR-88D and the WSR-88D Operational Support Facility. *Bull. Amer. Meteor. Soc.,* **74,** 1669–1687, doi:10.1175/1520-0477(1993)074<1669:TWATWO>2.0.CO;2.

Domingos, P., and M. Pazzani, 1997: Beyond independence: Conditions for the optimality of the simple Bayesian classifier. *Mach. Learn.,* **29,** 103–130, doi:10.1023/A:1007413511361.

Doswell, C. A., III, H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Wea. Forecasting,* **20,** 577–595, doi:10.1175/WAF866.1.

Goodman, S. J., R. Blakeslee, D. Boccippio, H. Christian, W. Koshak, and W. A. Petersen, 2006: GOES-R Lightning Mapper (GLM) research and applications risk reduction. Preprints, *Second Symp.: Toward a Global Earth Observation System of Systems—Future National Operational Environmental Satellite Systems,* New Orleans, LA, Amer. Meteor. Soc., P2.2. [Available online at https://ams.confex.com/ams/Annual2006/techprogram/paper_101645.htm.]

Greene, D. R., and R. A. Clark, 1972: Vertically integrated liquid water—New analysis tool. *Mon. Wea. Rev.,* **100,** 548–552, doi:10.1175/1520-0493(1972)100<0548:VILWNA>2.3.CO;2.

Guillot, E., T. Smith, V. Lakshmanan, K. Elmore, D. Burgess, and G. Stumpf, 2008: Tornado and severe thunderstorm warning forecast skill and its relationship to storm type. Preprints, *24th Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology,* New Orleans, LA, Amer. Meteor. Soc., 4A.3. [Available online at https://ams.confex.com/ams/pdfpapers/132244.pdf.]

Hartung, D. C., J. M. Sieglaff, L. M. Cronce, and W. F. Feltz, 2013: An intercomparison of UW cloud-top cooling rates with WSR-88D radar data. *Wea. Forecasting,* **28,** 463–480, doi:10.1175/WAF-D-12-00021.1.

Heidinger, A. K., and W. C. Straka III, 2013: ABI cloud mask. Version 3.0, NOAA/NESDIS/Center for Satellite Applications and Research (STAR) ATBD, 106 pp. [Available online at http://www.star.nesdis.noaa.gov/goesr/docs/ATBD/Cloud_Mask.pdf.]

ISU, cited 2013: IEM Valid Time Extent Code (VTEC) application. [Available online at mesonet.agron.iastate.edu/vtec/.]

Kain, J., and Coauthors, 2013: A feasibility study for probabilistic convection initiation forecasts based on explicit numerical guidance. *Bull. Amer. Meteor. Soc.,* **94,** 1213–1225, doi:10.1175/BAMS-D-11-00264.1.

Kelly, D. L., J. T. Schaefer, and C. A. Doswell III, 1985: Climatology of nontornadic severe thunderstorm events in the United States. *Mon. Wea. Rev.,* **113,** 1997–2014, doi:10.1175/1520-0493(1985)113<1997:CONSTE>2.0.CO;2.

Kossin, J. P., and M. Sitkowski, 2009: An objective model for identifying secondary eyewall formation in hurricanes. *Mon. Wea. Rev.,* **137,** 876–892, doi:10.1175/2008MWR2701.1.

Kuncheva, L. I., 2006: On the optimality of naïve Bayes with dependent binary features. *Pattern Recognit. Lett.,* **27,** 830–837, doi:10.1016/j.patrec.2005.12.001.

Lakshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *Atmos. Res.,* **67–68,** 367–380, doi:10.1016/S0169-8095(03)00068-1.

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

——, A. Fritz, T. Smith, K. Hondl, and G. Stumpf, 2007a: An automated technique to quality control radar reflectivity data. *J. Appl. Meteor. Climatol.,* **46,** 288–305, doi:10.1175/JAM2460.1.

——, T. Smith, G. Stumpf, and K. Hondl, 2007b: The Warning Decision Support System-Integrated Information. *Wea. Forecasting,* **22,** 596–612, doi:10.1175/WAF1009.1.

Lemon, L. R., and C. A. Doswell, 1979: Severe thunderstorm evolution and mesocyclone structure as related to tornadogenesis. *Mon. Wea. Rev.,* **107,** 1184–1197, doi:10.1175/1520-0493(1979)107<1184:STEAMS>2.0.CO;2.

Lubber, M., cited 2013: Wild weather a new normal and insurance companies must act. *Forbes.* [Available online at http://www.forbes.com/sites/mindylubber/2012/08/30/wild-weather-a-new-normal-and-insurance-companies-must-act/.]

Markowski, P., and Y. Richardson, 2010a: Convection initiation. *Mesoscale Meteorology in Midlatitudes,* Wiley-Blackwell, 185–199.

——, and ——, 2010b: Organization of isolated convection. *Mesoscale Meteorology in Midlatitudes,* P. Markowski and Y. Richardson, Eds., Series on Advancing Weather and Climate Science, Wiley-Blackwell, 201–244.

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.

——, P. D. Watts, and M. Koenig, 2011: Use of Meteosat Second Generation optimal cloud analysis fields for understanding physical attributes of growing cumulus clouds. *Atmos. Res.,* **102,** 175–190, doi:10.1016/j.atmosres.2011.06.023.

Menzel, W. P., and J. F. W. Purdom, 1994: Introducing GOES-I: The first of a new generation of Geostationary Operational Environmental Satellites. *Bull. Amer. Meteor. Soc.,* **75,** 757–781, doi:10.1175/1520-0477(1994)075<0757:IGITFO>2.0.CO;2.

Morgan, G. M., Jr., and P. W. Summers, 1982: Hailfall and hailstorm characteristics. *Thunderstorm Morphology and Dynamics,* E. Kessler, Ed., Vol. 2, *Thunderstorms: A Social, Scientific, and Technological Documentary,* U.S. Government Printing Office, 363–408.

Mueller, C., 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.

NSIDC, cited 2013: Mapx: Map Transformations Library. [Available online at http://geospatialmethods.org/mapx/.]

NWS, cited 2013: WFO severe weather products specification. National Weather Service Instruction 10-511. [Available online at http://www.weather.gov/directives/.]

Ortega, K. L., T. M. Smith, and G. J. Stumpf, 2006: Verification of multi-sensor, multi-radar hail diagnosis techniques. Preprints, *Symp. on the Challenges of Severe Convective Storms,* Atlanta, GA, Amer. Meteor. Soc., P1.1. [Available online at https://ams.confex.com/ams/pdfpapers/104885.pdf.]

——, ——, K. L. Manross, A. G. Kolodziej, K. A. Scharfenberg, A. Witt, and J. J. Gourley, 2009: The Severe Hazards Analysis and Verification Experiment. *Bull. Amer. Meteor. Soc.,* **90,** 1519–1530, doi:10.1175/2009BAMS2815.1.

Pavolonis, M. J., 2010a: GOES-R Advanced Baseline Imager (ABI) algorithm theoretical basis document for cloud type and cloud phase. Version 2, NOAA/NESDIS/Center for Satellite Applications and Research (STAR), 96 pp. [Available online at http://www.goes-r.gov/products/ATBDs/baseline/Cloud_CldType_v2.0_no_color.pdf.]

——, 2010b: Advances in extracting cloud composition information from spaceborne infrared radiances—A robust alternative to brightness temperatures. Part I: Theory. *J. Appl. Meteor. Climatol.,* **49,** 1992–2012, doi:10.1175/2010JAMC2433.1.

Peirce, C. S., 1884: The numerical measure of the success of predictions. *Science,* **4,** 453–454, doi:10.1126/science.ns-4.93.453-a.

Polger, P. D., B. S. Goldsmith, R. C. Przywarty, and J. R. Bocchieri, 1994: National Weather Service warning performance based on the WSR-88D. *Bull. Amer. Meteor. Soc.,* **75,** 203–214, doi:10.1175/1520-0477(1994)075<0203:NWSWPB>2.0.CO;2.

Rasmussen, E. N., 2003: Refined supercell and tornado forecast parameters. *Wea. Forecasting,* **18,** 530–535, doi:10.1175/1520-0434(2003)18<530:RSATFP>2.0.CO;2.

Reynolds, D. W., 1980: Observations of damaging hailstorms from geosynchronous satellite digital data. *Mon. Wea. Rev.,* **108,** 337–348, doi:10.1175/1520-0493(1980)108<0337:OODHFG>2.0.CO;2.

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.

Schmit, T. J., M. M. Gunshor, W. P. Menzel, J. J. Gurka, J. Li, and A. S. Bachmeier, 2005: Introducing the next-generation Advanced Baseline Imager on GOES-R. *Bull. Amer. Meteor. Soc.,* **86,** 1079–1096, doi:10.1175/BAMS-86-8-1079.

Schultz, C. J., W. A. Petersen, and L. D. Carey, 2011: Lightning and severe weather: A comparison between total and cloud-to-ground lightning trends. *Wea. Forecasting,* **26,** 744–755, doi:10.1175/WAF-D-10-05026.1.

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.

——, D. C. Hartung, W. F. Feltz, L. M. Cronce, and V. Lakshmanan, 2013: Development and application of a satellite-based convective cloud object-tracking methodology: A multipurpose data fusion tool. *J. Atmos. Oceanic Technol.,* **30,** 510–525, doi:10.1175/JTECH-D-12-00114.1.

——, L. M. Cronce, and W. F. Feltz, 2014: Improving satellite-based convective cloud growth monitoring with visible optical depth retrievals. *J. Appl. Meteor. Climatol.,* **53,** 506–520, doi:10.1175/JAMC-D-13-0139.1.

Stumpf, G. J., A. Witt, E. D. Mitchell, P. L. Spencer, J. T. Johnson, M. D. Eilts, K. W. Thomas, and D. W. Burgess, 1998: The National Severe Storms Laboratory mesocyclone detection algorithm for the WSR-88D. *Wea. Forecasting,* **13,** 304–326, doi:10.1175/1520-0434(1998)013<0304:TNSSLM>2.0.CO;2.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting,* **18,** 1243–1261, doi:10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2.

——, C. M. Mead, and R. Edwards, 2007: Effective storm-relative helicity and bulk shear in supercell thunderstorm environments. *Wea. Forecasting,* **22,** 102–115, doi:10.1175/WAF969.1.

Trapp, R. J., D. M. Wheatley, N. T. Atkins, R. W. Przybylinski, and R. Wolf, 2006: Buyer beware: Some words of caution on the use of severe wind reports in postevent assessment and research. *Wea. Forecasting,* **21,** 408–415, doi:10.1175/WAF925.1.

Vila, D. A., L. A. T. Machado, H. Laurent, and I. Velasco, 2008: Forecast and Tracking the Evolution of Cloud Clusters (ForTraCC) using satellite infrared imagery: Methodology and validation. *Wea. Forecasting,* **23,** 233–245, doi:10.1175/2007WAF2006121.1.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Elsevier, 627 pp.

Wilson, C. J., K. L. Ortega, and V. Lakshmanan, 2009: Evaluating multi-radar, multi-sensor hail diagnosis with high resolution hail reports. *Proc. 25th Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology,* Phoenix, AZ, Amer. Meteor. Soc., P2.9. [Available online at https://ams.confex.com/ams/pdfpapers/146206.pdf.]

Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell, and K. W. Thomas, 1998a: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting,* **13,** 286–303, doi:10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2.

——, ——, ——, E. D. Mitchell, J. T. Johnson, and K. W. Thomas, 1998b: Evaluating the performance of WSR-88D severe storm detection algorithms. *Wea. Forecasting,* **13,** 513–518, doi:10.1175/1520-0434(1998)013<0513:ETPOWS>2.0.CO;2.

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.