ProbSevere LightningCast: A Deep-Learning Model for Satellite-Based Lightning Nowcasting

JOHN L. CINTINEO,[a] MICHAEL J. PAVOLONIS,[a,b] AND JUSTIN M. SIEGLAFF[a]

[a] *Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, Wisconsin*
[b] *NOAA/NESDIS/Center for Satellite Applications and Research/Advanced Satellite Products Branch, Madison, Wisconsin*

(Manuscript received 27 January 2022, in final form 21 April 2022)

ABSTRACT: Lightning strikes pose a hazard to human life and property, and can be difficult to forecast in a timely manner. In this study, a satellite-based machine learning model was developed to provide objective, short-term, location-specific probabilistic guidance for next-hour lightning activity. Using a convolutional neural network architecture designed for semantic segmentation, the model was trained using *GOES-16* visible, shortwave infrared, and longwave infrared bands from the Advanced Baseline Imager (ABI). Next-hour *GOES-16* Geostationary Lightning Mapper data were used as the truth or target data. The model, known as LightningCast, was trained over the *GOES-16* ABI contiguous United States (CONUS) domain. However, the model is shown to generalize to *GOES-16* full disk regions that are outside of the CONUS. LightningCast provides predictions for developing and advecting storms, regardless of solar illumination and meteorological conditions. LightningCast, which frequently provides 20 min or more of lead time to new lightning activity, learned salient features consistent with the scientific understanding of the relationships between lightning and satellite imagery interpretation. We also demonstrate that despite being trained on data from a single geostationary satellite domain (GOES-East), the model can be applied to other satellites (e.g., GOES-West) with comparable specifications and without substantial degradation in performance. LightningCast objectively transforms large volumes of satellite imagery into objective, actionable information. Potential application areas are also highlighted.

SIGNIFICANCE STATEMENT: The outcome of this research is a model that spatially forecasts lightning occurrence in a 0–60-min time window, using only images of clouds from the GOES-R Advanced Baseline Imager. This model has the potential to provide early alerts for developing and approaching hazardous conditions.

KEYWORDS: Lightning; Satellite observations; Forecast verification/skill; Artificial intelligence; Machine learning

1. Introduction

Lightning is a frequent hazard to life and property in most of the world. Cloud-to-ground (CG) and in-cloud (IC) strikes are dangerous to surface and aviation interests, respectively. Furthermore, hazards such as large hail, severe wind gusts, and tornadoes are usually associated with thunderstorms.

In convective updrafts, noninductive charging theory states that charge separation occurs during collisions between unrimed ice particles (e.g., snow) and rimed ice particles (e.g., graupel, hail) that remain suspended in the mixed phase region by the updraft of a thunderstorm. Charged particles separate via three-dimensional wind shear and sedimentation. This creates a strong electric field and eventually, lightning (e.g., Takahashi 1978; Saunders and Peck 1999). Single-polarization radar reflectivity can be used to identify the regions of potential electrification within developing thunderstorms by showing where graupel and hail particles exhibit large reflectivity echoes (e.g., Shackford 1960; Hondl and Eilts 1994; Vincent et al. 2003; Mosier et al. 2011). This, when combined with environmental data such as isothermal heights, can be used to create products such as reflectivity at 0° or −20°C (e.g., Smith et al. 2016). Large reflectivity in this temperature range (within the mixed phase region) has been often used as an indicator of the onset of IC and CG lightning (e.g., Seroka et al. 2012; Mecikalski et al. 2013). Dual-polarization radar data has also been used to nowcast lightning. Namely, differential reflectivity is used to identify hail, graupel, and supercooled raindrops within the mixed phase region of updrafts (e.g., Woodward et al. 2012; Kumjian 2013). The explicit prediction of lightning has been attempted via parameterizing charging processes within cloud-resolving models (e.g., MacGorman et al. 2001; Fierro et al. 2012, 2015) and applying physics-informed concepts to calibrate the output of numerical weather models into a lightning forecast (e.g., McCaul et al. 2009; Lynn et al. 2012). While numerical weather models can provide good lightning forecasts at longer lead times than observation-based methods, they often are resource-intensive and can lack important timing and location specificity that observations can provide in the very near-term.

The launch of the first two satellites in the Geostationary Operational Environmental Satellite (GOES)-R series, *GOES-16* and *GOES-17*, has brought improved surveillance of atmospheric phenomena from the Advanced Baseline Imager (ABI). The ABI includes a 16-band radiometer, with 2 visible channels, 4 near-infrared channels, and 10 longwave infrared channels (Schmit et al. 2017), and has improved spatial and temporal resolutions from the previous GOES generation. By combining several channels into a single red–green–blue (RGB) composite image, humans are able to see

Corresponding author: John L. Cintineo, cintineo@wisc.edu

DOI: 10.1175/WAF-D-22-0019.1

atmospheric signals more clearly, such as moist or dry air masses, the thermodynamic phase of cloud tops, or the presence of volcanic ash or dust. Several such composites are pertinent to lightning forecasting and are in operational use (e.g., the daytime cloud-phase distinction RGB). Geostationary imagers have advantages over radar data in that they observe a larger domain and observe cloud particles prior to the onset of precipitation. Their main disadvantage relative to microwave-sensing radar is that they do not observe larger hydrometers within clouds that are pertinent to lightning formation.

Machine learning techniques have been successfully applied to observations of convection for short-term forecasting in a number of applications (e.g., Mecikalski et al. 2015; Ahijevych et al. 2016; Lagerquist et al. 2020; Cintineo et al. 2020a). Seroka et al. (2012) leveraged storm-tracking and thresholds in radar-based products to skillfully nowcast lightning, while La Fata et al. (2021) used a random forest model and environmental features to predict specifically CG lightning in the next hour over Italy. Karagiannidis et al. (2016) employed feature engineering to derive interest fields from geostationary satellite data to predict lightning in the next hour over Greece, with good performance.

Deep learning is a subset of machine learning that uses multilayer artificial neural networks to deliver high accuracy in a number of applications, including facial recognition and object detection for self-driving vehicles. It has been used for tropical cyclone intensity estimates (e.g., Wimmers et al. 2019), short-term tornado detection (Lagerquist et al. 2020), satellite-derived convective intensity (Cintineo et al. 2020b), and nowcasting radar echoes (Cuomo and Chandrasekar 2021; Ravuri et al. 2021). One advantage of deep learning over other machine learning (ML) models and techniques is the amount of feature engineering or extraction needed is greatly reduced or circumvented altogether. This not only may speed up the data collection and model training phases of development, but also prevents investigators imposing their own preconceptions of important features in the data, allowing the model to learn for itself what is and is not important to the problem.

Given that the spatial and multivariate context of images are used by forecasters to discern lightning potential, we chose to train a deep-learning model called a convolutional neural network (CNN), which is specifically designed for image inputs. Zhou et al. (2020) use such a CNN with satellite, ground-based lightning, and ground-based radar images as inputs to nowcast lightning in China. Another recent approach is that of Geng et al. (2021) and Zhou et al. (2022), where they use recent lightning, recent surface observations and NWP forecasts to predict lightning occurrence up to 6 h, using a spatiotemporal model suited for sequences of images (ConvLSTM). Geng et al. (2021) found that the inclusion of more data sources resulted in better performance over their evaluation domain of northern China. This research presents a similar method to Zhou et al. (2020), but is novel in the fact that it uses only spaceborne imager data as inputs, and is applied to the GOES-R ABI domain. As will be demonstrated, GOES-R ABI alone can provide an excellent short-term forecast of lightning potential, day and night, over land and sea, for lightning initiation as well as mature storms.

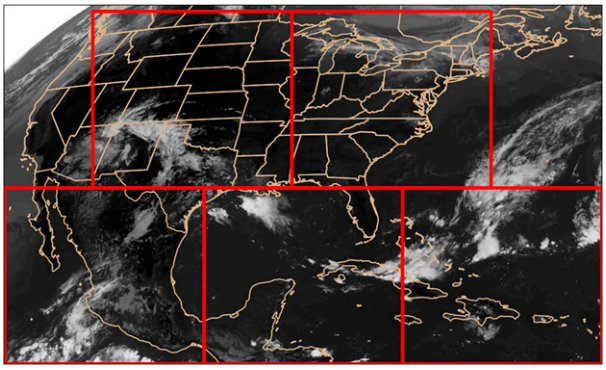

FIG. 1. An example of the *GOES-16* ABI CONUS sector domain with 10.3-μm brightness temperature background. The red boxes show the subdomains where patches were extracted from to train the LightningCast model.

2. Data and methods

a. *GOES-R data*

LightningCast was trained using multiple bands from *GOES-16* ABI (Schmit et al. 2017) as predictors. ABI provides frequent images of surface and atmospheric phenomena by passively observing reflected and/or emitted radiation in the visible, near-infrared, and infrared bands of the electromagnetic spectrum. The satellite is positioned at 75°W longitude over eastern North America and western South America. While there are several scanning strategies for *GOES-16*, this research used data from the contiguous U.S. (CONUS) scan sector for the training process, which covers the contiguous United States, Mexico, and adjacent oceanic regions every 5 min (see Fig. 1 for an example). While the CONUS scan sector was used for the training and validation datasets (explained herein), the model can be applied to any ABI domain, including mesoscale sectors, which have 1-min temporal resolution. Full-disk scan sectors (10-min resolution) from dates independent from the training and validation sets were used to generate performance statistics, while independent CONUS sector scenes were used for case studies and a lead-time analysis. Radiances were converted into reflectances (for the visible and near-infrared bands) or brightness temperatures (for the longwave infrared bands).

LightningCast uses four *GOES-16* bands as predictors: 1) 0.64-μm reflectance (the "red" portion of the visible spectrum) with a nadir resolution of 0.5 km, 2) 1.6-μm reflectance (shortwave infrared) with a nadir resolution of 1 km, 3) 10.3-μm brightness temperature (longwave infrared) with a nadir resolution of 2 km, and 4) 12.3-μm brightness temperature (longwave infrared) with a nadir resolution of 2 km. These bands were chosen because of their applicability to cloud-top phase discrimination (e.g., Elsenheimer and Gravelle 2019; Pavolonis 2010). Several other *GOES-16* channels were tested ad hoc (1.37-μm reflectance, 6.2-μm brightness temperature,

TABLE 1. The *GOES-16* ABI predictors and GLM target/truth field used to train LightningCast. The temporal resolution is given for the three partitions of the dataset.

Channels	Spatial resolution	Temporal resolution (training/validation/testing)
GOES-16 ABI 0.64-μm reflectance	0.5 km	5 min/5 min/10 min
GOES-16 ABI 1.6-μm reflectance	1 km	5 min/5 min/10 min
GOES-16 ABI 10.3-μm brightness temperature	2 km	5 min/5 min/10 min
GOES-16 ABI 12.3-μm brightness temperature	2 km	5 min/5 min/10 min
GOES-16 GLM flash-extent density (maximum over 60 min)	10 km; remapped to 2-km ABI fixed grid	1 min/1 min/1min

8.4-μm brightness temperature), but did not improve the model's performance on a validation dataset. We leave a thorough investigation of all of the ABI channels to future work.

The truth or target data were created from the Geostationary Lightning Mapper (GLM; Rudlosky et al. 2019; Goodman et al. 2013) aboard *GOES-16*. The GLM is an optical sensor that passively observes radiation in a single band at 777.4 nm, covering much of the *GOES-16* ABI full disk of observation. The GLM has a near-uniform spatial resolution of approximately 10 km and the temporal resolution of the GLM level-2 files is 20 s. The GLM instrument does not distinguish between IC and CG flashes. The GLM level-2 files contain flash data, which were processed using open-source software (glmtools; Bruning 2021). This software transformed point-based lightning flash, group, and event data into a gridded field of "flash-extent density," which depicts the number of flashes that traverse a given location within some unit of time (e.g., Bruning et al. 2019). The flash-extent density was also regridded to 2-km spatial resolution matching the resolution and domain of the 2-km ABI fixed grid.

Table 1 summarizes the data used to train the model. The flash-extent density was accumulated for 60 min after ABI scan times, taking the maximum value at every grid point over the 60-min interval. Despite its diminishing detection efficiency at higher viewing angles (e.g., Rudlosky et al. 2019), the GLM instrument was selected to provide the target field for the training of LightningCast because of the spatial depiction of total lightning (IC + CG) that it provides from geostationary orbit. *GOES-17* ABI and GLM data were used analogously as a secondary test set, to examine how well the model generalizes to a new spatial domain.

Due to graphical processing unit (GPU) memory limitations during training, the ABI CONUS sector was divided into five subdomains or patches (see Fig. 1). Each patch had horizontal dimensions of 2944 × 3328 pixels at the 0.5-km resolution (the 1- and 2-km inputs were oversampled to 0.5-km resolution). While six uniform patches could be extracted from the CONUS sector domain, we chose to use only five patches in order to exclude missing data in the space-look region of the domain (the northwest corner), which could have made the training process unstable. ABI data were extracted from the *GOES-16* ABI level-1b files, transformed into reflectance and brightness temperature, and reshaped into the necessary tensor shapes (i.e., multidimensional arrays) on-demand using a Python generator. *GOES-16* GLM level-2 data were transformed into flash-extent density over the CONUS sector domain beforehand using glmtools. The 60-min maximum flash-extent density patches were also extracted from the subdomains on-demand during model training time.

b. *Data labeling and partitioning*

For this model, given a single scene or scan of ABI predictors, we sought a probability of lightning within the next 60 min at each pixel. Since the problem we aimed to solve was a binary classification problem (lightning or no lightning), the accumulated flash-extent density was binarized at 1 flash. Thus, the final target field for each ABI scene was a mask of ones where there was one or more GLM-observed flashes in the 60 min following an ABI scan time, and a mask of zeros where no lightning was observed by GLM in the ensuing 60 min.

The dataset was broken down into the standard three parts for machine learning: training, validation, and testing sets. The training set contains scenes from where the model can learn spatial and multispectral features pertinent to lightning prediction; the validation set provides an independent assessment of a model during training, which is useful for hyperparameter selection and helps prevent overfitting on data from the training set; and the testing set is independent from the training and validation sets and provides a final assessment on model performance after training and hyperparameter tuning. The training and validation sets were selected from 2019 dates and the testing set was selected from 2020. The dates were selected arbitrarily, while ensuring the training set was substantially larger than the validation set (of the two sets used during training, the training set was approximately 75% of the total). See Table 2 and Fig. 2 for a summary of the data partitions.

An important design consideration while selecting dates and times for the training and validation sets was the fractional coverage of lightning in the patch. It is generally accepted that datasets balanced between classes train in a more stable fashion, since the learning algorithm will not be biased toward predicting the majority class. The two classes in this problem are pixels with lightning in the next 60 min and pixels with no lightning in the next 60 min. Since lightning is a relatively rare phenomenon in a geostationary satellite image, there is a natural imbalance in the partitioned sets. Instead of balancing the classes with weights, as is frequently performed in deep learning tasks (e.g., Ren et al. 2018), we decided to limit the imbalance in each patch (i.e., sample): for a patch to be added to the training or validation set, at least 2.5% of its pixels needed to exhibit the positive class (any lightning in the

TABLE 2. Summary of lightning pixel fractions in the three datasets. The training and validation set domains are from the *GOES-16* CONUS sector, while the testing sets are subsets of the *GOES-16* and *GOES-17* full-disk sectors.

	Training (*GOES-16* CONUS sector)	Validation (*GOES-16* CONUS sector)	Testing (*GOES-16/GOES-17* full-disk sectors)
Dates	2019	2019	2020
	19–21, 23–24 Jan	12–13 Feb	5–9 Jan
	19–21, 23–24, 27 Feb	12–15 Mar	5–9 Feb
	22–29 Mar	6–10 Apr	5–9 Mar
	15–30 Apr	3–8 May	5–9 Apr
	10–31 May	3–8 Jun	5–9 May
	10–30 Jun	3–8 Jul	5–9 Jun
	10–31 Jul	3–8 Aug	5–9 Jul
	10–31 Aug	3–8 Sep	5–9 Aug
	10–30 Sep	8 Oct	5–9 Sep
	10–11, 19–21, 25–26, 29, 31 Oct	1 Nov	5–9 Oct
	27–30 Nov	14–15, 18 Dec	5–9 Nov
	16–17 Dec		5–9 Dec
No. of samples (patches)	76 031	22 539	8488/8066
No. of unique date–times	31 802	9940	8488/8066
Percentage of pixels with ≥1 flash per 60 min	5.190%	5.344%	1.605%/0.184%

next 60 min). This helped ensure stability during training (i.e., convergence toward a solution) while also capturing enough non-lightning scenes under a wide range of conditions. Class weighting was also assessed, but the class weight experiments failed to produce a well-calibrated and skillful model, so that approach was abandoned.

An important goal of this work was to forecast the onset of lightning (i.e., lighting initiation in storms), so keeping a lower fractional lightning-coverage threshold appeared to help, since more lightning initiation examples were captured. Larger thresholds tended to skew the datasets toward scenes with too much mature and already-electrified convection. This threshold helped provide a balance between developing and developed convection, while being large enough to retain stability during the training process. One percent appeared to be too little (the model did not converge on a solution) and 10% appeared to be too high (not enough developing-convection scenes).

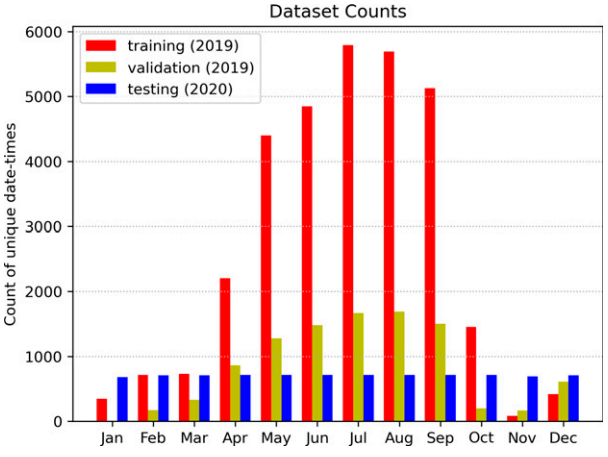

FIG. 2. The time sampling of the *GOES-16* training, validation, and testing datasets.

c. *Model design*

LightningCast uses a CNN designed for semantic segmentation, which performs classification and localization of the classes (i.e., determines where in the image each class resides). Ronneberger et al. (2015) pioneered this CNN model architecture in medical imagery. It is frequently referred to as a "U-net" because of how the model looks schematically (see Fig. 3).

The U-net contains two parts: the contracting path, which contains convolution layers and pooling layers; and the expanding path, which contains transposed convolution and convolution layers. The convolution and pooling layers transform input data into "feature maps," or abstractions of the inputs. The number of feature maps often increase with depth in the contracting path, increasing the number of features that may be learned. The pooling layers also reduce the spatial resolution of the feature maps, enabling the model to learn features at multiple spatial resolutions in the data. The convolution operator, as defined in Lagerquist et al. (2019), enables learning of features both spatially and in a multivariate fashion (i.e., combining input channels).

The expanding path uses transposed convolution layers, which perform an inverse convolution operation while "upsampling" the feature maps, that is, increasing the spatial resolution of the feature maps. The number of feature maps is also typically reduced as depth decreases in the expanding path. To better localize, high-resolution features from the contracting path are copied and concatenated (see gray arrows in Fig. 3) with the up-sampled output in the expanding path. The additional convolution layers can then learn to assemble a more precise output based on the concatenated high-resolution features (Ronneberger et al. 2015). Some U-nets have excluded

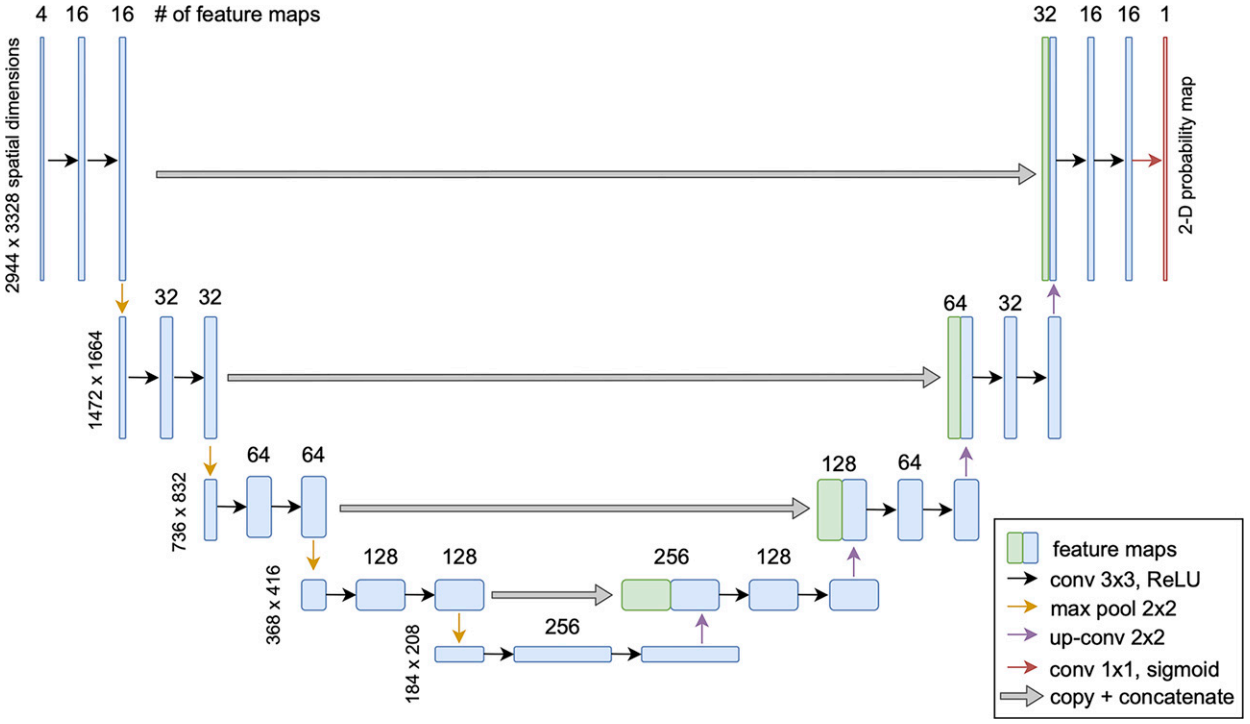

FIG. 3. A schematic of the U-net architecture for the LightningCast model.

the concatenation layers while retaining good performance (e.g., Hilburn et al. 2021).

In LightningCast, nonlinear rectified linear unit activations (ReLU; Nair and Hinton 2010) are used in each convolution layer (except the last convolution layer), to enable the model to learn nonlinear features. The last convolution layer uses a 1×1 pixel kernel with a sigmoid activation, forcing a probability of lightning, ranging from [0, 1]. Each ABI input channel with 1- or 2-km spatial resolution is up-sampled to 0.5-km resolution by repeating rows and columns by a factor of 2 and by 4, respectively, matching the 0.64-μm reflectance channel resolution (this up-sampling step is not shown in Fig. 3). While the output of this architecture has 0.5-km spatial resolution, we use a 4×4 maximum pooling to reduce the spatial resolution to 2 km, which decreases noise in the final output.

We sought to distinguish between the lightning and no-lightning classes, so binary cross-entropy was used as the loss function to minimize [Eq. (1)]. The p_i term is the probability of lightning in the ensuing 60 min at the ith pixel, y_i is the label (1 if lightning, 0 if no lightning) for the ith pixel, N is the number of pixels, and ε is the binary cross-entropy, ranging from $[0, \infty)$:

$$\varepsilon = \frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]. \qquad (1)$$

The Python computing language and Tensorflow-GPU library were used to perform the training on one NVIDIA Quadro RTX 6000 GPU. Please see the appendix for further model details.

3. Results

This section demonstrates output from the LightningCast model for several examples over the United States and adjacent oceanic regions, and presents verification statistics for the *GOES-16* test set. LightningCast was also applied to *GOES-17* data, to evaluate its performance on a geostationary scan domain that differed from the training and validation domain. Last, the attribution and relevance of predictions is investigated using several scenes from the examples in this section.

a. *Example scenes*

1) IOWA, 28 AUGUST 2020

A cold front swept through the state of Iowa during the afternoon of 28 August 2020. Each panel in Fig. 4 represents a different *GOES-16* ABI scan start time. The GLM flash-extent density is plotted over the 0.64-μm reflectance, and represents the number of accumulated GLM flashes at each point in the 5 min previous to the ABI scan start time, so that the lead time of the model predictions can be assessed (i.e., ideally GLM should first observe lightning, at a given location, after the model probability increases at that location). The colored contours overlaid on the imagery depict locations of LightningCast probability of lighting in the 0–60-min prediction window after the ABI scan.

Most of the vigorous convective development is along the cold front under clear skies with no overlapping cloud cover. The probability of lightning develops and increases over several different sections of the cold front. Over the span of

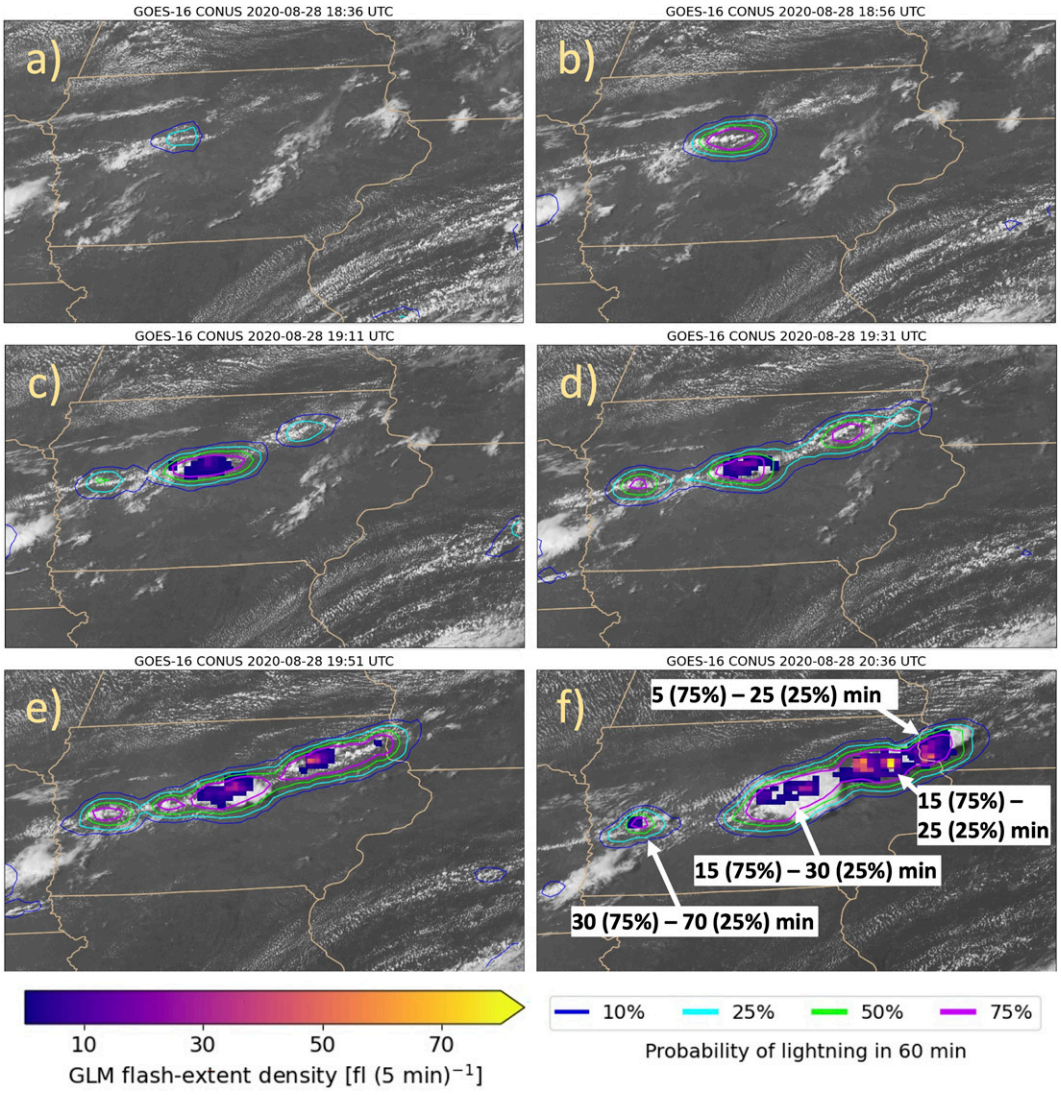

FIG. 4. (a)–(f) A select sequence of images depicting the evolution of LightningCast probabilities along a cold front in Iowa, superimposed on 0.64-μm reflectance (grayscale) and GLM flash-extent density from *GOES-16*. Lead times to the initial GLM flashes for several areas of interest are annotated in (f), showing lead times in minutes from both the 75% and 25% probability thresholds.

about an hour, four regions of the front show local maxima of lightning probability (Fig. 4d). Eventually, each high-probability region exhibits numerous GLM flashes. In this example, LightningCast results in variable lead time measured to the first GLM-observed flash for different regions of the front, reflecting the natural variability in the rate of thunderstorm development. The lead time varies between 5 and 30 min at the 75% probability threshold and 25–70 min at the 25% probability threshold (Fig. 4f).

2) INDIANA–MICHIGAN–OHIO, 14 SEPTEMBER 2021

On 14 September 2021, convective cells grew within the warm sector of a midlatitude cyclone near the Indiana/Michigan/Ohio nexus. In Figs. 5 and 6, the background image for the left-column panels is the daytime cloud-phase distinction RGB, created according to the method in Connell et al. (2021). In the left panel of each row of Figs. 5 and 6, cumulus clouds that change color from light blue to green to yellow indicate cloud-top glaciation, which is a prerequisite for lightning initiation in convective clouds. The right panel of each row shows the 0.64-μm reflectance from *GOES-16* ABI and the Multi-Radar Multi-Sensor reflectivity at −10°C (MRMS; Smith et al. 2016). The storm in northeast Indiana (the southwest-most cell in the sequence) exhibited a ≥50% probability of lightning contour (Fig. 5c) approximately 10 min before a 40-dBZ radar echo was evident at the −10°C level (Fig. 5d), and 30 min before the

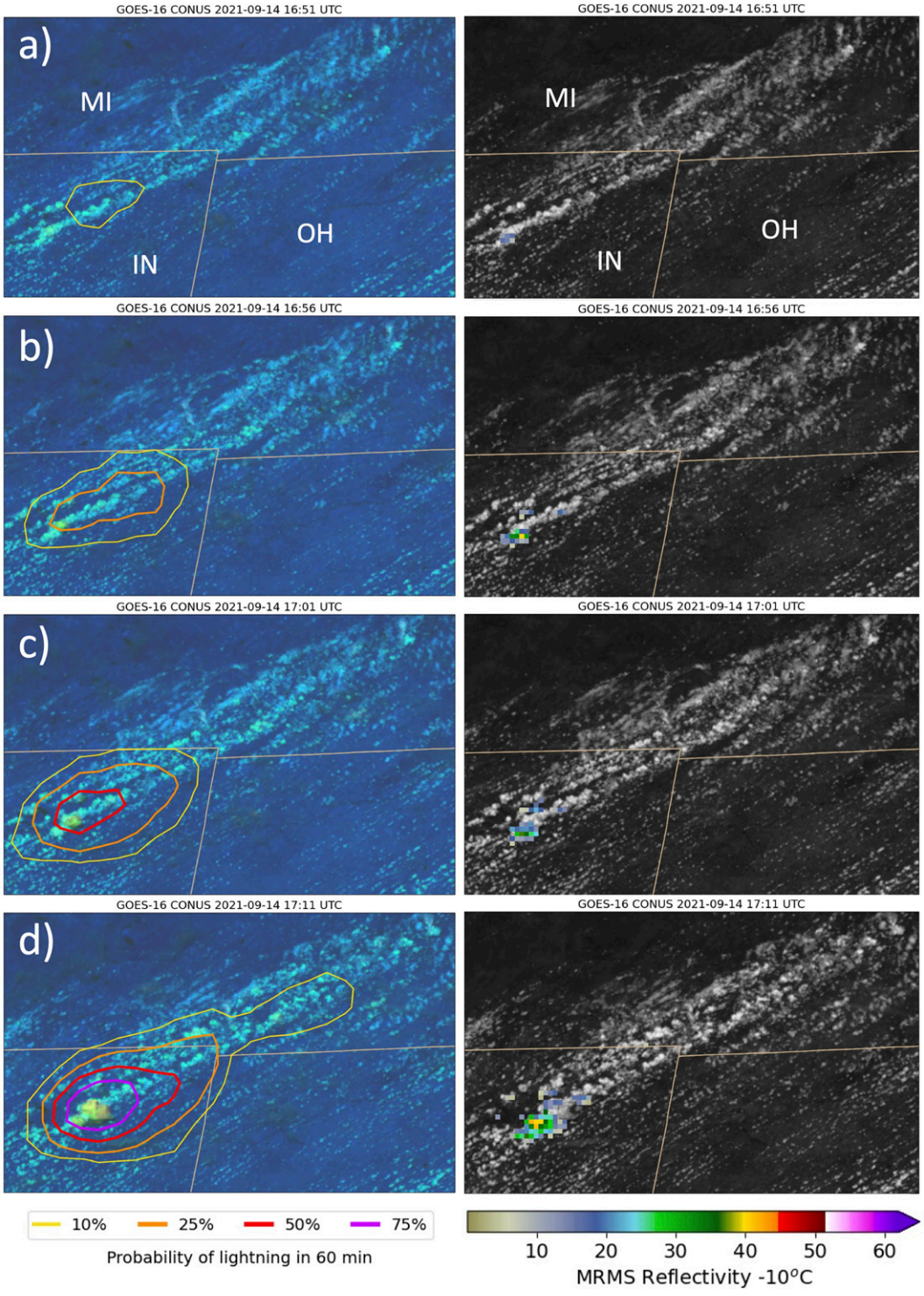

FIG. 5. (a)–(d) A select sequence of images depicting the evolution of LightningCast probabilities near the Indiana–Michigan–Ohio nexus, superimposed on the daytime cloud-phase distinction RGB and GLM flash-extent density from (left) *GOES-16* and (right) the 0.64-μm reflectance from *GOES-16* and the MRMS reflectivity at −10°C. U.S. state abbreviations are shown in (a).

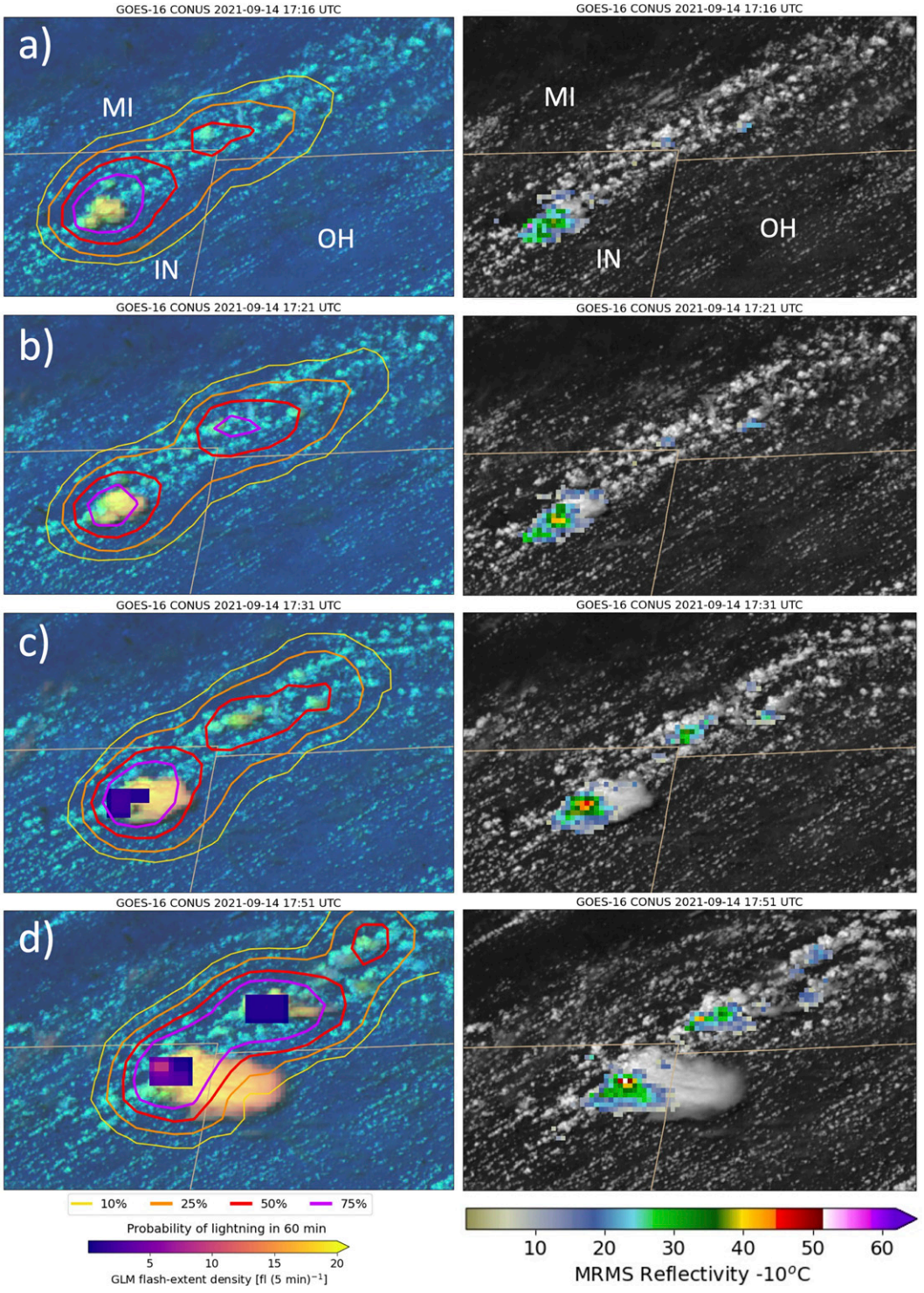

FIG. 6. As in Fig. 5, but for subsequent scan times.

initial GLM flash (Fig. 6c). The storm in southern Michigan exhibited ≥50% probability of lightning contour (Fig. 6a) approximately 10 min before a core of 30-dBZ was evident in the reflectivity at −10°C (time not shown), and 35 min before the initial GLM flash (Fig. 6d). The example demonstrates that LightningCast can complement radar-based methods of forecasting lightning, and potentially add lead time to lightning initiation in some instances.

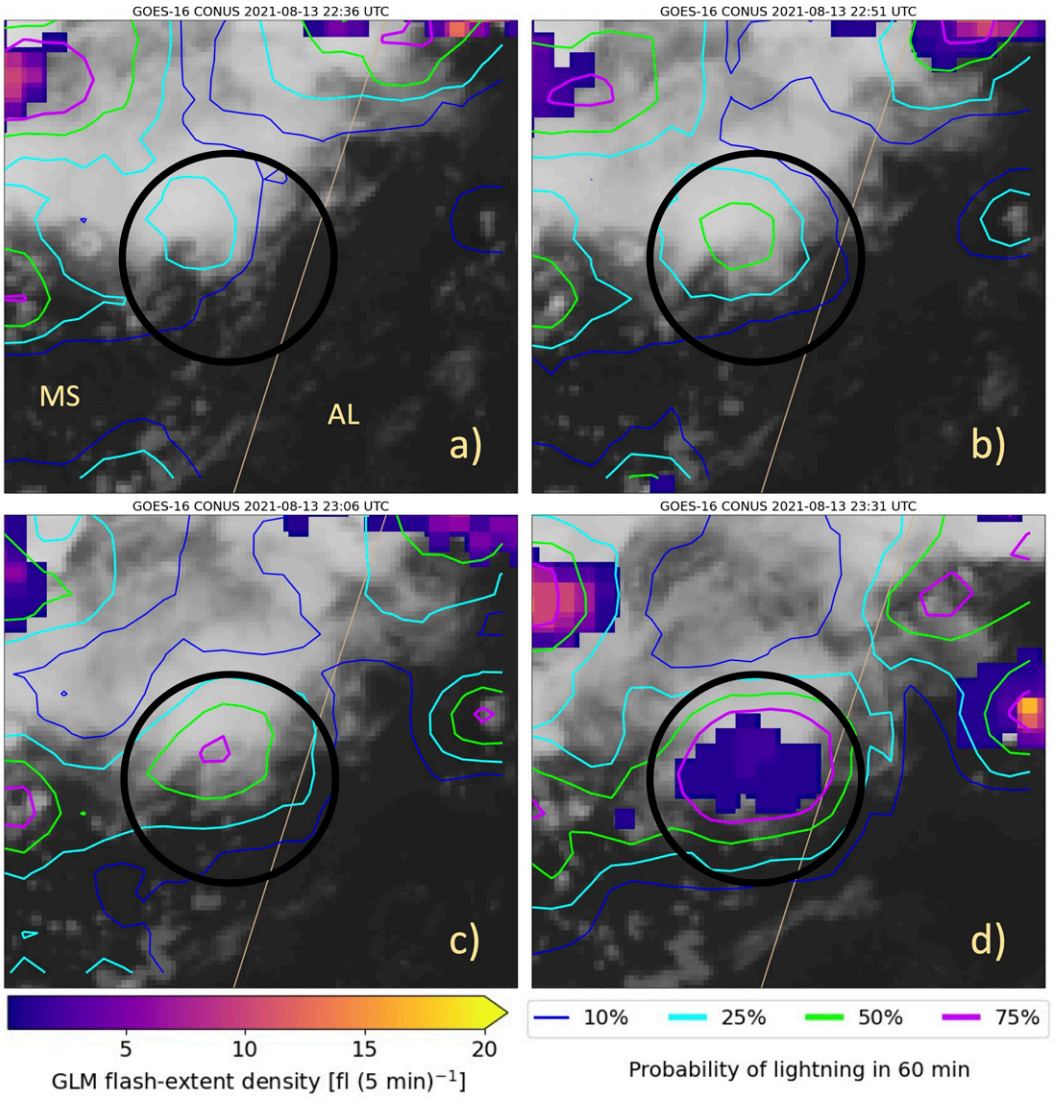

FIG. 7. (a)–(d) A select sequence of images depicting the evolution of LightningCast probabilities in eastern Mississippi, superimposed on the 10.3-μm brightness temperature (grayscale) and GLM flash-extent density from *GOES-16*. The black circle denotes an area of attention, where GLM observes several flashes of lightning. U.S. state abbreviations are shown in (a).

3) MISSISSIPPI, 12 AUGUST 2021

Thunderstorm activity was ongoing in eastern Mississippi during the early evening of 12 August 2021, in a regime characterized by a moist atmosphere (precipitable water $>$ 46 mm [$>$1.8 in.]), moderate convective available potential energy (CAPE; 1500–2000 J kg^{-1}) and weak flow throughout the region [850–300-hPa mean wind $<$ 5.1 m s^{-1} ($<$10 kt)]. Thick ice clouds can be seen in the 10.3-μm brightness temperature in Fig. 7. Some thick ice, or convective "blowoff" from other storms is common across east-central Mississippi. While difficult to see, cumuliform clouds begin to develop under this convective detritus. The LightningCast probability also increases during this time, in the same region (denoted by a black circle in Fig. 7). Numerous GLM flashes were observed between 2321 and 2345 UTC. In this area, LightningCast exhibited 50 min of lead time at the 25% threshold, 35 min at the 50% threshold, and 15 min at the 75% threshold, relative to the initial GLM-observed flash (at 2321 UTC). Thus, LightningCast may be able to provide early alerts of lightning formation to forecasters in busy situations or regions where lightning potential may be difficult to visually discern using standard analysis techniques.

4) EAST PACIFIC OCEAN, 13 DECEMBER 2021

In the wake of a cold front in the eastern Pacific Ocean, low-topped convection formed in a cellular-like pattern under a broad area of low pressure. LightningCast predicted widespread probabilities exceeding 25%, with a few regions

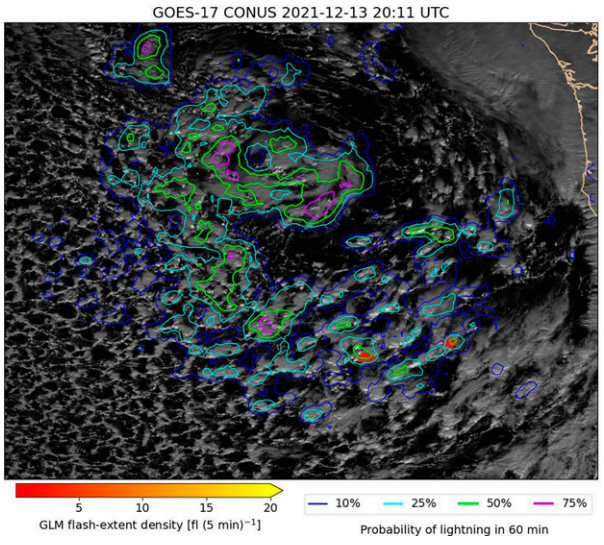

FIG. 8. LightningCast predictions (colored contours), superimposed on the 0.64-μm reflectance and GLM flash-extent density from *GOES-17*, for a scene in the eastern Pacific Ocean.

exceeding 75%. While a few intermittent GLM flashes were observed (e.g., red regions in Fig. 8), flashes were not observed for most of the predicted regions in the ensuing hour. This example illustrates one potential failure mode for the model. Convection is abundant, with clouds tops varying from approximately 7 to 9 km above sea level and minimum 10.3-μm brightness temperatures between 230 and 220 K. While increasing the seasonal diversity of the training set may improve model performance on such scenes, further investigation is needed on such failure modes.

b. GOES-16 *verification*

This section delves into model verification using the *GOES-16* test sets. The first test set uses a subset of the *GOES-16* full-disk domain, at 10-min resolution, for the 5th–9th of each month of 2020, resulting in 8488 unique image acquisitions (see Table 2). Recall that the training was performed using the *GOES-16* CONUS sector domain. This expanded domain enables us to examine the applicability of LightningCast to new regions, such as the northern portion of South America. A second *GOES-16* test set uses CONUS sector data (also for the 5th–9th of each month of 2020) to examine lead time to new lightning occurrences, which was an important goal in this work.

1) *GOES-16* FULL-DISK SUBSET

Since LightningCast produces probabilistic forecasts for a binary problem (lightning or no lightning), we can binarize the forecasts at any probability threshold p. For forecast probability p_i, $p_i \geq p$ is set to 1 and $p_i < p$ is set to 0, which creates binary forecasts. Based on the binary forecasts and binary observation of future lightning for given forecast (o_i), the following are defined:

- Hit or true positive (TP); when $p_i \geq p$ and $o_i = 1$
- False alarm or false positive (FP); when $p_i \geq p$ and $o_i = 0$
- Correct null or true negative (TN); when $p_i < p$ and $o_i = 0$
- Miss or false negative (FN); when $p_i < p$ and $o_i = 1$

The probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) at a given probability threshold p_i, at any given location (or pixel) in the domain can then be defined as

$$\mathrm{POD} = \frac{\sum \mathrm{TP}}{\sum \mathrm{TP} + \sum \mathrm{FN}}, \tag{2}$$

$$\mathrm{FAR} = \frac{\sum \mathrm{FP}}{\sum \mathrm{TP} + \sum \mathrm{FP}}, \tag{3}$$

$$\mathrm{CSI} = \frac{\sum \mathrm{TP}}{\sum \mathrm{TP} + \sum \mathrm{FN} + \sum \mathrm{FP}}. \tag{4}$$

POD, FAR, and CSI range from [0, 1], with 1 (0) being perfect for POD and CSI (FAR). The CSI is the accuracy of the model neglecting true negatives, and is a better representation of model performance than binary accuracy for rare-occurring events (the test set contained 1.6% lightning samples).[1] The probability threshold that achieved the highest CSI on the validation set was p_i = 36% (we use the approximation p_i = 35% in Fig. 11).

Overall, for the full-disk-subset test dataset, the predicted probability threshold maximizing the CSI was 45%, but 35%–55% exhibited similar CSI, of approximately 0.4 (Fig. 9). Probabilistic predictions should also be skillfully calibrated, that is, the forecast probability and frequency of lightning should be the same for a given probability. The attributes diagram (Fig. 10) shows an overprediction bias in each probability bin, but with each bin contributing skillfully to the Brier skill score when measured against the frequency of lightning in the test dataset (i.e., "climatology"). The Brier score measures the accuracy of probabilistic predictions and is equivalent to the mean squared error for binary classification (Wilks 2006), while the Brier skill score measures improvement over climatology, in this case. Performance on the validation set indicated a higher (i.e., better) Brier skill score than the test set. The overpredictions could be a result of an expanded spatial domain or a more uniform seasonal distribution in the test set, or a combination of both, resulting in cloud types or regimes that were underrepresented during the training process. While further investigation is needed to understand and correct the overprediction, it is promising that the probabilistic predictions are skillful on new data.

The POD, FAR, and CSI were computed on a per-pixel basis over on the *GOES-16* test set. The POD was generally 0.6–0.8 and the FAR was 0.3–0.5 for most locations. This yielded CSI scores of 0.35–0.55 for most regions (Fig. 11) and includes lightning initiation and developed thunderstorms. The CSI of the model appears to be better over land than

[1] A model that forecasted 0% all the time would achieve an accuracy of 0.984, with 1 being perfect.

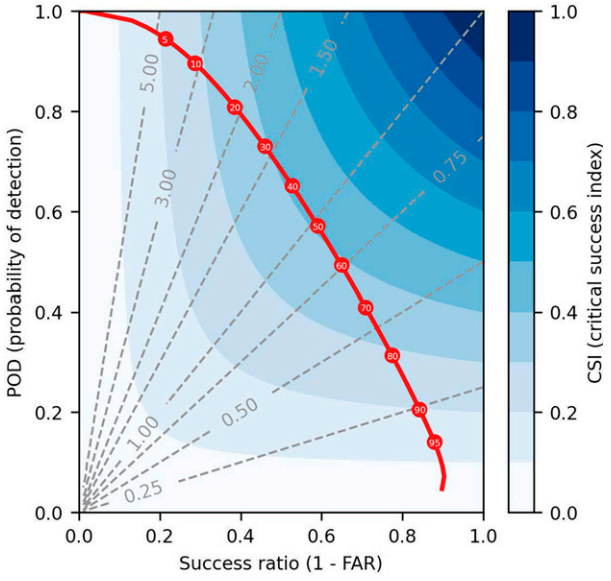

FIG. 9. A performance diagram for the LightningCast model on a test set, which was a subset of the *GOES-16* ABI full-disk sector. The dashed gray lines show frequency bias and the filled blue color shows CSI. Red circles represent select probability thresholds (from top left to bottom right): 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 95%.

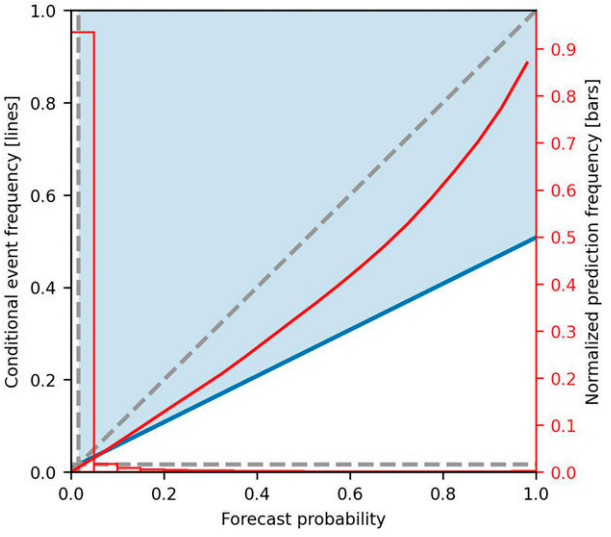

FIG. 10. An attributes diagram for the LightningCast model, for the same test dataset as in Fig. 9. The red line represents the frequency of next-hour lightning, given a certain forecast probability. The diagonal dashed gray 1-to-1 line represents perfect reliability or forecast calibration. The vertical dashed gray line is the climatology $\bar{y}$, or frequency of lightning for the test dataset. The horizontal dashed gray line is the no-resolution line (produced by always forecasting $\bar{y}$). The diagonal blue line represents the line of "no skill" with respect to climatology, or where the Brier skill score is zero. This line separates the area where forecasts contribute positively to the Brier skill score (shaded blue) and where forecasts contribute negatively to the Brier skill score (white). The red-outlined bars show the prediction frequency for each probability bin, normalized between 0 and 1.

over water. This may be a result of the fact that maritime storms have lower vertical velocities and lower liquid water fluxes, in general, when compared to continental storms (e.g., Lucas et al. 1994). Solomon and Baker (1998) suggest that maritime storms exhibit increased coalescence, removing much of the liquid water before it reaches the freezing level, diminishing lightning production. Maritime storms also typically have lower aerosol particle concentrations, reducing cloud condensation nuclei, and thus reducing the vertical liquid water fluxes of storms (e.g., Thornton et al. 2017). This suggests that such processes (e.g., liquid water fluxes) are more difficult to discern for the model over water than over land, using only satellite imager inputs. Nonetheless, the CSI performance over water still appears to be quite good, overall.

The FAR is rather high in several oceanic regions, exceeding 0.8 (e.g., portions of the east Pacific and North Atlantic). While many of these locations have very little lightning activity (evidenced by the event count image in Fig. 11), making it difficult to draw conclusions, some locales have an adequate event count, such as the east Pacific Ocean (an event is the occurrence of maximum GLM flash-extent density of at least one flash over an ensuing 60-min period). Our hypothesis for the lower performance in this region is meteorological in nature—storms in this region, while still deep convection, tend to exhibit less lightning overall due to decreased vigor in updrafts compared to continental updrafts (e.g., LeMone and Zipser 1980; Lucas et al. 1994; Solomon and Baker 1998; Thornton et al. 2017).

In regions where lightning is very rare, this analysis will also be more sensitive to missed GLM detections, so we may expect reduced scores in such regions. The GLM's detection efficiency is not perfect—for example, optical extinction due to a large cloud water path can greatly reduce the number of observed flashes (Rutledge et al. 2020; Bateman and Mach 2020). Thus, a missed detection in a region where there is only one flash in the ensuing 60 min could erroneously result in a false alarm.

Seasonally, LightningCast shows very consistent performance during the months of April through September, with maximum CSI ranging between 0.4 and 0.5, with a 35%–40% probability threshold range (Fig. 12). Comparatively, the months of January–March, and October show a reduced performance, while November and December have a CSI of about 0.25–0.3. The best CSI occurs at probability thresholds of 55%–65% for November, December, and January. During these three months, the FAR is higher for most probability thresholds (not shown), which indicates some overprediction for these months (i.e., the observed frequency of lightning is less than the predicted frequency at a given probability level). This reduced performance may be due to relatively more prevalent tropical and subtropical convection, but we posit that this is a result of the relative dearth of training samples from these months, due to sample-patch selection criteria favoring a larger spatial coverage of lightning, relative to patches with sporadic observations of lightning (see section 2b). Increasing the number of training samples in these months could produce a sizeable improvement in performance, as there

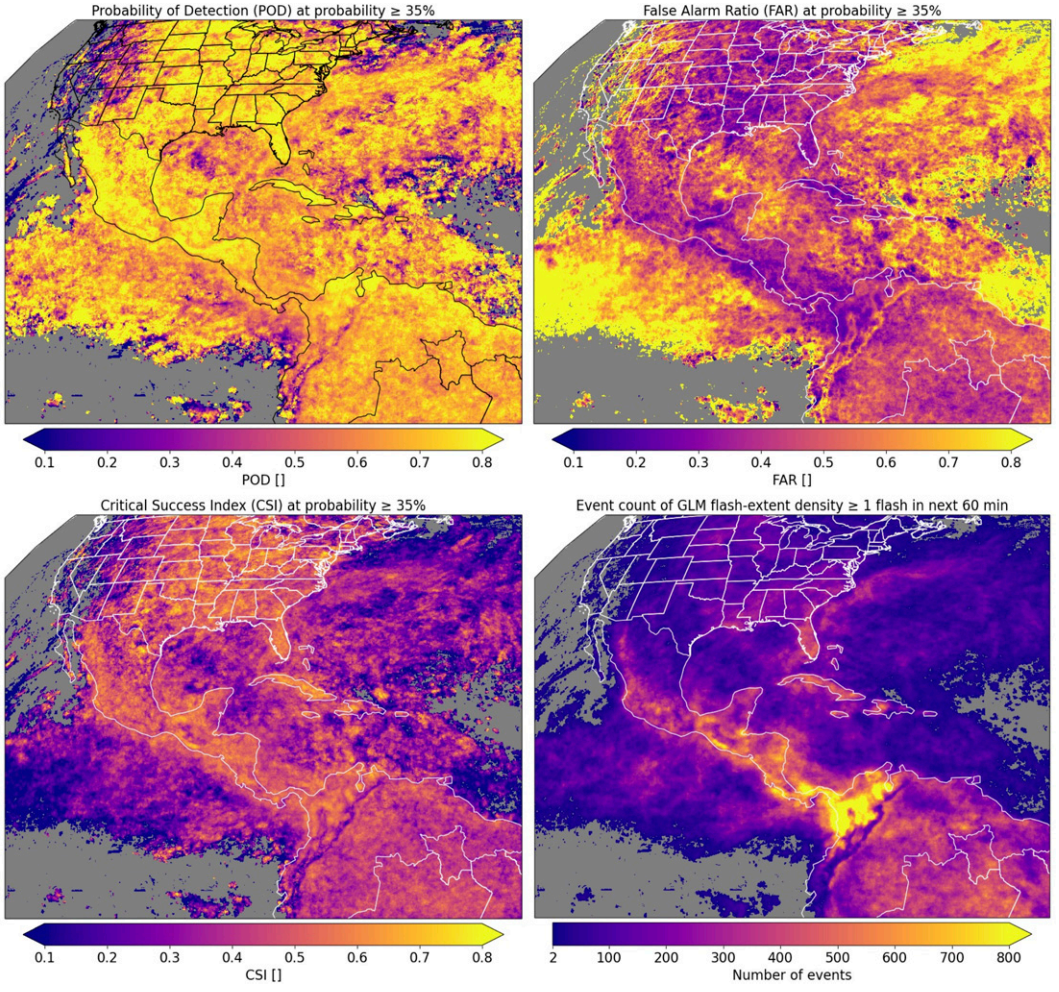

FIG. 11. The (top left) POD, (top right) FAR, (bottom left) CSI, and (bottom right) event count for the *GOES-16* test set (full-disk subset).

was a reduced number of training and validation samples in January–March and November–December, relative to the months of April–October (see Fig. 2).

Overall, LightningCast performs best during the day, with a little reduced performance during nighttime (it was determined to be "daytime" at a location if its solar-zenith angle was ≤88°). This improved daytime performance is small but fairly consistent during the warm season (May through September), when thunderstorms are most numerous in North and Central America. This improved performance is most stark during February–April, though it is unclear why. December shows an inverse of this, with nighttime performance being better.

2) *GOES-16* CONUS SECTOR

To investigate the lead time offered by LightningCast, we used the *GOES-16* CONUS sector instead of the full-disk sector; the former has 5-min temporal resolution and the latter 10-min resolution. Thus, the higher temporal resolution provides a more accurate estimate of lead-time expectation. Similar to the full-disk test set, the dates for this set were also the 5th–9th of each month of 2020. In an effort to isolate lead time to lightning initiation, a few conditions were introduced. The purpose of these conditions was to capture only lightning initiation or new storms advecting into a location. It is important to provide lead time to initial lightning flashes so that we may diagnose if the model provides adequate warning for taking mitigating action. Providing lead time to advecting lightning activity is also important, and can approximate the time of arrival of lightning, particularly when the trend of probability is plotted as a function of time. These two conditions facilitated the analysis:

1) The set of points to evaluate, P, were selected at approximately every 80 km over the CONUS domain, striding the image in both spatial dimensions. Let each point be denoted as P_{ij}.
2) Each point P_{ij} was spatially buffered by approximately 10 km in both spatial dimensions, denoted as b. Let the set of points within buffer b of P_{ij} be denoted as P_b.

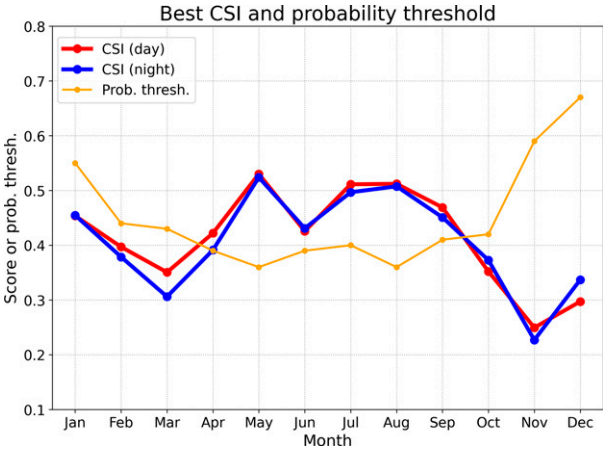

FIG. 12. The maximum CSI and associated probability threshold as a function of month and day or night.

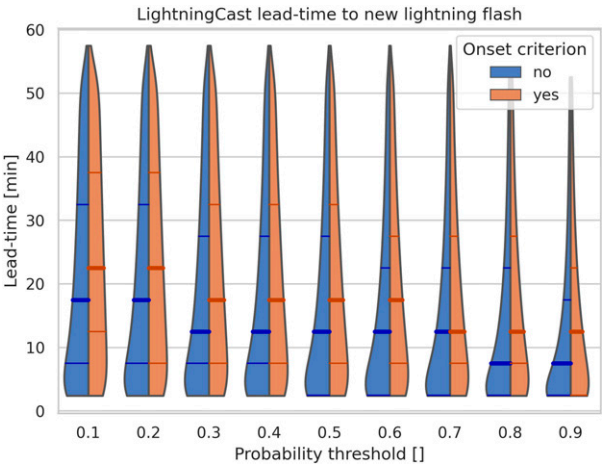

FIG. 13. Distributions of lead time to the first lightning flashes, aggregated across the *GOES-16* ABI CONUS sector test set, for both when the lightning onset criterion was applied (orange) and when it was not (blue). Thick horizontal lines indicate median lead-time values for each distribution, while thinner horizontal lines indicate the 75th and 25th percentiles.

Striding each image to generate the set of points P reduced the redundancy of samples among neighboring pixels. The spatial buffer criterion of 10 km helps capture the predictions and lightning observed "nearby" point P_{ij}. After the probability of lightning was computed for the entire domain at time t_0, the maximum probability and GLM flash-extent density within P_b were recorded.

To measure lead time to the next lightning observation within P_b and within the ensuing 60 min (from t_{0+5} to t_{0+60}), if any lightning was observed, two different criteria were evaluated: 1) if no lightning was observed within P_b at t_0 only, but lightning was observed within t_{0+5} to t_{0+60}, then lead time could be measured at select probability thresholds, and 2) if no lightning was observed within P_b and within the *previous* 30 min, between t_0 and t_{0-30}, then lead time could also be measured to the next lightning observation within 60 min (t_{0+60}). The second criterion, the lightning "onset criterion," reduces the occurrence of storms with pulsing or intermittent lightning at P_b, that is, storms that may exhibit one or more 5-min periods without lightning within P_b. Using the onset criterion better captures new lightning activity at a location, including lightning initiation.

This analysis used CONUS scans at 5-min temporal resolution, and thus we also used corresponding GLM flash-extent density at 5-min resolution, with the flash-extent density max-accumulated over 5-min increments. Thus, this quantizes the measured lead time into 5-min bins. Since we do not know when in the 5-min increment that lightning was observed, 2.5 min was subtracted from each lead-time measurement to provide an unbiased estimate. For example, if the probability of lightning was p at t_0 and there was no lightning at t_0, but lightning was observed 5 min in the future, the recorded lead time is 2.5 min at threshold p.

By screening out intermittent convection with the onset criterion, we can ascertain lead time for storms initiating lightning at a given location. Aggregating over the set of points P and all times in this test set, the median lead time to the first flash is 17.5 min at the 30% and 40% thresholds (Fig. 13), which contained the most-skillful range for the validation set. Without the lightning onset criterion, the median was reduced to 12.5 min in this probability range.

Thus, for the most skillful probability range with the lightning onset criterion (which is more applicable for new lightning at a location), lead time of at least 17.5 min to the first flash was attained half of the time, while a quarter of the true-positive predictions measured 32.5 min or more of lead time. These results demonstrate that LightningCast can provide ample lead time for a number of applications.

c. GOES-17 *application*

The ABI's infrared detectors must be very cold to accurately measure thermal radiation from Earth (NOAA/NESDIS 2021). Unfortunately, the loop heat pipe (LHP) on *GOES-17*, which regulates the temperatures of the infrared detectors, proved faulty soon after the checkout period for the satellite. This makes certain ABI bands particularly noisy and unusable during certain periods of the year. One of LightningCast's predictors, the 12.3-μm brightness temperature, is severely affected by the LHP problem. Thus, in order to evaluate LightningCast on the *GOES-17* domain, a very similar model was trained, using only the 0.64- and 1.6-μm reflectances and the 10.3-μm brightness temperature on *GOES-16*.[2] We applied the three-channel version of LightningCast, trained using *GOES-16*, to *GOES-17* in order to investigate how well the model generalizes to new spatial domains and viewing angles. This evaluation uses a subset of the *GOES-17* full-disk sector for the 5th–9th of each month of 2020.

Using the best-performing probability threshold from the validation set (35%), we see that the model has good performance in certain regions, but less so in others. In particular, from

[2] The three-channel model applied to the *GOES-16* validation dataset decreased maximum CSI by 0.011.

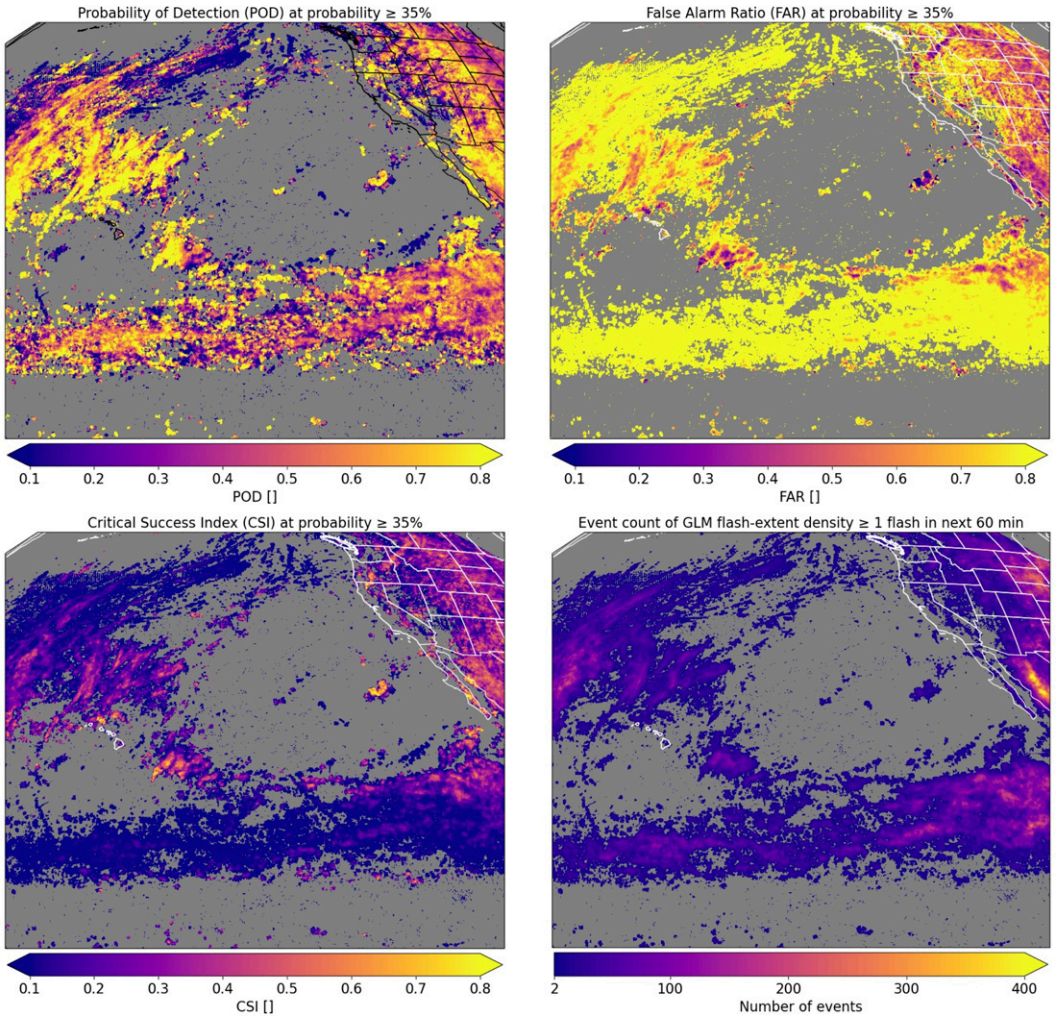

FIG. 14. The (top left) POD, (top right) FAR, (bottom left) CSI, and (bottom right) event count for the *GOES-17* test set (full-disk subset), using a three-channel version of LightningCast.

Fig. 14, the CSI is between 0.3 and 0.5 in the vicinity of Hawaii, the western United States, and parts of the northern belt of convection along the intertropical convergence zone (ITCZ). This is generally comparable to the *GOES-16* performance in many locations. However, low CSI ($\leq$0.2) is apparent in much of the ITCZ belt of convection and in the northern Pacific Ocean, owing to high FAR ($\geq$0.8). Part of this reduced performance may be due to the exclusion of the 12.3-μm brightness temperature, but we attribute much of it to clouds in meteorological regimes very different from the training and validation data, as well as potentially missed detections of infrequent lightning from GLM [see discussion in section 3b(1)]. The *GOES-17* domain contains mostly oceanic regions, and while oceanic storms have less vigor than their continental counterparts (e.g., Lucas et al. 1994; Solomon and Baker 1998), the *GOES-16* test set exhibited good performance in many oceanic regions. In the North Pacific, differences could be partially attributable to differences in the oceanic currents. For example, the more energetic Gulf Stream of the Atlantic may be able to provide better forcing for convection that produces lightning, compared to the North Pacific Drift current. Climatological observations reveal the fact that storms in the central Pacific Ocean exhibit much more intermittent and less frequent lightning than those in much of the western Atlantic Ocean (e.g., Orville and Henderson 1986; Cecil et al. 2014).

In the tropics, it is well established that convective updrafts are generally weaker over oceans than over land due to a number of reasons, but chiefly "skinnier" CAPE vertical profiles (e.g., LeMone and Zipser 1980; Lucas et al. 1994), leading to less lightning over ocean (e.g., Cecil et al. 2005; Liu and Zipser 2005; Liu et al. 2012). However, tropical maritime updrafts are still deep enough to develop cumulonimbus clouds to the anvil stage, with cold cloud tops. Tall, cold clouds that produce infrequent or no lightning appear to cause overprediction of lightning in the deep tropics. Training a model using *GOES-17* data may improve performance, but more analysis is needed to determine if additional training samples from this domain alone will help, or if additional predictors will increase

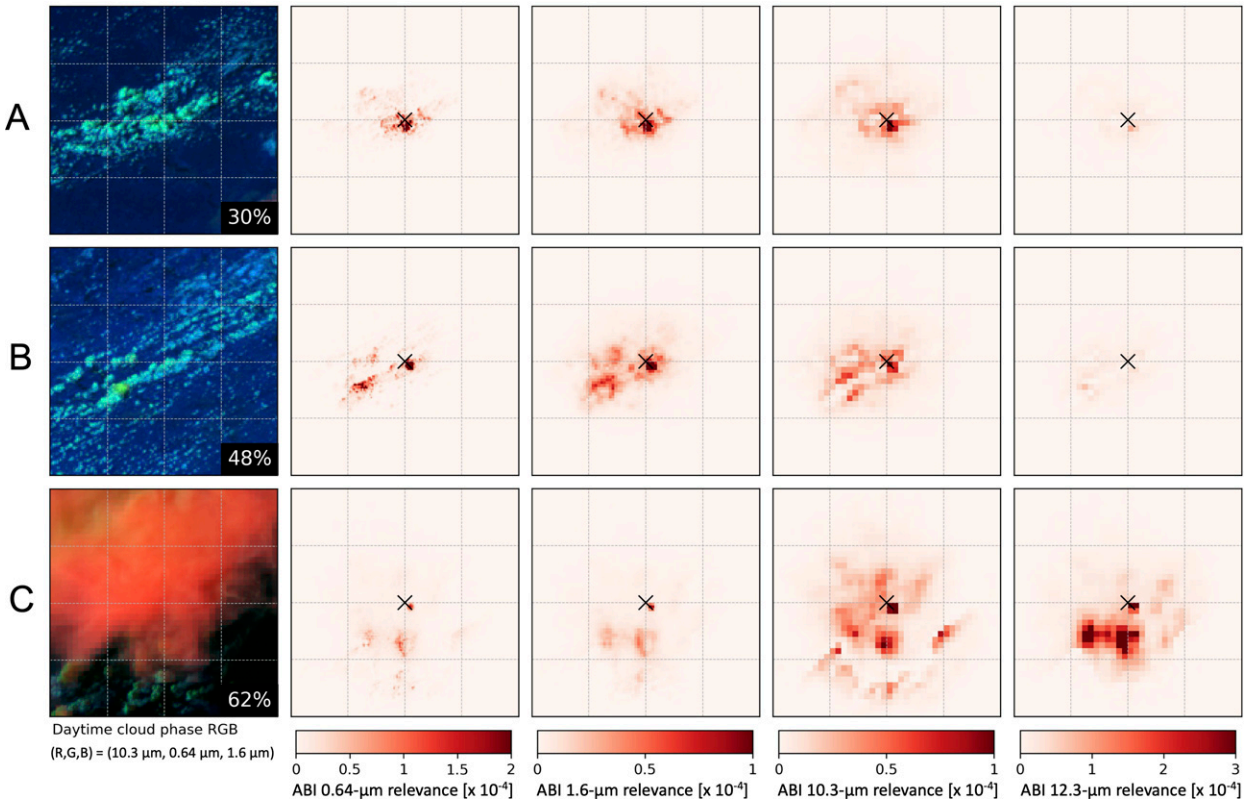

FIG. 15. Relevance plots for the input channels of LightningCast for three examples. The first column shows the daytime cloud-phase distinction RGB for the three examples, with the predicted probability in the lower-right corner. That probability is valid for the pixel immediately to the southeast of the black "X" found in each relevance plot (remaining columns, from left to right). Sample A is from northeast Iowa at 1911 UTC 28 Aug 2020 (see Fig. 4c), sample B is from northeast Indiana at 1656 UTC 14 Sep 2021 (see Fig. 5b), and sample C is from eastern Mississippi at 2251 UTC 13 Aug 2021 (see Fig. 7b).

the performance in these regions. Regardless, this analysis demonstrates some success in portability of LightningCast to different domains, and end users of *GOES-17* data can still benefit from the three-channel LightningCast model.

d. Prediction attribution and relevance

This section explores the attribution of LightningCast predictions for several of the image times shown in Figs. 4, 5, and 7. We employ a method called layer-wise relevance propagation (LRP; Alber et al. 2019), to find the "relevance" or importance of each pixel in each predictor for a given prediction made by LightningCast. LRP is a process that aims to track relevance backward from an output neuron (i.e., pixel) to each channel/predictor of the input image. The process progressively tracks backward in the network, finding which neurons in the prior layer were most responsible for the values of a neuron in the later layer [see Alber et al. (2019), Montavon et al. (2019), and Toms et al. (2020) for details and implementation of LRP]. We used an implementation of the LRP-Z Plus rule (Bach et al. 2015), which approximates positive attribution for a prediction at a given pixel, locating regions in the input images that contributed to the predicted probability.[3]

The three samples in Fig. 15 can be found in the examples presented previously—row A of Fig. 15 is from northeast Iowa (see Fig. 4c), row B is from northeast Indiana (see Fig. 5b), and row C is from eastern Mississippi (see Fig. 7b). The pixel of interest in each row is marked by a black "X" in the relevance columns (second through fifth columns), while the predicted probability of lightning is found in the lower right corner of the first column, given as a percentage. The first columns show the daytime cloud-phase distinction RGB for each sample.

In samples A and B, several features stand out as important for the predicted probability of lightning. Pixels with high 0.64-μm reflectance in cumulus clouds near the pixel of interest contribute to the probability. These clouds also appear to be either glaciated or beginning to glaciate (notice the greenish tint of the clouds in the RGBs). The spatial gradients of the 1.6-μm reflectance channel around cumulus clouds are also flagged as important. These gradients can be most easily seen in sample B, to the north and east of the glaciated cloud

[3] Please see appendix B in Hilburn et al. (2021) for further details on implementing LRP for a U-Net model.

that is southwest of the pixel of interest. Close inspection of the 1.6-μm relevance in samples A and B reveals other areas where the spatial gradient of the reflectance is important. The spatial gradient of the 10.3-μm brightness temperature is also found to be highly relevant. Evident in both samples A and B, but particularly in sample B, strong brightness temperature gradients around the main line of cumulus clouds are highlighted as very relevant. In both samples, high relevance is also found in pixels very close to the pixel of interest, exhibiting high 0.64-μm reflectance in clouds that appear to be almost glaciating. The 12.3-μm brightness temperature in both samples A and B appears to have low relevance, compared to the other three channels.

Sample C is very different from samples A and B, with anvil clouds enveloping the pixel of interest. Relevance for the two reflective channels (0.64 and 1.6 μm) is more muted compared to that of samples A and B. However, cold, glaciated clouds at the pixel of interest are still deemed relevant, as well as some cumulus clouds underneath the thick ice, to the south of the pixel of interest. The 10.3-μm relevance also highlights this feature, as well as the cold clouds in the general vicinity of the pixel of interest. Several features to the south are relevant to a lesser extent, including some brightness-temperature gradients around cumulus clouds. Most interesting about sample C is the very high relevance found in the 12.3-μm brightness temperature. In a scene where the influence of the reflective bands is subdued, the 12.3-μm brightness temperature appears to be greatly contributing to the probability. The main features highlighted to the south of the pixel of interest in the 12.3-μm relevance plot appear to be cumulus clouds (or gradients around them) beneath the thick ice of the anvil.

The practice of pixel-level attribution of artificial intelligence models is still maturing and there is no standard set of rules or methods for determining relevance or channel importance in such models. While different interpretation methods may show some different results, we believe the analysis on the three aforementioned cases shows: 1) reflective, glaciating clouds are important for increasing the probability of lightning; and 2) strong spatial gradients around clouds are important for lightning probability, possibly indicating key spatial patterns in typical scenes of lightning production (e.g., glaciating cumulus fields). Further investigation is needed to determine additional patterns or modes in satellite imager data that are relevant to lightning production. Nevertheless, the complex spectral and spatial patterns that are relevant to nowcasting lightning would be difficult to capture without the use of deep learning designed for computer vision problems.

4. Summary and discussion

We developed a convolutional neural network, LightningCast, that uses GOES-R ABI channels to predict the probability of lightning in the next 60 min for every Earth-located pixel in a given satellite domain. Using the 0.64-μm reflectance (0.5-km resolution), the 1.6-μm reflectance (1-km resolution), the 10.3-μm brightness temperature (2-km resolution), and the 12.3-μm brightness temperature (2-km resolution) as predictors and next-hour GLM flash-extent density as the truth or target field, LightningCast was found to make skillful predictions (CSI ~ 0.4) day and night, over land and sea, for both advecting and developing storms. For a seasonally balanced *GOES-16* test dataset covering the contiguous United States, Mexico, northern South America, and adjacent oceanic regions, the model exhibited very consistent performance for the months of April through September, when lightning was most prevalent, with degraded, but useful skill in other months. Daytime predictions were slightly more accurate than nighttime predictions, while continental predictions were generally more accurate than maritime predictions. To test the limits of generalization, we created a second test set using *GOES-17* data over the western United States and eastern Pacific Ocean. The *GOES-17* LightningCast critical success index (CSI) over the western United States was comparable to *GOES-16*, but was relatively low over the Pacific Ocean, where lightning is less common. Further investigation is needed to determine if geographic and seasonal performance gaps could be addressed with additional training or additional model input features. Regardless, LightningCast performs well in many locations and meteorological regimes, even outside of the domain of the training dataset.

When isolating new convection in localized regions, we found that for half of true-positive forecasts, LightningCast yields nearly 20 min or more of lead time to the initial flash at a location (as observed by GLM) relative to the most skillful probability thresholds (30%–40%). LightningCast can complement data from ground-based radar networks (when available) by extending lead time and increasing forecaster situational awareness. LightningCast helps to quantify the lightning potential information intrinsic in the qualitative multispectral satellite imagery, representing a truly new capability for lightning nowcasting without a reliance on radar.

There are a number of direct applications of LightningCast. The U.S. National Weather Service provides decision-support services. These services include providing rapidly updating lightning threat guidance for outdoor events. LightningCast could be used directly by forecasters to assist in providing objective guidance to these users in a timely manner. LightningCast could aid forecasters at NOAA's Ocean Prediction Center or Aviation Weather Center as they issue guidance and routine products for mariners and aircraft to avoid convection. Model output could potentially be used to give emergency managers early guidance for lightning potential in fire-prone areas, or help forecasters determine the timing and location of convective initiation.

At the time of this publication, LightningCast is running in near-real time at the University of Wisconsin–Madison for several *GOES-16* and *GOES-17* sectors.[4] LightningCast will also be evaluated in several NWS testbeds during 2022 to test its applicability in NWS operations. While a satellite-inputs-only model may always be useful because of the large domain it can cover, there a number of ways to derive more tailored models, which we hope to explore. These include using NWP data and Multi-Radar Multi-Sensor data to both extend lead

[4] Near-real-time output is available at https://cimss.ssec.wisc.edu/severe_conv/pltg.html.

time and to provide more precise spatial forecasts of lightning. Using multiple time frames of data may improve the skill of the model (e.g., Zhou et al. 2022), but at the cost of a more expensive training phase due to increased model complexity and training multiple models for different temporal resolutions of satellite images (e.g., 1-, 5-min, etc.). Simply using the current lightning observations to predict the next-hour lightning could improve performance; however, doing so may also decrease lead time to the first flash, which would be very undesirable, and thus should be carefully examined. While a nowcast of one flash or more is important from a lightning initiation standpoint, several operational forecasters have noted that forecasts of more substantial lightning [e.g., 10 or 20 flashes per 5 min] would also be valuable.

Since LightningCast only depends on one snapshot in time, it can be applied to nonpermanent scan sectors, such as the 1-min mesoscale sectors from ABI. In practice, LightningCast takes 1–3 s to compute probabilities for the mesoscale sectors and 20–30 s for the CONUS sector, using modest CPU-based computing. The imager aboard Japan's Himawari satellites (AHI) is very similar to ABI. Furthermore, Europe's third-generation geostationary satellites, Meteosat Third Generation (MTG), will also provide imagery similar to the ABI. Both satellites present an opportunity to test the applicability of LightningCast on similar instruments but very different domains. While some retraining may be needed to optimize performance, near global lightning nowcasts are likely possible.

Acknowledgments. Funding for this work was provided by the NOAA/NESDIS GOES-R Program Office. The authors would like to acknowledge Satellite Data Services at the University of Wisconsin–Madison for aid in staging and accessing GOES-R ABI and GLM data, as well as Kyle Hilburn of the Cooperative Institute for Research in the Atmosphere, for guidance in using the iNNvestigate API for the LightningCast model. The authors thank Derrick Herndon of the University of Wisconsin–Madison, for productive conversations regarding the model's performance. Lastly, the authors thank three anonymous reviewers for constructive reviews that improved the quality of this paper. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect those of NOAA or the Department of Commerce.

Data availability statement. The *GOES-16* and *GOES-17* data used in this study can be freely obtained from NOAA's Comprehensive Large Array Data Stewardship System (CLASS; online at https://www.class.noaa.gov/).

APPENDIX

Model Training Details

The LightningCast model is a convolutional neural network described in section 2c and Fig. 3. We provide further model details here. We used the Adam optimizer, which is a popular stochastic gradient descent method based on adaptive estimation of first-order and second-order moments (Kingma and Ba 2014). The minibatch size was 1, since that was the maximum number of samples or patches that could be stored in our GPU memory. The samples were randomly shuffled before each epoch. We normalized the input data by subtracting the mean and dividing by the standard deviation on a per-channel basis, which were found a priori on a subset of the full dataset. We trained for 15 epochs, stopping early when validation loss did not increase after 4 epochs. Thus, the epoch with the lowest validation loss was epoch 11, which was the checkpoint model used in this paper (a model was saved after each epoch if validation loss was a new minimum). The initial learning rate was 0.0003, decreasing by a factor of 10 after each epoch where validation loss did not decrease. All feature maps were zero-padded such that the size of the output feature maps was the same as the size of the input feature maps.

REFERENCES

Ahijevych, D., J. O. Pinto, J. K. Williams, and M. Steiner, 2016: Probabilistic forecasts of mesoscale convective system initiation using the random forest data mining technique. *Wea. Forecasting*, **31**, 581–599, https://doi.org/10.1175/WAF-D-15-0113.1.

Alber, M., and Coauthors, 2019: iNNvestigate neural networks! *J. Mach. Learn. Res.*, **20**, 1–8.

Bach, S., A. Binder, G. Montavon, F. Klauschen, K. R. Muller, and W. Samek, 2015: On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLOS ONE*, **10**, e0130140, https://doi.org/10.1371/journal.pone.0130140.

Bateman, M., and D. Mach, 2020: Preliminary detection efficiency and false alarm rate assessment of the geostationary lightning mapper on the *GOES-16* satellite. *J. Appl. Remote Sens.*, **14**, 032406, https://doi.org/10.1117/1.JRS.14.032406.

Bruning, E. C., 2021: glmtools. GitHub, accessed 15 May 2021, https://github.com/deeplycloudy/glmtools.

——, and Coauthors, 2019: Meteorological imagery for the geostationary lightning mapper. *J. Geophys. Res. Atmos.*, **124**, 14 285–14 309, https://doi.org/10.1029/2019JD030874.

Cecil, D. J., S. J. Goodman, D. J. Boccippio, E. J. Zipser, and S. W. Nesbitt, 2005: Three years of TRMM precipitation features. Part I: Radar, radiometric, and lightning characteristics. *Mon. Wea. Rev.*, **133**, 543–566, https://doi.org/10.1175/MWR-2876.1.

——, D. E. Buechler, and R. J. Blakeslee, 2014: Gridded lightning climatology from TRMM-LIS and OTD: Dataset description. *Atmos. Res.*, **135**, 404–414, https://doi.org/10.1016/j.atmosres.2012.06.028.

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, L. Cronce, and J. Brunner, 2020a: NOAA ProbSevere v2.0-ProbHail, ProbWind, and ProbTor. *Wea. Forecasting*, **35**, 1523–1543, https://doi.org/10.1175/WAF-D-19-0242.1.

——, ——, ——, A. Wimmers, J. Brunner, and W. Bellon, 2020b: A deep-learning model for automated detection of intense midlatitude convection using geostationary satellite images. *Wea. Forecasting*, **35**, 2567–2588, https://doi.org/10.1175/WAF-D-20-0028.1.

Connell, B., E. Dagg, and M. Bowlan, 2021: Day cloud phase distinction RGB quick guide. NOAA/NASA, 2 pp., https://rammb.cira.colostate.edu/training/visit/quick_guides/QuickGuide_DayCloudPhaseDistinction_final_v2.pdf.

Cuomo, J., and V. Chandrasekar, 2021: Use of deep learning for weather radar nowcasting. *J. Atmos. Oceanic Technol.*, **38**, 1641–1656, https://doi.org/10.1175/JTECH-D-21-0012.1.

Elsenheimer, C. B., and C. M. Gravelle, 2019: Introducing lightning threat messaging using the *GOES-16* day cloud phase distinction RGB composite. *Wea. Forecasting*, **34**, 1587–1600, https://doi.org/10.1175/WAF-D-19-0049.1.

Fierro, A. O., E. R. Mansell, C. L. Ziegler, and D. R. MacGorman, 2012: Application of a lightning data assimilation technique in the WRF-ARW model at cloud-resolving scales for the tornado outbreak of 24 May 2011. *Mon. Wea. Rev.*, **140**, 2609–2627, https://doi.org/10.1175/MWR-D-11-00299.1.

——, A. J. Clark, E. R. Mansell, D. R. MacGorman, S. R. Dembek, and C. L. Ziegler, 2015: Impact of storm-scale lightning data assimilation on WRF-ARW precipitation forecasts during the 2013 warm season over the contiguous United States. *Mon. Wea. Rev.*, **143**, 757–777, https://doi.org/10.1175/MWR-D-14-00183.1.

Geng, Y. A., and Coauthors, 2021: A deep learning framework for lightning forecasting with multi-source spatiotemporal data. *Quart. J. Roy. Meteor. Soc.*, **147**, 4048–4062, https://doi.org/10.1002/qj.4167.

Goodman, S. J., and Coauthors, 2013: The GOES-R geostationary lightning mapper (GLM). *Atmos. Res.*, **125–126**, 34–49, https://doi.org/10.1016/j.atmosres.2013.01.006.

Hilburn, K. A., I. Ebert-Uphoff, and S. D. Miller, 2021: Development and interpretation of a neural-network-based synthetic radar reflectivity estimator using GOES-R satellite observations. *J. Appl. Meteor. Climatol.*, **60**, 3–21, https://doi.org/10.1175/JAMC-D-20-0084.1.

Hondl, K. D., and M. D. Eilts, 1994: Doppler radar signatures of developing thunderstorms and their potential to indicate the onset of cloud-to-ground lightning. *Mon. Wea. Rev.*, **122**, 1818–1836, https://doi.org/10.1175/1520-0493(1994)122<1818:DRSODT>2.0.CO;2.

Karagiannidis, A., K. Lagouvardos, and V. Kotroni, 2016: The use of lightning data and Meteosat infrared imagery for the nowcasting of lightning activity. *Atmos. Res.*, **168**, 57–69, https://doi.org/10.1016/j.atmosres.2015.08.011.

Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. arXiv, 1412.6980, https://arxiv.org/pdf/1412.6980v9.pdf.

Kumjian, M. R., 2013: Principles and applications of dual-polarization weather radar. Part 1: Descriptions of the polarimetric radar variables. *J. Oper. Meteor.*, **1**, 226–242, https://doi.org/10.15191/nwajom.2013.0119.

La Fata, A., F. Amato, M. Bernardi, M. D'Andrea, R. Procopio, and E. Fiori, 2021: Cloud-to-ground lightning nowcasting using machine learning. *35th Int. Conf. on Lightning Protection (ICLP) and XVI Int. Symp. on Lightning Protection (SIPDA)*, Vol. 1, Colombo, Sri Lanka, IEEE, 1–6, https://doi.org/10.1109/ICLPandSIPDA54065.2021.9627428.

Lagerquist, R., A. McGovern, and D. J. Gagne, 2019: Deep learning for spatially explicit prediction of synoptic-scale fronts. *Wea. Forecasting*, **34**, 1137–1160, https://doi.org/10.1175/WAF-D-18-0183.1.

——, ——, C. R. Homeyer, D. J. Gagne, and T. Smith, 2020: Deep learning on three-dimensional multiscale data for next-hour tornado prediction. *Mon. Wea. Rev.*, **148**, 2837–2861, https://doi.org/10.1175/MWR-D-19-0372.1.

LeMone, M. A., and E. J. Zipser, 1980: Cumulonimbus vertical velocity events in GATE. 1. Diameter, intensity and mass flux. *J. Atmos. Sci.*, **37**, 2444–2457, https://doi.org/10.1175/1520-0469(1980)037<2444:CVVEIG>2.0.CO;2.

Liu, C. T., and E. J. Zipser, 2005: Global distribution of convection penetrating the tropical tropopause. *J. Geophys. Res.*, **110**, D23104, https://doi.org/10.1029/2005JD006063.

——, D. J. Cecil, E. J. Zipser, K. Kronfeld, and R. Robertson, 2012: Relationships between lightning flash rates and radar reflectivity vertical structures in thunderstorms over the tropics and subtropics. *J. Geophys. Res.*, **117**, D062129, https://doi.org/10.1029/2011JD017123.

Lucas, C., E. J. Zipser, and M. A. Lemone, 1994: Vertical velocity in oceanic convection off tropical Australia. *J. Atmos. Sci.*, **51**, 3183–3193, https://doi.org/10.1175/1520-0469(1994)051<3183:VVIOCO>2.0.CO;2.

Lynn, B. H., Y. Yair, C. Price, G. Kelman, and A. J. Clark, 2012: Predicting cloud-to-ground and intracloud lightning in weather forecast models. *Wea. Forecasting*, **27**, 1470–1488, https://doi.org/10.1175/WAF-D-11-00144.1.

MacGorman, D. R., J. M. Straka, and C. L. Ziegler, 2001: A lightning parameterization for numerical cloud models. *J. Appl. Meteor.*, **40**, 459–478, https://doi.org/10.1175/1520-0450(2001)040<0459:ALPFNC>2.0.CO;2.

McCaul, E. W., S. J. Goodman, K. M. LaCasse, and D. J. Cecil, 2009: Forecasting lightning threat using cloud-resolving model simulations. *Wea. Forecasting*, **24**, 709–729, https://doi.org/10.1175/2008WAF2222152.1.

Mecikalski, J. R., X. L. Li, L. D. Carey, E. W. McCaul, and T. A. Coleman, 2013: Regional comparison of GOES cloud-top properties and radar characteristics in advance of first-flash lightning initiation. *Mon. Wea. Rev.*, **141**, 55–74, https://doi.org/10.1175/MWR-D-12-00120.1.

——, J. K. Williams, C. P. Jewett, D. Ahijevych, A. Leroy, and J. R. Walker, 2015: Probabilistic 0-1-h convective initiation nowcasts that combine geostationary satellite observations and numerical weather prediction model data. *J. Appl. Meteor. Climatol.*, **54**, 1039–1059, https://doi.org/10.1175/JAMC-D-14-0129.1.

Montavon, G., A. Binder, S. Lapuschkin, W. Samek, and K. Müller, 2019: Layer-wise relevance propagation: An overview. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, Lecture Notes in Computer Science, W. Samek et al., Eds., Vol. 11700, Springer, 193–209.

Mosier, R. M., C. Schumacher, R. E. Orville, and L. D. Carey, 2011: Radar nowcasting of cloud-to-ground lightning over Houston, Texas. *Wea. Forecasting*, **26**, 199–212, https://doi.org/10.1175/2010WAF2222431.1.

Nair, V., and G. Hinton, 2010: Rectified linear units improve restricted Boltzmann machines. *Proc. 27th Int. Conf. on Machine Learning*, Haifa, Israel, International Machine Learning Society, 807–814.

NOAA/NESDIS, 2021: *GOES-17* loop heat pipe fact sheet. NOAA/NESDIS, 1 p., https://www.goes-r.gov/imagesContent/newsItems/NESDISfactsheetLoopheatpipe.pdf.

Orville, R. E., and R. W. Henderson, 1986: Global distribution of midnight lightning - September 1977 to August 1978. *Mon. Wea. Rev.*, **114**, 2640–2653, https://doi.org/10.1175/1520-0493(1986)114<2640:GDOMLS>2.0.CO;2.

Pavolonis, M. J., 2010: 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, https://doi.org/10.1175/2010JAMC2433.1.

Ravuri, S., and Coauthors, 2021: Skilful precipitation nowcasting using deep generative models of radar. *Nature*, **597**, 672–677, https://doi.org/10.1038/s41586-021-03854-z.

Ren, M. Y., W. Y. Zeng, B. Yang, and R. Urtasun, 2018: Learning to reweight examples for robust deep learning. *Proc. 35th Int. Conf. on Machine Learning (ICML)*, Vol. 80, Stockholm, Sweden, PMLR, 4334–4343.

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Lecture Notes in Computer Science, N. Navab et al., Eds., Vol. 9351, Springer, 234–241.

Rudlosky, S. D., S. J. Goodman, K. S. Virts, and E. C. Bruning, 2019: Initial Geostationary Lightning Mapper observations. *Geophys. Res. Lett.*, **46**, 1097–1104, https://doi.org/10.1029/2018GL081052.

Rutledge, S. A., K. A. Hilburn, A. Clayton, B. Fuchs, and S. D. Miller, 2020: Evaluating Geostationary Lightning Mapper flash rates within intense convective storms. *J. Geophys. Res. Atmos.*, **125**, e2020JD032827, https://doi.org/10.1029/2020JD032827.

Saunders, C. P. R., E. E. Avila, S. L. Peck, N. E. Castellano, and G. G. A. Varela, 1999: A laboratory study of the effects of rime ice accretion and heating on charge transfer during ice crystal graupel collisions. *Atmos. Res.*, **51**, 99–117, https://doi.org/10.1016/S0169-8095(99)00023-X.

Schmit, T. J., P. Griffith, M. M. Gunshor, J. M. Daniels, S. J. Goodman, and W. J. Lebair, 2017: A closer look at the ABI on the GOES-R series. *Bull. Amer. Meteor. Soc.*, **98**, 681–698, https://doi.org/10.1175/BAMS-D-15-00230.1.

Seroka, G. N., R. E. Orville, and C. Schumacher, 2012: Radar nowcasting of total lightning over the Kennedy Space Center. *Wea. Forecasting*, **27**, 189–204, https://doi.org/10.1175/WAF-D-11-00035.1.

Shackford, C. R., 1960: Radar indications of a precipitation-lightning relationship in New-England thunderstorms. *J. Meteor.*, **17**, 15–19, https://doi.org/10.1175/1520-0469(1960)017<0015:RIOAPL>2.0.CO;2.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

Solomon, R., and M. Baker, 1998: Lightning flash rate and type in convective storms. *J. Geophys. Res.*, **103**, 14 041–14 057, https://doi.org/10.1029/97JD03323.

Takahashi, T., 1978: Riming electrification as a charge generation mechanism in thunderstorms. *J. Atmos. Sci.*, **35**, 1536–1548, https://doi.org/10.1175/1520-0469(1978)035<1536:REAACG>2.0.CO;2.

Thornton, J. A., K. S. Virts, R. H. Holzworth, and T. P. Mitchell, 2017: Lightning enhancement over major oceanic shipping lanes. *Geophys. Res. Lett.*, **44**, 9102–9111, https://doi.org/10.1002/2017GL074982.

Toms, B. A., E. A. Barnes, and I. Ebert-Uphoff, 2020: Physically interpretable neural networks for the geosciences: Applications to Earth system variability. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002002, https://doi.org/10.1029/2019MS002002.

Vincent, B. R., L. D. Carey, D. Schneider, K. Keeter, and R. Gonski, 2003: Using WSR-88D reflectivity for the prediction of cloud-to-ground lightning: A North Carolina study. *Natl. Wea. Dig.*, **27**, 35–44.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.

Wimmers, A., C. Velden, and J. H. Cossuth, 2019: Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery. *Mon. Wea. Rev.*, **147**, 2261–2282, https://doi.org/10.1175/MWR-D-18-0391.1.

Woodward, C. J., L. D. Carey, W. A. Peterson, and W. P. Roeder, 2012: Operational utility of dual-polarization variables in lightning initiation forecasting. *Electron. J. Oper. Meteor.*, **13**, 79–102, http://nwafiles.nwas.org/ej/pdf/2012-EJ6.pdf.

Zhou, K. H., Y. G. Zheng, W. S. Dong, and T. B. Wang, 2020: A deep learning network for cloud-to-ground lightning nowcasting with multisource data. *J. Atmos. Oceanic Technol.*, **37**, 927–942, https://doi.org/10.1175/JTECH-D-19-0146.1.

Zhou, X. Y., Y. Geng, H. Yu, Q. Li, L. Xu, W. Yao, D. Zheng, and Y. Zhang, 2022: LightNet+: A dual-source lightning forecasting network with bi-direction spatiotemporal transformation. *Appl. Intell.*, in press, https://doi.org/10.1007/s10489-021-03089-5.