

Deep Learning–Based Summertime Turbulence Intensity Estimation Using Satellite Observations

YOONJIN LEE¹,^a SOO-HYUN KIM,^b YOO-JEONG NOH,^a AND JUNG-HOON KIM^b

^a *Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado*

^b *School of Earth and Environmental Sciences, Seoul National University, Seoul, South Korea*

(Manuscript received 5 December 2022, in final form 30 May 2023, accepted 1 September 2023)

ABSTRACT: Turbulence is what we want to avoid the most during flight. Numerical weather prediction (NWP) model-based methods for diagnosing turbulence have offered valuable guidance for pilots. NWP-based turbulence diagnostics show high accuracy in detecting turbulence in general. However, there is still room for improvements such as capturing convectively induced turbulence. In such cases, observation data can be beneficial to correctly locate convective regions and help provide corresponding turbulence information. Geostationary satellite data are commonly used for upper-level turbulence detection by utilizing its water vapor band information. The *Geostationary Operational Environmental Satellite (GOES)-16* carries the Advanced Baseline Imager (ABI), which enables us to observe further down into the atmosphere with improved spatial, temporal, and spectral resolutions. Its three water vapor bands allow us to observe different vertical parts of the atmosphere, and from its infrared window bands, convective activity can be inferred. Such multispectral information from ABI can be helpful in inferring turbulence intensity at different vertical levels. This study develops U-Net based machine learning models that take ABI imagery as inputs to estimate turbulence intensity at three vertical levels: 10–18, 18–24, and above 24 kft (1 kft \approx 300 m). Among six different U-Net-based models, U-Net3+ model with a filter size of three showed the best performance against the pilot report (PIREP). Two case studies are presented to show the strengths and weaknesses of the U-Net3+ model. The results tend to be overestimated above 24 kft, but estimates of 10–18 and 18–24 kft agree well with the PIREP, especially near convective regions.

SIGNIFICANCE STATEMENT: Turbulence is directly related to aviation safety as well as cost-effective aircraft operation. To avoid turbulence, turbulence diagnostics are calculated from numerical weather prediction (NWP) model outputs and are provided to pilots. The goal of this study is to develop a satellite data–driven machine learning model that estimates turbulence intensity in three different vertical layers to provide additional information along with the NWP-based turbulence diagnostics. Validation results against pilot reports show that the machine learning model performs comparable to NWP-based turbulence diagnostics. Furthermore, results with different channel selections reveal that using multiple water vapor channels can help extract additional information for estimating turbulence intensity at lower levels.

KEYWORDS: Turbulence; Satellite observations; Artificial intelligence; Machine learning; Neural networks

1. Introduction

Atmospheric turbulence is one of the most important factors in aviation safety along with icing. The use of reliable information of turbulence location helps prevent accidents with possible injuries as well as passenger comfort and cost-effective aircraft operations (Sharman and Lane 2016). Turbulence can be generated from various sources: clear-air turbulence (CAT), mountain-wave turbulence, convectively induced turbulence (CIT), and low-level turbulence (Sharman and Lane 2016). Although computational resources have been improved, there are still limitations to explicitly predict turbulence. In this regard, various numerical weather prediction (NWP)-based turbulence diagnostics have been developed based on the assumption of downscale cascade process (e.g., Sharman et al. 2006; Sharman and Pearson 2017; Kim et al. 2019, 2021). Current operational turbulence

forecasting methods, Graphical Turbulence Guidance (GTG; Sharman et al. 2006; Sharman and Pearson 2017; Pearson and Sharman 2017) and Korean Turbulence Guidance (KTG; Kim and Chun 2012; Lee et al. 2022), estimate turbulence potential using multiple NWP-based turbulence diagnostics. Recently machine learning (ML) techniques have also been applied for turbulence forecasting (Muñoz-Esparza et al. 2020). As NWP models have been rapidly advanced in both accuracy and spatial resolution, turbulence predictions that highly depend on NWP model outputs have improved accordingly in recent years. However, NWP models still have spatial and temporal mismatches of weather systems, and they can lead to inaccurate turbulence estimation. In particular, summertime convection is hard to simulate in NWP models both spatially and temporally, and the accuracy of its turbulence estimates will be largely affected by that. In such cases, satellite observation can help correctly locate the weather system and infer corresponding turbulence intensity.

Remote detection of turbulence has been conducted with airborne radar, radiosonde, in situ measurements, and satellites. Satellite observations have offered global information on aviation weather conditions including turbulence. In particular, geostationary satellite (GEO) data have practical uses for aviation

¹ Denotes content that is immediately available upon publication as open access.

Corresponding author: Yoonjin Lee, yoonjin.lee@colostate.edu

operations with consistent time and wide spatial coverage, and especially their water vapor channels are commonly used to estimate CAT in the upper atmosphere (Ellrod 1989). One of the notable features of CAT that can be inferred from GEO imagery is tropopause folding. Tropopause folding is an intrusion of stratospheric air into the troposphere induced in regions with high baroclinicity. Strong gradients in water vapor imagery are a good indicator of the tropopause folding (Wimmers and Moody 2004). The water vapor channel traditionally with center wavelengths around 6–6.5 μm is also essential in detecting jet stream, gravity waves, frontal system, and upper troughs that are features associated with CAT (Ellrod and Pryor 2019). Longwave infrared window band is also useful in detecting convection which is often associated with severe turbulence, and the Federal Aviation Administration (FAA) recommends pilots avoid and remain 20 mi (32.2 km horizontally) away from severe convection regions (Federal Aviation Administration 2017). Features of convective clouds that are observed from GEOs are decreases in brightness temperature (Mecikalski et al. 2010; Sieglaff et al. 2011; Monette and Sieglaff 2014), associated transverse cirrus bands (Lenz et al. 2009), enhanced-V signature (Brunner et al. 2007), or overshooting tops (Bedka et al. 2010). Since turbulence diagnostics used in current turbulence forecasting systems are mostly related to jet streams or fronts and not as effective in predicting convection-related turbulence, any information on convective regions obtained from GEOs with high temporal resolution can be beneficial to the aviation community by contributing to improved predictions of CIT.

The next-generation GEO satellites such as Japanese *Himawari-8* Advanced Himawari Imager (AHI; Bessho et al. 2016), U.S. National Oceanic and Atmospheric Administration (NOAA) Geostationary Operational Environmental Satellite (GOES-R) series carrying the Advanced Baseline Imager (ABI; Schmit et al. 2017), Chinese Fengyun-4 series (FY-4) with the Advanced Geosynchronous Radiation Imager (AGRI; Yang et al. 2017), and South Korean *Geostationary Korea Multi-Purpose Satellite-2A* (*GEO-KOMPSAT-2A* or *GK-2A*) with the Advanced Meteorological Imager (AMI; D. Kim et al. 2021) have provided global cloud observations in very high spatial and temporal resolutions. Recently launched European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT)'s *Meteosat Third Generation* (*MTG-II* or *Meteosat-12*) with a Flexible Combined Imager (FCI; Just et al. 2014) also joined the global GEO constellation. But the information from traditional passive radiometers based on visible and infrared channels is mostly biased toward the cloud top due to natural limitations of the wavelengths and is still known to be inefficient to derive the vertical structure of the atmosphere. However, recent studies show the potential to use GEO data to get more information below cloud top. With increased channel numbers in GEOs and with the help of ML methods, innovative approaches have been introduced in low cloud detection (Haynes et al. 2022), detecting severe weather systems (Hilburn et al. 2021; Lagerquist et al. 2021; Lee et al. 2021), or precipitation estimation (Hayatbini et al. 2019).

Inferring turbulence from GEO observations has been mostly focused on detecting certain features in the upper

atmosphere, i.e., tropopause folding, overshooting tops, or transverse cirrus bands. In this study, we attempt to develop an ML model that estimates summertime turbulence intensity in different vertical layers by taking advantage of both advanced GEO data and high-resolution NWP model output which has been operationally improved over decades.

In this study, inputs of the developed ML models are five brightness temperatures of *GOES-16* ABI channels 8, 9, 10, 11, and 13, and terrain heights, and outputs of the ML models are NWP-based turbulence diagnostics in terms of the cube root of eddy dissipation rate (EDR). As for the output data, the hourly outputs from the High Resolution Rapid Refresh (HRRR) model (Dowell et al. 2022), which is NOAA's operational model over the contiguous United States (CONUS), are used to calculate turbulence diagnostics in 3-km horizontal resolution. During the training, however, synthetic brightness temperatures are used as inputs instead of observed *GOES-16* ABI brightness temperatures to avoid feeding wrong information to the ML model. Although current NWP models have improved significantly, there can be discrepancies between observed and simulated weather. If observed brightness temperatures were used during training, and the HRRR model, which is used to compute turbulence data, may not simulate exactly the same scene, it can mislead the ML model to learn wrong relationships between brightness temperatures and turbulence estimates, and can be detrimental during training. To prevent this, the HRRR-based multiturbulence diagnostics and HRRR-based synthetic brightness temperature are used for training ML models, while once it is trained, the actual ABI brightness temperatures are used to get turbulence estimates. Synthetic brightness temperatures are simulated using the Community Radiative Transfer Model (CRTM) with the HRRR model outputs. Since this is an image-to-image translation problem, where inputs and outputs are both images, U-Net-based models are used. Six U-Net models using different filter sizes are trained and the results are compared between the models, and experiments using different ABI channel selections are conducted.

The goal of this study is to develop a satellite data-driven ML model as an additional source of remote detection of turbulence intensity at different vertical levels. In this study, we train the model mainly using turbulence cases induced by summertime convection, which include CIT as well as CAT.

2. Data and methodology

When it comes to training an ML model, the most important thing is to construct a reliable dataset in which its inputs contain enough spatial and spectral information to predict outputs, and its output needs to be as accurate as possible. It would be the best to train the ML model against pilot reports (PIREPs), which are the conventional observations of turbulence encounters, but their coverage is limited to flight tracks, and PIREPs can have potential errors in spatiotemporal information of turbulence encounters due to pilot's subjectivity (Schwartz 1996; Cormman et al. 2004). Therefore, this study uses NWP-based turbulence diagnostics as the truth data for ML model training which can provide spatiotemporally homogeneous turbulence estimates. NWP-based turbulence diagnostics are calculated at

TABLE 1. Dates used for ML model training, validation, and testing datasets.

Training	Validation	Testing
3, 4, 14, 15, 20–24 May 2020	25, 26 May 2020	26–28 May 2021
3–7, 10 Jun 2020	21, 22 Jun 2020	11, 13 Jun 2021
6–8, 10–12 Jul 2020	17–19 Jul 2020	2, 10, 28 Jul 2021

three of conventional flight-level-based vertical layers (10–18, 18–24, and above 24 kft; 1kft \approx 300 m). The three vertical layers are chosen based on five vertical layers (surface–5, 5–10, 10–18, 18–24, and above 24 kft) that are provided by the NOAA NWS Operational Advisory Team (Li and Heidinger 2021) and are operationally used for the routine route forecast issued by the National Weather Service (available online at <https://forecast.weather.gov/product.php?site=CRH&issuedby=KSF&product=RFR&format=CI&version=1&glossary=1>). Among the five vertical layers, the first two bottom vertical layers are excluded from the training to focus on upper-level turbulence prediction, whose infrared data are most useful. Considering that operational NWP-based turbulence forecast is available at every 2 kft (available online at <https://www.aviationweather.gov/turbulence/gtg>), the three vertical layers used in this study are rather coarse. However, since using infrared data from a GEO can be an underconstrained problem, we decided to start exploring using the simplest but widely used vertical layers as a proof-of-concept study.

During training the ML model, simulated brightness temperatures are used as inputs, and HRRR-based multiturbulence diagnostics in terms of EDR (hereafter, NWP-based EDR) are used as outputs, but once the model is trained, the actual brightness temperature data from *GOES-16* ABI are used as inputs to estimate EDR-scale turbulence.

Training of the ML model requires three independent datasets: training, validation, and testing. Training and validation datasets are used during the training, and the testing dataset is used to validate the ML model after the training. Training and validation data are chosen from the days that had convective activities based on storm reports in 2020 provided by the Storm Prediction Center (SPC; available online at <https://www.spc.noaa.gov/exper/archive/>), as this study focuses on summertime

turbulence, especially induced by convection. To be completely independent of the training and to validate the result properly, testing data are chosen from storm reports in 2021. The dates selected for this study include various weather cases: clear day, deep convection, tornadoes, or hail-producing storms. Table 1 shows the dates selected for each dataset. From the cases listed in Table 1, a total number of 8676, 2484, and 2928 images are collected for training, validation, and testing dataset, respectively.

The training data are carefully selected so that they include turbulence cases over the entire CONUS domain, and there are sufficient number of images with severe weather as well as images with clear sky. Among 8694 training images, 2723 images had at least one SPC report of either tornado, hail, or wind, and PIREPs during the training period in Fig. 1 shows that the training data contain turbulence cases all over CONUS. As with the training data, one-third of validation and testing data (709 of 2484 validation images and 1036 of 3312 testing images) include at least one SPC report.

a. PIREP

The PIREP is the pilot's verbal report of turbulence encounters. It is an observation that reflects what pilots actually encounter, and in this study, we used publicly available PIREP from the Aviation Weather Center (additional information is available on web page at <https://www.aviationweather.gov/>). PIREPs provide information about the location (latitude, longitude, and altitude) and intensity of turbulence. Turbulence intensity is classified in nine categories: 1 = negligible, 2 = smooth into light, 3 = light, 4 = light to moderate, 5 = moderate, 6 = moderate to severe, 7 = severe, 8 = severe to extreme, and 9 = extreme. Although it provides valuable information in real time, the data are not provided in a homogeneous manner and rather sparse even over the CONUS, and PIREPs include spatial and temporal uncertainties in turbulence information (Schwartz 1996; Cornman et al. 2004; Sharman et al. 2014). Therefore, in this study, PIREPs are only used to validate the ML results, and not used during training. We matched PIREPs with the closest point from the ML model results and conducted evaluations.

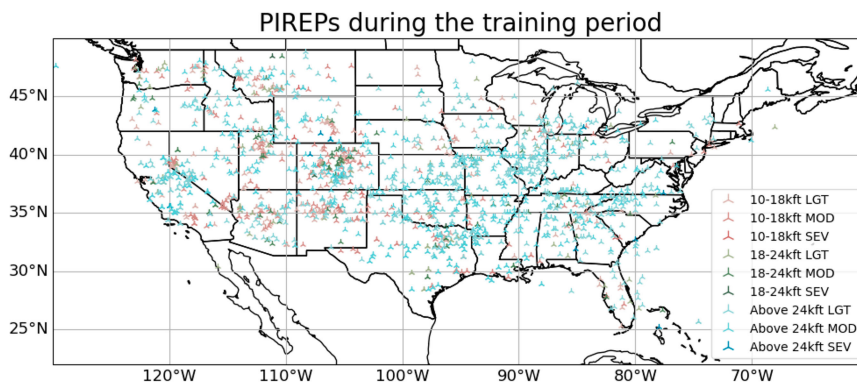


FIG. 1. Distributions of PIREPs during the training period.

TABLE 2. Description of five turbulence diagnostics used in this study.

Turbulence diagnostics	Description	References
Ellrod3	Ellrod–Knapp turbulence index with divergence tendency term	Ellrod and Knapp (1992), Ellrod and Knox (2010)
NGM1	Nested grid model turbulence index, which is wind speed multiplied by the absolute value of flow deformation	Reap (1996)
$ w ^2/Ri$	Square of vertical velocity (w) divided by gradient Richardson number (Ri)	Sharman and Pearson (2017)
UBF/Ri	Unbalanced flow diagnostic divided by Ri	Knox (1997), McCann (2001), Sharman et al. (2006), Sharman and Pearson (2017)
NCSU2/Ri	North Carolina State University turbulence index 2 (NCSU2), which is the cross product between the vertical vorticity gradient and gradient of the Montgomery streamfunction on isentropic surface, divided by Ri	Kaplan et al. (2004), Sharman and Pearson (2017)

b. HRRR model data for turbulence diagnostics

The HRRR model is a regional model developed at the NOAA (additional information is available on web page at <http://ruc.noaa.gov/hrrr>), which has a 3-km horizontal resolution and 50 vertical levels. Its analysis and forecast data are archived every hour in (available online at <https://console.cloud.google.com/storage/browser/high-resolution-rapid-refresh>). In this study, HRRR 1-h forecast data (f01), instead of the analysis data (f00) are used to avoid possible spinup in the data for diagnosing turbulence intensity. Native-level data are used to utilize both number concentration and mixing ratio for each hydrometeors provided by the double-moment Thompson scheme in the CRTM simulation.

The NWP-based turbulence diagnostics considered in this study have been used in NWP-based turbulence forecasting systems (Sharman and Pearson 2017; Kim et al. 2018). Even though these diagnostics are CAT indices mostly related to upper-level fronts and jets, they have been used to infer CIT as well because they can partially capture turbulence generation due to convective clouds developed along large-scale disturbance (S.-H. Kim et al. 2021). Among these diagnostics, five turbulence diagnostics that are related to CIT are selected in this study, and they are listed in Table 2. Standard outputs from the HRRR model such as wind, air temperature, and specific humidity are used to compute turbulence diagnostics. These turbulence diagnostics are calculated at every grid point and vertical level using outputs of the HRRR model. They are then remapped to the EDR scale using lognormal mapping technique developed by Sharman and Pearson (2017). Given that each turbulence diagnostic has a different physical meaning and unit, each diagnostic should be normalized to a common scale such as the EDR in the range between 0 and 1. The lognormal mapping, expressed in Eq. (1), was designed to establish a correspondence between NWP-based turbulence diagnostic and turbulence observation, considering the lognormal property of atmospheric turbulence:

$$\ln(D^*) = \ln(\epsilon^{1/3}) = a + b \ln(D), \quad (1)$$

where D^* is the EDR value corresponding to a raw turbulence diagnostic value D , $\epsilon^{1/3}$ is the EDR, and a and b are

remapping coefficients obtained using the expectation operator and standard deviation operator of the probability distribution function of D and EDR observations.

Five turbulence diagnostics chosen in this study (Table 2) are remapped into the EDR scale using Eq. (1). For each turbulence diagnostic, the maximum EDR value within each of three selected vertical layers (10–18, 18–24, and above 24 kft) is computed for each grid point, and then, the maximum value among the five diagnostics is computed for each vertical layer and used as the outputs to train the ML models. It is common to use a weighted or unweighted mean of multiple diagnostics to produce the final output for NWP-based turbulence forecasts (e.g., Sharman et al. 2006; Kim et al. 2011, 2018; Sharman and Pearson 2017; Pearson and Sharman 2017; S.-H. Kim et al. 2021; Lee et al. 2022). However, in this study, we adopt taking the maximum value from multiple turbulence diagnostics rather than the mean value to focus on increasing hit rates of turbulence. As stated in S.-H. Kim et al. (2021), there are two aspects of CIT forecasting: forecast quality and value as identified by Murphy (1993). Since CIT, which is the main focus of this study, needs to be avoided as much as possible, we decided to take the maximum of the variables. Although this study considered only the maximum value of NWP-based turbulence diagnostics for simplicity, the impacts of the use of the mean of turbulence diagnostics to construct the ML model would be conducted in the future study.

Figure 2 shows how NWP-based EDR evolves as convection is initiated and how it is reflected in observed brightness temperatures. Figures 2a and 2b show HRRR-based EDR above 24 kft at 0200 and 0300 UTC 10 June 2020, respectively, while Figs. 2c and 2d show corresponding channel 13 brightness temperature, and Figs. 2e and 2f channel 8 brightness temperature. As convection is initiated in the purple box between 0200 and 0300 UTC (Figs. 2c,d), cloud-top temperature in the region decreases over time, and NWP-based EDR increases around convective clouds. Turbulence occurring near convective regions in the purple box is well captured in the NWP-based EDR by virtue of HRRR model's cloud resolving capability, which enables to reflect changes in large-scale flows caused by convection. On the other hand, in the red box region, there were PIREPs of light- and moderate-intensity

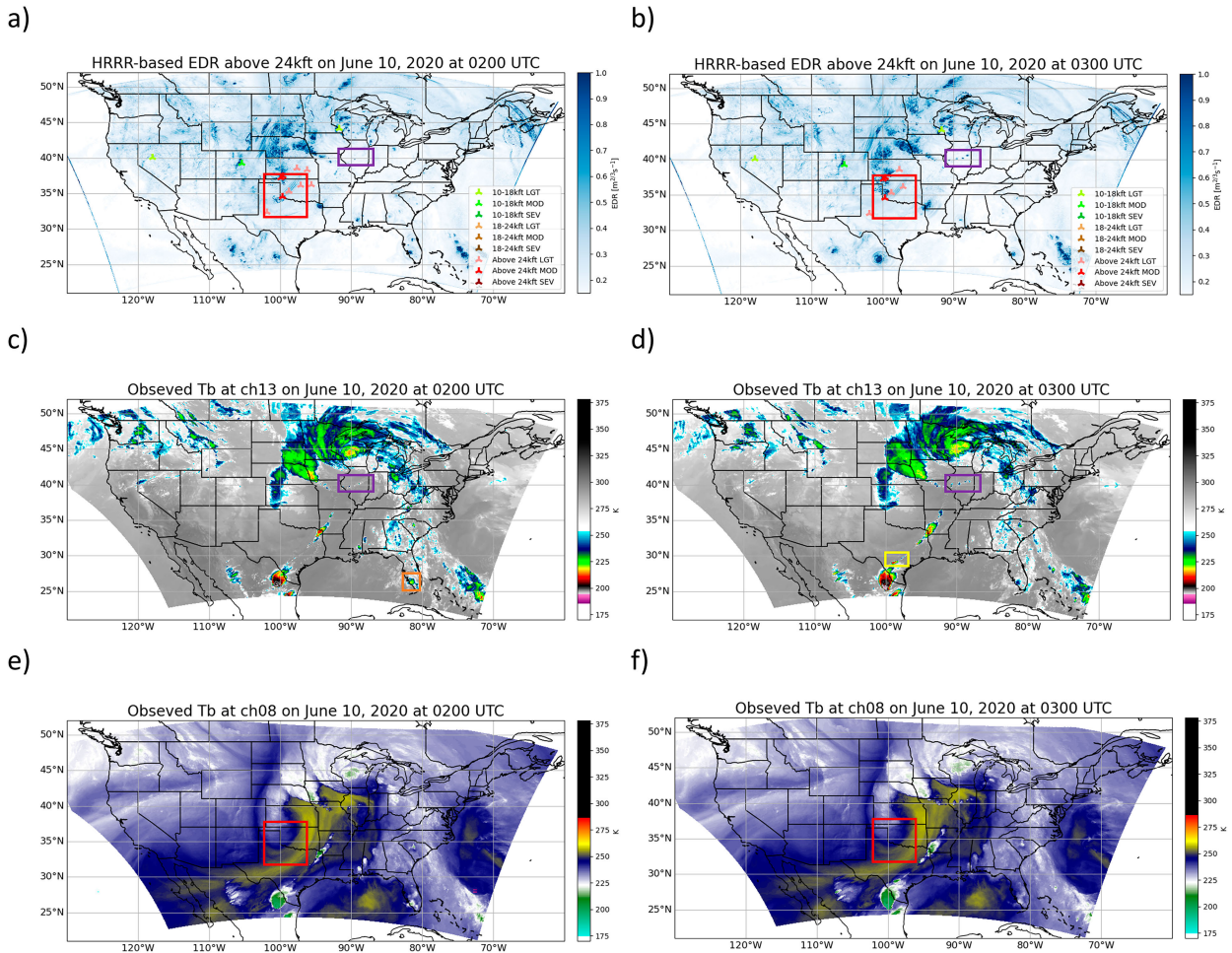


FIG. 2. A case study on 10 Jun 2020 is provided to show how HRRR-based EDR evolves along with convective initiation, as well as observed brightness temperature. HRRR-based EDR above 24 kft at (a) 0200 and (b) 0300 UTC. Observed brightness temperature at channel 13 at (c) 0200 and (d) 0300 UTC. Observed brightness temperature at channel 8 at (e) 0200 and (f) 0300 UTC.

turbulence under cloud-free air. According to upper-air data from SPC, high wind speed is observed at upper level (~ 300 hPa) in this region, which suggests that this case can be considered as a conventional type of CAT possibly associated with the shear or inertial instability. The presence of CAT in this region can be inferred from high gradients in water vapor band imagery in Fig. 2e or Fig. 2f. This example shows the ability of NWP-based CAT diagnostics to capture CIT as well, and how turbulence intensity can be inferred from a satellite imagery. The use of turbulence diagnostics explicitly developed for the CIT (e.g., Kim et al. 2019, 2021) for the ML model will be considered in the future.

c. GOES-16 ABI

GOES-16 is NOAA's current operational GEO that views the eastern part of the CONUS and the Atlantic Ocean. It carries the ABI that has 16 channels ranging from visible to infrared. Its infrared channels have a spatial resolution of 2 km, and CONUS sector data have a temporal resolution of five minutes. Among 16 channels, 9 channels that are infrared are used in this study. Three water vapor channels each centered at 6.2, 6.9, and

7.3 μm are sensitive to water vapor at different vertical levels, and thus, can observe vertical distributions of water vapor. The three water vapor channels have been used to detect a jet stream which is associated with turbulence generation, as well as a tropopause folding feature which is a notable feature of turbulence in clear sky regions (Wimmers and Moody 2004). Infrared longwave clean and dirty window bands at 8.5, 10.3, 11.2, and 12.3 μm that have different sensitivities to cloud waters are used to infer cloud properties such as cloud-top height or cloud-top phase, while channels at 9.6 and 13.3 μm are sensitive to ozone and carbon dioxide, respectively. For detecting turbulence induced by convective clouds, channels around longwave infrared window bands can be useful because brightness temperatures from these channels exhibit features of convective clouds, such as enhanced-V signatures or overshooting tops (Brunner et al. 2007; Bedka et al. 2010). All nine infrared channels are used to form a baseline model, but through sensitivity test, five channels (channels 8, 9, 10, 11, and 13; 6.2, 6.9, 7.3, 8.5, and 10.3 μm) are selected as inputs to the final ML model to reduce the input data size.

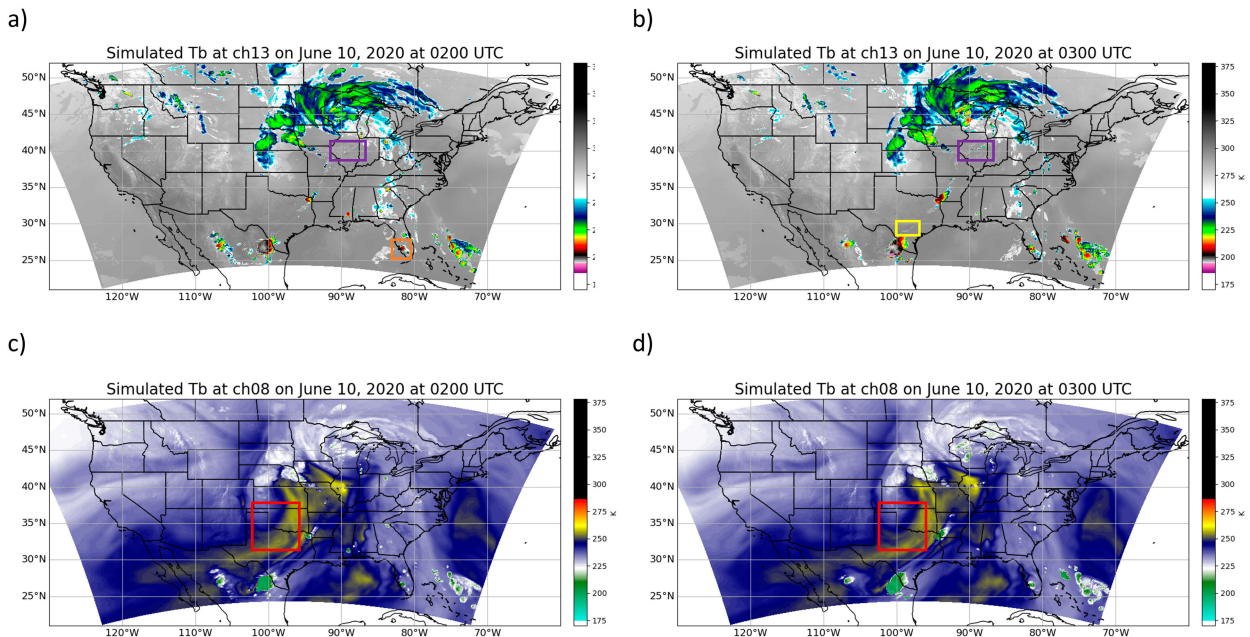


FIG. 3. Synthetic channel 13 brightness temperature at (a) 0200 and (b) 0300 UTC and synthetic channel 8 brightness temperature at (c) 0200 and (d) 0300 UTC are shown to compare with the observed brightness temperatures in Fig. 2.

d. Synthetic brightness temperature data using HRRR data and CRTM

As briefly mentioned in the introduction, we use synthetic brightness temperature during training not to confuse the model with mismatched weather features between *GOES-16* observation and HRRR model output which is used to calculate turbulence diagnostics. CRTM developed at the Joint Center for Satellite Data Assimilation (Weng 2007; Chen et al. 2011; Liu et al. 2012) is one of the most commonly used radiative transfer models, and its version 2.1.3 is used in this study. It can simulate more than 100 sensors including all the channels of *GOES-16* ABI. Synthetic brightness temperatures for channels 8, 9, 10, 11, and 13 of *GOES-16* ABI are simulated using HRRR model data as input for CRTM.

To justify the approach of using synthetic brightness temperature and its ability to simulate evolution of convection, synthetic brightness temperatures at channel 13 and 8 for the same case study presented in section 2b are shown in Fig. 3 for comparison. Synthetic and observed brightness temperatures shown in Figs. 2 and 3 look very similar in general. The upper-level low pressure system in the red box region as well as multiple convective cells developing in the purple box region are correctly simulated in the synthetic brightness temperature (Fig. 3). Although synthetic brightness temperatures are generally in good agreement with the observed data, location, timing, or intensity of convection are slightly different and sometimes not well represented in the synthetic map due to inevitable problems both from NWP model and radiative transfer model simulations. Convection developing in the yellow box of Figs. 2d and 3b is an example of mismatch of time in the

NWP model. Convection in the yellow box region is not observed from the synthetic brightness temperature map at 0300 UTC (Fig. 3b), thereby missing CIT by NWP-based EDR (Fig. 2b). Convective cloud in the orange box in Figs. 2c and 3a shows an overestimation of hydrometeors by the NWP model, thus exhibiting lower brightness temperature in the synthetic map. Such mismatches between NWP variables and observed brightness temperatures can lead to wrong learning during the training. Therefore, although further improvements in simulating brightness temperature are required, synthetic data are used in the training of the ML model to provide correct location of the weather system.

3. Machine learning model design

Machine learning models are designed to estimate turbulence at three different levels (10–18, 18–24, and 24 kft and above) using *GOES-16* ABI data.

a. Inputs and outputs for the model

Input to the models is $512 \times 512 \times 6$ image of five synthetic brightness temperature images (ABI channels 8, 9, 10, 11, and 13; 6.2, 6.9, 7.3, 8.4, and $10.3 \mu\text{m}$, respectively) and geopotential height map, and output is a 512×512 image of NWP-based EDR. Since the spatial resolution of ABI (either synthetic data used for training or observed data used for validation after training) is 2 km whereas that of HRRR model is 3 km, ABI data are interpolated into the HRRR model grid space before training. Brightness temperature values are normalized to have values from 0 to 1 based on the minimum and maximum channel values used in Hayatbini et al. (2019), and terrain heights are also normalized to range between 0 and 1.

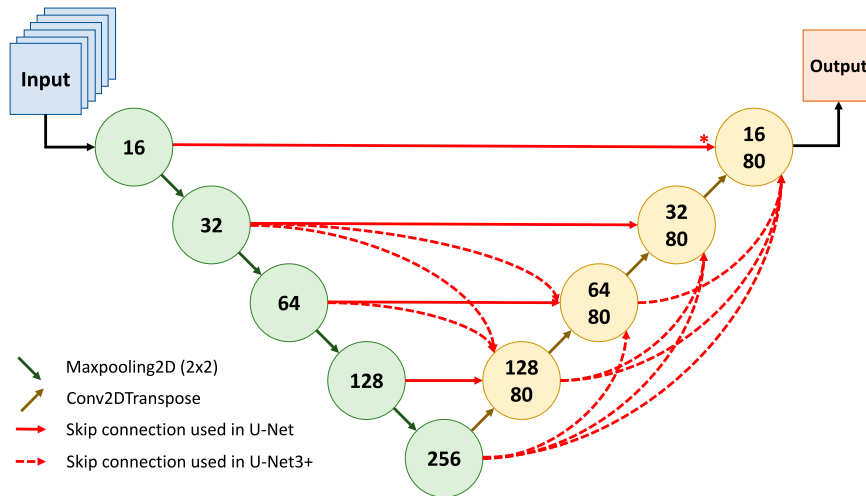


FIG. 4. U-Net and U-Net3+ models are described. Each circle consists of Conv2D, Batch-Normalization, and ReLU activation layers. Green circles represent the encoder part while yellow circles represent the decoder part, and the numbers in the circles are numbers of filters. The upper number in the yellow circle is the number of filters for the U-Net model, and the lower number is for the U-Net3+ model. Red solid arrows represent skip connections used in the U-Net model, while red dashed arrows represent additional skip connections used in the U-Net3+ model. Note that the skip connection from the first encoder unit, which has an asterisk, is not used in the U-Net3+ model.

b. Model architectures

U-Net is one of the most commonly used model architectures for an image-to-image translation problem where both inputs and outputs are images. U-Net was first introduced by Ronneberger et al. (2015), and is a fully convolutional network with skip connections. Skip connections help prevent from losing fine-scale features during the upsampling path. One benefit of using U-Net-based models is that it provides turbulence information at each pixel with 3-km resolution. Since the U-Net model was developed by Ronneberger et al. (2015), many variations of U-Net have been developed, one of which is U-Net3+ developed by Huang et al. (2020). Unlike U-Net which only has connections between encoder and decoder, U-Net3+ adds connections within the decoder. With this additional intraconnection between the decoder layers, each decoder layer in U-Net3+ integrates small-scale feature maps from encoder layers as well as large-scale feature maps from decoder layers. In this study, six different U-Net based models are tested following the code available online at <https://github.com/dopplerchase/keras-unet-collection>: three U-Net models and three U-Net3+ models using filter sizes of 3, 5, and 7. The bigger filter sizes (5 and 7) are tested as upper-level features conducive to turbulence generation such as tropopause folding are rather large. Details of the model architecture are shown in Fig. 4. Since using regularizations did not seem to learn well, batch normalization is used to prevent overfitting, and the model is compiled with a loss function of mean-squared error (MSE) and RMSprop optimizer. Each model is run for 100 epochs, and the model is saved every 20 epochs. All the saved models are used to compare results.

4. Statistical model results

In Fig. 5, MSEs on the validation dataset are compared between six different models to evaluate model performance during training. Figure 5 shows MSE over 100 epochs for each model. All the U-Net3+ models tend to have lower MSE than the U-Net models. The lowest MSE of 0.005 is achieved using a filter size of 3 in the U-Net3+ model, but there is not much difference in MSE between the U-Net3+ models.

Even though MSE appears to be small enough, this is only against the model diagnostic which is also not perfect. Therefore, further analysis is conducted by applying the models to GOES-16 ABI observations and comparing the results against PIREPs. To account for possible delays in the report time, PIREPs within the 15-min window around the evaluation time are collected and combined for validation. The forecasting performance of ML-based turbulence estimate is evaluated using probability-of-detection (POD) statistics. For comparison, NWP-based EDR is also validated against PIREPs. In the current study, null (NIL) and moderate-or-greater (MOG)-intensity turbulence events are used in the performance evaluation to mitigate uncertainties in reported turbulence intensity of light-level turbulence (e.g., Sharman et al. 2006; Kim et al. 2018; S.-H. Kim et al. 2021). Therefore, the POD “yes” (PODY) and POD “no” (PODN) are computed for the MOG and NIL-level turbulence, respectively.

Each PODY–POFD pair, where POFD is a probability of false detection ($=1 - \text{PODN}$) is obtained for a given threshold, and relative operating characteristic (ROC; Mason and Graham 1999; Marzban 2004) curves can be constructed by applying forty different EDR values as the threshold. It is noted that the area under the ROC curves (AUC) can be

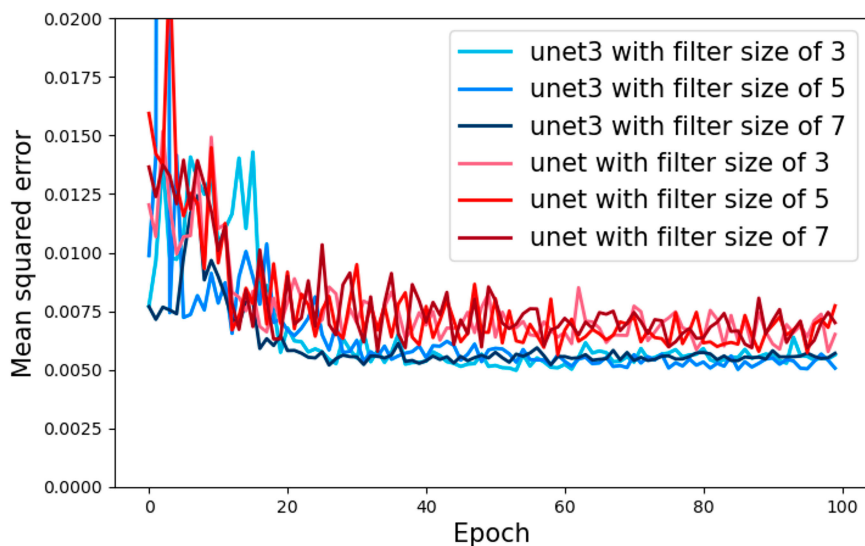


FIG. 5. MSE of six different models on the validation dataset along all the epochs during training. Blue colors represent MSE for U-Net3+ models with filter sizes of 3, 5, and 7, and red colors represent MSE for U-Net models with filter sizes of 3, 5, and 7.

used as a measure of the performance skill of each forecasting method (Mason 2003). Details of this procedure can be found in S.-H. Kim et al. (2021).

Figures 6 and 7 show the ROC curve with the AUC values for NWP-based EDR and ML-based EDR estimates (derived from U-Net3+ and U-Net model) at three vertical levels. From Fig. 5, models seem to converge after 40 epochs, and thus, each model with filter sizes of 3, 5, and 7 at 40, 60, and 80 epochs is plotted in one figure to compare performances and pick the best model. Numbers on the label indicate AUCs for each model. AUC value of 1 represents a perfect skill for turbulence intensity estimation. The blue line represents the NWP-based EDR. Above 24 kft, there is relatively less difference in AUC between the models, all of which are close to the maximum AUC value of 0.7. However, there is a

large difference in results at the two lower vertical layers (10–18 and 18–24 kft). U-Net3+ models tend to perform better than U-Net models at 10–18 and 18–24 kft, achieving the maximum AUCs of 0.79 and 0.73, respectively. Among nine U-Net3+ models, models with a filter size of 3 that are stopped at 60 and 80 epochs seem to be the best two models. Performance skills of all ML-based EDRs appear to be lower than the NWP-based EDR, and it can be related to the fact that the NWP-based EDR is used as the true dataset to train the model. Considering the spatial uncertainty of PIREPs (e.g., an average horizontal distance error of 46 km; Sharman et al. 2014), the performance evaluation using the in situ EDR data should also be conducted in the future study.

Although five channels are selected for these models, further experiments are conducted to explore how different

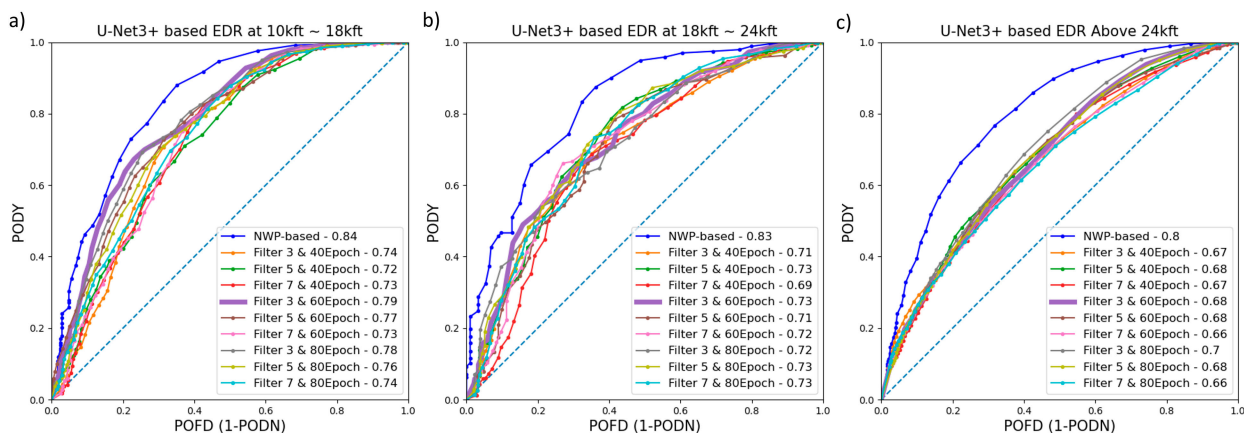


FIG. 6. ROC curve of nine different U-Net3+ model EDR estimates at three vertical layers: (a) 10–18, (b) 18–24, and (c) above 24 kft. NWP-based EDR is presented in a blue line for comparison. Numbers on the right in the legends are AUC values for each model. The thick purple line (filter 3 and epoch 60) is the one with the best overall performance, and it is used for case study results in section 5.

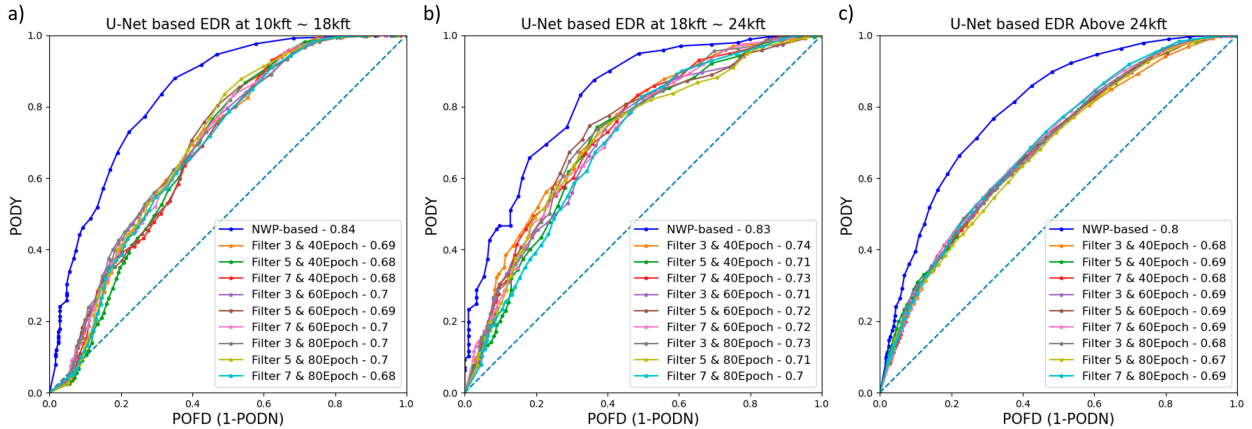


FIG. 7. As in Fig. 6, but using nine different U-Net models.

channel selection impacts the ML result. ABI has nine infrared channels in total, and thus, experiments are conducted using the whole *GOES-16* spectral bands to compare with results using five channels. Currently, there are several satellites over the globe that carry similar but different channels, and thus, similar ML models can be applied to several satellites. For example, FCI does not contain a midlevel water vapor channel ($6.9 \mu\text{m}$), and the Visible Infrared Imaging Radiometer Suite (VIIRS) aboard NOAA’s Joint Polar Satellite System (JPSS) satellites (Goldberg et al. 2013) lacks the whole water vapor bands. Two additional experiments are further conducted using channels that FCI and VIIRS have. For a direct comparison, additional experiments using nine ABI channels, FCI-like channels, VIIRS-like channels are conducted using the same model architecture (U-Net3+ model with the filter size of 3). Results using five channels (the purple line in Fig. 6), using nine ABI channels, using four channels (FCI-like channels), and using two channels (VIIRS-like channels) are presented in Table 3 in

terms of AUC as well as PODY, PODN, and true skill statistics (TSS) using a 0.22 threshold value as done in many previous studies (Pearson and Sharman 2017; S.-H. Kim et al. 2021). The performance of NWP-based EDR which is used as truth in this study is provided on the top, and it is followed by an experiment using all nine channels from *GOES-16* ABI, which is named as “baseline,” because it includes all channels available. All experiments show similar AUC above 24 kft, while different skills are observed at lower layers. The baseline model that uses the most spectral information shows detrimental skill (low AUC) at 10–18 kft layer which might indicate that some information might be redundant, although the baseline model shows the best performance at 18–24 kft. It seems that having the full set of water vapor channels helps better estimate turbulence at lower levels, although there might be some redundant information from other spectral bands. However, it does not seem to have impacts on the upper-level turbulence estimation much. Moreover, when we additionally conducted experiments excluding channel

TABLE 3. AUC, PODY, PODN, and TSS from experiments using nine ABI channels (baseline), five ABI channels (final), FCI-like channels, and VIIRS-like channels.

		AUC	PODY	PODN	TSS
10–18 kft	NWP-based EDR	0.836	0.473	0.869	0.342
	Nine channels (baseline)	0.756	0.778	0.578	0.356
	Five channels (final)	0.785	0.769	0.594	0.363
	FCI-like channels (without $6.9 \mu\text{m}$)	0.785	0.886	0.457	0.343
	VIIRS-like channels (8.4 and $10.3 \mu\text{m}$)	0.766	0.794	0.549	0.343
18–24 kft	NWP-based EDR	0.828	0.684	0.745	0.429
	Nine channels (baseline)	0.760	0.775	0.610	0.385
	Five channels (final)	0.730	0.838	0.421	0.259
	FCI-like channels (without $6.9 \mu\text{m}$)	0.705	0.861	0.429	0.289
	VIIRS-like channels (8.4 and $10.3 \mu\text{m}$)	0.721	0.819	0.490	0.309
Above 24 kft	NWP-based EDR	0.797	0.920	0.400	0.320
	Nine channels (baseline)	0.696	0.843	0.338	0.181
	Five channels (final)	0.682	0.916	0.227	0.143
	FCI-like channels (without $6.9 \mu\text{m}$)	0.681	0.854	0.287	0.141
	VIIRS-like channels (8.4 and $10.3 \mu\text{m}$)	0.679	0.661	0.518	0.179

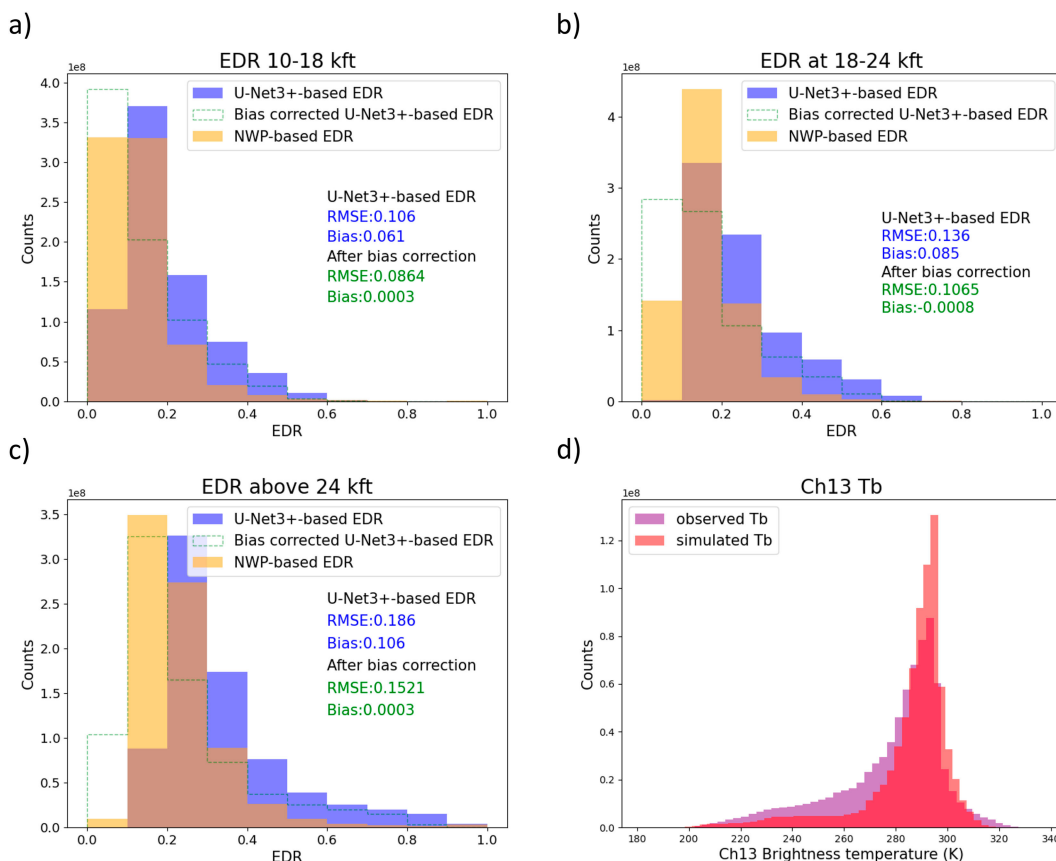


FIG. 8. Histograms of NWP-based EDR (orange) and U-Net3+-based EDR (purple) at (a) 10–18, (b) 18–24, and (c) above 24 kft are shown along with RMSE and bias. (d) Observed (purple) and synthetic (pink) brightness temperature distributions are presented to explain possible cause of overestimation in U-Net3+-based EDR.

13, which is most sensitive to clouds, it is found that the performance skill is degraded (not shown).

The U-Net3+-based model with a filter size of three using five channels that shows good model performance is chosen as the final model to conduct further analysis. Figure 8 shows histograms of NWP-based EDR (orange) and U-Net3+-based EDR (purple) along with root-mean-squared error (RMSE) and bias at each vertical layer. Positive bias meaning overestimation in U-Net3+-based EDR is shown at all vertical layers, and it tends to be larger as turbulence intensity is larger. Such overestimation can be partially explained by the brightness temperature histogram in Fig. 8d. Observed brightness temperatures tend to be more spread out, while simulated brightness temperatures have a higher peak around 288 K. This suggests that there are more or wider clouds in the observation, which can lead to the overestimation (positive bias) in the predicted turbulence intensity. Since it seems that systematic biases exist in the U-Net3+-based EDR results, bias obtained at each vertical layer is subtracted from the ML model output. After the bias correction, distributions of U-Net3+-based EDR plotted in green dashed lines in Fig. 2 agree better with those of NWP-based EDR, and their RMSE and bias are reduced at all vertical layers as shown in Fig. 2. For the case study results in the next section, bias correction is applied using biases obtained from this analysis.

5. Case study results

In this section, forecasts from the best U-Net3+ model using five channels (purple line in Fig. 6 with bias correction) are examined for two real cases that are selected from the testing datasets. Although the model is trained with small patches of 512×512 images, a map that covers the whole CONUS domain is obtained by applying the weights of the trained model to the whole CONUS image.

a. A case with convections accompanying hail and tornado (2200 UTC 10 July 2021)

Applying the best model to the case on 10 July 2021 is presented in this section. This is a case that had many reports of high winds, hail, and tornadoes across the United States as shown in Fig. 9.

Figure 10 shows the horizontal distributions of NWP-based EDR (Figs. 10a,c,e) and U-Net3+-based EDR (Figs. 10b,d,f) for the three vertical layers. Turbulence encounters included in PIREPs are indicated as triangles [null (NIL): black; light (LGT): green; moderate (MOD): orange; severe (SEV): red]. At 10–18 kft, there is turbulence over mountainous regions in the midwestern United States, which may be related to mountain waves. Both the NWP-based EDR and U-Net3+-based

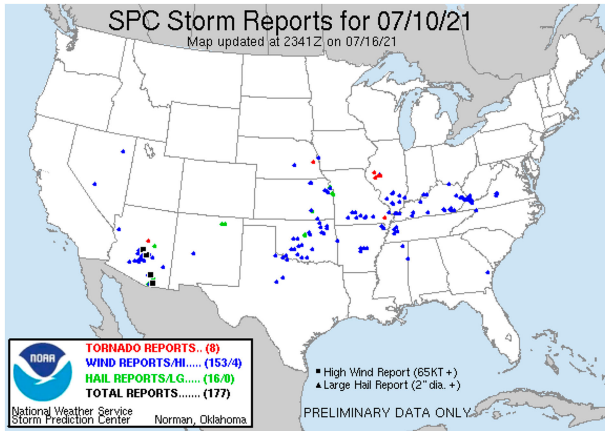


FIG. 9. Storm reports on 10 Jul 2021 from the NOAA Storm Prediction Center.

EDR values are relatively high near the region with MOD-level turbulence (orange triangle). In the two light turbulence regions (green triangle) over the eastern part of CONUS, NWP-based EDR and U-Net3+-based EDR show similar

values and patterns. NWP-based EDR and U-Net3+-based EDR for the left green triangle region are 0.12 and $0.10 \text{ m}^{2/3} \text{ s}^{-1}$, respectively, and for the right green triangle region, they are 0.30 and $0.33 \text{ m}^{2/3} \text{ s}^{-1}$, respectively. Both EDR values in the two regions are close to what is considered as light turbulence ($0.15\text{--}0.22 \text{ m}^{2/3} \text{ s}^{-1}$). At 18–24 kft, the overall spatial patterns seem to agree well between NWP-based EDR and U-Net3+-based EDR. However, light turbulence over Texas is missed by the ML-based EDR, but well captured by the NWP-based EDR. Above 24 kft, high EDR values are observed due to deep convection, which affects upper-level turbulence. U-Net3+-based EDR agrees well with PIREPs in terms of location, especially in three red box regions of Fig. 10f where NWP-based EDR does not exhibit strong (e.g., moderate-or-greater) turbulence. Especially in two moderate turbulence regions (with orange triangles) of the bottom red box, U-Net3+-based EDR exhibits high EDR value while NWP-based EDR does not exhibit.

To explain overestimation in U-Net3+-based EDR to some extent, maps of synthetic and observed brightness temperatures at channel 13 are shown in Fig. 11. Although synthetic and observed brightness temperatures look similar in

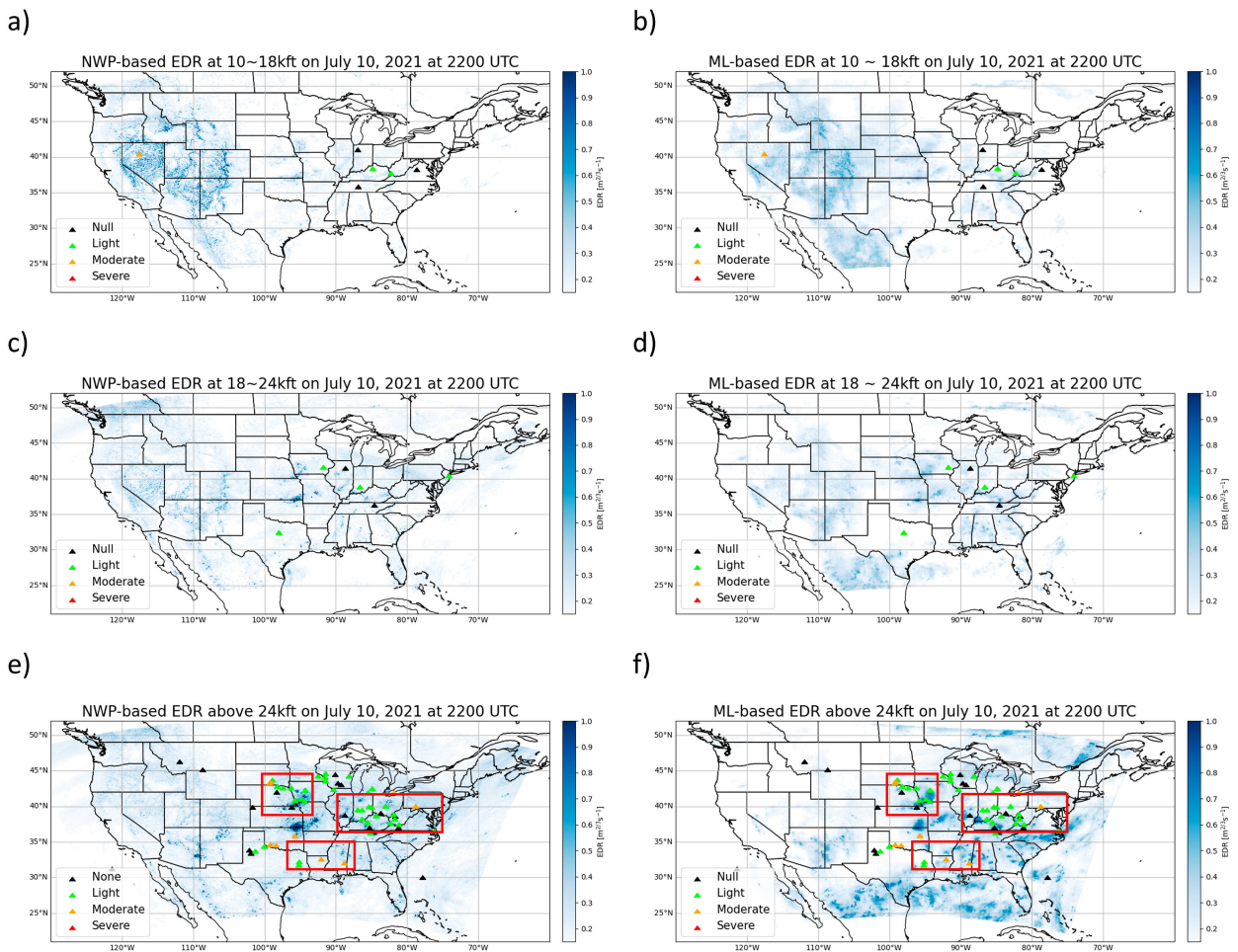


FIG. 10. NWP-based EDR at 2200 UTC 10 Jul 2021 at (a) 10–18, (c) 18–24, and (e) above 24 kft are shown, as well as U-Net3+-based EDR estimates at (b) 10–18, (d) 18–24, and (f) above 24 kft. Note that only the EDR values greater than $0.15 \text{ m}^{2/3} \text{ s}^{-1}$ are shown.

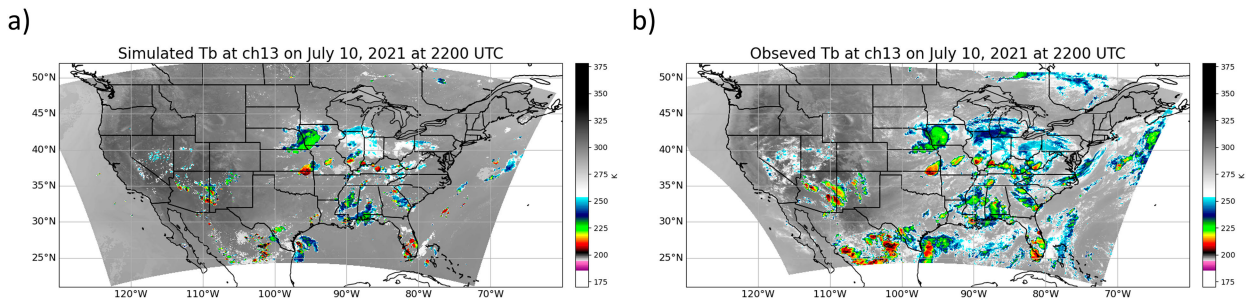


FIG. 11. (a) Synthetic and (b) observed channel 13 brightness temperatures for the case study at 2200 UTC 10 Jul 2021.

general, a lot of high clouds are not shown in the synthetic map, which means that the HRRR model did not simulate some cirrus or deep convective clouds, and there still is a difference in terms of location or size of convective clouds. Such less clouds in HRRR model simulations might have missed CIT detection in NWP-based EDR (e.g., the bottom red box in Fig. 10), while U-Net3+-based EDR, which is obtained using observed brightness temperature as inputs to the trained ML model, might appear to have more CIT. Also note that U-Net-based EDR estimates look blurrier, which is expected with less skip connection (not shown).

Although U-Net3+-based EDR seems to do better overall in this case, it tends to overestimate turbulence compared to NWP-based EDR. The overestimation can be attributed to several things. As NWP-based turbulence diagnostics used in this study were developed by assuming the downscale cascade from forcing such as fronts (Sharman and Pearson 2017), they tend not to predict turbulence due to small-scale convection well (e.g., S.-H. Kim et al. 2021). There can be a problem in the HRRR model itself such as not simulating convection with the correct intensity in the area where it should be. Last, there are biases between observed and simulated brightness temperatures as previously shown in Fig. 8, which can affect both during the training and predicting with observed brightness temperature.

b. A case with clear-air turbulence (1100 UTC 2 July 2021)

Another case on 2 July 2021 is presented in Fig. 12 to address the weakness of U-Net3+-based EDR. This is a case that reported strong CAT event above 24 kft over the central United States and convection over the southern and eastern regions. At 10–18 kft (Figs. 12a,b), U-Net3+-based EDR correctly shows moderate turbulence in the area with orange triangle. However, NWP-based EDR does not exhibit turbulence in the orange triangle spot for the same reason in the previous case study (10 July 2021) that the HRRR model did not simulate high clouds (not shown). On the other hand, a red box region in Figs. 12e and 12f, which shows EDR above 24 kft, has large areas with CAT, which seems to be due to shear or inertial instability caused by upper-level jet streams. The ML model struggles to capture CAT in this region as shown in Fig. 12f. Figure 13 shows channels 8 and 13 brightness temperature maps, which are two of the six input images used to estimate turbulence. A

horizontal gradient in water vapor imagery is shown in the red box region of the channel 8 map (Fig. 13a), which can be an indicator of turbulence, but it is not very clear to infer turbulence intensity from this figure. This shows the limitations of using just satellite images only. Nevertheless, U-Net3+-based EDR shows good agreements in convective regions at all three vertical levels.

6. Discussion

The U-Net3+-based model developed in this study can provide valuable information that the NWP-based EDR might miss, although at the current stage, it might not outperform NWP-based EDR. Nevertheless, the U-Net3+-based model shows overall good agreement with the NWP-based EDR and good performance against PIREPs. However, there is still much room for improvements because the ML model developed in this study has some limitations. The U-Net3+-based model estimates turbulence intensity based on real-time satellite observation, which is not forecast data. However, the ML model can still offer a useful means of monitoring the current weather situation on a large scale and can provide essential information over data-sparse regions. This study primarily aims to show the potential of using satellite imagery as inputs to the ML model for inferring turbulence intensity. With recent efforts to introduce a nowcasting ability to satellite observations by adding wind data, future study could be conducted to produce more sophisticated turbulence data utilizing satellite data for short-term forecasting.

This study uses the maximum values between five HRRR-based turbulence diagnostics across each of the three relatively broad vertical layers. Using a broad vertical layer can cause uncertainties when comparing with PIRPEs, especially if turbulence intensity varies significantly within a vertical layer. Since this study demonstrated the feasibility of generating turbulence intensity information from GEO data, the future research can be expanded to estimating turbulence intensity at a finer vertical resolution with more carefully weighted turbulence diagnostics or with additional CIT diagnostics.

7. Conclusions

GEO data have provided useful weather observations to the aviation community with broad spatial and consistent time coverage. The data have also proven useful for detecting and estimating turbulence of the upper-level atmosphere, but not as

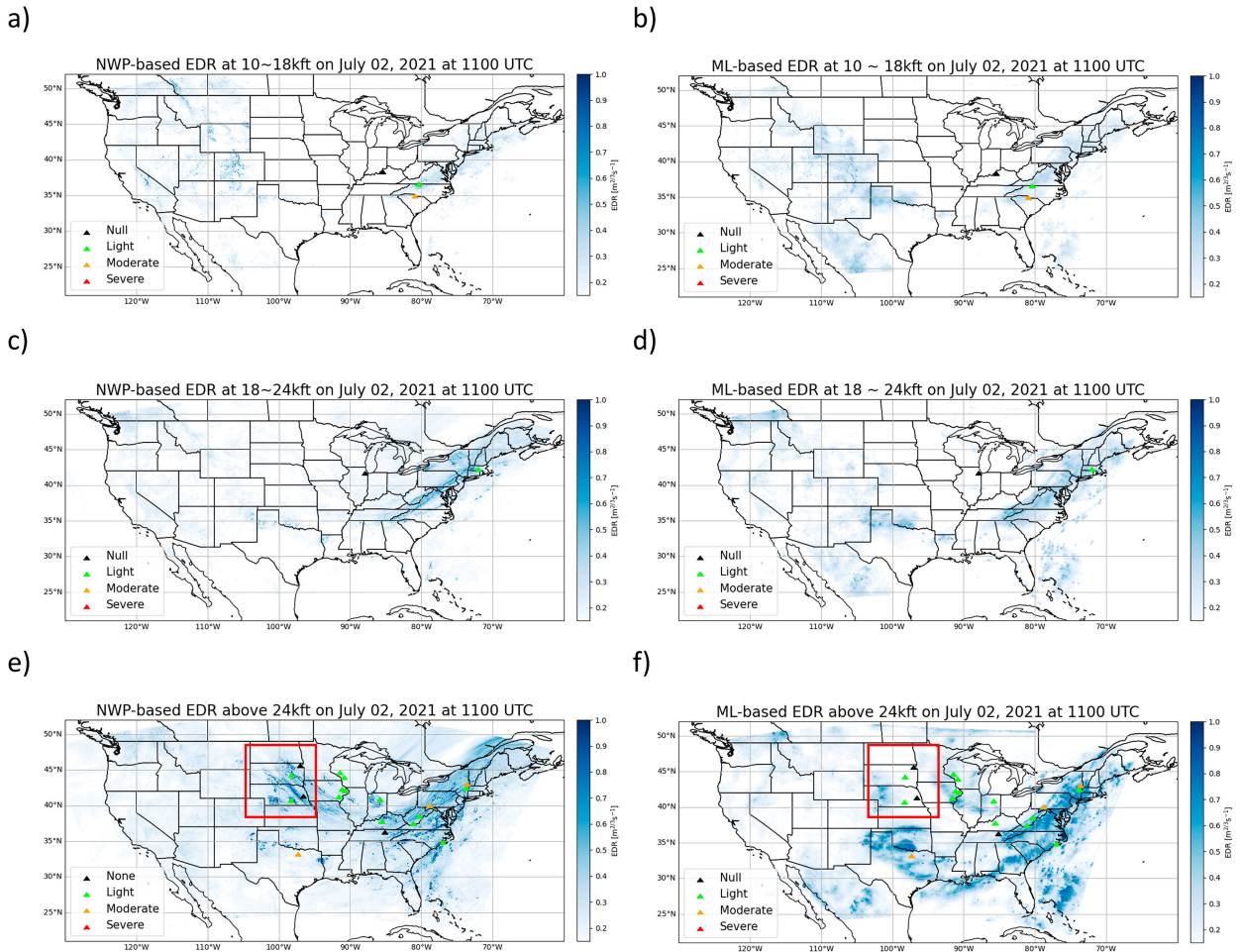


FIG. 12. As in Fig. 10, but at 1100 UTC 2 Jul 2021.

much in the lower atmosphere due to lack of vertical information from conventional passive radiometer sensors. Although it is still challenging to get the whole vertical structure of turbulence solely from satellite data, this study suggests that more vertical information can be extracted using machine learning techniques. Machine learning techniques can help find nonlinear relationships between the multispectral satellite images and extract useful features to estimate turbulence intensity at different vertical levels.

In this study, we developed U-Net-based machine learning models to produce turbulence intensity estimates at three different vertical levels using brightness temperature data from *GOES-16* ABI. Three water vapor channels, two longwave infrared channels, and terrain height are used as inputs to provide turbulence intensity estimates at three different vertical levels: 10–18, 18–24, and above 24 kft. Since PIREPs are only available at a few grid points, HRRR model outputs are used to generate turbulence diagnostics in EDR to be used as outputs for the

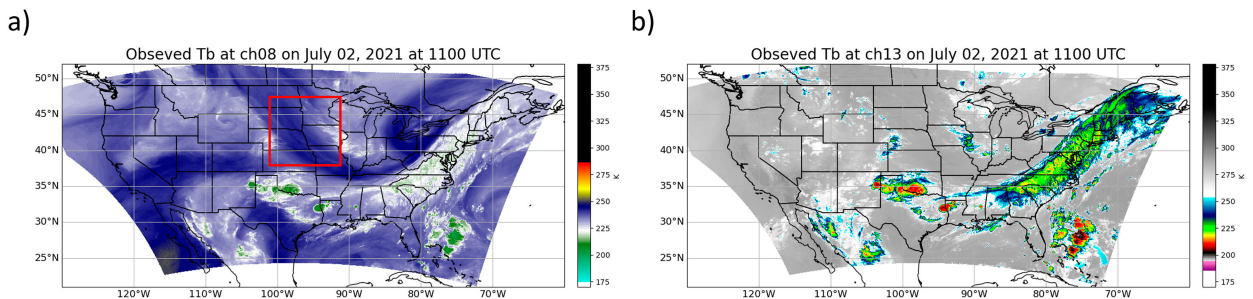


FIG. 13. Observed brightness temperatures at (a) channel 8 ($6.19 \mu\text{m}$) and (b) channel 13 ($10.35 \mu\text{m}$) at 1100 UTC 2 Jul 2021.

machine learning model. To maintain consistency between input and output during training, brightness temperatures are simulated using the CRTM with HRRR model outputs, and synthetic brightness temperatures are used as inputs during training. Three U-Net and U-Net3+-based models (each with a filter size of 3, 5, and 7, respectively) are tested to examine the effects of additional skip connections and different filter sizes. Among the six models, U-Net3+ model with a filter size of 3 performs best. Above 24 kft, it estimates high turbulence intensity near convective regions reasonably well, but it tends to overestimate compared to the NWP-based EDR. After bias correction, RMSE and biases are lowered at all three vertical layers, and the ML-based EDR shows better agreement with NWP-based EDR. However, in the case of CAT where signals observed from infrared images might not be obvious, ML-based EDR can miss CAT generation. At lower levels, however, the U-Net3+ model shows similar skills as the NWP-based EDR. Additional experiments using different channel selections revealed that having water vapor channels was indeed beneficial for estimating turbulence intensity at lower levels.

This study shows some potential of developing a purely satellite observation-based model to estimate turbulence intensity and help turbulence detection in convective regions. However, since this is a proof-of-concept study, there are still limitations, and thus, further studies are needed in the future. As the first attempt of only using multispectral information from *GOES-16* ABI to estimate turbulence at different vertical levels, we focused on NWP-based EDR over 10 kft. Even though turbulence information below 10 kft is critical for small-size aircraft, we excluded turbulence below 10 kft in this study because NWP-based turbulence diagnostics used as ground truth for ML model training in this study are considered optimal for upper-level turbulence prediction. Muñoz-Esparza and Sharman (2018) developed a specialized method for low-level turbulence, and such methods can be included in a future study. In terms of improving the ML model accuracy, more complex ML models such as an attention-based model can be tested. Nevertheless, this study shows that ML-based estimates using *GOES-16* ABI can still be beneficial as a supplementary product to support NWP model-based turbulence prediction, and such satellite-based turbulence product would be useful for nowcasting turbulence. Furthermore, since there are several satellites that carry sensors with similar channels, ABI-based models can be retrained with data such as AHI or AMI, and global turbulence estimates can be provided from these satellites.

Acknowledgments. This work was supported by the National Oceanic and Atmospheric Administration under Grant NA19OAR4320073. J.-H. Kim and S.-H. Kim were funded by the Korea Meteorological Administration Research and Development Program under Grant KMI2022-00310. S.-H. Kim was also supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education Grant NRF-2022R111A1A01071708. We greatly appreciate Dr. Robert Sharman at NCAR for providing EDR conversion coefficients of NWP-based turbulence diagnostics.

Data availability statement. HRRR data are downloaded from <https://console.cloud.google.com/storage/browser/high-resolution-rapid-refresh>. PIREPs can be obtained from the Aviation Weather Center (<https://www.aviationweather.gov/>). *GOES-16* data are publicly available at <https://registry.opendata.aws/noaa-goes> (last accessed on 30 November 2022). U-Net and U-Net3+ models used in this study are based on the keras-unet-collection library codes publicly accessible at <https://github.com/dopplerchase/keras-unet-collection>.

REFERENCES

- 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, <https://doi.org/10.1175/2009JAMC2286.1>.
- Bessho, K., and Coauthors, 2016: An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. *J. Meteor. Soc. Japan*, **94**, 151–183, <https://doi.org/10.2151/jmsj.2016-009>.
- Brunner, J. C., S. A. Ackerman, A. S. Bachmeier, and R. M. Rabin, 2007: A quantitative analysis of the enhanced-V feature in relation to severe weather. *Wea. Forecasting*, **22**, 853–872, <https://doi.org/10.1175/WAF1022.1>.
- Chen, Y., Y. Han, Q. Liu, P. Van Delst, and F. Weng, 2011: Community Radiative Transfer Model for Stratospheric Sounding Unit. *J. Atmos. Oceanic Technol.*, **28**, 767–778, <https://doi.org/10.1175/2010JTECHA1509.1>.
- Cornman, L., G. Meymaris, and M. Limber, 2004: An update on the FAA Aviation Weather Research Program's in situ turbulence measurement and reporting system. *11th Conf. on Aviation, Range, and Aerospace Meteorology*, Hyannis, MA, Amer. Meteor. Soc., P4.3, <https://ams.confex.com/ams/11aram22sls/webprogram/Paper81622.html>.
- Dowell, D. C., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. *Wea. Forecasting*, **37**, 1371–1395, <https://doi.org/10.1175/WAF-D-21-0151.1>.
- Ellrod, G. P., 1989: A decision tree approach to clear air turbulence analysis using satellite and upper air data. NOAA Tech. Memo. NESDIS 23, 26 pp., <https://repository.library.noaa.gov/view/noaa/19299>.
- , and D. I. Knapp, 1992: An objective clear-air turbulence forecasting technique: Verification and operational use. *Wea. Forecasting*, **7**, 150–165, [https://doi.org/10.1175/1520-0434\(1992\)007<0150:AOCATF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1992)007<0150:AOCATF>2.0.CO;2).
- , and J. A. Knox, 2010: Improvements to an operational clear-air turbulence diagnostic index by addition of a divergence trend term. *Wea. Forecasting*, **25**, 789–798, <https://doi.org/10.1175/2009WAF2222290.1>.
- , and K. Pryor, 2019: Applications of geostationary satellite data to aviation. *Pure Appl. Geophys.*, **176**, 2017–2043, <https://doi.org/10.1007/s00024-018-1821-1>.
- Federal Aviation Administration, 2017: Safety of flight. Aeronautical Information Manual: Official Guide to Basic Flight Information and ATC Procedures, Federal Aviation Administration Transportation Dept. Doc., https://www.faa.gov/air_traffic/publications/.
- Goldberg, M. D., H. Kilcoyne, H. Cikanek, and A. Mehta, 2013: Joint Polar Satellite System: The United States next

- generation civilian polar-orbiting environmental satellite system. *J. Geophys. Res. Atmos.*, **118**, 13463–13475, <https://doi.org/10.1002/2013JD020389>.
- Hayatbini, N., and Coauthors, 2019: Conditional generative adversarial networks (cGANs) for near real-time precipitation estimation from multispectral GOES-16 satellite imagery—PERSIANN-cGAN. *Remote Sens.*, **11**, 2193, <https://doi.org/10.3390/rs11192193>.
- Haynes, J. M., Y.-J. Noh, S. D. Miller, K. D. Haynes, I. Ebert-Uphoff, and A. Heidinger, 2022: Low cloud detection in multilayer scenes using satellite imagery with machine learning methods. *J. Atmos. Oceanic Technol.*, **39**, 319–334, <https://doi.org/10.1175/JTECH-D-21-0084.1>.
- 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>.
- Huang, H., and Coauthors, 2020: UNet 3+: A full-scale connected UNet for medical image segmentation. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Barcelona, Spain, IEEE, 1055–1059, <https://doi.org/10.1109/ICASSP40776.2020.9053405>.
- Just, D., R. Gutiérrez, F. Roveda, and T. Steenbergen, 2014: Meteosat Third Generation imager: Simulation of the Flexible Combined Imager instrument chain. *Proc. SPIE*, **9241**, 92410E, <https://doi.org/10.1117/12.2066872>.
- Kaplan, M. L., and Coauthors, 2004: Characterizing the severe turbulence environments associated with commercial aviation accidents: A Real-Time Turbulence Model (RTTM) designed for the operational prediction of hazardous aviation turbulence environments. *Meteor. Atmos. Phys.*, **94**, 235–270, <https://doi.org/10.1007/s00703-005-0181-4>.
- Kim, D., M. Gu, T.-H. Oh, E.-K. Kim, and H.-J. Yang, 2021: Introduction of the advanced meteorological imager of Geostationary Operational Environmental Satellite-2a: In-orbit tests and performance validation. *Remote Sens.*, **13**, 1303, <https://doi.org/10.3390/rs13071303>.
- Kim, J.-H., and H.-Y. Chun, 2012: Development of the Korean Aviation Turbulence Guidance (KTG) system using the Operational Unified Model (UM) of the Korea Meteorological Administration (KMA) and pilot reports (PIREPs). *J. Korean Soc. Aviat. Aeronaut.*, **20**, 76–83, <https://doi.org/10.12985/ksaa.2012.20.4.076>.
- , —, R. D. Sharman, and T. L. Keller, 2011: Evaluations of upper-level turbulence diagnostics performance using the Graphical Turbulence Guidance (GTG) system and pilot reports (PIREPs) over East Asia. *J. Appl. Meteor. Climatol.*, **50**, 1936–1951, <https://doi.org/10.1175/JAMC-D-10-05017.1>.
- , R. Sharman, M. Strahan, J. W. Scheck, C. Bartholomew, J. C. H. Cheung, P. Buchanan, and N. Gait, 2018: Improvements in nonconvective aviation turbulence prediction for the World Area Forecast System. *Bull. Amer. Meteor. Soc.*, **99**, 2295–2311, <https://doi.org/10.1175/BAMS-D-17-0117.1>.
- Kim, S.-H., H.-Y. Chun, R. D. Sharman, and S. B. Trier, 2019: Development of near-cloud turbulence diagnostics based on a convective gravity wave drag parameterization. *J. Appl. Meteor. Climatol.*, **58**, 1725–1750, <https://doi.org/10.1175/JAMC-D-18-0300.1>.
- , —, D.-B. Lee, J.-H. Kim, and R. D. Sharman, 2021: Improving numerical weather prediction-based near-cloud aviation turbulence forecasts by diagnosing convective gravity wave breaking. *Wea. Forecasting*, **36**, 1735–1757, <https://doi.org/10.1175/WAF-D-20-0213.1>.
- Knox, J. A., 1997: Possible mechanisms of clear-air turbulence in strongly anticyclonic flows. *Mon. Wea. Rev.*, **125**, 1251–1259, [https://doi.org/10.1175/1520-0493\(1997\)125<1251:PMOCAT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1251:PMOCAT>2.0.CO;2).
- Lagerquist, R., J. Q. Stewart, I. Ebert-Uphoff, and C. Kumler, 2021: Using deep learning to nowcast the spatial coverage of convection from Himawari-8 satellite data. *Mon. Wea. Rev.*, **149**, 3897–3921, <https://doi.org/10.1175/MWR-D-21-0096.1>.
- Lee, D.-B., H.-Y. Chun, S.-H. Kim, R. D. Sharman, and J.-H. Kim, 2022: Development and evaluation of global Korean aviation turbulence forecast systems based on an operational numerical weather prediction model and in situ flight turbulence observation data. *Wea. Forecasting*, **37**, 371–392, <https://doi.org/10.1175/WAF-D-21-0095.1>.
- Lee, Y., C. D. Kummerow, and I. Ebert-Uphoff, 2021: Applying machine learning methods to detect convection using Geostationary Operational Environmental Satellite-16 (GOES-16) Advanced Baseline Imager (ABI) data. *Atmos. Meas. Tech.*, **14**, 2699–2716, <https://doi.org/10.5194/amt-14-2699-2021>.
- Lenz, A., K. Bedka, W. F. Feltz, and S. A. Ackerman, 2009: Convectively induced transverse band signatures in satellite imagery. *Wea. Forecasting*, **24**, 1362–1373, <https://doi.org/10.1175/2009WAF2222285.1>.
- Li, Y., and A. Heidinger, 2021: AWG Cloud Cover Layer algorithm (CCL). NOAA Algorithm Theoretical Basis Doc., version 1.0, 30 pp., https://www.star.nesdis.noaa.gov/jps/documents/ATBD/ATBD_EPS_Cloud_CCL_v1.0.pdf.
- Liu, Q., and Coauthors, 2012: Community Radiative Transfer Model for radiance assimilation and applications. *2012 IEEE Int. Geoscience and Remote Sensing Symp.*, Munich, Germany, IEEE, 3700–3703, <https://doi.org/10.1109/IGARSS.2012.6350612>.
- Marzban, C., 2004: The ROC curve and the area under it as performance measures. *Wea. Forecasting*, **19**, 1106–1114, <https://doi.org/10.1175/825.1>.
- Mason, I. B., 2003: Binary events. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, I. T. Jolliffe and D. B. Stephenson, Eds., Wiley, 37–76.
- Mason, S. J., and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels. *Wea. Forecasting*, **14**, 713–725, [https://doi.org/10.1175/1520-0434\(1999\)014<0713:CPROCA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2).
- McCann, D. W., 2001: Gravity waves, unbalanced flow, and aircraft clear air turbulence. *Natl. Wea. Dig.*, **25** (1–2), 3–14, <http://nwafiles.nwas.org/digest/papers/2001/Vol25No12/Pg3-McCann.pdf>.
- Mecikalski, J. R., W. M. MacKenzie Jr., M. Koenig, and S. Muller, 2010: Cloud-top properties of growing cumulus prior to convective initiation as measured by Meteosat Second Generation. Part I: Infrared fields. *J. Appl. Meteor. Climatol.*, **49**, 521–534, <https://doi.org/10.1175/2009JAMC2344.1>.
- Monette, S. A., and J. M. Sieglaff, 2014: Probability of convectively induced turbulence associated with geostationary satellite-inferred cloud-top cooling. *J. Appl. Meteor. Climatol.*, **53**, 429–436, <https://doi.org/10.1175/JAMC-D-13-0174.1>.
- Muñoz-Esparza, D., and R. Sharman, 2018: An improved algorithm for low-level turbulence forecasting. *J. Appl. Meteor. Climatol.*, **57**, 1249–1263, <https://doi.org/10.1175/JAMC-D-17-0337.1>.
- , —, and W. Deierling, 2020: Aviation turbulence forecasting at upper levels with machine learning techniques based on regression trees. *J. Appl. Meteor. Climatol.*, **59**, 1883–1899, <https://doi.org/10.1175/JAMC-D-20-0116.1>.

- Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, [https://doi.org/10.1175/1520-0434\(1993\)008<0281:WIAGFA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2).
- Pearson, J. M., and R. D. Sharman, 2017: Prediction of energy dissipation rates for aviation turbulence. Part II: Nowcasting convective and nonconvective turbulence. *J. Appl. Meteor. Climatol.*, **56**, 339–351, <https://doi.org/10.1175/JAMC-D-16-0312.1>.
- Reap, R. M., 1996: Probability forecasts of clear-air-turbulence for the contiguous US. NWS Tech. Procedures Bulletin 430, 18 pp., https://www.nws.noaa.gov/mdl/pubs/Documents/TechProcBulls/TPB_430.pdf.
- Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention*, N. Navab et al., Eds., Lecture Notes in Computer Science, Vol. 9351, Springer, 234–241.
- Schmit, T. J., P. Griffith, M. M. Gunshor, J. M. Daniels, S. J. Goodman, and W. J. Lebar, 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>.
- Schwartz, B., 1996: The quantitative use of PIREPs in developing aviation weather guidance products. *Wea. Forecasting*, **11**, 372–384, [https://doi.org/10.1175/1520-0434\(1996\)011<0372:TQUOPI>2.0.CO;2](https://doi.org/10.1175/1520-0434(1996)011<0372:TQUOPI>2.0.CO;2).
- Sharman, R., and T. Lane, 2016: *Aviation Turbulence: Processes, Detection, Prediction*. Springer International, 523 pp.
- , and J. M. Pearson, 2017: Prediction of energy dissipation rates for aviation turbulence. Part I: Forecasting nonconvective turbulence. *J. Appl. Meteor. Climatol.*, **56**, 317–337, <https://doi.org/10.1175/JAMC-D-16-0205.1>.
- , C. Tebaldi, G. Wiener, and J. Wolff, 2006: An integrated approach to mid- and upper-level turbulence forecasting. *Wea. Forecasting*, **21**, 268–287, <https://doi.org/10.1175/WAF924.1>.
- , L. B. Cornman, G. Meymaris, J. Pearson, and T. Farrar, 2014: Description and derived climatologies of automated in situ eddy-dissipation-rate reports of atmospheric turbulence. *J. Appl. Meteor. Climatol.*, **53**, 1416–1432, <https://doi.org/10.1175/JAMC-D-13-0329.1>.
- Sieglauff, J. M., L. M. Crouce, 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, <https://doi.org/10.1175/2010JAMC2496.1>.
- Weng, F., 2007: Advances in radiative transfer modeling in support of satellite data assimilation. *J. Atmos. Sci.*, **64**, 3799–3807, <https://doi.org/10.1175/2007JAS2112.1>.
- Wimmers, A. J., and J. L. Moody, 2004: Tropopause folding at satellite-observed spatial gradients: 1. Verification of an empirical relationship. *J. Geophys. Res.*, **109**, D19306, <https://doi.org/10.1029/2003JD004145>.
- Yang, J., Z. Zhang, C. Wei, F. Lu, and Q. Guo, 2017: Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4. *Bull. Amer. Meteor. Soc.*, **98**, 1637–1658, <https://doi.org/10.1175/BAMS-D-16-0065.1>.