

Toward Operational Real-Time Identification of Frontal Boundaries Using Machine Learning

ANDREW D. JUSTIN,^a COLIN WILLINGHAM,^a AMY MCGOVERN,^a AND JOHN T. ALLEN^b

^a *School of Meteorology, University of Oklahoma, Norman, Oklahoma*

^b *Department of Earth and Atmospheric Sciences, Central Michigan University, Mount Pleasant, Michigan*

(Manuscript received 8 July 2022, in final form 22 April 2023, accepted 4 May 2023)

ABSTRACT: We present and evaluate a deep learning first-guess front-identification system that identifies cold, warm, stationary, and occluded fronts. Frontal boundaries play a key role in the daily weather around the world. Human-drawn fronts provided by the National Weather Service's Weather Prediction Center, Ocean Prediction Center, Tropical Analysis and Forecast Branch, and Honolulu Forecast Office are treated as ground-truth labels for training the deep learning models. The models are trained using ERA5 data with variables known to be important for distinguishing frontal boundaries, including temperature, equivalent potential temperature, and wind velocity and direction at multiple heights. Using a 250-km neighborhood over the contiguous U.S. domain, our best models achieve critical success index scores of 0.60 for cold fronts, 0.43 for warm fronts, 0.48 for stationary fronts, 0.45 for occluded fronts, and 0.71 using a binary classification system (front/no front), whereas scores over the full unified surface analysis domain were lower. For cold and warm fronts and binary classification, these scores significantly outperform prior baseline methods that utilize 250-km neighborhoods. These first-guess deep learning algorithms can be used by forecasters to locate frontal boundaries more effectively and expedite the frontal analysis process.

SIGNIFICANCE STATEMENT: Fronts are boundaries that affect the weather that people experience daily. Currently, forecasters must identify these boundaries through manual analysis. We have developed an automated machine learning method for detecting cold, warm, stationary, and occluded fronts. Our automated method provides forecasters with an additional tool to expedite the frontal analysis process.

KEYWORDS: Fronts; Artificial intelligence; Deep learning; Neural networks

1. Introduction

Identifying frontal boundaries, or boundaries between air masses, is important as fronts can lead to the creation of hazardous weather conditions for a variety of end users (Schultz and Vaughan 2011; Maddox et al. 1980; Childs and Schumacher 2019). The National Weather Service (NWS) forecasts fronts over much of the western Northern Hemisphere, spanning from the Pacific Ocean to the Atlantic, and extending to the edge of Europe. The NWS Weather Prediction Center (WPC), Ocean Prediction Center (OPC), Tropical Analysis and Forecast Branch (TAFB), and Honolulu Forecast Office (HFO) release real-time weather analysis maps every 6 h over this full domain to facilitate forecasting for end users ranging from ocean-going vessels to weather forecast offices. Frontal analysis is currently performed by hand, where forecasters use their best interpretations of a variety of data sources to draw lines on the leading edges of the fronts' thermal and wind gradients. In this work, we take a step toward creating an automated machine learning product that will provide real-time first-guess guidance to

the human forecasters. The goal of this work is to significantly reduce the workload of the human forecasters, freeing them to focus on other decision support tasks.

The approach developed here builds on the work of Lagerquist et al. (2019), who used a convolutional neural network (CNN; Lecun et al. 1998), a type of deep learning algorithm, to identify warm and cold fronts over the contiguous United States (CONUS). While Lagerquist et al. (2019) used a standard CNN architecture with an output of class probabilities for just one pixel, we used the UNet 3+ architecture (Huang et al. 2020), a more complex model structure that utilizes image segmentation to output an entire image of pixelwise predictions. The ability to generate one image with pixelwise predictions as opposed to entirely separate predictions for each pixel in the image drastically reduces the computational resources and time needed to generate predictions across larger domains, helping to meet time-sensitive operational needs. In addition to saving time and resources, the UNet 3+ contains a variety of "skip connections" that help to retain features that may be lost in deeper levels of the model where image sizes can be much smaller than the original input. Using a deep UNet 3+ architecture, we aimed to extract features that can assist with the identification of four types of frontal boundaries: cold, warm, stationary, and occluded.

Frontal boundaries are baroclinic zones that separate two air masses and are identified with a distinct thermal gradient (Renard and Clarke 1965). In general, frontal boundaries are associated with up to 90% of extreme precipitation events in

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/AIES-D-22-0052.s1>.

Corresponding author: Andrew D. Justin, andrewjustin@ou.edu

the midlatitudes of the Northern Hemisphere (Catto and Pfahl 2013), and different frontal types can create various hazards. For example, as a cold front propagates it forces the warmer air ahead of the boundary to rise, creating conditions ideal for convective development (Smith and Reeder 1988; Hobbs et al. 1990; Catto and Pfahl 2013) and secondary frontal cyclogenesis (Zhang et al. 1999). Frontal boundaries are also areas of enhanced relative vorticity, and in eastern Colorado this vorticity helps to generate landspout tornadoes in developing thunderstorms along stationary fronts (Childs and Schumacher 2019). Supercell thunderstorms can become tornadic when interacting with a front, where relative vorticity is also enhanced (Maddox et al. 1980; Markowski et al. 1998). Occluded fronts form when a cold front runs into a warm front and are associated with enhanced precipitation bands (Schultz and Vaughan 2011).

Different analysts have numerous ways of interpreting surface data when drawing frontal boundaries. Renard and Clarke (1965) showed that the variety of interpretations affects the decisions of the forecasters around the globe by gathering analyses from 16 international and national weather centers that contained the location of fronts for the same time step. It was determined that the placement of frontal boundaries could differ by as much as 300 n mi (~555 km) between forecast offices. Uccellini et al. (1992) showed the disagreement that can exist among forecasters through a surface analysis by participants at a National Meteorological Center workshop. Cold fronts usually have significant thermal gradients and wind shifts, whereas a warm front is less likely to show such a significant gradient or wind shift. Sanders (1999) suggested that wind shifts without stark temperature gradients should not be plotted as fronts but rather baroclinic troughs. Stationary fronts are often associated with wind convergence and the presence of an occluded front can be identified with the vertical stacking of geopotential heights. However, if these identifiers are difficult to establish or not significant enough, this can mean that fronts are sometimes not analyzed by forecasters. Machine learning methods could be used to shrink this gap in uncertainty and provide more objectively based predictions; however, the human labels used to train such methods are subjective and may influence the resulting models.

Numerical frontal analysis (NFA) methods provide the advantage of removing human subjectivity in the identification of frontal boundaries, but the process of selecting the rules for each of the methods is subjective in nature. Berry et al. (2011) used the field of wet-bulb potential temperature θ_w at the 850-hPa level to locate frontal boundaries. Berry et al. (2011) found that using θ_w at this level results in fronts associated with pressure troughs, a feature commonly found near frontal boundaries. Simmonds et al. (2012) used the change in 10-m wind over the course of 6 h (in our case, two time steps) to locate fronts. Wind shifts are generally located near pressure troughs due to the changing orientation of the isobars, and results from the θ_w method used by Berry et al. (2011) support the idea of using wind shifts to locate fronts. Schemm et al. (2015) used gradients of equivalent potential temperature θ_E at the 850-hPa level along with the same wind shift method implemented by Simmonds et al. (2012) to identify frontal boundaries. Schemm

et al. (2015) decided to use θ_E instead of θ_w as they claim θ_E gradients are able to better separate air masses of differing thermodynamic characteristics. Hewson (1998) developed a thermal gradient method that builds upon the work of Renard and Clarke (1965) and defines three types of fronts of varying complexity, with a “type 1” front being a straight boundary without a thermal gradient parallel to the front, a “type 2” front being a straight boundary with a thermal gradient parallel to the front, and a “type 3” front being a curved boundary that may or may not have a thermal gradient parallel to the front.

Schemm et al. (2015) compared the wind shift method from Simmonds et al. (2012) and the thermal method from Hewson (1998) to attribute strengths and weaknesses to each method; the wind shift method was not able to reliably locate warm fronts but succeeded in locating cold fronts. However, the wind shift method alone was unable to discern between cold and warm fronts. Schemm et al. (2015) also found that fronts identified by the thermal method typically had a larger zonal component, while those identified by the wind shift method had more of a meridional orientation. Hope et al. (2014) compared six frontal detection methods to study winter rainfall over the central wheatbelt of the southwest of Western Australia—wind shifts, temperature gradients, and θ_w gradients at 850 hPa, as well as a self-organizing map, a manual synoptic technique, and a “pattern matching” method that involved using data from manual synoptic analyses to produce mean patterns for individual types of fronts; they noted that wind shifts at 850 hPa revealed a frontal count time series that had the highest correlation with rainfall, while temperature and θ_w gradients at 850 hPa produced the largest number of fronts across all days analyzed. Thomas and Schultz (2019) found that fronts defined with potential temperature θ resulted in climatologies with larger seasonal variations in frontal frequency over continents in the Northern Hemisphere than climatologies defined by other thermodynamic quantities. These methods only focus on a few variables and their rates of change with space and time, so they do not fully capture the cohesive vertical structures of the atmosphere. This is one of the motivations for research into new methods for frontal detection, particularly with machine learning algorithms.

We build on prior work on developing deep learning approaches for frontal identification. Supervised deep learning algorithms automatically learn how to recognize features within provided input data and require minimal human intervention outside of the initial setup of the algorithm, removing much of the subjectivity associated with human-created rules for objective methods like those described above and returning more consistent answers. Lagerquist et al. (2019) trained CNNs to detect cold and warm fronts by using a combination of thermodynamic variables at the surface and various pressure levels. Biard and Kunkel (2019) trained a 2D CNN with fields of temperature, specific humidity, pressure, and zonal and meridional wind velocities as input. Their CNN was able to detect up to 90% of analyzed fronts over North America for the 2003–15 period. Recent work by Niebler et al. (2022) showed success in detecting cold, warm, stationary, and occluded fronts using a 2D U-Net (Ronneberger et al. 2015). Niebler et al. (2022) trained on nine pressure levels with five input variables; the

TABLE 1. Variables provided in the ERA5 datasets.

Variable	Heights/levels
Geopotential height	1000, 950, 900, and 850 hPa
Temperature	2 m AGL and 1000, 950, 900, and 850 hPa
Dewpoint temperature	2 m AGL
Specific humidity	1000, 950, 900, and 850 hPa
Surface pressure	Surface
u wind	10 m AGL and 1000, 950, 900, and 850 hPa
v wind	10 m AGL and 1000, 950, 900, and 850 hPa

variables were the same as Biard and Kunkel (2019) but with pressure replaced with vertical velocity. In our approach to automated frontal detection, we use the UNet 3+ architecture (Huang et al. 2020), a successor of the U-Net that utilizes a more complex structure to extract features from an input image. We trained UNet 3+ models with 2D and 3D convolution kernels to see which model structure (if any) would be superior.

We use three sets of CNNs implementing the UNet 3+ architecture to predict frontal boundaries. More specifically, one model set predicts the locations of cold and warm fronts, another predicts the locations of stationary and occluded fronts, and the third set simply predicts the locations of any of the four types of fronts via binary classification. In the case of the third set, every cold, warm, stationary, and occluded front is labeled as a “front” rather than using the conventional labels. Assigning the same label to all boundaries allows the models to locate fronts more easily by generalizing all gradients associated with the four frontal types, analogous to earlier frontal detection methods that were not type specific (e.g., Berry et al. 2011; Schemm et al. 2015; Simmonds et al. 2012), but does not allow the models to classify the type of front. This is the motivation behind the setups that classify individual frontal types; classifying the type of front can reduce the time needed for operational forecasters to determine what type of front to draw.

2. Data and methods

a. Data preprocessing

1) PREDICTOR VARIABLES

The predictor variables come from ERA5 data (Hersbach et al. 2018) provided by the European Centre for Medium-Range Weather Forecasts on a $0.25^\circ \times 0.25^\circ$ grid at 3-h time

steps (0000, 0300, ..., 2100 UTC) for the period 2008–20. We accessed the Climate Data Store to gather data for predictor variables listed in Table 1. Geopotential height and pressure were chosen because troughs can be used to locate baroclinic zones at the surface (Payer et al. 2011). The u wind and v wind were chosen because previous studies support the idea of using wind shifts to locate frontal boundaries (Berry et al. 2011; Schemm et al. 2015; Simmonds et al. 2012; Schultz 2005), and a wind shift collocated with a temperature gradient can confirm the presence of a boundary separating two air masses (Schultz 2005). Dewpoint and specific humidity can help to reveal moisture gradients that aid in the identification of warm fronts in the eastern Pacific (Lagerquist et al. 2020), which is partially covered by our CONUS domain [see section 2a(2) below].

Using the variables from Table 1, we calculated variables listed in Table 2 (Bolton 1980; Davies-Jones 2008; Stull 2011); θ_E was added because Schemm et al. (2015) found it to be useful in locating warm fronts, and the finding by Berry et al. (2011) that most thermodynamic fields were able to produce reasonable results against manual synoptic analyses motivated us to calculate the remaining predictor variables (60 variables total). The collinearity between many of the variables allows us to leverage the nonlinearities in the UNet 3+ created by the activation functions, which are functions within convolutional layers that transform a given input or set of inputs, usually a sum of weighted inputs, to generate an output known as a feature map (Chase et al. 2022). However, the inputs to the activation functions in our U-Nets are sets of normalized feature maps (see the UNet 3+ architecture in section 2b).

2) FRONTAL OBJECTS

Data for frontal boundaries, which are treated as the ground truth for the deep learning models, were generated by forecasters at the WPC, OPC, TAFB, and HFO offices (NOAA 2023). The WPC and TAFB datasets contain fronts across North America, including the CONUS, and are generated every 3 h at the same time steps as the retrieved ERA5 datasets for thermodynamic variables. The unified surface analyses are generated at synoptic hours (0000, 0600, 1200, and 1800 UTC) and encompass a domain whose extent ranges from the equator to 80°N and from 130°E eastward to 10°E (Fig. 1). The unified surface analysis domain contains fronts drawn by forecasters from the four offices listed above and covers a much larger portion of the Northern Hemisphere than

TABLE 2. Variables calculated from the provided variables in Table 1.

Variable	Heights/levels
Dewpoint temperature	1000, 950, 900, and 850 hPa
Equivalent potential temperature	2 m AGL and 1000, 950, 900, and 850 hPa
Mixing ratio	2 m AGL and 1000, 950, 900, and 850 hPa
Relative humidity	10 m AGL and 1000, 950, 900, and 850 hPa
Specific humidity	2 m AGL
Virtual temperature	2 m AGL and 1000, 950, 900, and 850 hPa
Wet-bulb temperature	2 m AGL and 1000, 950, 900, and 850 hPa
Wet-bulb potential temperature	2 m AGL and 1000, 950, 900, and 850 hPa

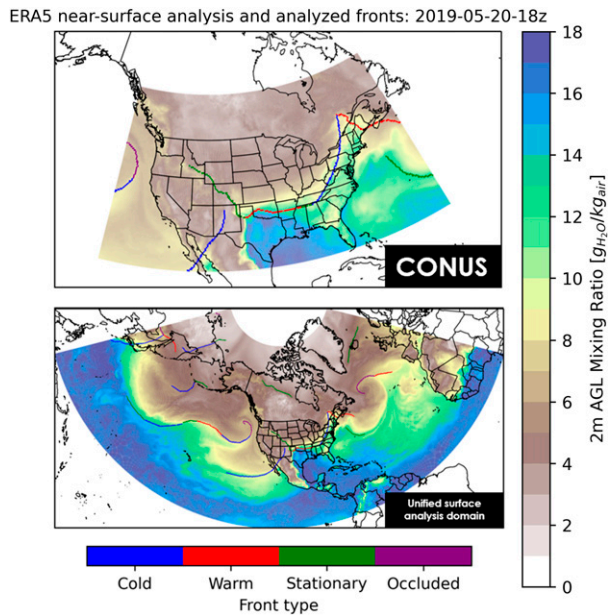


FIG. 1. The (top) CONUS domain ranges from 25° to 56.75° N and from 132° to 60.25° W, and the (bottom) unified surface analysis domain ranges from the equator to 80° N and from 130° E eastward to 10° E. The domains are shown with fronts from the unified surface analysis plotted on top of mixing ratio at 2 m AGL at 1800 UTC 20 May 2019.

the Coded Surface Bulletin (National Weather Service 2019) used in automated front-detection methods from Niebler et al. (2022) and Lagerquist et al. (2019). Using a larger domain provided the opportunity to train our models on fronts with more variability in the predictor variables.

A limitation of the data in this format is that the fronts are marked vertices along the drawn boundaries, and these vertices may be separated by several hundred kilometers. To create more continuous frontal boundaries, we interpolated these points every 25 km and translated the interpolated fronts to a $0.25^{\circ} \times 0.25^{\circ}$ grid to match the grid of the ERA5 data. The interpolation process started with translating the coordinates of the vertices to an x/y grid using the haversine great circle formula, where the components of the resulting vertices had units of kilometers. Points were then added in 25-km increments along straight lines connecting pairs of the original vertices and

translated back to a latitude–longitude coordinate system by inverting the haversine formula. As infinitely thin fronts are physically inconsistent with the atmospheric process (Sanders 1999), we expanded the interpolated fronts by 25 km in all directions. Expanding the fronts provided the models with a larger area to identify various hypergradients and helped to alleviate any data displacements that resulted from the ERA5 data assimilation process or inconsistencies in the locations of analyzed frontal boundaries. 3D labels were created for our 3D models by taking the interpolated fronts and duplicating their locations at every pressure level to help the 3D models identify features above the surface.

The CONUS datasets have a domain that ranges from 25° to 56.75° N and from 132° to 60.25° W (Fig. 1), which works out to 288×128 pixels on the $0.25^{\circ} \times 0.25^{\circ}$ grid. We designed our models for images with sizes of 128×128 , so there are 161 unique images that can be subsampled from each CONUS dataset. Fronts from WPC and TAFB are included in our CONUS datasets.

Another challenge that we faced with the frontal objects was dealing with the varying sample sizes of the frontal types (see Table 3). Although there is an increase in the size or number of the fronts over time, the change is relatively small and it is unknown as to why these increases in sample sizes are observed as the forecast offices do not monitor the amount of fronts being analyzed. Figure 2 shows the climatology of all fronts in our data; cold fronts were the most common frontal type, followed by stationary, warm, and occluded fronts. Figure 2 also shows cutoffs in frontal frequencies that mark the boundaries between the offices' individual domains. We discuss how we handle these boundaries during training in the section on model structure and hyperparameters (section 2c). To prevent models from becoming too biased toward the “no front” category, we did not train the models on images that had fewer than five pixels with frontal boundaries. The dropped images accounted for approximately 5% of the training and validation datasets.

b. UNet 3+ architecture

The UNet 3+ is a CNN designed by Huang et al. (2020) for image segmentation. The UNet 3+ builds upon the U-Net (Ronneberger et al. 2015) and the UNet++ (Zhou et al. 2018) by adding more connections while at the same time reducing the number of parameters in the model. Details on these connections can be found later in this section.

TABLE 3. Datasets and frontal counts for the models across the CONUS domain. The frontal sample sizes represent the percent of pixels containing each specific frontal type after the interpolated fronts have been expanded by 25 km.

Model frontal types	Dataset (yr)	Images	No front	Cold	Warm	Stationary	Occluded
Cold and warm	Training (2008–16)	3 600 121	98.75%	0.96%	0.29%	—	—
	Validation (2017–18)	825 930	98.55%	1.10%	0.34%	—	—
	Test (2019–20)	941 528	98.12%	1.41%	0.47%	—	—
Stationary and occluded	Training (2008–16)	3 834 054	99.07%	—	—	0.79%	0.13%
	Validation (2017–18)	820 456	98.91%	—	—	0.96%	0.13%
	Test (2019–20)	941 528	98.63%	—	—	1.21%	0.16%
All (front/no front)	Training (2008–16)	3 660 657	97.86%	0.93%	0.29%	0.78%	0.13%
	Validation (2017–18)	845 250	97.52%	1.06%	0.34%	0.94%	0.13%
	Test (2019–20)	941 528	96.78%	1.40%	0.46%	1.19%	0.16%

Frequency of front types: 2008-2020

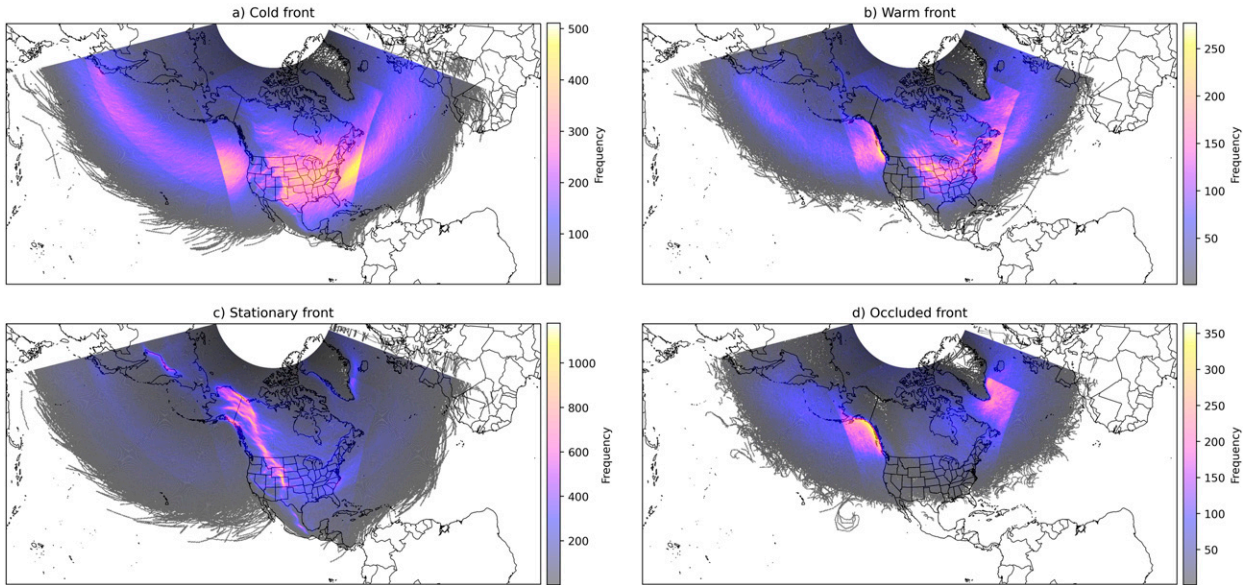


FIG. 2. Frequency of fronts from 2008 to 2020: (a) cold fronts, (b) warm fronts, (c) stationary fronts, and (d) occluded fronts.

The general structure of our UNet 3+ models can be visualized with Fig. 3. The 2D models were designed for images with sizes of 128×128 grid boxes, or $32^\circ \times 32^\circ$, and the 3D models were designed for images with $128 \times 128 \times 5$ grid boxes, with the third dimension representing the number of levels. Six encoder nodes and five decoder nodes were used in all model architectures. The number of parameters in the three model structures can be found in Table 4. In each encoder node, the images are passed five “modules,” where we define a module to be a convolution layer followed by a batch

normalization layer and a ReLU activation function layer. Since convolutions normally reduce the dimensions of the input images, we implemented zero padding in the convolution layers to preserve the configuration of the images throughout the models. Zero padding involves padding the borders of an image such that the output of a convolution layer has the same dimensionality as the input (O’Shea and Nash 2015). Figure 4 illustrates how image size can be preserved with zero padding. The images are passed through 64 (16 for 3D) convolution filters in the convolution layers within the first encoder node,

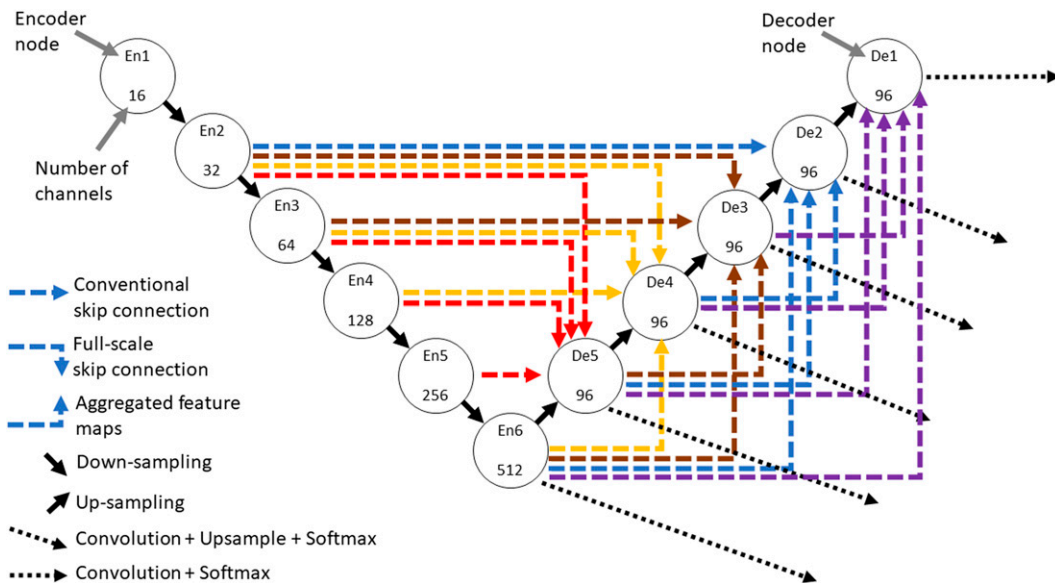


FIG. 3. UNet 3+ model for identifying frontal boundaries. Note the number of channels doubling with each subsequent encoder node and a constant number of channels throughout the decoder nodes.

TABLE 4. Number of parameters in the three model structures.

Structure	No. of parameters
2D: 3×3 kernel	268.8 million
3D: $3 \times 3 \times 3$ kernel	50.5 million
3D: $5 \times 5 \times 5$ kernel	233.5 million

and the number of filters doubles with each subsequent encoder node. The ReLU activation function is defined as

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}. \quad (1)$$

The batch normalization layers stabilize the UNet 3+ by keeping the mean of the inputs close to 0 and the standard deviation close to 1. After passing through the five modules, the images undergo downsampling, which involves passing the images through a max-pooling layer with a pool size of 2×2 ($2 \times 2 \times 1$ for 3D). The module and downsampling process repeats throughout the left side of the UNet 3+ until reaching the sixth encoder node at the bottom of the model where the image size is 4×4 ($4 \times 4 \times 5$ for 3D). We used a pool size of $2 \times 2 \times 1$ in the 3D max-pooling layers rather than $2 \times 2 \times 2$ because the third dimension of the 3D images that represents the pressure levels consists of just five pixels, and implementing a $2 \times 2 \times 2$ pool size would result in losing data in the pressure levels and a UNet 3+ with limited depth.

The UNet 3+ contains conventional and full-scale skip connections along with aggregated feature maps that help to preserve features lost during the downsampling and upsampling operations. Conventional skip connections pass the output from each encoder node through one module, with the convolution layers containing 64 (16 for 3D) filters, and connect it the respective decoder node on the same level of the UNet 3+. Full-scale skip connections connect encoder nodes to decoder nodes at lower levels in the model and send the output from the encoder nodes through max-pooling operations followed by one module with the convolution layers containing 64 (16 for 3D) filters. The pool size in the max-pooling operations in the full-scale skip connections depends on the levels

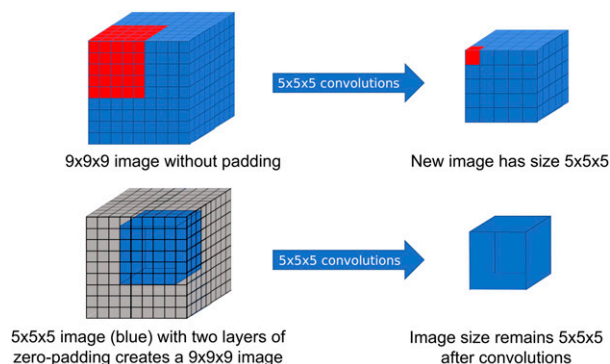


FIG. 4. Diagrams of (top) $5 \times 5 \times 5$ convolutions performed on a $9 \times 9 \times 9$ with no padding and (bottom) $5 \times 5 \times 5$ convolutions performed on a $5 \times 5 \times 5$ image with two layers of zero padding.

of the encoder and decoder nodes. Aggregated feature maps connect the bottom encoder node and decoder nodes to decoder nodes located at higher levels on the right side of the UNet 3+. Like full-scale skip connections, these feature maps pass an image through a module where the convolution layers contain 64 (16 for 3D) filters, but upsampling is used in place of max-pooling and the pool size of each upsampling layer is dependent on the levels of the nodes that are connected to the feature maps (Huang et al. 2020).

The right side of the UNet 3+ contains decoder nodes where images from the upsampling operations, aggregated feature maps, and the conventional and full-scale skip connections are concatenated and sent through five modules. All convolution layers in the decoder nodes have 384 (96 for 3D) filters. The output of the sixth encoder node and each decoder node is passed through a convolution layer with a number of filters equal to the number of classes for the model, followed by a Softmax (Bridle 1989) activation function layer. During training, the outputs of all Softmax layers in the model are summed to produce a loss value. The Softmax function in the final decoder node (denoted as De1) will produce an image with each grid point containing individual probabilities ranging from 0 to 1 for each type of front.

c. Model structure and hyperparameters

Three sets of UNet 3+ models were used to detect frontal boundaries—one set predicted cold and warm fronts (CF/WF), another predicted stationary and occluded fronts (SF/OF), and the last set predicted any frontal type by assigning all types the same label such that the model performs binary classification [front/no front (F/NF)]. These setups were chosen because cold and warm fronts were used by Lagerquist et al. (2019) and are often related during the evolution of surface cyclones (Bjerknes 1919), and we believed that limiting the number of classes to two (excluding the no-front class) would allow the models to better generalize features associated with each type of front and avoid the need for class balancing. We tested three structures for the three sets of models (total of nine individual models): 2D models with kernel sizes of 3×3 , 3D models with kernel sizes of $3 \times 3 \times 3$, and 3D models with kernel sizes of $5 \times 5 \times 5$.

All models were trained with a batch size of 32 and 20 steps per epoch over the CONUS domain. The models struggled to learn from training over the full unified surface analysis domain, possibly due to fronts over the ocean having different thermodynamic properties than those over land (see section 3) and the fact that the unified surface analyses are done every 6 h, whereas WPC and TAFB perform analyses over North America every 3 h, leading to a lower density of fronts outside of the WPC/TAFB domain that encompasses our CONUS domain (see Figs. 1 and 2). Validation took place every epoch with a batch size of 32 and 20 steps. The models continued training and validation until they completed 6000 epochs or reached a point where the losses did not improve for 1000 epochs, whichever came

first. The optimizer used was Adam (Kingma and Ba 2014) with a learning rate of 1×10^{-4} .

We used the fractions skill score (FSS) as the loss function for the models. Originally designed by Roberts (2008), FSS serves as a spatial verification measure that takes into account surrounding pixels when evaluating a forecast. FSS is defined by Roberts (2008) as

$$\text{FSS} = 1 - \frac{\text{FBS}}{\text{FBS}_{\text{worst}}}. \quad (2)$$

FBS is the fractions Brier score, and $\text{FBS}_{\text{worst}}$ is the worst possible forecast or the largest possible FBS. FSS is equal to 1 with a perfect forecast, and a value of 0 indicates that the model has made the worst possible forecast. Using the original definitions of FBS and $\text{FBS}_{\text{worst}}$, we can write Eqs. (3)–(6) to represent how FSS is calculated in our models:

$$\begin{aligned} \text{FBS}_{2\text{D}} &= \frac{1}{N_{\text{lat}}} \frac{1}{N_{\text{lon}}} \sum_{\text{lat}} \sum_{\text{lon}} [F_{\text{pred}}(\text{lat}, \text{lon}) \\ &\quad - F_{\text{true}}(\text{lat}, \text{lon})]^2, \end{aligned} \quad (3)$$

$$\begin{aligned} \text{FBS}_{3\text{D}} &= \frac{1}{N_{\text{lat}}} \frac{1}{N_{\text{lon}}} \frac{1}{N_z} \sum_{\text{lat}} \sum_{\text{lon}} \sum_z [F_{\text{pred}}(\text{lat}, \text{lon}, z) \\ &\quad - F_{\text{true}}(\text{lat}, \text{lon}, z)]^2, \end{aligned} \quad (4)$$

$$\begin{aligned} \text{FBS}_{\text{worst}(2\text{D})} &= \frac{1}{N_{\text{lat}}} \frac{1}{N_{\text{lon}}} \left[\sum_{\text{lat}} \sum_{\text{lon}} F_{\text{pred}}^2(\text{lat}, \text{lon}) \right. \\ &\quad \left. + \sum_{\text{lat}} \sum_{\text{lon}} F_{\text{true}}^2(\text{lat}, \text{lon}) \right], \quad \text{and} \end{aligned} \quad (5)$$

$$\begin{aligned} \text{FBS}_{\text{worst}(3\text{D})} &= \frac{1}{N_{\text{lat}}} \frac{1}{N_{\text{lon}}} \frac{1}{N_z} \left[\sum_{\text{lat}} \sum_{\text{lon}} \sum_z F_{\text{pred}}^2(\text{lat}, \text{lon}, z) \right. \\ &\quad \left. + \sum_{\text{lat}} \sum_{\text{lon}} \sum_z F_{\text{true}}^2(\text{lat}, \text{lon}, z) \right]. \end{aligned} \quad (6)$$

To use FSS as a loss function that represents the model's error in making predictions, we can use the following formula:

$$\text{FSS}_{\text{loss}} = 1 - \text{FSS}. \quad (7)$$

Our FSS loss function [Eq. (7)] behaves opposite to that of the FSS metric [Eq. (2)], so FSS_{loss} will be 0 when the model makes a perfect forecast and equal to 1 when the worst possible forecast is made. We used a mask size of 3×3 ($75 \text{ km} \times 75 \text{ km}$) for the 2D models and $3 \times 3 \times 3$ ($75 \text{ km} \times 75 \text{ km} \times 3$ levels) for the 3D models. Readers are directed to Roberts (2008) and Ebert-Uphoff et al. (2021) for more information on the FSS loss function.

d. Evaluation and postprocessing

To evaluate the performance of our models, we first tested them over both the CONUS and the full domain. The entire

testing set (2019–20) was used during the evaluations over the CONUS, but only one-half of the dataset covering the synoptic hours (0000, 0600, 1200, and 1800 UTC) was used over the full domain as these are the hours when the unified surface analyses over the full domain are available. Since the models output images of size 128×128 along the horizontal dimensions, we needed to generate three images for each time step to cover the 288×128 CONUS domain. The process of generating these predictions varies between the 2D and 3D models [see more below in section 2d(1)]. For the CF/WF models, we evaluate the critical success index (CSI, described in more detail below) of each model and compare it with a baseline method implemented by Lagerquist et al. (2019) over North America. We compare our F/NF models with another baseline method implemented by Niebler et al. (2022) over the NWS domain. To our knowledge there are no objective NFA methods for stationary or occluded fronts, so no baselines are available for either frontal type.

1) CREATING PREDICTIONS

To create our CONUS predictions, the models made three separate predictions over the domain. The three predictions were made with extents of 132° – 100.25°W , 112° – 80.25°W , and 92° – 60.25°W . The extents of the predictions imply that parts of the images would overlap one another and a grid point where two images overlap would see the first image “behind” the second image, such that the second image would overwrite any data in the first image. To prevent this, the maximum probability for each frontal type was taken at each grid point where the images overlapped. This means that if two images overlap at a particular grid point and the first image has a probability of 0.80 for a cold front and the second has a probability of 0.50 for a cold front, the final prediction will show 0.80 at that grid point for the probability of a cold front. This procedure also helps to mitigate a well-known issue with CNNs structure with zero padding: pixels near the edge of an output image from a model can be unrealistic as the model can become “confused” by the layers of zeroes surrounding the input image (Liu et al. 2018). This process was repeated for the full domain, where 24 images were stitched together to create the final prediction across the domain. We also evaluated them over the full domain with 90 images to see if larger overlap between images would create a prediction with better results. Using additional images to create a final prediction results in larger overlap between the images, and the extra overlap means that pixels closer to the centers of the images are used to stitch the images together, helping to alleviate the near-edge issues brought on by zero padding as described above. For the 3D models we needed to transform the predictions to a 2D spatial grid before taking the maximum of overlapping pixels (see above). To turn the 3D predictions into 2D images, we took the maximum probability of each frontal type over all pressure levels at every point across the domain.

2) CSI

After making predictions over the whole CONUS domain for each available time step, we counted the number of true

positives (TP), false positives (FP), and false negatives (FN) associated with each frontal type at probability thresholds from 0.01 to 1 in steps of 0.01 (0.01, 0.02, 0.03, ..., 1) to calculate the CSI (Donaldson et al. 1975) for each frontal type at each probability threshold. CSI is defined as

$$CSI = \frac{TP}{TP + FP + FN} = \frac{1}{\frac{1}{POD} + \frac{1}{SR} - 1}, \quad (8)$$

where POD is the probability of detection and SR is the success ratio:

$$POD = \frac{TP}{TP + FN} \quad \text{and} \quad (9)$$

$$SR = \frac{TP}{TP + FP}. \quad (10)$$

When evaluating the CSI pixel by pixel, any fronts that the model predicts that are offset from their true locations will be marked as FP and will bring down the CSI score. To prevent the model from being penalized for “close” predictions, we calculated the CSI using 50-, 100-, 150-, 200-, and 250-km neighborhoods. This process is illustrated in Fig. 5.

For each model architecture, we calculated the 95% confidence interval of the CSI for each frontal type at 50, 100, 150, 200, and 250 km at each probability threshold (0.01, 0.02, 0.03, ..., 1) through bootstrapping. We resampled the statistics with replacement 10000 times, each time obtaining a set of performance statistics containing a number of time steps equal to the number of time steps in each test dataset (5848) and calculating the probability of detection [POD; Eq. (9)] and success ratio [SR; Eq. (10)] of each individual time step (58.48 million total samples). The 2.5th and 97.5th percentiles were obtained from the POD and SR statistics calculated from the resampled datasets to obtain the final confidence intervals for each model architecture and frontal type at every neighborhood and probability threshold.

3) FB

Another metric for assessing model performance is known as frequency bias (FB), as defined in Eq. (11) below. In an ideal situation, a value of 1 for FB would be preferred. When FB is equal to 1, the number of false positives and false negatives are equal, or there is one false positive for every false negative or missed target. If FB is greater than 1, there are more false positives than false negatives, and vice versa when FB is less than 1:

$$FB = \frac{TP + FP}{TP + FN}. \quad (11)$$

3. Results

a. Model performance

The results from the model evaluations across the testing set (2019–20; see Table 3) are summarized in Tables 5 and 6

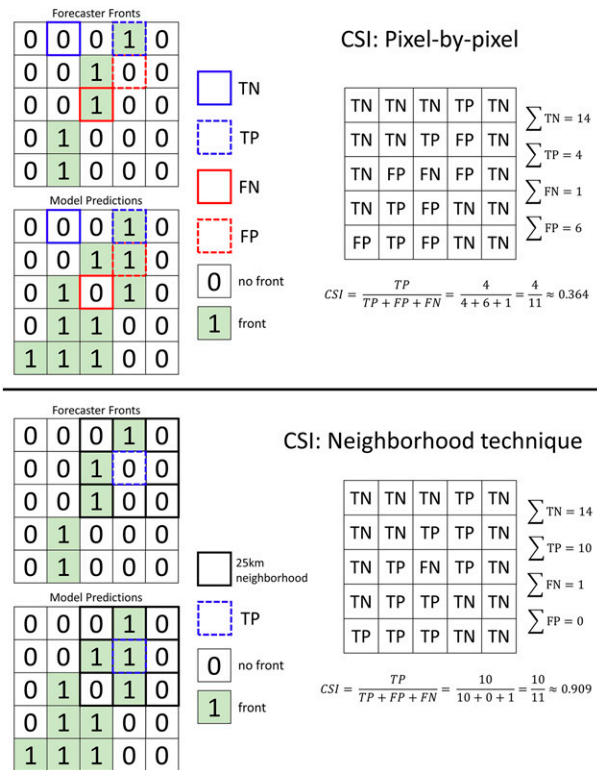


FIG. 5. Comparison of CSI calculation methods. Using the neighborhood technique, the model is not penalized for predicting a front that is slightly offset from its true location and results in a higher CSI (TN = true negative, TP = true positive, FN = false negative, and FP = false positive).

and in Fig. 6, Figs. A1–A14 in the appendix, and Figs. S1–S30 in the online supplemental material. We have provided the figure for 3×3 cold-frontal performance over the CONUS in the results because cold fronts are the most common type of weather front; the rest of the 3×3 figures are provided in the appendix. Figures for the 3D architectures are provided as online supplemental material because they showed overall weaker performance than the 2D models. Overall, all models saw CSI scores increase as the neighborhood grew larger, which was expected and consistent with the findings of Lagerquist et al. (2019) and Niebler et al. (2022).

TABLE 5. The 100-km (first three rows) and 250-km (second three rows) CSI scores across the CONUS. Boldface numbers represent the highest CSI score for each frontal type or model type.

Model architecture	Cold	Warm	Stationary	Occluded	Binary (F/NF)
2D (3×3)	0.48	0.30	0.33	0.33	0.51
3D ($3 \times 3 \times 3$)	0.35	0.20	0.24	0.19	0.47
3D ($5 \times 5 \times 5$)	0.44	0.26	0.30	0.29	0.53
2D (3×3)	0.60	0.43	0.48	0.45	0.71
3D ($3 \times 3 \times 3$)	0.48	0.32	0.37	0.27	0.66
3D ($5 \times 5 \times 5$)	0.55	0.36	0.44	0.39	0.71

TABLE 6. The 250-km CSI scores across the full domain with 24 images (first three rows) and 90 images (second three rows). Boldface numbers represent the highest CSI score for each frontal type.

Model architecture	Cold	Warm	Stationary	Occluded	Binary (F/NF)
2D (3 × 3)	0.48	0.38	0.27	0.41	0.48
3D (3 × 3 × 3)	0.39	0.29	0.25	0.27	0.55
3D (5 × 5 × 5)	0.50	0.35	0.26	0.30	0.54
2D (3 × 3)	0.49	0.40	0.29	0.42	0.51
3D (3 × 3 × 3)	0.43	0.29	0.26	0.29	0.57
3D (5 × 5 × 5)	0.54	0.37	0.27	0.31	0.56

1) CONUS

The 3D models with 3 × 3 × 3 convolutions performed worse than the other model architectures over the CONUS across all frontal types and neighborhoods. This was an unexpected finding as the authors anticipated that the 3D convolutions would outperform the 2D convolutions of the same kernel size. We have two possible explanations for this finding. First, the 3D models with 3 × 3 × 3 kernels had less than one-fifth of the number of parameters of the 2D models with 3 × 3 kernels (see Table 4). Fewer parameters can introduce the problem of underfitting, which occurs when a model does

not have enough complexity to describe given sets of data. Second, the 3 × 3 × 3 kernels only allowed the model to process images from three levels at once (e.g., 1000, 950, and 900 hPa), so it seems probable that a much larger dataset is needed to train the 3D models with 3 × 3 × 3 kernels to capture more feature variance associated with each frontal type. Although the 2D models do not process volumetric data with 3D convolutions, it is possible the large number of parameters helps the 2D structure outperform the 3D models with 3 × 3 × 3 kernels.

The 2D CF/WF model outperformed the 3D architectures with cold and warm fronts at all neighborhood sizes. The vertical depth of warm fronts can change by as little as 1 km over a distance of 100 km (Heymfield 1979), and cold fronts can often be identified through wind shifts (Simmonds et al. 2012), so perhaps 3D convolutions processing volumetric data are not necessary to identify cold or warm fronts. It is possible that the models struggle to recognize features in the vertical structure of the shallower warm fronts, but they still show skill in warm-frontal detection. The small sample size of warm fronts than of cold fronts (see Table 3) could also explain warm fronts’ weaker performance relative to cold fronts.

Stationary and occluded fronts saw the best performance with the 2D 3 × 3 architecture. As mentioned earlier, the 3 × 3 × 3 models performed worse with all frontal types, but

2D U-Net 3+ (3x3 kernel): Cold fronts over CONUS domain (3 images per map)

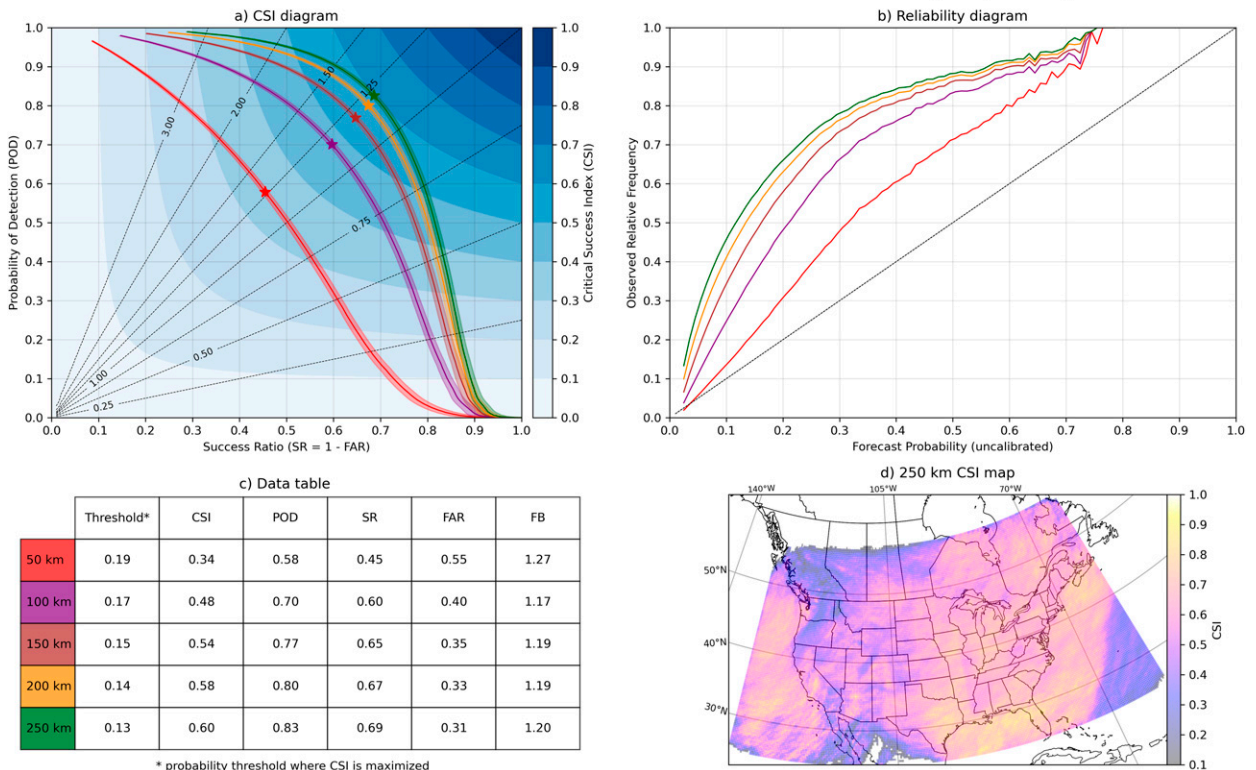


FIG. 6. (a) CSI diagram (dashed lines = frequency bias), (b) reliability diagram (dashed line = perfect reliability), (c) data table, and (d) 250-km CSI map for cold fronts using the 2D CF/WF model with 3 × 3 convolutions over the CONUS. Colored lines in (a) and (b) correspond to neighborhoods indicated by shaded cells in (c).

performance was especially low for the $3 \times 3 \times 3$ SF/OF model. The 250-km CSI scores for stationary and occluded fronts over the CONUS were 0.37 and 0.27, and the 2D model achieved scores of 0.48 and 0.45. The smaller convolutions in the 3D $3 \times 3 \times 3$ architecture (as opposed to the $5 \times 5 \times 5$ architecture) may not be able to reliably extract features in stationary and occluded fronts, but it is not clear as to why the $3 \times 3 \times 3$ SF/OF model underperformed to such a degree. These findings are consistent with results from Niebler et al. (2022), who achieved CSI scores of 0.45 and 0.49 for stationary and occluded fronts over the NWS domain when the NWS domain was also used as the training region for the model. The 250-km CSI scores for stationary fronts were highest over the Rockies (Fig. A2d in the appendix and Figs. S3d and S18d in the online supplemental material), where the frequency of stationary fronts was highest (Fig. 2c). Occluded fronts performed best over portions of the United States and Canada just east of the Rockies (Fig. A3d in the appendix and Figs. S4d and S19d in the online supplemental material), which makes sense given that mature cyclones are often seen at these locations following lee cyclogenesis (Bannon 1992).

The any fronts setup (F/NF) showed excellent performance with the 2D 3×3 and 3D $5 \times 5 \times 5$ models (Figs. A4a,c in the appendix and Figs. S20a,c in the online supplemental material). Performance did not decrease as much with the $3 \times 3 \times 3$ architecture on the F/NF setup as it did with the individual frontal types in the CF/WF and SF/OF models (Figs. S1a,c–S5a,c in the online supplemental material). Since all fronts are given the same label, it is likely that the $3 \times 3 \times 3$ convolutions were able to better capture gradients associated with the binary fronts than the individual frontal types. The best F/NF models (3×3 and $5 \times 5 \times 5$) both achieved a CSI score of 0.71 at 250 km across our CONUS domain, matching the performance of the U-Net used by Niebler et al. (2022) with a CSI of 0.67 across the NWS domain and outperforming the baseline 250-km CSI of 0.22 from Niebler et al. (2022). The binary-front performance from our models was relatively consistent across the CONUS domain, with the exception of the West Coast of the United States.

The 2D 3×3 CF/WF architecture achieved a 100-km CSI score of 0.44 across the CONUS domain. This final CSI score is heavily biased to cold fronts; warm fronts are far less prevalent in the testing data than cold fronts (see Table 3). When performing evaluations with a 250-km neighborhood, the CSI score reached 0.57. The small change in CSI scores indicates that the models derive most of their skill from the smaller evaluation neighborhoods. Our CF/WF results are consistent with the findings of Lagerquist et al. (2019) and Niebler et al. (2022). The CNN used by Lagerquist et al. (2019) to detect cold and warm fronts achieved an overall CSI score of 0.52 using a 250-km neighborhood over North America, while the U-Net used by Niebler et al. (2022) achieved 250-km CSI scores of 0.56 and 0.37 for cold and warm fronts, respectively. The CF/WF models drastically outperform the baseline NFA method used by Lagerquist et al. (2019) over North America that achieved a 250-km CSI score of 0.23. Cold and warm fronts both performed worse over the Rockies than other parts of the CONUS domain. The frequency of cold and

warm fronts over the Rockies is much lower than the rest of the CONUS domain (Figs. 2a,b), so the CF/WF models may not have been able to fully capture the structure of fronts over the Rockies.

Looking at the reliability diagrams for the 2D models (Fig. 6b; Figs. A1b–A4b in the appendix), it is clear that the 2D models underpredicted all frontal types across the CONUS. The 3D models with $3 \times 3 \times 3$ kernels were closer to matching the forecast probabilities and target frequencies than the 2D models but tended to overpredict fronts using 50-km neighborhoods with the exception of stationary fronts (see Figs. S1b–S5b in the online supplemental material). The $5 \times 5 \times 5$ models suffered from larger overpredictions than the $3 \times 3 \times 3$ models (see Figs. S16b–S20b in the online supplemental material), indicating that incorporating more spatial features into the convolutions via larger kernel sizes may result in greater confidence and higher probabilities from the models. All of the models and frontal types had FB values greater than 1 using all neighborhoods and show that a majority of incorrect predictions from the models are false alarms as opposed to missed fronts. However, these false alarms can be attributed to inconsistencies within the provided frontal data as Lagerquist et al. (2019) illustrated how WPC labels can disappear, reappear, or change type between successive time steps.

2) FULL DOMAIN

All models performed worse over the unified surface domain than over the CONUS domain. This was not an unexpected result, but we were surprised by the degree to which the models underperformed over this expanded domain. We eventually learned from an operational forecaster that fronts drawn over the ocean outside of the WPC domain often rely on satellite imagery to be located as observations from ships and buoys are relatively sparse. Therefore, we believe it is likely that some of the boundaries plotted over the ocean based on real-time satellite observations are displaced from where they would otherwise be identifiable in the ERA5 data; this could be an explanation for the weaker performance across the entire unified surface domain. Minimal performance increases were observed when the number of images was increased from 24 to 90 (see Table 6), indicating that the models are likely not experiencing issues near the edges of predictions as is commonly observed with CNNs utilizing zero padding (Liu et al. 2018).

All F/NF model architectures had relatively similar performance (Figs. A9a,c and A14a,c in the appendix and Figs. S10a,c, S15a,c, S25a,c, and S30a,c in the online supplemental material), but the F/NF model with $3 \times 3 \times 3$ convolution kernels outperformed the 2D 3×3 architecture. Both 3D architectures achieved 250-km CSI scores that lie within each other's 95% confidence intervals while the $5 \times 5 \times 5$ architecture achieved better performance with smaller neighborhoods. The fact that all fronts are given the same label likely helped the $3 \times 3 \times 3$ F/NF architecture learn features and hypergradients associated with all fronts as opposed to individual types. The 250-km CSI maps [panel (d) in the performance figures (Fig. 6, Figs. A1–A14, and Figs. S1–S30 in the

online supplemental material)] show that the models were able to skillfully locate fronts across most of the full domain, with the lowest scores being north of the Arctic Circle and west of the Rocky Mountains over the Pacific states and British Columbia, Canada. We initially assumed that low scores over the full domain were due to the models not performing well over the oceans, but the models are able to locate fronts over much of the Atlantic and Pacific Oceans just as well as the CONUS.

Cold fronts and warm fronts performed best with 3×3 and $5 \times 5 \times 5$ convolutions over the full unified surface domain, as was the case over the CONUS (Figs. A5a,c, A6a,c, A10a,c, and A11a,c in the appendix and Figs. S6a,c, S7a,c, S11a,c, S12a,c, S21a,c, S22a,c, S26a,c, and S27a,c in the online supplemental material). However, the difference in 250-km CSI between the CONUS and the full domain was much larger with cold fronts. In an analysis of a cold front over the eastern Atlantic Ocean, Wakimoto and Murphey (2008) found a prominent virtual potential temperature gradient θ_v across the cold front that was strongest at 2 km above the surface, which contradicts previous studies such as Wakimoto and Cai (2002) and Sanders (1955) who found that the temperature gradients of cold fronts are maximized at the surface. The findings from these studies seem to suggest that the vertical structure of a cold front can vary considerably, so the inclusion of θ_v in our list of predictors should be considered, and perhaps a static land/ocean parameter is needed. Because 850 hPa is our highest pressure level and is typically located at 1.5 km AGL, data at higher pressure levels (e.g., 700 hPa) could perhaps help to identify features found aloft in maritime cold fronts. Wakimoto and Bosart (2001) analyzed observations of an oceanic warm front and found that it was also more well defined aloft than at the surface and was characterized by sloped θ_v and θ_E isopleths, further suggesting that θ_v is a thermodynamic variable that should be included in our list of predictors. Cold and warm fronts both performed poorly over the Rocky Mountains, the Sierra Madre range, and north of the Arctic Circle. These regions all had low cold- and warm-frontal frequencies (Figs. 2a,b). We think that separate models trained exclusively over complex terrain and areas of low frontal frequency could result in better performance with cold and warm fronts in these regions.

Stationary fronts performed significantly worse over the full domain, with false alarm rates near 70% at 250 km (Figs. A7a,c and A12a,c in the appendix and Figs. S8a,c, S13a,c, S23a,c, and S28a,c in the online supplemental material). We discovered that the models tend to identify the intertropical convergence zone (ITCZ) as a large stationary front, leading to an exceptionally large number of false positives. We believe the wind convergence associated with the ITCZ causes the model to misinterpret the ITCZ as a stationary front. Since the domain used for model training did not include the ITCZ, the models were inherently biased to put a greater emphasis on wind convergence to identify stationary fronts, perhaps implying that we need to model delineation between the tropics and the midlatitudes. Similar to binary fronts, the stationary-front performance was highest over areas of the Rocky Mountains, northwestern Canada, Alaska, and the Sierra Madre in Mexico. The CSI maxima over the Sierra Madre

and the Rocky Mountains show that the SF/OF models are able to identify stationary fronts over mountainous terrain.

Occluded-front performance decreased considerably with the 3×3 and $5 \times 5 \times 5$ architectures but did not decrease much with the $3 \times 3 \times 3$ architecture as it located more fronts but had a higher false alarm rate (Figs. A8a,c and A13a,c in the appendix and Figs. S9a,c, S14a,c, S24a,c, and S29a,c in the online supplemental material). The models seemed to struggle in properly identifying occluded fronts that were interpreted by forecasters to be wrapped around the center of mature cyclones (see Fig. 3f from Reed and Albright 1997). When the models located occluded fronts, the sections of the fronts highlighted were usually attached to the triple point where the occluded fronts intersected with the cyclones' respective cold and warm fronts; occluded fronts may not be as easily detectable near the triple point due to mixed features over short distances within the convolutions. With the exception of the western Pacific, there were no local maximums in occluded-front CSI scores that stood out as having clear significance. Occluded fronts are the rarest frontal type in our datasets (see Table 3), so it seems probable that the models struggled to fully learn the vertical structure of occluded fronts. We think that more training or including all frontal types in one singular model may improve occluded-front performance, especially over areas where these boundaries are frequently analyzed by forecasters.

b. Variable importance

Variable importance was determined by performing permutation studies (Lakshmanan et al. 2015; McGovern et al. 2019) with the 3D models structured with $5 \times 5 \times 5$ convolution kernels using the 200-km boundary. Despite the 2D models performing slightly better, we performed the permutation studies on the best-performing 3D models since we believe that the 3D structure that processes volumetric images has the most potential to be improved upon for better reliability in detecting fronts. Two types of permutations were performed—individual permutations and grouped permutations.

Individual permutations involve running evaluations with a model while randomizing the values of one predictor per evaluation. Since 60 variables were used in the models, 60 evaluations were performed with all models, and one variable was randomized for every run. If a variable is to be deemed “important” for detecting fronts, a drop in the CSI will be observed when its values are randomized. Likewise, a variable that does not help the model detect fronts will cause the CSI to increase when its values are randomized.

Grouped permutations are similar to individual permutations, except that multiple predictors are randomized per run as opposed to just one. A total of 12 runs were performed, with each run having 1 of the 12 variables randomized at the five levels (see Table 1). For example, when temperature data are randomized, all temperature data at the surface and at 1000, 950, 900, and 850 hPa are randomized at the same time. The grouped permutation method allowed us to visualize overall variable importance as opposed to assigning importance based on the level where the data reside.

TABLE 7. Grouped variable importance by frontal type, ranked from 1 to 12 with 1 and 12 being the most and least important variables, respectively. Cells with numbers in boldface or italics indicate that the variables respectively helped or diminished performance of the respective frontal type.

Variable	Cold	Warm	Stationary	Occluded	Binary (F/NF)
Dewpoint	6	6	6	11	<i>11</i>
Equivalent potential temperature	8	9	<i>11</i>	6	4
Geopotential height + pressure	3	1	1	1	2
Mixing ratio	10	10	5	7	7
Relative humidity	6	6	4	7	<i>12</i>
Specific humidity	10	10	6	9	7
Temperature	4	5	6	4	1
<i>u</i> wind	1	3	3	3	6
<i>v</i> wind	2	2	2	2	3
Virtual temperature	4	4	<i>11</i>	5	4
Wet-bulb potential temperature	9	8	<i>10</i>	<i>12</i>	7
Wet-bulb temperature	<i>12</i>	10	6	9	7

Results from grouped and individual permutations can be viewed in Table 7 as well as in Table A1 of the appendix. Surface pressure, geopotential height, *u* wind, and *v* wind were consistently ranked as some of the most important variables for front detection. In other words, randomizing these variables resulted in the largest drops in the 200-km CSI scores. Frontal boundaries are commonly associated with pressure troughs (Berry et al. 2011), thus it was unsurprising to see surface pressure and geopotential height to be prioritized by the model for detecting fronts. 10-m wind components were exceptionally important in identifying cold fronts, which is consistent with findings from Simmonds et al. (2012) who showed that wind shift method was successful in locating cold fronts and contrasted by lower importance for warm fronts. Variables such as θ_w , used by Berry et al. (2011), and θ_E , used by Schemm et al. (2015), showed little utility in detecting frontal boundaries with the UNet 3+ models. This was an unexpected finding but could be a result of the collinearity that exists between the numerous thermodynamic variables we included as inputs to the models. Geopotential height showed to be particularly important in the detection of occluded fronts, which was expected as the vertical stacking of geopotential heights can indicate the presence of an occluded front in a mature cyclone. Nearly half of the variables in the individual and grouped permutations showed to have negative effects in the detection of stationary fronts, so it is possible that stationary fronts can be detected with fewer variables than the other frontal types. Interestingly, using the binary setup (front/no front), the most important variable in the individual permutation studies was relative humidity at 2 m AGL. Given that all frontal types are being identified with the same label in the binary setup, the authors believe that the $5 \times 5 \times 5$ F/NF model placed a greater emphasis on thermal gradients instead of wind shifts and leading to relative humidity becoming an important variable in frontal detection. However, relative humidity had a negative effect on the model in the grouped permutation studies, suggesting that relative humidity may only be useful at certain pressure levels. Wet-bulb temperature and specific humidity had almost no effect on the performance with any frontal type in grouped permutations, with the possible exception being occluded fronts.

c. Case studies

We chose two case studies, one over the CONUS and the other over the full unified surface analysis domain, to highlight some of the differences and similarities between the human-drawn fronts and predictions from the $5 \times 5 \times 5$ models.

Case 1 can be viewed in Figs. 7 and 8. This study was chosen for two reasons; the corridors of higher probabilities indicated by the SF/OF and F/NF models of an apparent stationary front extending from southeastern New Mexico northwestward into western Colorado and Utah, and a warm front drawn over western Indiana that was determined by the CF/WF model to most likely exist up to 200 mi (300+ km) to the northeast of its analyzed location.

As highlighted in Table 7, *u* wind and *v* wind are two of the three most important variables for detecting stationary fronts and also play a role in identifying fronts with the binary F/NF model. Wind convergence at the 900-hPa level is indicated by wind barbs along the corridors of probabilities from the SF/OF and F/NF models near the Four Corners region where no boundary was analyzed by forecasters (Fig. 7), and the ERA5 data also show wind convergence at the surface along these same corridors. However, there does not appear to be wind convergence at the surface when looking at surface observations from ASOS stations, possibly explaining the lack of an analyzed stationary front in the Four Corners region.

The warm front over Indiana was not identified by the CF/WF nor the F/NF model, but a possible warm front was found over eastern Michigan. The ERA5 data did not show any prominent wind shift or gradients of temperature, dewpoint, or virtual temperature at the locations of both the analyzed warm front and the warm front predicted by the CF/WF model. These variables were determined to be important for detecting warm fronts (see Table 7 and appendix). However, a weak trough in geopotential heights at the 950- and 900-hPa levels exists to the northeast of the predicted warm front, which suggests the presence of a boundary (Berry et al. 2011).

Case 2 (Figs. 9 and 10) highlights model predictions over the full unified surface analysis domain. The SF/OF model identifies much of the ITCZ as a stationary front, leading to significantly lower CSI scores over the full domain as mentioned

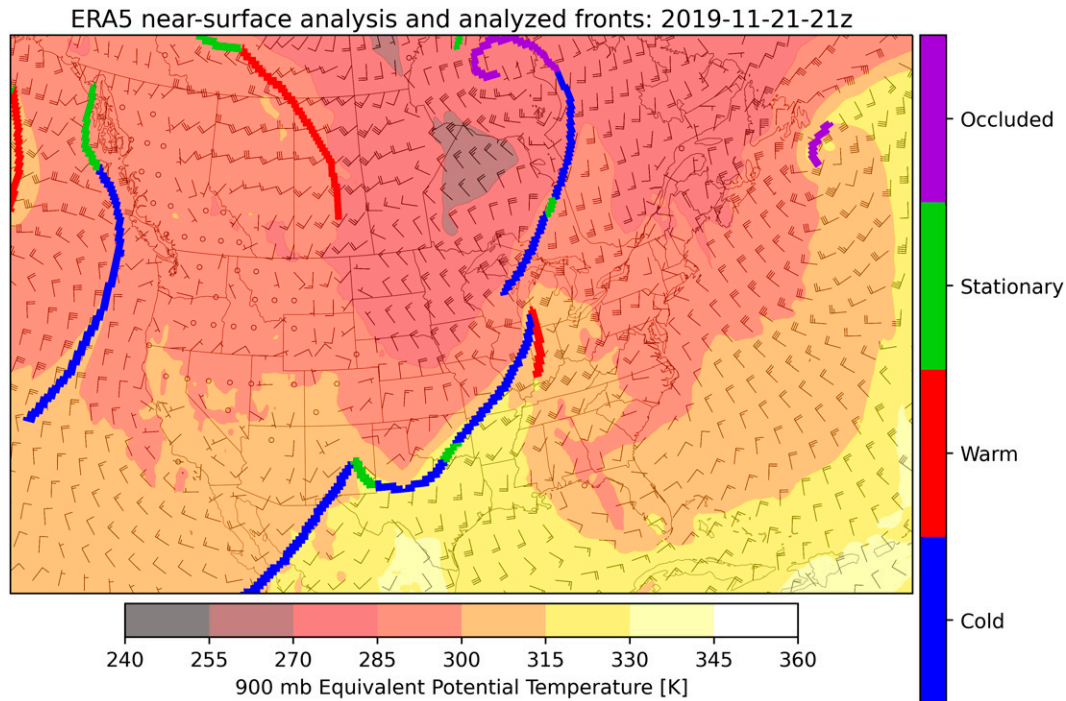


FIG. 7. Case 1 with WPC fronts plotted on top of 900-hPa θ_E and wind barbs for 2100 UTC 21 Nov 2019.

in section 3a (model performance). The CF/WF model showed cold-front probabilities exceeding 90% extending south of an analyzed warm front over the central Atlantic Ocean, though there was not a strong thermal gradient associated with the model's prediction. However, a pressure trough is present over the area of the highest cold-front probabilities, which is consistent with results from our permutation studies that showed pressure variables along with u wind and v wind being the most important variables for detecting cold fronts (see Table 7 and appendix). The CF/WF model also suggested the analyzed warm front over the central Atlantic Ocean is a cold front, but we attribute this to be an issue with the model prioritizing wind over any thermal variable as a clear temperature gradient exists along the warm front and supports the forecasters' interpretation of this boundary. The SF/OF model shows good skill in detecting occluded fronts near the center of mature cyclones. The F/NF model was able to identify nearly all boundaries that were plotted by the forecasters with high probabilities, highlighting the F/NF model's ability to generalize the thermal gradients, wind shifts, and pressure troughs that are present with the four frontal types.

We also noticed some fronts that seemed to be either missed or mislabeled by the forecasters, including an area highlighted by the CF/WF model south of the Aleutian Islands with probabilities exceeding 80% for the existence of a warm front. This area had a baroclinic trough analyzed in the region where a warm front is indicated by the CF/WF model. This baroclinic trough is not shown in Fig. 10 but the surface analysis for this time step can be found in the WPC surface analysis archive. Since troughs are not included in any of the models, it seems probable that the models were overpredicting the existence of fronts where baroclinic troughs have been analyzed by forecasters. Baroclinic

troughs have sharp wind shifts but do not have prominent thermal gradients (Sanders 1999); the models are likely too dependent on u wind and v wind when predicting the locations of fronts as they lack knowledge of the existence of troughs.

4. Discussion and future work

Our deep learning models were shown to be effective at detecting cold, warm, stationary, and occluded fronts and significantly outperformed prior baseline methods that aimed to objectively locate frontal boundaries. We demonstrated that both the 2D 3×3 and 3D $5 \times 5 \times 5$ architectures are able to generalize properties associated with different frontal types over the CONUS and the unified surface analysis domain.

We cannot conclude that any of the three structures used (2D with 3×3 kernels, 3D with $3 \times 3 \times 3$ kernels, or 3D with $5 \times 5 \times 5$ kernels) is the superior model structure for detecting fronts. Overall, we noticed that performance with a particular frontal type was positively correlated to its sample size (see Table 3 for sample sizes among the datasets), so manipulating class weights and other model parameters may be necessary to account for the varying sample sizes.

Pressure variables, u wind, and v wind consistently rank among the most important variables for detecting fronts and are supported by previous studies and frontal detection methods (Berry et al. 2011; Payer et al. 2011; Schemm et al. 2015; Simmonds et al. 2012; Schultz 2005). With the exception of stationary fronts, temperature and virtual temperature were the thermodynamic variables that showed the most utility with identifying different frontal types after pressure. Models appear to place too much emphasis on locating fronts with

2019-11-21-21z analyzed fronts and model predictions

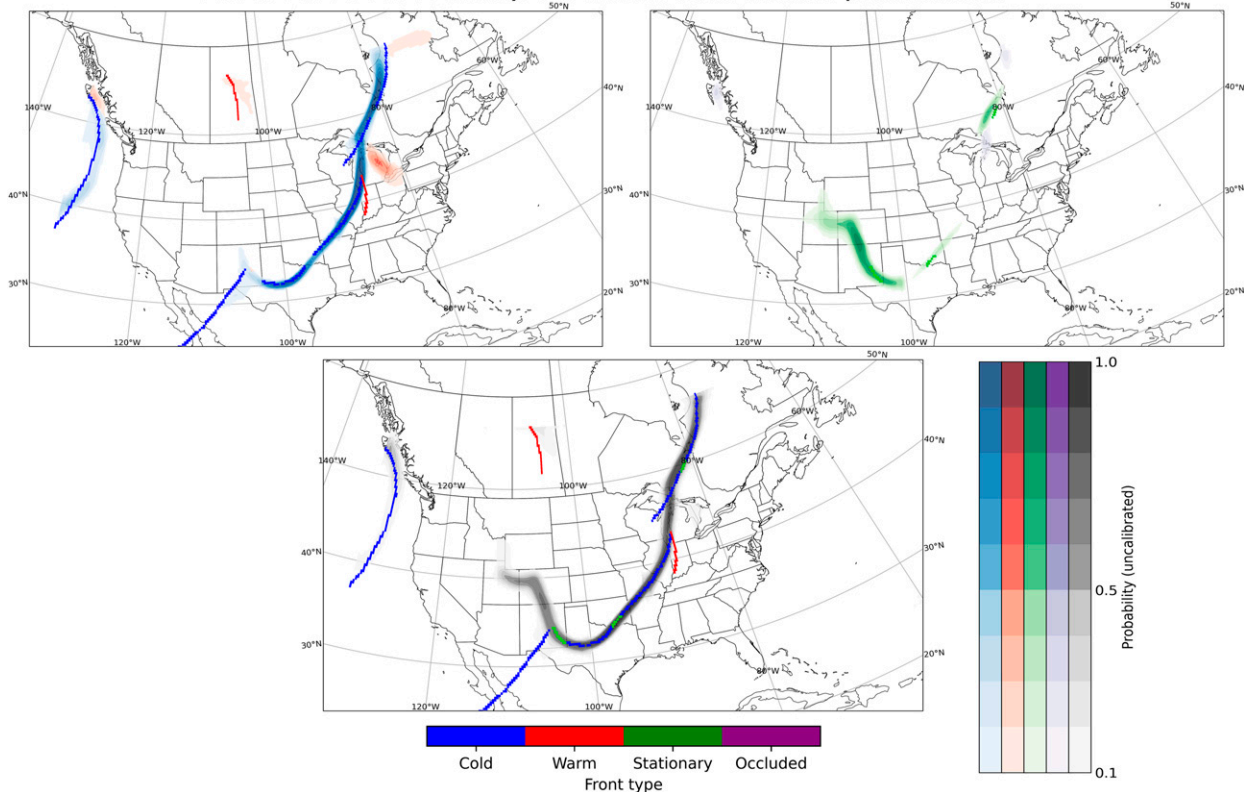


FIG. 8. Case-1 model predictions over the CONUS for 2100 UTC 21 Nov 2019 using the $5 \times 5 \times 5$ (top left) WF/WF, (top right) SF/OF, and (bottom) F/NF models. The fronts analyzed by the forecasters (solid lines) are plotted on top of model predictions (contours).

wind shifts and do not rely enough on thermal gradients, resulting in many baroclinic troughs being misinterpreted as frontal boundaries by the models (see case 2). The dependence on u wind and v wind likely explains why many thermal

variables do not have a substantial effect on the models' predictions. We believe that training the models to also identify baroclinic troughs will help limit the models' tendencies to identify frontal boundaries without thermal gradients.

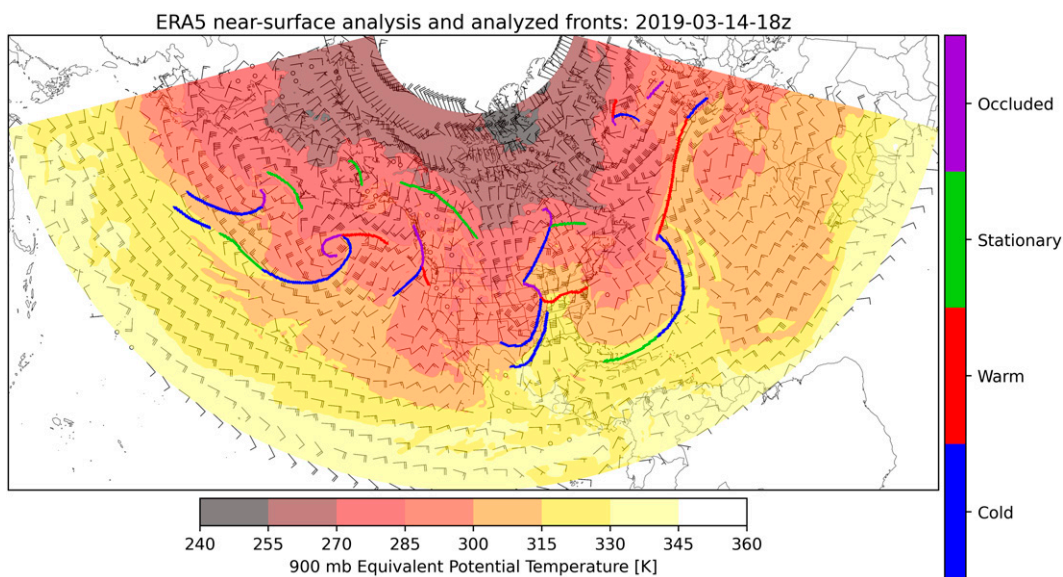


FIG. 9. Case 2 with unified surface analysis fronts plotted on 900-hPa θ_E and wind barbs for 1800 UTC 14 Mar 2019.

2019-03-14-18z analyzed fronts and model predictions

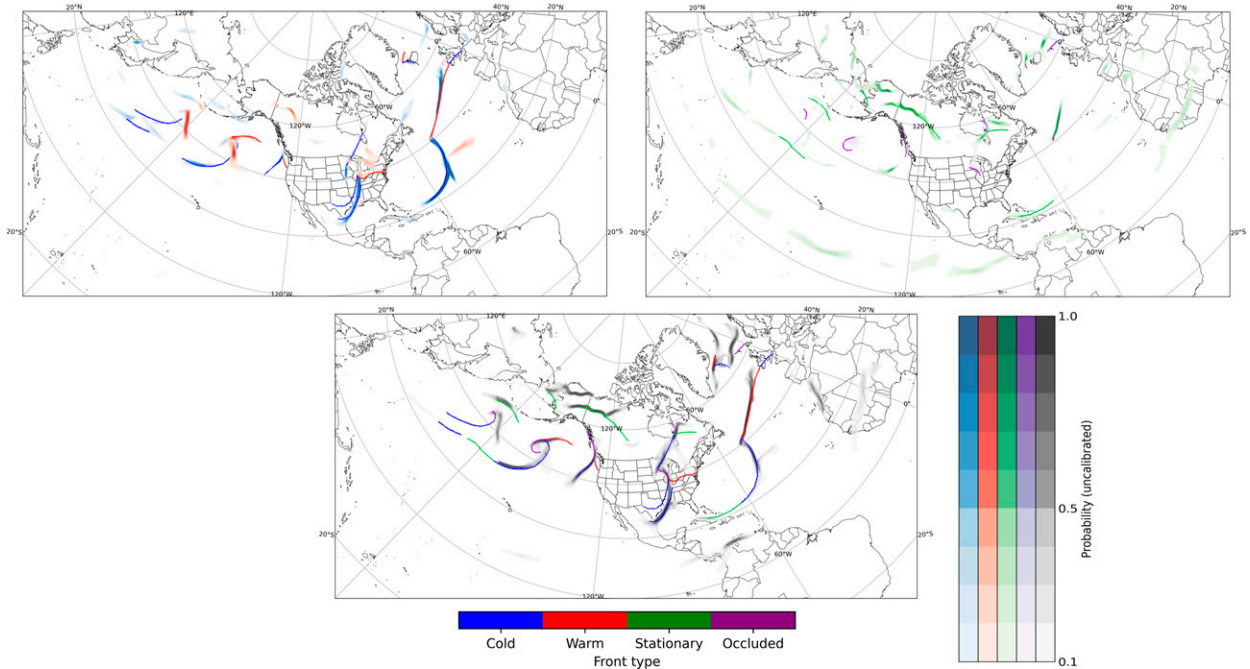


FIG. 10. Case-2 model predictions over the CONUS for 1800 UTC 14 Mar 2019 using the $5 \times 5 \times 5$ (top left) CF/WF, (top right) SF/OF, and (bottom) F/NF models. The fronts analyzed by the forecasters (solid lines) are plotted on top of model predictions (contours).

Proper calibration of our models is needed so that the forecast probabilities are consistent with the actual frequencies of the different frontal types. This can be achieved through various forms of regression and will give forecasters confidence that the probabilities output by the models are accurate representations of how often fronts have been historically analyzed with similar sets of data. Calibration will need to be performed individually with each neighborhood so that the probabilities closely resemble the frequency of fronts located within the given neighborhoods.

In future work, we will improve upon our models by including data from higher pressure levels, with a particular focus on 700-hPa data. This should help the models locate more frontal boundaries at higher elevations in the Rocky Mountains and other regions of complex terrain. Our future experiments will also exclude variables that have been highlighted in this paper as having a net-negative or net-zero effect on identifying specific types of boundaries.

To make more accurate classifications of the different frontal types, another direction of future exploration will be a set of models that predict the four frontal types examined in this paper—cold, warm, stationary, and occluded. Since any of our current models can only identify up to two frontal types at once, one model set that can identify all four types may help reduce the number of fronts that are labeled incorrectly as a result of insufficient labels to ensure class discrimination. It is possible that an entirely separate set of models will be needed to identify oceanic fronts as they may have different thermodynamic properties than those over land.

To assess the efficacy and utility of these models, we conducted interviews with forecasters from WPC, OPC, and TAFB on the 3D models with $5 \times 5 \times 5$ kernels. The forecasters had overwhelmingly positive reactions to the models' predictions in relation to the fronts from the unified surface analyses. Ongoing work is exploring the evaluation of the utility of the frontal analysis first-guess tool by WPC, OPC, and TAFB forecasters through the use of a web-based interface.

We are working to validate the performance of our U-Nets using data from the Global Forecast System (GFS) and Global Data Assimilation System (GDAS). These were chosen because NOAA's operational forecasters use them in real time in the NAWIPS system to assist in locating surface boundaries. In our preliminary tests, the deep learning models appeared to locate frontal boundaries with GFS and GDAS data with similar degrees of accuracy as ERA5 data at the same time steps on which the U-Nets were tested. Our preliminary results give us confidence that operational forecasters can use our models as another tool to expedite the frontal analysis process. Aiding this transition, the models can simply be stored on a local machine and run using Python code from our GitHub repository (<https://github.com/ai2es/fronts>).

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grant ICER-2019758. This material is also based upon work supported by the National Oceanic and Atmospheric Administration under Grant NA20OAR4590347. We thank the OPC, WPC, and TAFB forecasters who have assisted us in

developing this approach, with special appreciation to Amanda Reinhart at TAFB for providing us with invaluable assistance throughout the course of this project, and Greg Carbin who provided Python code and an initial form of the WPC analysis during this project. The authors would also like to thank Christopher Bailey for fixing our frontal data. The results contain modified Copernicus Climate Change Service information 2021. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains.

Data availability statement. ERA5 data on single and pressure levels were downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store and can be found via [Hersbach et al. \(2018\)](#). Frontal data derived from the

WPC analyses can be found via [NOAA \(2023\)](#). Python code used in this project is available in our GitHub repository at <https://github.com/ai2es/fronts>.

APPENDIX

Additional Results

[Table A1](#) is similar to [Table 7](#) but shows the ranking of importance for variables by frontal type for individual permutations. The appendix figures show performance diagrams for 2D models with 3×3 convolutions. [Figures A1–A4](#) apply to the CONUS domain and have three images per map. [Figures A5–A9](#) apply to the full domain and have 24 images per map. [Figures A10–A14](#) apply to the full domain and have 90 images per map.

TABLE A1. As in Table 7, but for individual permutations.

Variable	Cold	Warm	Stationary	Occluded	Binary (F/NF)
Dewpoint (2 m AGL)	43	52	18	47	17
Dewpoint (1000 hPa)	28	29	42	50	32
Dewpoint (950 hPa)	21	12	48	50	25
Dewpoint (900 hPa)	24	12	52	29	25
Dewpoint (850 hPa)	25	16	30	39	32
Equivalent potential temperature (2 m AGL)	41	39	42	29	25
Equivalent potential temperature (1000 hPa)	32	39	52	16	4
Equivalent potential temperature (950 hPa)	35	33	51	20	8
Equivalent potential temperature (900 hPa)	32	48	42	13	13
Equivalent potential temperature (850 hPa)	37	46	27	20	20
Surface pressure	2	39	3	7	2
Geopotential height (1000 hPa)	38	15	11	8	40
Geopotential height (950 hPa)	7	2	2	1	3
Geopotential height (900 hPa)	3	1	1	1	4
Geopotential height (850 hPa)	15	9	6	3	20
Mixing ratio (2 m AGL)	38	16	40	29	20
Mixing ratio (1000 hPa)	43	33	30	28	32
Mixing ratio (950 hPa)	48	33	17	29	47
Mixing ratio (900 hPa)	57	48	13	24	52
Mixing ratio (850 hPa)	57	56	27	29	50
Relative humidity (2 m AGL)	10	10	7	4	1
Relative humidity (1000 hPa)	12	5	13	16	7
Relative humidity (950 hPa)	55	24	22	29	9
Relative humidity (900 hPa)	49	29	30	54	12
Relative humidity (850 hPa)	41	59	19	10	25
Specific humidity (2 m AGL)	43	24	36	39	13
Specific humidity (1000 hPa)	28	16	36	39	16
Specific humidity (950 hPa)	43	33	22	39	25
Specific humidity (900 hPa)	59	60	15	29	36
Specific humidity (850 hPa)	60	52	19	24	51
Temperature (2 m AGL)	28	52	54	50	52
Temperature (1000 hPa)	15	52	58	20	55
Temperature (950 hPa)	9	24	59	13	47
Temperature (900 hPa)	17	22	56	13	25
Temperature (850 hPa)	25	33	48	20	36
<i>u</i> wind (10 m AGL)	18	19	40	58	13
<i>u</i> wind (1000 hPa)	5	14	42	60	9
<i>u</i> wind (950 hPa)	1	4	15	59	4
<i>u</i> wind (900 hPa)	4	6	10	29	9
<i>u</i> wind (850 hPa)	10	11	11	16	25
<i>v</i> wind (10 m AGL)	55	6	22	24	58
<i>v</i> wind (1000 hPa)	6	3	8	9	57
<i>v</i> wind (950 hPa)	8	8	4	4	36
<i>v</i> wind (900 hPa)	20	33	5	4	40
<i>v</i> wind (850 hPa)	25	56	9	10	46
Virtual temperature (2 m AGL)	32	48	52	39	54
Virtual temperature (1000 hPa)	13	48	57	24	59
Virtual temperature (950 hPa)	13	22	59	16	60
Virtual temperature (900 hPa)	19	24	55	10	56
Virtual temperature (850 hPa)	21	29	42	29	47
Wet-bulb potential temperature (2 m AGL)	38	56	36	47	40
Wet-bulb potential temperature (1000 hPa)	23	44	50	50	40
Wet-bulb potential temperature (950 hPa)	28	20	34	54	40
Wet-bulb potential temperature (900 hPa)	35	21	27	56	40
Wet-bulb potential temperature (850 hPa)	43	39	22	56	36
Wet-bulb temperature (2 m AGL)	49	29	19	47	32
Wet-bulb temperature (1000 hPa)	49	24	22	39	20
Wet-bulb temperature (950 hPa)	49	39	30	39	20
Wet-bulb temperature (900 hPa)	49	46	36	29	17
Wet-bulb temperature (850 hPa)	49	44	34	39	17

2D U-Net 3+ (3x3 kernel): Warm fronts over CONUS domain (3 images per map)

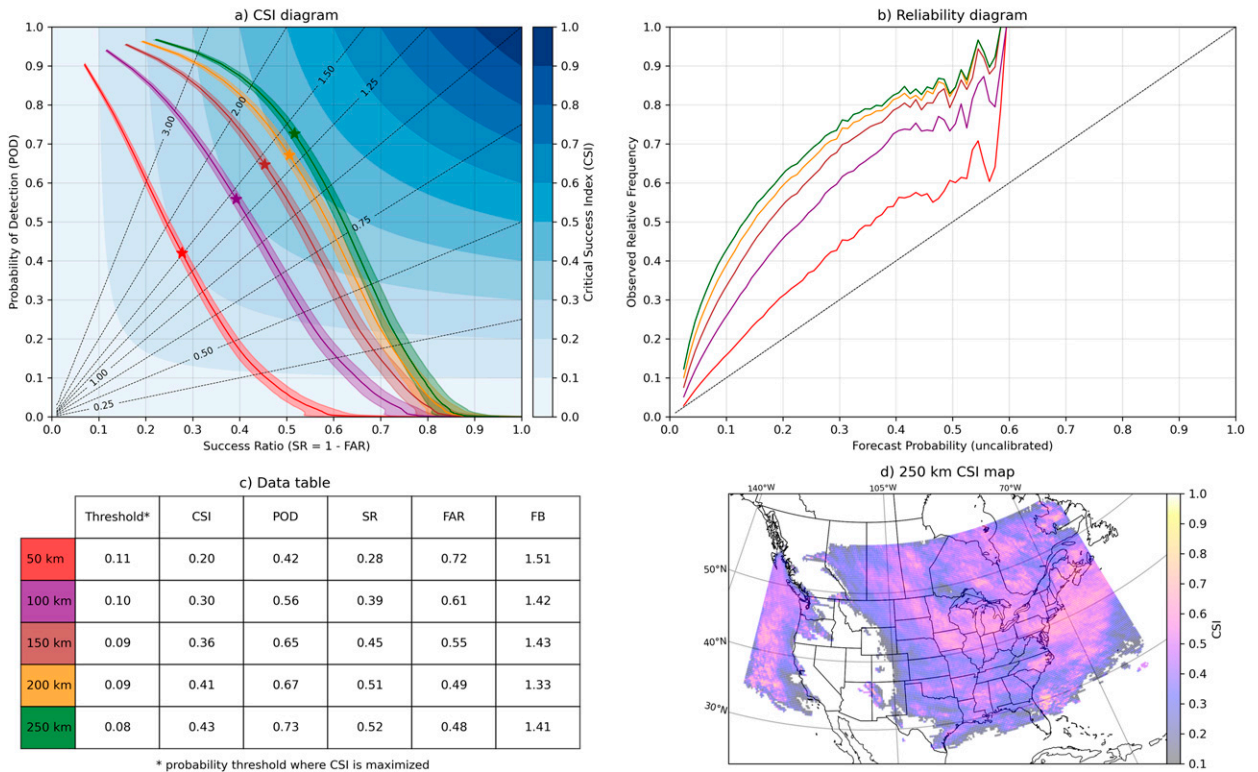


FIG. A1. As in Fig. 6, but for warm fronts.

2D U-Net 3+ (3x3 kernel): Stationary fronts over CONUS domain (3 images per map)

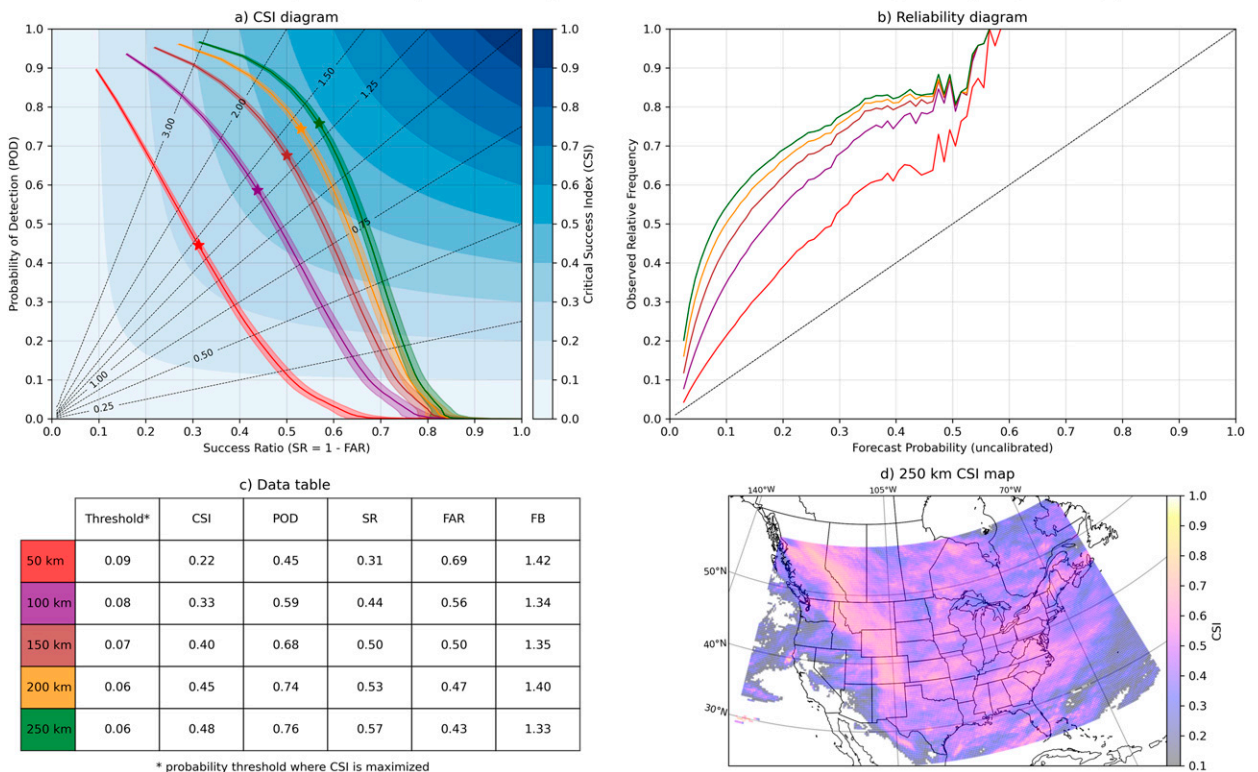


FIG. A2. As in Fig. 6, but for stationary fronts with the 2D SF/OF model with 3 × 3 convolutions.

2D U-Net 3+ (3x3 kernel): Occluded fronts over CONUS domain (3 images per map)

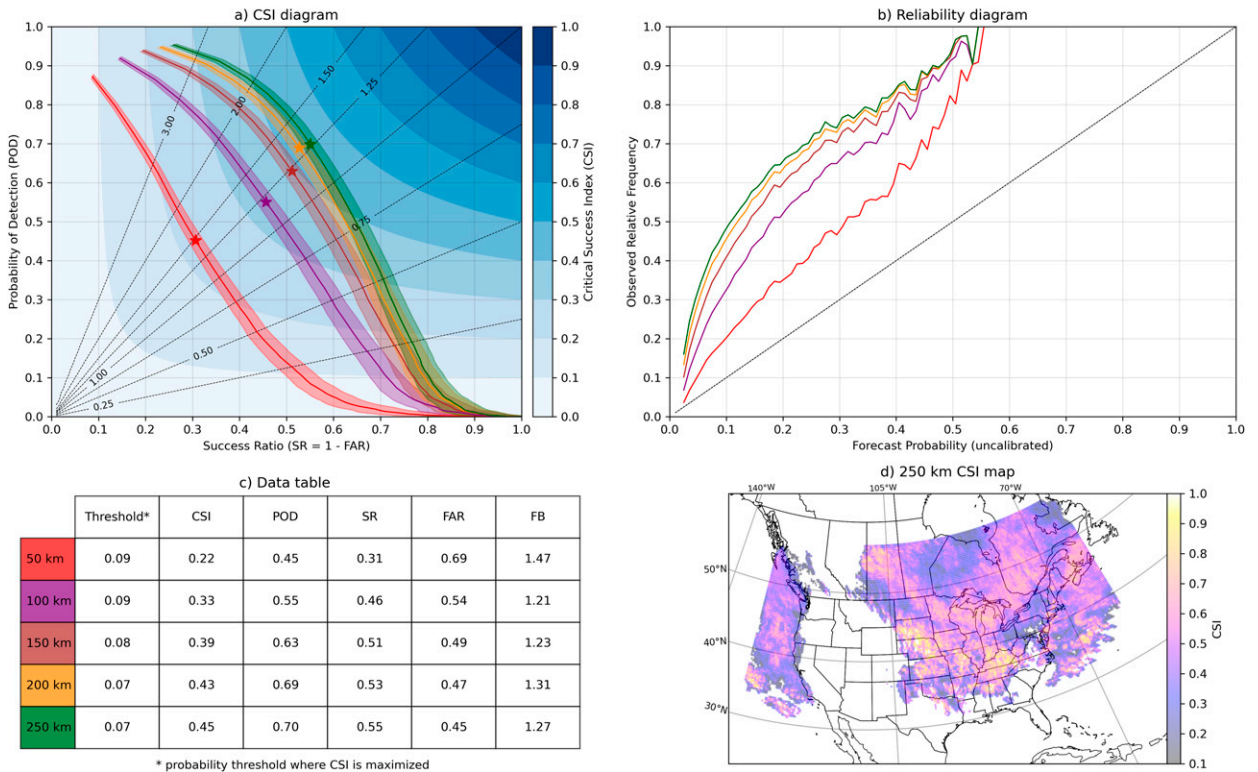


FIG. A3. As in Fig. A2, but for occluded fronts.

2D U-Net 3+ (3x3 kernel): Binary fronts (front / no front) over CONUS domain (3 images per map)

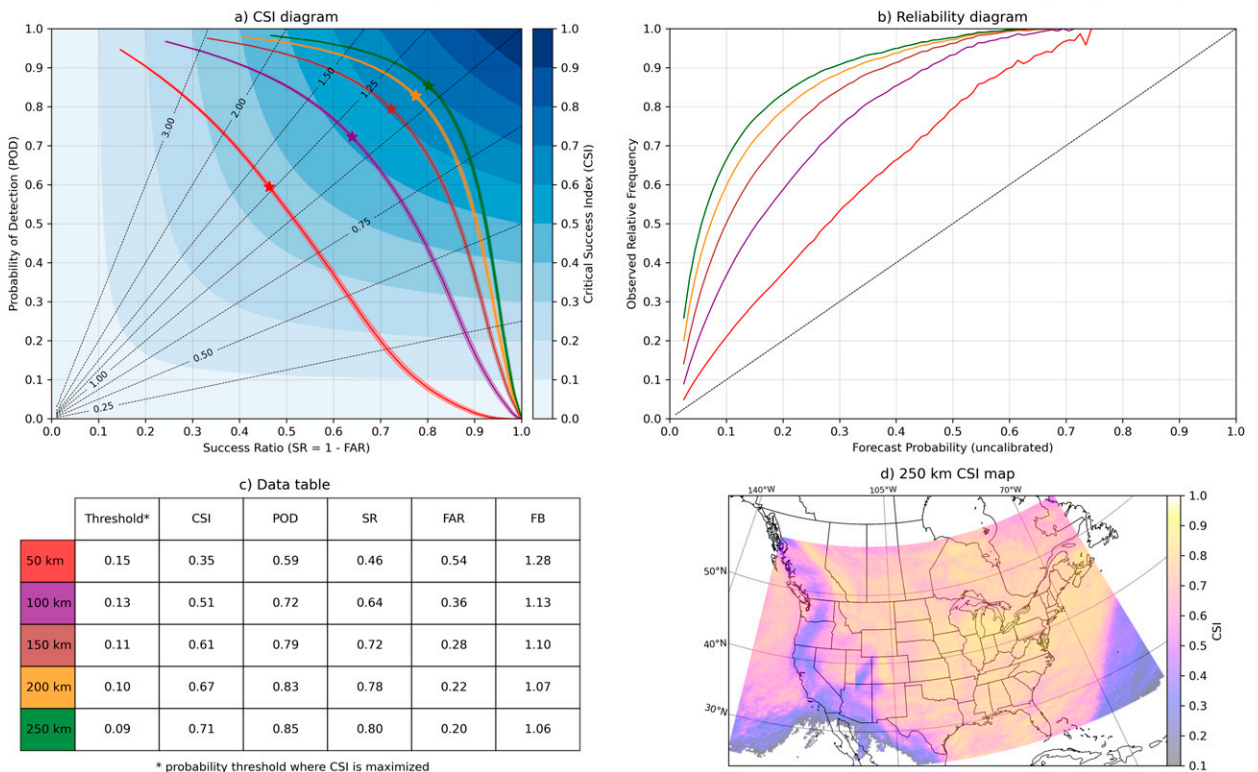


FIG. A4. As in Fig. A2, but for the 2D F/NF model with 3 × 3 convolutions.

2D U-Net 3+ (3x3 kernel): Cold fronts over full domain (24 images per map)

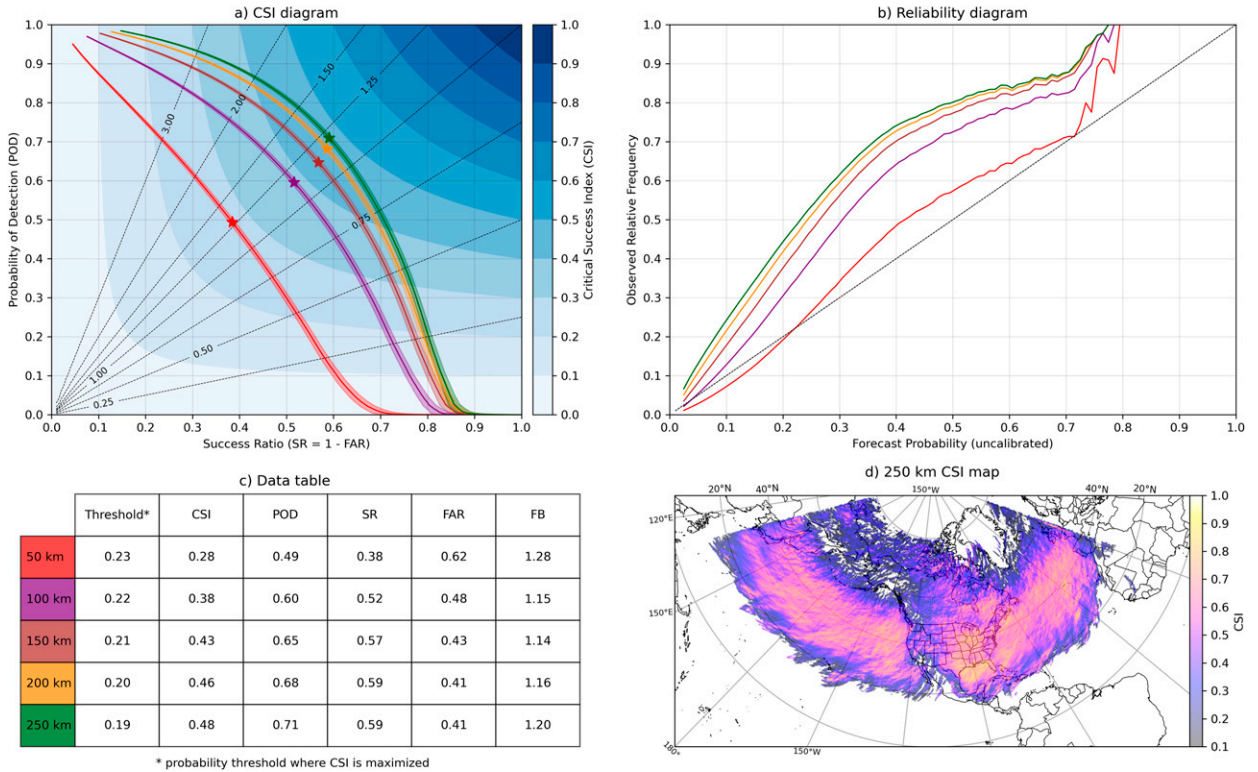


FIG. A5. As in Fig. 6, but over the full domain with 24 images covering the domain.

2D U-Net 3+ (3x3 kernel): Warm fronts over full domain (24 images per map)

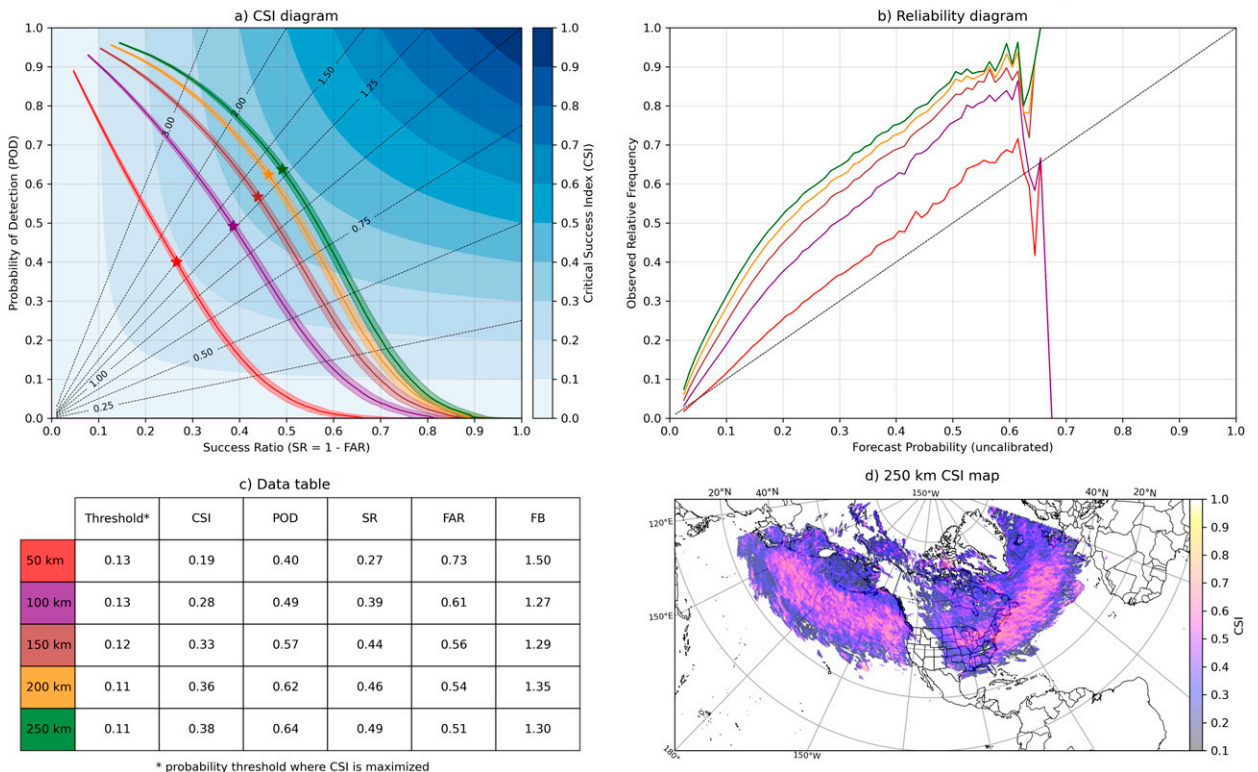


FIG. A6. As in Fig. A5, but for warm fronts.

2D U-Net 3+ (3x3 kernel): Stationary fronts over full domain (24 images per map)

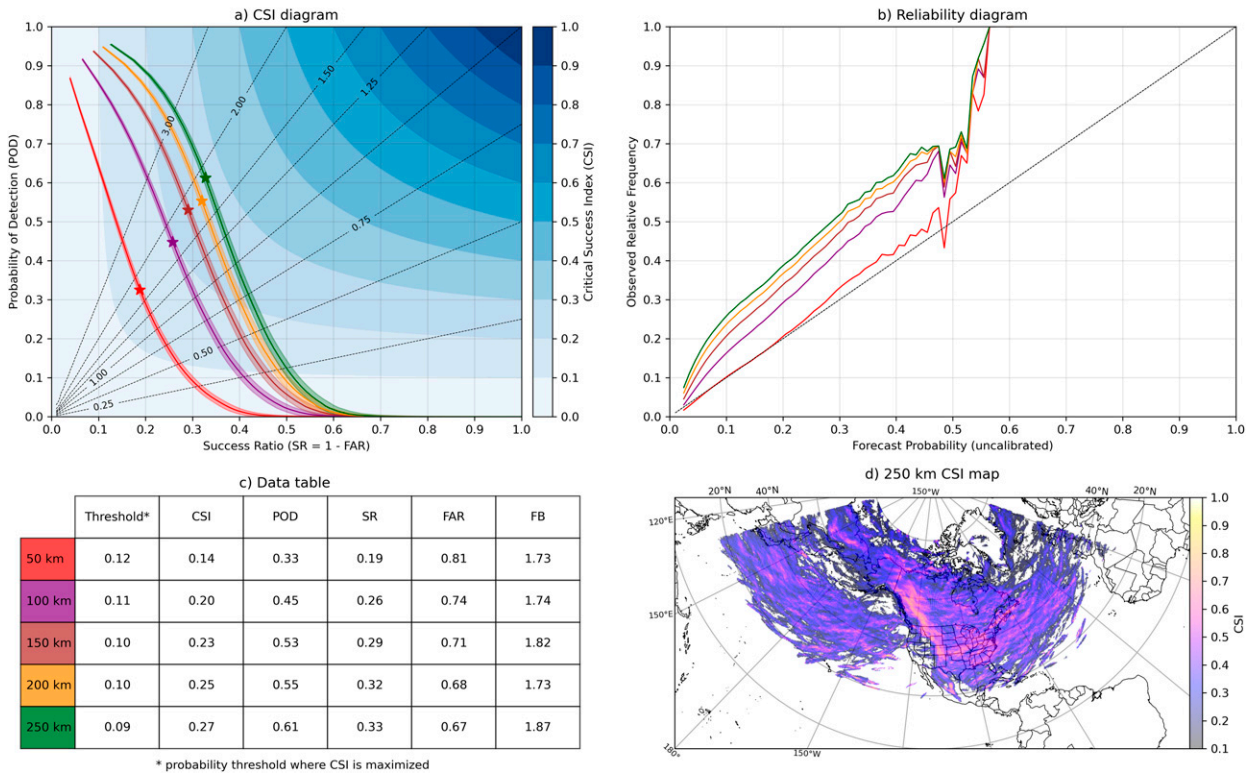


FIG. A7. As in Fig. A5, but for stationary fronts with the 2D SF/OF model with 3×3 convolutions.

2D U-Net 3+ (3x3 kernel): Occluded fronts over full domain (24 images per map)

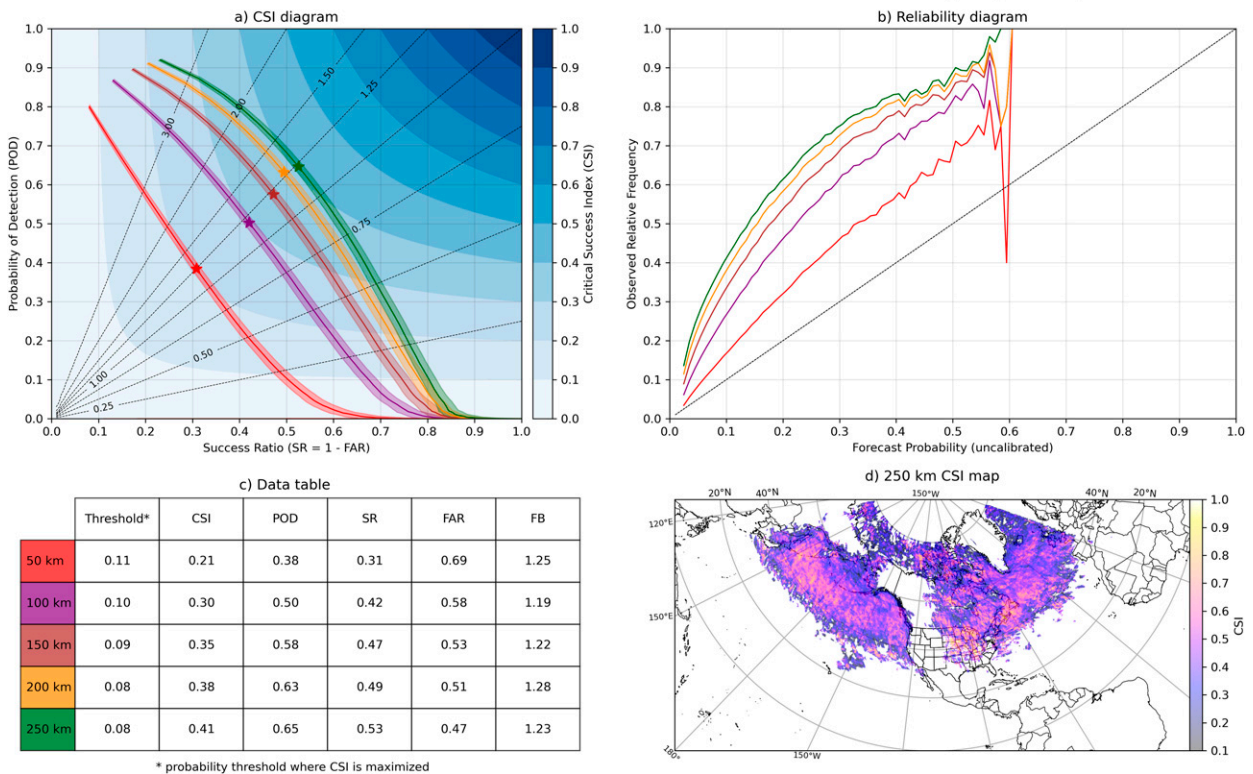


FIG. A8. As in Fig. A7, but for occluded fronts.

2D U-Net 3+ (3x3 kernel): Binary fronts (front / no front) over full domain (24 images per map)

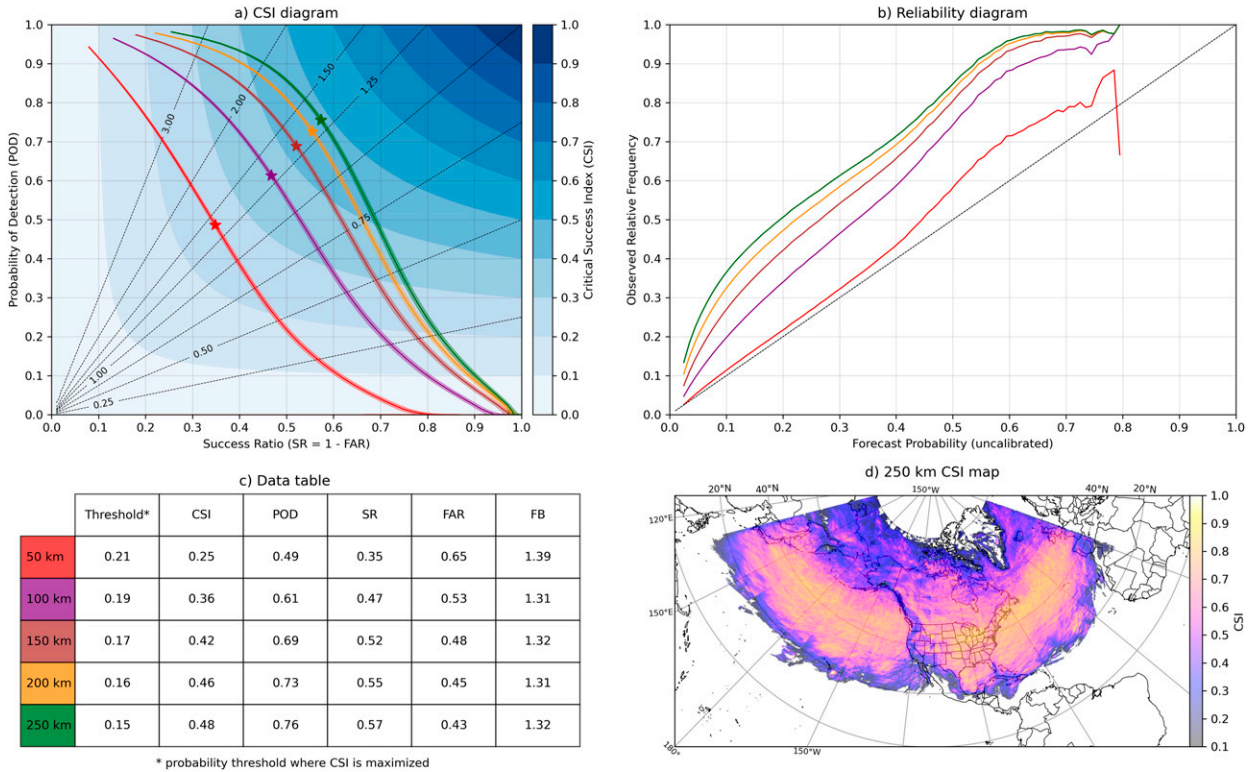


FIG. A9. As in Fig. A7, but for the 2D F/NF model with 3 × 3 convolutions.

2D U-Net 3+ (3x3 kernel): Cold fronts over full domain (90 images per map)

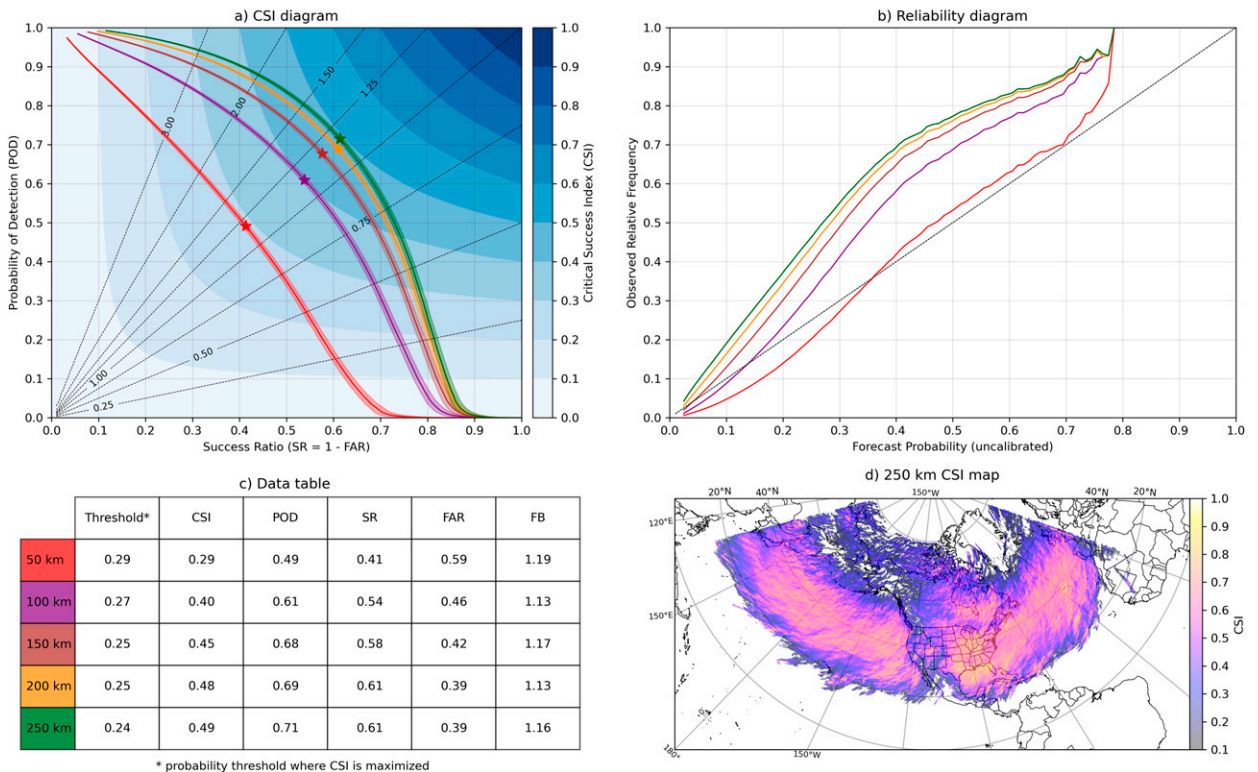


FIG. A10. As in Fig. A5, but with 90 images covering the domain.

2D U-Net 3+ (3x3 kernel): Warm fronts over full domain (90 images per map)

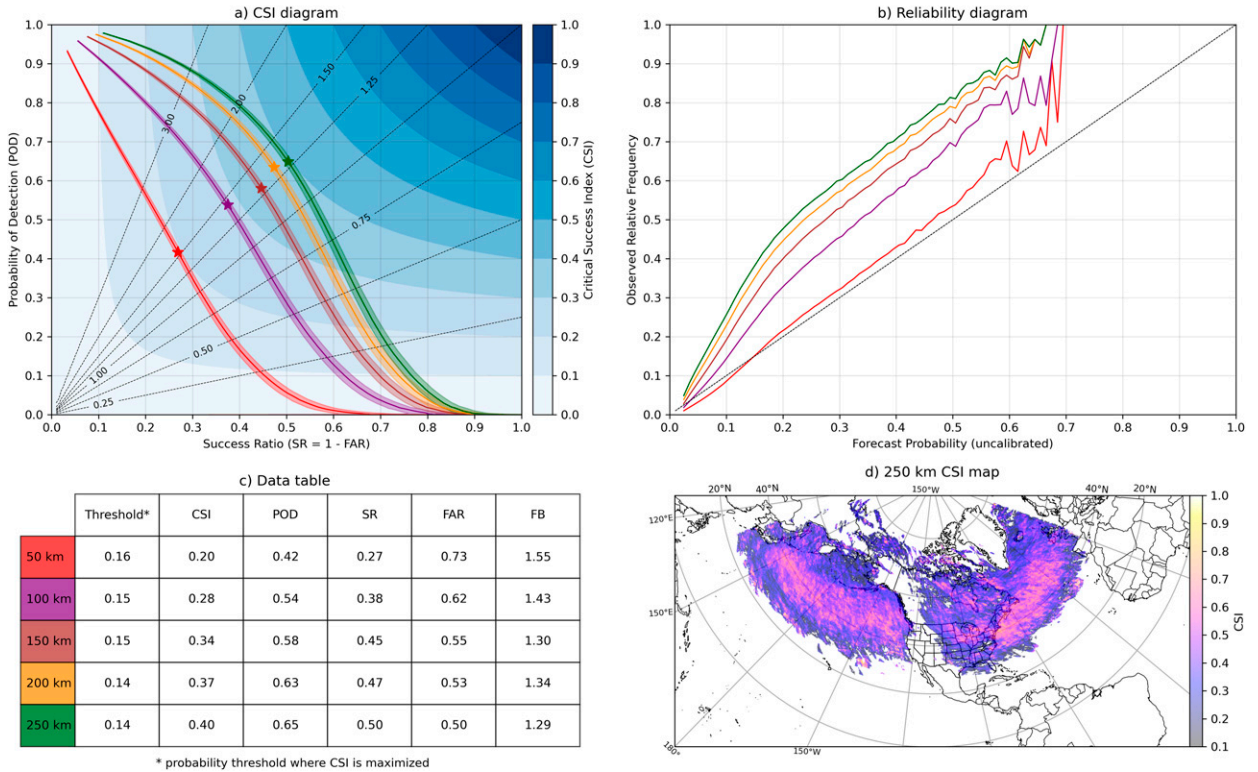


FIG. A11. As in Fig. A10, but for warm fronts.

2D U-Net 3+ (3x3 kernel): Stationary fronts over full domain (90 images per map)

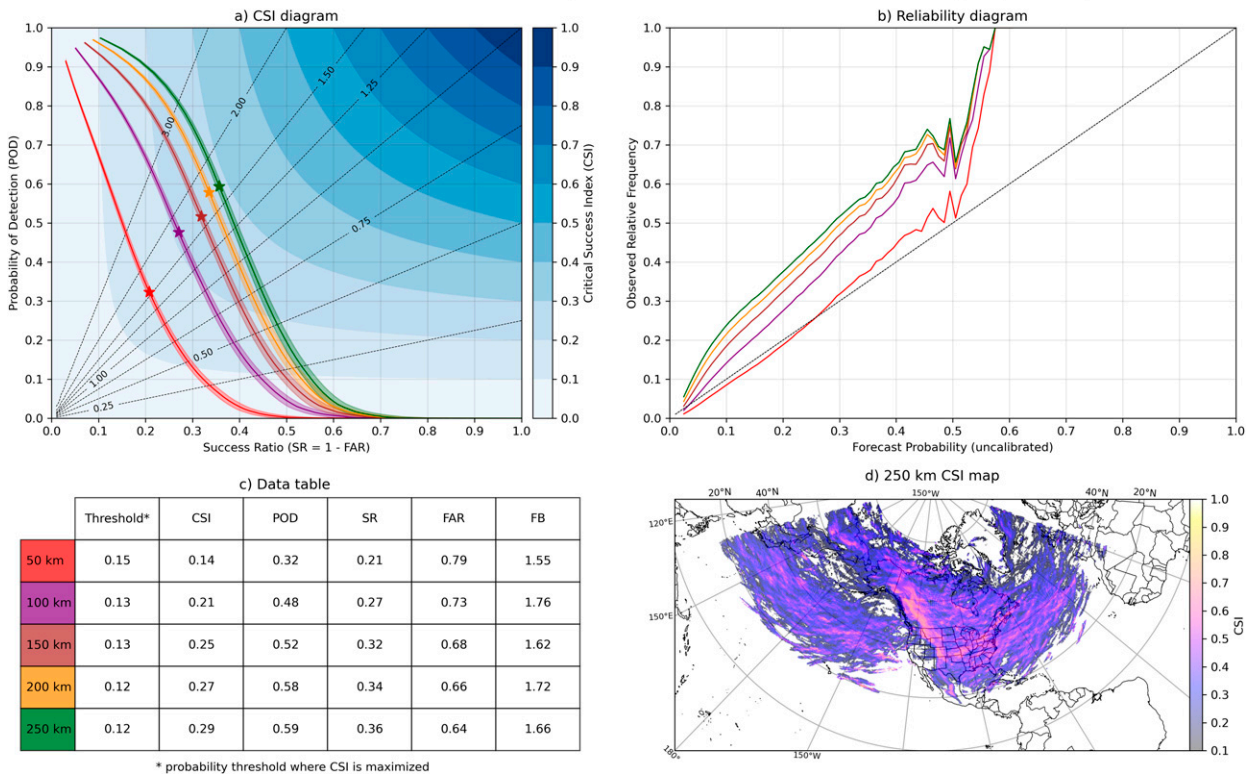


FIG. A12. As in Fig. A10, but for stationary fronts with the 2D SF/OF model with 3 × 3 convolutions.

2D U-Net 3+ (3x3 kernel): Occluded fronts over full domain (90 images per map)

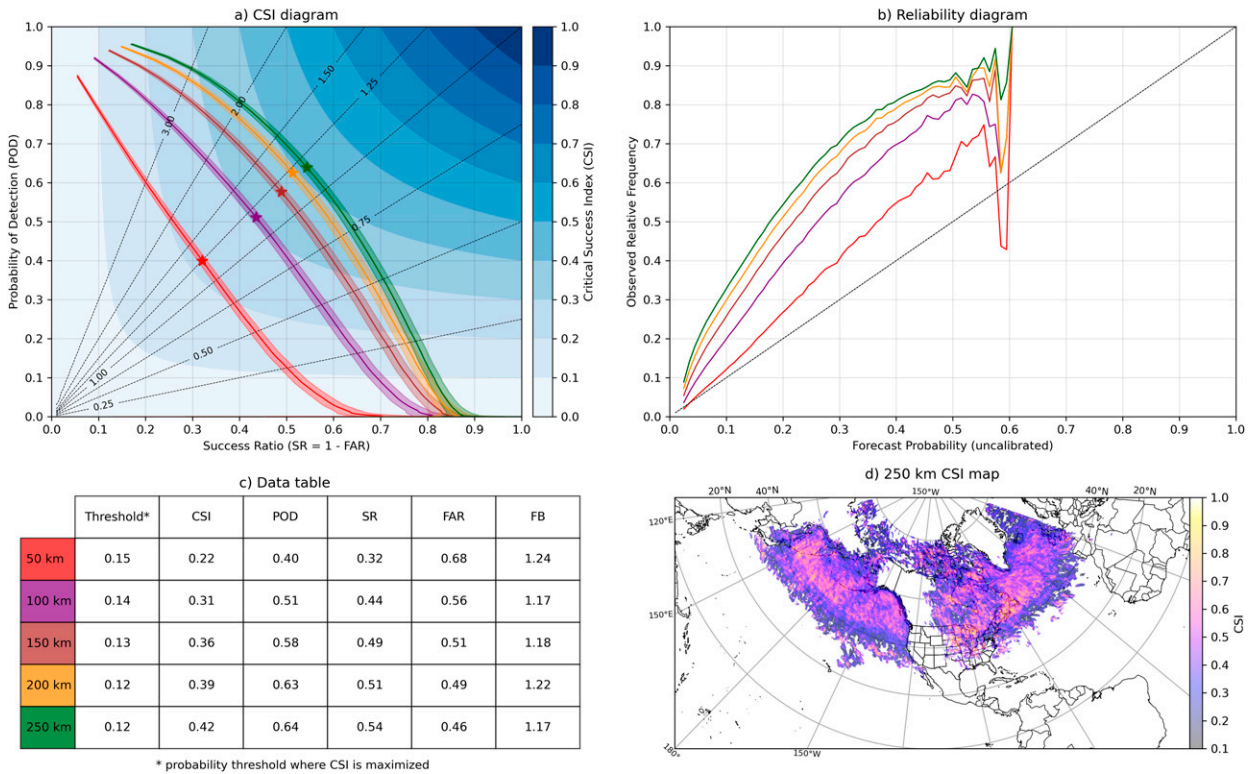


FIG. A13. As in Fig. A12, but for occluded fronts.

2D U-Net 3+ (3x3 kernel): Binary fronts (front / no front) over full domain (90 images per map)

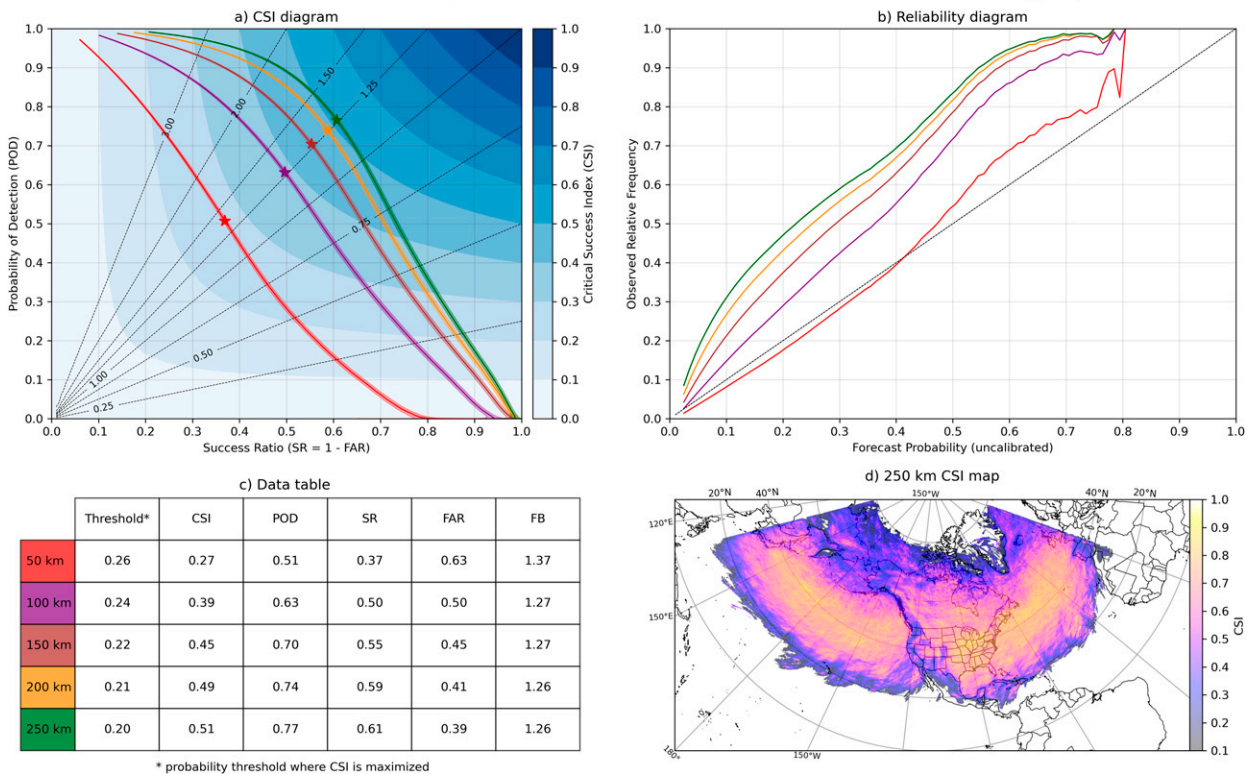


FIG. A14. As in Fig. A12, but for the 2D F/NF model with 3 × 3 convolutions.

REFERENCES

- Bannon, P. R., 1992: A model of Rocky Mountain lee cyclogenesis. *J. Atmos. Sci.*, **49**, 1510–1522, [https://doi.org/10.1175/1520-0469\(1992\)049<1510:AMORML>2.0.CO;2](https://doi.org/10.1175/1520-0469(1992)049<1510:AMORML>2.0.CO;2).
- Berry, G., M. J. Reeder, and C. Jakob, 2011: A global climatology of atmospheric fronts. *Geophys. Res. Lett.*, **38**, L04809, <https://doi.org/10.1029/2010GL046451>.
- Biard, J. C., and K. E. Kunkel, 2019: Automated detection of weather fronts using a deep learning neural network. *Adv. Stat. Climatol. Meteor. Oceanogr.*, **5**, 147–160, <https://doi.org/10.5194/asemo-5-147-2019>.
- Bjerknes, J., 1919: On the structure of moving cyclones. *Mon. Wea. Rev.*, **47**, 95–99, [https://doi.org/10.1175/1520-0493\(1919\)47<95:OTSOMC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1919)47<95:OTSOMC>2.0.CO;2).
- Bolton, D., 1980: The computation of equivalent potential temperature. *Mon. Wea. Rev.*, **108**, 1046–1053, [https://doi.org/10.1175/1520-0493\(1980\)108<1046:TCOEPT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2).
- Bridle, J. S., 1989: Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. *NIPS'89: Proceedings of the 2nd International Conference on Advances in Neural Information Processing Systems*, Vol. 2, MIT Press, 211–217, <https://dl.acm.org/doi/10.5555/2969830.2969856>.
- Catto, J. L., and S. Pfahl, 2013: The importance of fronts for extreme precipitation. *J. Geophys. Res. Atmos.*, **118**, 10791–10801, <https://doi.org/10.1002/jgrd.50852>.
- Chase, R. J., D. R. Harrison, G. Lackmann, and A. McGovern, 2022: A machine learning tutorial for operational meteorology, part II: Neural networks and deep learning. arXiv, 2211.00147v2, <https://doi.org/10.48550/arxiv.2211.00147>.
- Childs, S. J., and R. S. Schumacher, 2019: An updated severe hail and tornado climatology for eastern Colorado. *J. Appl. Meteor. Climatol.*, **58**, 2273–2293, <https://doi.org/10.1175/JAMC-D-19-0098.1>.
- Davies-Jones, R., 2008: An efficient and accurate method for computing the wet-bulb temperature along pseudoadiabats. *Mon. Wea. Rev.*, **136**, 2764–2785, <https://doi.org/10.1175/2007MWR2224.1>.
- Donaldson, R., R. M. Dyer, and R. M. Kraus, 1975: An objective evaluator of techniques for predicting severe weather events. Preprints, *Ninth Conf. on Severe Local Storms*, Norman, OK, Amer. Meteor. Soc., 321–326.
- Ebert-Uphoff, I., R. Lagerquist, K. Hilburn, Y. Lee, K. Haynes, J. Stock, C. Kumler, and J. O. Stewart, 2021: CIRA guide to custom loss functions for neural networks in environmental sciences—Version 1. arXiv, 2106.09757v1, <https://doi.org/10.48550/arXiv.2106.09757>.
- Hersbach, H., and Coauthors, 2018: ERA5 hourly data on pressure levels from 1940 to present. Copernicus Climate Change Service Climate Data Store, accessed 15 March 2021, <https://doi.org/10.24381/cds.bd0915c6>.
- Hewson, T. D., 1998: Objective fronts. *Meteor. Appl.*, **5**, 37–65, <https://doi.org/10.1017/S1350482798000553>.
- Heymysfield, G. M., 1979: Doppler radar study of a warm frontal region. *J. Atmos. Sci.*, **36**, 2093–2107, [https://doi.org/10.1175/1520-0469\(1979\)036<2093:DRSOAW>2.0.CO;2](https://doi.org/10.1175/1520-0469(1979)036<2093:DRSOAW>2.0.CO;2).
- Hobbs, P. V., J. D. Locatelli, and J. E. Martin, 1990: Cold fronts aloft and the forecasting of precipitation and severe weather east of the Rocky Mountains. *Wea. Forecasting*, **5**, 613–626, [https://doi.org/10.1175/1520-0434\(1990\)005<0613:CFAATF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1990)005<0613:CFAATF>2.0.CO;2).
- Hope, P., and Coauthors, 2014: A comparison of automated methods of front recognition for climate studies: A case study in southwest Western Australia. *Mon. Wea. Rev.*, **142**, 343–363, <https://doi.org/10.1175/MWR-D-12-00252.1>.
- Huang, H., and Coauthors, 2020: UNet 3+: A full-scale connected UNet for medical image segmentation. arXiv, 2004.08790v1, <https://doi.org/10.48550/arxiv.2004.08790>.
- Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. arXiv, 1412.6980v9, <https://doi.org/10.48550/arxiv.1412.6980>.
- Lagerquist, R., A. McGovern, and D. J. Gagne II, 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>.
- , J. T. Allen, and A. McGovern, 2020: Climatology and variability of warm and cold fronts over North America from 1979 to 2018. *J. Climate*, **33**, 6531–6554, <https://doi.org/10.1175/JCLI-D-19-0680.1>.
- Lakshmanan, V., C. Karstens, J. Krause, K. Elmore, A. Ryzhkov, and S. Berkseth, 2015: Which polarimetric variables are important for weather/no-weather discrimination? *J. Atmos. Oceanic Technol.*, **32**, 1209–1223, <https://doi.org/10.1175/JTECH-D-13-00205.1>.
- Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner, 1998: Gradient-based learning applied to document recognition. *Proc. IEEE*, **86**, 2278–2324, <https://doi.org/10.1109/5.726791>.
- Liu, G., K. J. Shih, T.-C. Wang, F. A. Reda, K. Sapra, Z. Yu, A. Tao, and B. Catanzaro, 2018: Partial convolution based padding. arXiv, 1811.11718v1, <https://doi.org/10.48550/arxiv.1811.11718>.
- Maddox, R. A., L. R. Hoxit, and C. F. Chappell, 1980: A study of tornadic thunderstorm interactions with thermal boundaries. *Mon. Wea. Rev.*, **108**, 322–336, [https://doi.org/10.1175/1520-0493\(1980\)108<0322:ASOTTI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<0322:ASOTTI>2.0.CO;2).
- Markowski, P. M., J. M. Straka, E. N. Rasmussen, and D. O. Blanchard, 1998: Variability of storm-relative helicity during VORTEX. *Mon. Wea. Rev.*, **126**, 2959–2971, [https://doi.org/10.1175/1520-0493\(1998\)126<2959:VOSRHD>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<2959:VOSRHD>2.0.CO;2).
- McGovern, A., R. Lagerquist, D. J. Gagne II, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199, <https://doi.org/10.1175/BAMS-D-18-0195.1>.
- National Weather Service, 2019: National Weather Service Coded Surface Bulletins, 2003–. Zenodo, accessed 7 July 2020, <https://doi.org/10.5281/zenodo.2642801>.
- Niebler, S., A. Miltenberger, B. Schmidt, and P. Spichtinger, 2022: Automated detection and classification of synoptic-scale fronts from atmospheric data grids. *Wea. Climate Dyn.*, **3**, 113–137, <https://doi.org/10.5194/wcd-3-113-2022>.
- NOAA, 2023: NOAA unified surface analysis fronts. Zenodo, accessed 5 January 2023, <https://doi.org/10.5281/zenodo.7505022>.
- O’Shea, K., and R. Nash, 2015: An introduction to convolutional neural networks. arXiv, 1511.08458v2, <https://doi.org/10.48550/arxiv.1511.08458>.
- Payer, M., N. F. Laird, R. J. Maliawco Jr., and E. G. Hoffman, 2011: Surface fronts, troughs, and baroclinic zones in the Great Lakes region. *Wea. Forecasting*, **26**, 555–563, <https://doi.org/10.1175/WAF-D-10-05018.1>.
- Reed, R. J., and M. D. Albright, 1997: Frontal structure in the interior of an intense mature ocean cyclone. *Wea. Forecasting*, **12**, 866–876, [https://doi.org/10.1175/1520-0434\(1997\)012<0866:FSITIO>2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012<0866:FSITIO>2.0.CO;2).

- Renard, R. J., and L. C. Clarke, 1965: Experiments in numerical objective frontal analysis. *Mon. Wea. Rev.*, **93**, 547–556, [https://doi.org/10.1175/1520-0493\(1965\)093<0547:EINOFA>2.3.CO;2](https://doi.org/10.1175/1520-0493(1965)093<0547:EINOFA>2.3.CO;2).
- Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15**, 163–169, <https://doi.org/10.1002/met.57>.
- Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional networks for biomedical image segmentation. arXiv, 1505.04597v1, <https://doi.org/10.48550/arxiv.1505.04597>.
- Sanders, F., 1955: An investigation of the structure and dynamics of an intense surface frontal zone. *J. Meteor.*, **12**, 542–552, [https://doi.org/10.1175/1520-0469\(1955\)012<0542:AIOTSA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1955)012<0542:AIOTSA>2.0.CO;2).
- , 1999: A proposed method of surface map analysis. *Mon. Wea. Rev.*, **127**, 945–955, [https://doi.org/10.1175/1520-0493\(1999\)127<0945:APMOSM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<0945:APMOSM>2.0.CO;2).
- Schemm, S., I. Rudeva, and I. Simmonds, 2015: Extratropical fronts in the lower troposphere—global perspectives obtained from two automated methods. *Quart. J. Roy. Meteor. Soc.*, **141**, 1686–1698, <https://doi.org/10.1002/qj.2471>.
- Schultz, D. M., 2005: A review of cold fronts with prefrontal troughs and wind shifts. *Mon. Wea. Rev.*, **133**, 2449–2472, <https://doi.org/10.1175/MWR2987.1>.
- , and G. Vaughan, 2011: Occluded fronts and the occlusion process: A fresh look at conventional wisdom. *Bull. Amer. Meteor. Soc.*, **92**, 443–466, <https://doi.org/10.1175/2010BAMS3057.1>.
- Simmonds, I., K. Keay, and J. A. T. Bye, 2012: Identification and climatology of Southern Hemisphere mobile fronts in a modern reanalysis. *J. Climate*, **25**, 1945–1962, <https://doi.org/10.1175/JCLI-D-11-00100.1>.
- Smith, R. K., and M. J. Reeder, 1988: On the movement and low-level structure of cold fronts. *Mon. Wea. Rev.*, **116**, 1927–1944, [https://doi.org/10.1175/1520-0493\(1988\)116<1927:OTMALL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1988)116<1927:OTMALL>2.0.CO;2).
- Stull, R., 2011: Wet-bulb temperature from relative humidity and air temperature. *J. Appl. Meteor. Climatol.*, **50**, 2267–2269, <https://doi.org/10.1175/JAMC-D-11-0143.1>.
- Thomas, C. M., and D. M. Schultz, 2019: Global climatologies of fronts, airmass boundaries, and airstream boundaries: Why the definition of “front” matters. *Mon. Wea. Rev.*, **147**, 691–717, <https://doi.org/10.1175/MWR-D-18-0289.1>.
- Uccellini, L. W., S. F. Corfidi, N. W. Junker, P. J. Kocin, and D. A. Olson, 1992: Report on the surface analysis workshop held at the National Meteorological Center 25–28 March 1991. *Bull. Amer. Meteor. Soc.*, **73**, 459–472.
- Wakimoto, R. M., and B. L. Bosart, 2001: Airborne radar observations of a warm front during FASTEX. *Mon. Wea. Rev.*, **129**, 254–274, [https://doi.org/10.1175/1520-0493\(2001\)129<0254:AROOAW>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0254:AROOAW>2.0.CO;2).
- , and H. Cai, 2002: Airborne observations of a front near a col during FASTEX. *Mon. Wea. Rev.*, **130**, 1898–1912, [https://doi.org/10.1175/1520-0493\(2002\)130<1898:AOOAFN>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1898:AOOAFN>2.0.CO;2).
- , and H. V. Murphey, 2008: Airborne Doppler radar and sounding analysis of an oceanic cold front. *Mon. Wea. Rev.*, **136**, 1475–1491, <https://doi.org/10.1175/2007MWR2241.1>.
- Zhang, D.-L., E. Radeva, and J. Gyakum, 1999: A family of frontal cyclones over the western Atlantic Ocean. Part I: A 60-h simulation. *Mon. Wea. Rev.*, **127**, 1725–1744, [https://doi.org/10.1175/1520-0493\(1999\)127<1725:AFOFCO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<1725:AFOFCO>2.0.CO;2).
- Zhou, Z., M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, 2018: UNet++: A nested U-Net architecture for medical image segmentation. arXiv, 1807.10165v1, <https://doi.org/10.48550/arxiv.1807.10165>.