UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

EXPLAINABLE FRONTAL BOUNDARY PREDICTIONS FOR APPLICATIONS IN OPERATIONAL ENVIRONMENTS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

ANDREW D. JUSTIN
Norman, Oklahoma
2024

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/340396

EXPLAINABLE FRONTAL BOUNDARY PREDICTIONS FOR APPLICATIONS IN OPERATIONAL ENVIRONMENTS

A THESIS APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Amy McGovern, Chair

Dr. John Allen

Dr. Aaron Hill

Dr. Steven Cavallo

©Copyright by ANDREW D. JUSTIN 2024
All Rights Reserved.

Acknowledgements

The work and accomplishments highlighted in this thesis would not have come to fruition without the support of my friends and colleagues, in particular my advisor Amy McGovern and close collaborator John Allen. During my junior year of college, I started to develop an interest in research and decided to reach out to Amy about an open URA position. Amy and John have since fully supported me throughout my time as a URA and GRA, offering assistance whenever I stumbled across a problem that seemed impossible to solve. I would like to thank my numerous colleagues in the OU IDEA Lab group, who are always willing to chime in and help out whenever I have a research related question. I extend my appreciation to my committee members who were willing to take time out of their schedules to assist me in preparing this thesis and offering further guidance and insight to improve my work. I would also like to thank all the NOAA forecasters at WPC, OPC, and TAFB for taking the time to review the products I developed alongside Amy and John and providing invaluable feedback and encouragement on our machine learning models. OU's Supercomputing Center for Education and Research provided many of the resources I used to develop products from this research. I would also like to thank my parents and extended family for their unwavering support and encouragement throughout this project. This work would not have been possible without the support of my family.

This material is based upon work supported by the National Science Foundation under Grant No. ICER-2019758 and the National Oceanic and Atmospheric Administration under Grant No NA20OAR4590347.

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.

Table of Contents

Acknowledgements **iv**

List Of Tables **vi**

List Of Figures **vii**

Abstract **x**

1 Introduction **1**

2 Literature Review **5**

3 Data and Methods **24**
3.1 UNET3+ Architecture 24
3.2 Datasets and Preprocessing 28
3.3 Training and Hyperparameters 32
3.4 Evaluation 34
3.4.1 Permutation Studies 35
3.4.2 Saliency Maps 36

4 Results and Discussion **37**
4.1 Case Study: 2023 Christmas Winter Storm 65
4.2 Case Study: Saliency Maps 72

5 Conclusions and Future Work **81**

Reference List **85**

List Of Tables

2.1 Strengths and weaknesses of non-ML-based automated frontal detection methods discussed in this literature review. 13
2.2 Strengths and weaknesses of ML-based automated frontal detection methods discussed in this literature review. 23

3.1 Predictor variables used as input to FCM. Variables or levels in **boldface** were derived from variables directly retrieved from ERA5, which are not in boldface. 29
3.2 Fraction of pixels containing each front type in the training, validation, and testing datasets. The extents of the CONUS and USAD domains can be found in Figure 3.3. CF = cold front, WF = warm front, SF = stationary front, OF = occluded front, DL = dryline. 30
3.3 Hyperparameter choices for FCM. 33

4.1 FCM performance over CONUS (top) and USAD (bottom). The three values in each cell represent scores using 50, 100, and 250 km neighborhoods. 37

List Of Figures

2.1 a) 850-hPa Thermal Front Parameter (using θ field) and b) 850-hPa Minimum Maximum Locator (using T field) over USAD when applied to ERA5 reanalysis data for 2100 UTC May 20 2019. 8

2.2 2×2 max pooling operation performed on an image (left). The output (right) has its dimensions reduced by a factor equal to the sizes of the pooling filter dimensions. In this case, each dimension is divided by 2. Numbers on the right represent the maximum values contained within each respective filter. 15

2.3 2D convolution operation with a 3×3 kernel (yellow) applied to a 5×5 image (green) with zero padding. The convolved feature (red) has the same shape as the input, in this case 5×5. If zero-padding was not used, the convolved feature would have a shape of 3×3, represented by the non-boldface numbers in the convolved feature. **Boldface** numbers on the convolved feature represent feature elements introduced by zero-padding. 16

3.1 Architecture of the UNET3+ model used to predict cold, warm, stationary, and occluded fronts and drylines. This example shows an input size of 288×128×5×10, where the third (vertical) dimension is unmodified until just prior to deep supervision. Note that the bottom node is both an encoder and decoder node. 25

3.2 a) ReLU and GELU activation functions, and b) derivatives of the ReLU and GELU activation functions. 27

3.3 CONUS domain (red) and NOAA's USAD (blue). The bounds of the CONUS domain are 132°W, 60.25°W, 25°N, 56.75°N (288×128 pixels on the 0.25° ERA5 grid), while the USAD has bounds of 130°E, 10°E, 0°N, 80°N (960×320 pixels after 1-pixel truncation along each dimension). . 32

4.1 Cold front results over CONUS; a) CSI diagram (dashed lines = frequency bias), b) reliability diagram (dashed line = perfect reliability), c) data table with upper and lower performance bounds indicated with superscripts and subscripts, and d) spatial CSI diagram using a 100 km neighborhood. CSI scores lower than 0.1 are not shown on the spatial diagram. 40

4.2 Same as Figure 4.1 but for cold fronts over USAD. 41

4.3 Cold front permutation results over the CONUS domain for a) grouped variables, b) grouped vertical levels, and c) variables on single variables ranked from 1-60 with 1 (60) being the most (least) important variable and level combination. 42

4.4 Same as Figure 4.3 but for cold fronts over USAD. 43
4.5 Frequency of a) cold fronts, b) warm fronts, c) stationary fronts, d) occluded fronts, and e) drylines drawn by NWS forecasters over USAD for the period 2008-2022 at synoptic hours. Non-synoptic hours are not shown as WPC only draws over North America for these timesteps and frequencies are significantly higher over the WPC domain (see Fig. 2 from Justin et al., 2023) 45
4.6 Same as Figure 4.1 but for warm fronts over CONUS. 46
4.7 Same as Figure 4.2 but for warm fronts over USAD. 47
4.8 Same as Figure 4.3 but for warm fronts over CONUS. 48
4.9 Same as Figure 4.4 but for warm fronts over USAD. 49
4.10 Same as Figure 4.1 but for stationary fronts over CONUS. 51
4.11 Same as Figure 4.2 but for stationary fronts over USAD. 52
4.12 FCM predictions for 1800 UTC Jan 2 2019 over USAD. 52
4.13 Same as Figure 4.3 but for stationary fronts over CONUS. 53
4.14 Same as Figure 4.4 but for stationary fronts over USAD. 54
4.15 Same as Figure 4.1 but for occluded fronts over CONUS. 56
4.16 Same as Figure 4.2 but for occluded fronts over USAD. 57
4.17 Same as Figure 4.3 but for occluded fronts over CONUS. 58
4.18 Same as Figure 4.4 but for occluded fronts over USAD. 59
4.19 Same as Figure 4.1 but for drylines over CONUS. 61
4.20 Same as Figure 4.2 but for drylines over USAD. 62
4.21 Same as Figure 4.3 but for drylines over CONUS. 63
4.22 Same as Figure 4.4 but for drylines over USAD. 64
4.23 FCM predictions over CONUS using GFS model data: a) 0000 UTC Dec 26 2023, b) 1200 UTC Dec 26 2023, c) 0000 UTC Dec 27 2023, d) 1200 UTC Dec 27 2023. Note that all predictions use forecast hour 0 for the respective initialization times and are calibrated to a 100 km neighborhood with filled contours at 10% intervals. (blue = cold front, red = warm front, green = stationary front, purple = occluded front) . 66
4.24 GOES-16 mid-level water vapor imagery (band 9) showing a cyclone over CONUS: a) 0001 UTC Dec 26 2023, b) 1201 UTC Dec 26 2023, c) 0001 UTC Dec 27 2023, d) 1201 UTC Dec 27 2023. 67
4.25 WPC surface analyses over CONUS: a) 0000 UTC Dec 26 2023, b) 1200 UTC Dec 26 2023, c) 0000 UTC Dec 27 2023, d) 1200 UTC Dec 27 2023. 68
4.26 Objectively analyzed surface maps from the Storm Prediction Center: a) 0000 UTC Dec 26 2023, b) 1200 UTC Dec 26 2023, c) 0000 UTC Dec 27 2023, d) 1200 UTC Dec 27 2023. 69
4.27 Objectively analyzed 850 hPa maps from the Storm Prediction Center: a) 0000 UTC Dec 26 2023, b) 1200 UTC Dec 26 2023, c) 0000 UTC Dec 27 2023, d) 1200 UTC Dec 27 2023. 70

4.28 FCM cold front predictions over CONUS contoured at 10% intervals (a) and saliency maps at b) the surface, c) 1000 hPa, d) 950 hPa, e) 900 hPa, and f) 850 hPa at 1800 UTC Feb 2 2019. 73
4.29 Same as Figure 4.28 but for warm fronts. 75
4.30 Same as Figure 4.28 but for stationary fronts. 76
4.31 Same as Figure 4.28 but for occluded fronts. 78
4.32 Same as Figure 4.28 but for drylines. 79
4.33 Cold, warm, stationary, and occluded front and dryline positions in the WPC analysis for 1800 UTC Feb 4 2019. 80

Abstract

Frontal boundaries drive many high-impact weather events around the globe. Identifying fronts through various thermodynamic fields increases predictability of hazardous weather phenomena. Frontal analysis is still primarily done by human forecasters, often implementing their own subjectivity rules and criteria for determining frontal positions and types. Subjective placements of fronts can result in various solutions by different forecasters when given identical sets of data. Numerous studies have attempted to make frontal analysis more consistent through numerical frontal analysis, using sets of rules and thresholds with thermodynamic fields to locate and classify fronts. In recent years, machine learning algorithms have gained more popularity in meteorology due to their ability to learn complex relationships within large quantities of atmospheric data. We present a novel machine learning algorithm that predicts five different types of frontal boundaries - cold, warm, stationary, and occluded fronts and drylines. The algorithm was able to locate 76-86% and 70-81% of fronts over CONUS and NOAA's Unified Surface Analysis domain, respectively, on an independent testing dataset. We applied two Explainable Artificial Intelligence methods to the model - permutation studies and saliency maps. Permutation studies allowed us to determine variable importance for each frontal type. Saliency maps for the selected case study gave us insight as to how the model output can change as the ambient environment is modified. While more work needs to be done to improve the algorithm, we have demonstrated that machine learning can be used to develop an accurate and efficient model for detecting frontal boundaries.

Chapter 1

Introduction

Fronts are boundaries and density contrasts that separate two air masses (Bjerknes, 1919; Thomas and Schultz, 2019) and influence the weather experienced around the globe daily. For example, fronts are the leading cause of extreme precipitation events in the United States (Kunkel et al., 2012) and can influence tornado development (Childs and Schumacher, 2019; Maddox et al., 1980). These hazards drive forecasters to identify these boundaries with a reasonable level of accuracy to better prepare the public for future weather events that pose threats to life and property.

Forecasters at the National Weather Service (NWS) draw fronts every three hours over North America and every six hours over the Unified Surface Analysis Domain (USAD) which stretches from 130°E eastward to 10°E and from the equator northward to 80°N. Even among trained forecasters, the subjective nature of drawing fronts leads to a wide variety of solutions for the same data, a well-illustrated example of which was presented by Uccellini et al. (1992). Subjective interpretation can create inconsistencies between contiguous analyses, affecting the quality of the archived frontal data. There is also a significant workload associated with drawing fronts over the entire USAD.

In this thesis, we present a novel machine learning (ML) algorithm to efficiently locate five types of frontal boundaries - cold, warm, stationary, and occluded fronts and drylines. We discuss prior work with automated frontal detection, applications of ML in frontal detection, and highlight how our algorithm improves upon previous work. Explainable Artificial Intelligence (XAI) methods are applied to our new algorithm to

understand what variables are important in detecting different frontal types and how model predictions change as functions of the inputs.

Our contributions to objective frontal analysis methods include developing two different models trained to detect cold, warm, stationary, and occluded fronts over NOAA's Unified Surface Analysis domain (USAD). Both models use a deep learning approach based on the UNET3+ architectures (Huang et al., 2020), which will be described in greater detail in Section 3.1.

In our original work (Justin et al., 2023), three sets of models were trained - one predicted cold and warm fronts (CFWF), another predicted stationary and occluded fronts (SFOF), while the third set applied the same label to the four aforementioned frontal types and performed binary classification (front / no front, FNF). Data used in this previous study covered the period 2008-2020. We used 12 variables at the surface, 1000, 950, 900, and 850 hPa sourced and/or derived from ERA5 reanalysis data (Hersbach et al., 2018). Frontal positions are hand-drawn by NWS forecasters and treated as ground truth for the models. These were obtained from the NWS and contained fronts drawn over North America every three hours and over USAD every six hours. The models were trained by randomly selecting 128×128 grid point (32°×32° on the ERA5 lat/lon grid) patches over our pre-defined CONUS domain (USAD and CONUS domains can be seen in Figure 3.3) and passing them into the models. Outside of slicing the data and performing min-max normalization, no preprocessing or augmentation was applied to the training data. The same process was applied to the validation data after every training epoch. Nine years of data were chosen for training, two years were used for validation, and the two remaining years were used for testing.

Overall, the original models showed good skill at detecting all frontal types over CONUS and CSI scores were similar to those achieved by Niebler et al. (2022, hereafter N22), though direct comparisons should be made with caution as our testing

dataset covered a different period than N22. We hypothesized that cold fronts would be the best-performing front type, which ended up being the case for both CONUS and USAD. Warm fronts were the worst performing front type, consistent with findings from Lagerquist et al. (2019) and N22. An exceptionally high FAR with stationary fronts was found over USAD, which we eventually attributed to the models falsely identifying the Intertropical Convergence Zone (ITCZ) as a large stationary front due to the wind convergence. We also performed permutation studies to determine which variables and levels were important to the models' predictions (Breiman, 2001; McGovern et al., 2019).

In this thesis, we address some of the notable flaws with our previous model. First, the models were exceptionally large, some exceeding 233 million parameters. Second, the models had a fixed input shape, meaning that input images larger than 128×128 pixels could not be passed into the models. This forced us to resort to image stitching, where multiple 128×128 images would be stitched together over a larger domain to generate a final prediction. The image stitching combined with the excessively large models meant that the models would be computationally intensive and would likely not be very efficient when deployed in operational environments. Finally, the separation of cold and warm fronts from stationary and occluded fronts prevented the CFWF and SFOF models from knowing about each other's frontal types (e.g., CFWF models were given no information about the existence of stationary and occluded fronts). In practice, this could result in the CFWF models predicting a 70% chance of a cold front at a location where the SFOF models predict a 70% chance of a stationary front. This situation makes no mathematical or physical sense and is the main driver of motivating us to develop a new, single algorithm that predicts cold, warm, stationary, and occluded fronts. We did not continue to pursue a new FNF algorithm as binary classification turned out to not be useful for Weather Prediction Center (WPC) forecasters who

preferred having models that classified fronts by type. We also explored the idea of adding drylines into our algorithm, as they possess characteristics that are unique from the other frontal types (e.g., drylines are defined by their often extreme moisture gradients, see Pietrycha and Rasmussen, 2004; Hoch and Markowski, 2005).

This thesis details further contributions to automated frontal detection through the use of a new machine learning algorithm. With the inclusion of five frontal types in a single model, we expected to see the new model perform better than our previous model system outside of CONUS with lower FAR scores for all frontal types, especially stationary fronts. We expected that dryline performance would be comparable to cold fronts given their unique thermodynamic properties. We applied two XAI methods to our algorithm, exploring case studies to give deeper insight into how the model behaves and responds to changes in the environment. The overarching goal of this work is to present an efficient and accurate frontal detection algorithm that can assist forecasters in the labor-intensive process of frontal analysis.

Chapter 2

Literature Review

Fronts encompass the transition zone and density contrast between two air masses with differing thermodynamic characteristics. Depending on the specific characteristics of the air masses and the direction of movement of the front, the front can be classified in multiple ways. Cold fronts mark the leading edge of a moving cold and often dry air mass. Following the passage of a cold front, a sharp drop in temperature and moisture content and a sudden shift in the wind direction can often be noted (Schultz, 2005). The translation of a cold front forces warm, moist air ahead of the boundary to rise, which under certain conditions can encourage convection (Smith and Reeder, 1988; Hobbs et al., 1990; Catto and Pfahl, 2013) and secondary frontal cyclogenesis (Zhang et al., 1999).

Warm fronts separate a moving warm air mass from a colder air mass. Warm fronts are usually associated with weaker wind shifts and thermal gradients than cold fronts, making them more difficult to locate on a surface chart. While warm fronts often have more subtle vertical slopes, they can influence hazardous weather and may even encourage tornadogenesis (Maddox et al., 1980).

Stationary fronts are the equatorward edges of slow-moving density discontinuities (National Weather Service, 2013). Stationary fronts are often characterized by parallel or convergent flow on either side of the boundaries. Despite the name, stationary fronts are not necessarily stationary, but rather are labeled stationary when they are moving slower than a specified threshold (usually 10 knots, see National Weather

Service, 2013), leading many to prefer the term "quasi-stationary" instead. Stationary fronts most commonly reside near terrain features (see Figure 4.5). Enhanced relative vorticity along stationary boundaries can encourage landspout tornado formation in developing thunderstorms (Childs and Schumacher, 2019).

Occluded fronts form when an extratropical cyclone has reached peak maturity and a thermal ridge above the surface starts to elongate and wrap around the cyclone center (Martin, 1999; Schultz and Vaughan, 2011). Like cold and warm fronts, occluded fronts are commonly associated with temperature gradients and wind shifts. Occluded fronts are commonly associated with enhanced precipitation (Schultz and Vaughan, 2011).

Drylines mark the transition between a hot, dry air mass and a cooler, moist air mass. They are commonly defined by sloped terrain, and the moisture gradients across drylines can be rather extreme (Schaefer, 1986; Pietrycha and Rasmussen, 2004; Hoch and Markowski, 2005). NWS forecasters usually look for a 14-17°C change in the dewpoint across the boundary to label it as a dryline (National Weather Service, 2013). Sharp changes in mixing ratio or specific humidity can also indicate the presence of a dryline.

Numerous studies have attempted to automate the frontal analysis process to improve accuracy and consistency in the depicted positions of fronts. Strengths and weaknesses of aforementioned non-ML-based methods can be found in Table 2.1. Renard and Clarke (1965, hereafter RC65) found that front positions could vary by up to 300 nautical miles (555.6 km) among various analysts and forecast centers. RC65 used potential temperature (θ) and its derivatives to locate frontal boundaries, eventually settling on a parameter that would later become known as the thermal front parameter [TFP; Equation (2.1)]:

$$TFP = -\frac{\nabla|\nabla\tau| \cdot \nabla\tau}{|\nabla\tau|}, \tag{2.1}$$

where τ is any thermodynamic variable with first and second order derivatives. Since RC65 define fronts as the warm air boundaries in synoptic-scale baroclinic zones, fronts are drawn along TFP ridges with this method. Limited comparisons to surface analyses resulted in RC65 performing analyses at the 850 hPa level, and the sloped nature of fronts (e.g., Smith and Reeder, 1988) means that the front analyzed at 850 hPa will not always be aligned with the front analyzed at the surface. Building on the TFP, Clarke and Renard (1966) developed the Minimum Maximium Locator [MML; Equation (2.2)]:

$$MML = \nabla\tau \cdot \frac{\nabla(TFP(\tau))}{|\nabla(TFP(\tau))|}, \tag{2.2}$$

where τ is any thermodynamic variable with first, second, and third order derivatives. Using temperature (T) at 850 hPa, MML was concluded by Hewson (1998, hereafter H98) to perform worse than TFP ($\tau = \theta$) at locating fronts, receiving relative accuracy ratings of "Fair" and "Fair/Poor", respectively; however it should be noted that the ratings assigned by H98 are subjective. Regardless, the use of second and third order derivatives in these parameters has the downside of producing noisy results, especially with high-resolution weather data. As is discussed later in this section, this is one of the motivations behind ML-based frontal detection; some ML methods are robust when working with noisy data (Krause et al., 2016). To illustrate this issue, TFP and MML were applied to ERA5 reanalysis data over USAD for 2100 UTC May 20 2019 (Figure 2.1). Both parameters resulted in noisy outputs but were able to detect some prominent fronts, though could not differentiate their types. H98 argued that an ideal frontal plotting technique should be entirely automatic as this would remove the human-driven subjectivity from frontal analysis, though the techniques are still partly subjective, as humans are choosing the parameters to use and how they are applied to thermodynamic fields.

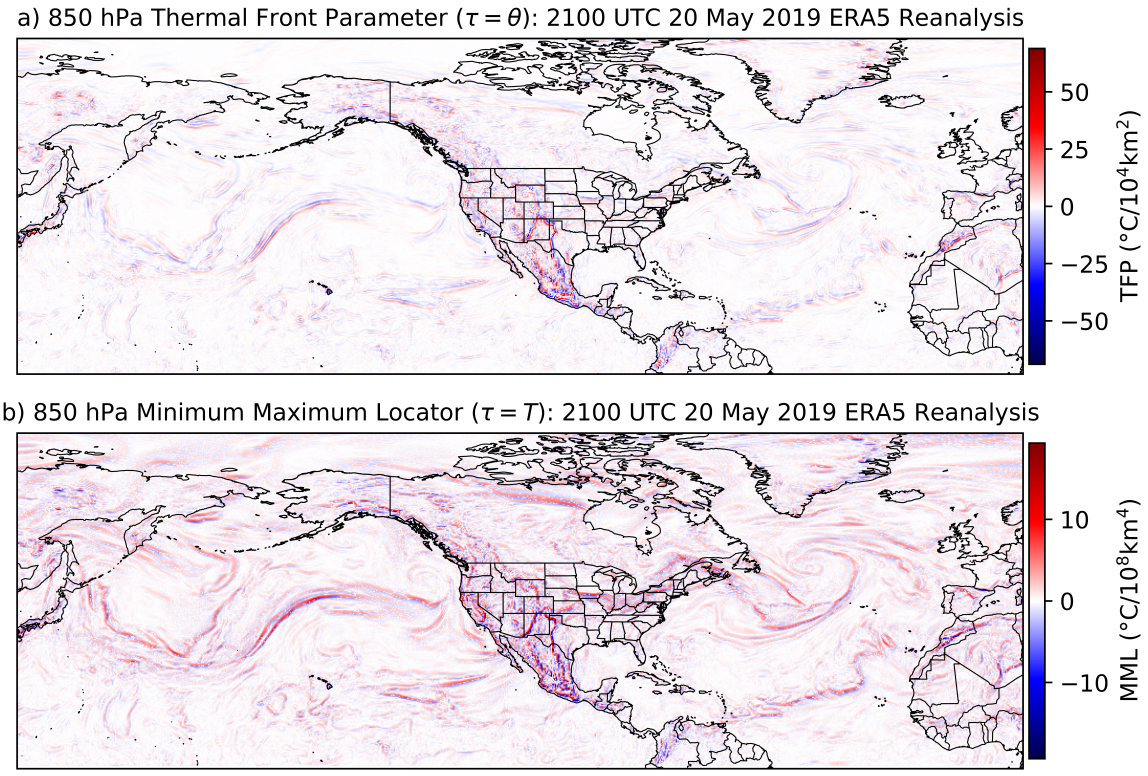


Figure 2.1: a) 850-hPa Thermal Front Parameter (using θ field) and b) 850-hPa Minimum Maximum Locator (using T field) over USAD when applied to ERA5 reanalysis data for 2100 UTC May 20 2019.

Building upon findings from H98, Berry et al. (2011, hereafter B11) performed a 44-year climatology of fronts across the entire globe. While B11 determined that most thermodynamic variables yielded TFP fields that were well correlated with the manual frontal analyses used in their climatology, they decided to use θ_w at 850 hPa as input to TFP as it most closely resembled the frontal analyses. TFP values less than $-8 \times 10^{-12} K/m^2$ were masked out in order to remove weaker frontal zones. B11 then separated the fronts into cold, warm, and quasi-stationary categories based on each front's movement speed perpendicular to the orientation of their θ_w contours, using a frontal speed parameter from H98 (V_{front}; Equation (2.3)):

$$V_{front} = \frac{\overrightarrow{\boldsymbol{V}} \cdot \nabla|\nabla\theta_w|}{|\nabla|\nabla\theta_w||} \tag{2.3}$$

where $\overrightarrow{\boldsymbol{V}}$ is the 2-D wind field. Fronts with V_{front} exceeding 1.5 m/s were classified as warm fronts, while fronts with V_{front} less than -1.5 m/s (speed still exceeding 1.5 m/s) were classified as cold fronts. Quasi-stationary boundaries were defined as fronts having speeds less than 1.5 m/s. In H98, negative (positive) V_{front} values were associated with cold (warm) fronts. B11 found that their methodology of categorizing cold, warm, and quasi-stationary fronts produced similar results to subjective analyses and front frequencies matched conceptual models of mid-latitude cyclones.

Simmonds et al. (2012, hereafter S12) created a climatology of fronts in the Southern Hemisphere using temporal changes in the meridional component of the wind (v-wind). Fronts were identified in pixels that experienced a northwest-to-southwest wind shift with a 2 m/s increase of 10 m v-wind over the course of six hours. They concluded that their wind shift method allowed them to locate nearby fronts with reasonable accuracy; however, since they used ERA-interim data (Dee et al., 2011) with a grid resolution of 1.5°, it is unknown if the wind shift method is a viable option for detecting fronts on higher-resolution data such as ERA5 reanalysis (Hersbach et al., 2018), which is on a

0.25° grid. This appears to be an issue with other front detection methods originally applied on low-resolution data (e.g., TFP and MML) that produce noisy outputs when applied to ERA5 reanalysis data (see Figure 2.1).

In a case study that aimed to identify fronts in Southwestern Australia, Hope et al. (2014) compared the wind method from S12 and B11's method that used θ_w at 850 hPa. They also used a self-organizing map (SOM; Kohonen, 1982), a type of unsupervised ML method, to organize fields of three variables (mean sea level pressure, 500 hPa geopotential height, thickness) into 20 groups, with each group having characteristics determined by the SOM to be distinct from the other groups. Hope et al. (2014) found that the SOM and both methods from S12 and B11 had lower interannual variability than manual hand analyses. Methods from B11 and S12 tended to identify more fronts than the manual analyses, while the SOM identified fewer fronts. The three methods all identified more fronts on days with rain than days without, however the SOM identified fewer fronts than the other two methods and was less correlated with the manual analyses. It was noted that the SOM sometimes struggled to differentiate between fronts and other synoptic features (e.g., cutoff lows), a likely source of the low correlation with the manual analyses.

Schemm et al. (2015, hereafter S15) used two methods to identify fronts, one being a thermal (TH) method and the other being a wind (WND) method. Implementing the TFP developed by Renard and Clarke (1965), S15 settled on using equivalent potential temperature (θ_e) at 850 hPa as the thermodynamic variable for TFP calculations in the TH method due to its conservation properties and inclusion of moisture content. θ_e gradients smaller than 4 K $(100\ km)^{-1}$ were masked out to remove weak thermal gradients that may not be representative of frontal zones. S15 drew fronts where $TFP = 0$, whereas B11 drew fronts where $\nabla TFP = 0$. S15 found that their TH method typically placed fronts in the middle of surface precipitation patterns, while the

method from B11 generally placed fronts on the leading edge of precipitation patterns and frontal zones. The WND method used by S15 was very similar to Simmonds et al. (2012); the only difference being S15 looked for southwest-to-northwest wind shifts as opposed to northwest-to-southwest wind shifts since they were analyzing the Northern Hemisphere. They found that the WND method was able to identify cold fronts and performed well with weaker baroclinic zones, while the TH method was good at detecting both cold and warm fronts with strong thermal gradients. However, S15 notes that these methods should be used with caution given the large number of user-selected parameters and thresholds. While the parameters can be tuned for specific applications and regional climatologies, such fine-tuning can introduce the risk of overfitting the methods to unique situations that cannot be applied to other geographical regions where frontal structures may differ.

The methods described above remove subjectivity from frontal analysis by providing objective criteria for detecting fronts, however the criteria for these methods are still subjectively chosen by humans and therefore are subjective in nature. The decades-old problem of subjective frontal analysis has spurred motivation to research the potential for applying ML algorithms in the frontal analysis process. ML is a broad term encompassing a range of algorithms whose parameters update automatically when given more data (Janiesch et al., 2021). In recent years, the use of ML in meteorological applications has grown exponentially (Chase et al., 2022), including for frontal analysis (e.g., Biard and Kunkel, 2019; Bochenek et al., 2021; Clark et al., 2015; Dagon et al., 2022; Justin et al., 2023; Lagerquist et al., 2019; Matsuoka et al., 2019; Niebler et al., 2022). Automated front detection with ML provides the benefit of having an algorithm(s) whose parameters update without human intervention after the training process begins, making for an objective frontal detection method. Unlike

the traditional automated methods mentioned above, ML algorithms can learn the various structures associated with different frontal types using data covering multi-year periods.

Method	Strengths	Weaknesses
Renard and Clarke (1965)	• Fronts are commonly located in troughs • Locates fronts within 0.4° of manual analyses over US	• Noisy outputs on high-resolution data • Overpredicts compared to manual analyses
Clarke and Renard (1966)	• Locates two-thirds of fronts within 2° of ground truth	• Low correlation with manual hand analyses • High false alarm rate • Noisy outputs on high-resolution data
Berry et al., (2011)	• Utilizes thermal gradients and wind shifts • Fronts are found close to troughs and cloud features • Can discern between cold, warm, and stationary fronts • Applied to entire globe	• Coarse dataset (2.5° ERA-40 reanalysis; Uppala et al., 2005) • Prediction of front type not overly reliable
Simmonds et al., (2012)	• Tracking algorithm predicts front locations 6 hours in advance • Can determine the length, intensity, and meridional slopes of fronts	• Limited to Southern Hemisphere • No thermal gradients explicitly used • Cannot determine front type
Hope et al., (2014)	• Can discern between cold, warm, and stationary fronts • Identifies fronts that bring precipitation • May also detect prefrontal troughs	• Limited to Southwest Western Australia • Coarse dataset (2.5° NCAR-NCEP reanalyses; Kalnay et al., 2018) • No compelling evidence for using wind or thermal method only
Clark et al., (2015)	• Only objective dryline algorithm in literature (to our best knowledge) • 90% of drylines detected within 15 km	• High false alarm rate • Labels some cold fronts and outflow boundaries as drylines
Schemm et al., (2015)	• Thermal method corresponded with precipitation patterns • Good performance with weak and strong baroclinic zones	• Wind method is sensitive to parameter choices • Methods may overfit to local regions with parameter choices

Table 2.1: Strengths and weaknesses of non-ML-based automated frontal detection methods discussed in this literature review.

One popular ML algorithm used to detect frontal boundaries is the UNET (Ronneberger et al., 2015, hereafter R15), a type of convolutional neural network (CNN; LeCun et al., 1989) designed for image segmentation and object recognition (Denker and Burges, 1995). CNNs perform convolution operations on input images using kernels or filters (terms used interchangeably) to extract features that can be used to identify objects within the images. The features are then passed through an activation function to enable the model to learn relationships between features and variables. Non-linear activation functions enable the model to learn non-linear relationships, and the absence of non-linear activation functions will restrict the model to learning linear relationships (Lederer, 2021). The UNET has an "encoder" and "decoder" in the architecture. The encoder extracts features with convolutions, passes the features through a user-defined activation function, then max pooling operations shrink the features down so more low-level features can be learned by the model (see **??** for max pooling illustration). This process is usually repeated until the features are a sufficiently small size, however this size is subjectively chosen. The decoder enlarges low-level features via upsampling and retrieves features from the encoder nodes through "skip connections". Skip connections help to maintain spatial information within an image that would otherwise be lost through the max pooling operations in the encoder portion of the UNET. The UNET outputs pixel-wise probabilities between 0 and 1 (non-inclusive ends) for each class, enabling the UNET to be used for object detection problems. Identifying frontal boundaries can be interpreted as an object detection problem; many fronts of various types can be in close proximity to one another, making the UNET a good candidate for frontal detection.

Strengths and weaknesses of ML-based frontal detection methods discussed in this section can be found in Table 2.2. Clark et al. (2015, hereafter C15) used two algorithms to identify drylines - one algorithm was not ML-based, while the other was a

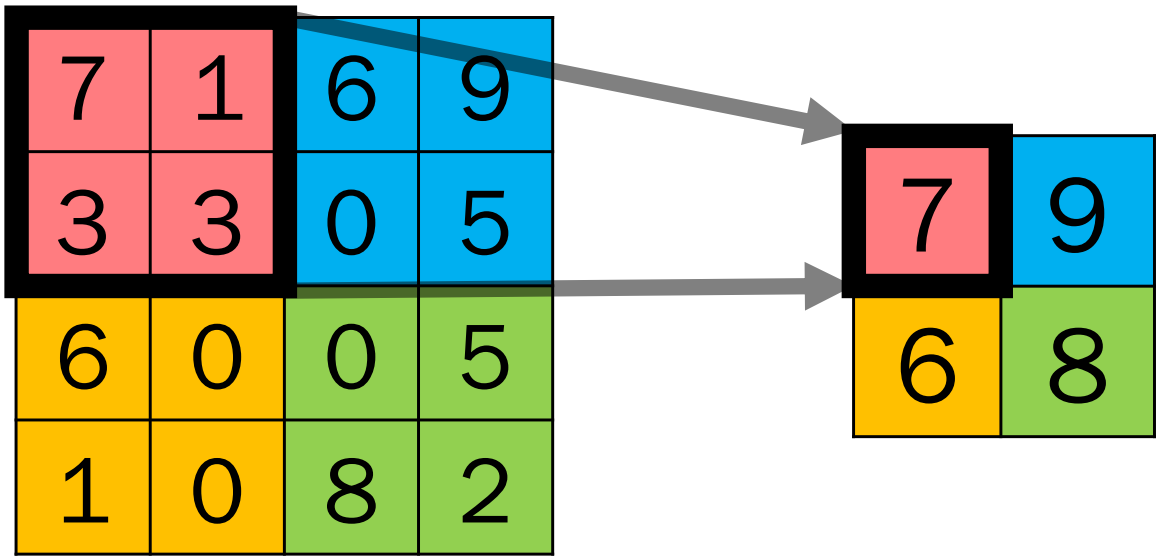


Figure 2.2: 2×2 max pooling operation performed on an image (left). The output (right) has its dimensions reduced by a factor equal to the sizes of the pooling filter dimensions. In this case, each dimension is divided by 2. Numbers on the right represent the maximum values contained within each respective filter.

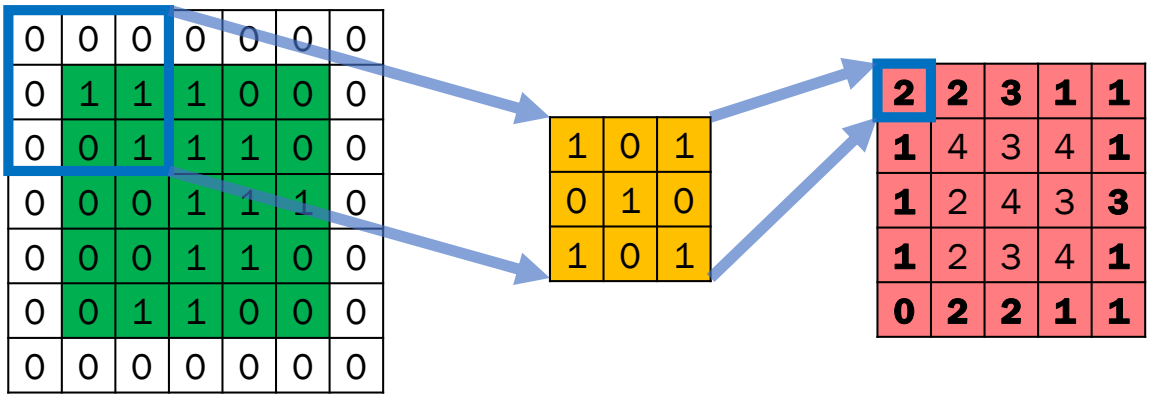


Figure 2.3: 2D convolution operation with a 3×3 kernel (yellow) applied to a 5×5 image (green) with zero padding. The convolved feature (red) has the same shape as the input, in this case 5×5. If zero-padding was not used, the convolved feature would have a shape of 3×3, represented by the non-boldface numbers in the convolved feature. **Boldface** numbers on the convolved feature represent feature elements introduced by zero-padding.

random forest (RF Breiman, 2001), a type of ML model. RFs consist of numerous randomly-generated, interlaced decision trees that work together to make predictions. The algorithms were compared against 145[1] manually identified drylines from Apr-June 2007-2012. The non-ML-based algorithm involved transforming smoothed fields of 2 m dewpoint and specific humidity to objectively locate drylines. When only considering drylines detected by the algorithm that come within 30 km of a manually identified dryline, 75 and 90 percent of the objective drylines fall within 10 and 15 km of manually identified drylines, respectively. However, when considering all objective drylines, the algorithm identifies many more drylines than indicated by the manual analyses used for comparison. The false alarm rate (FAR) was 60-70% with a 40 km neighborhood. The highest objective dryline frequencies match up well with those of manually identified drylines, as can be seen in Figure 6 from C15. FAR dropped below 30% with the introduction of a random forest that assisted the algorithm, albeit with the tradeoff of a 5-10% decrease in POD. While the RF decreased the FAR, the performance from the non-ML-based model suggests that ML may not be needed to produce an algorithm that can effectively locate drylines.

Lagerquist et al. (2019, hereafter L19) trained multiple CNNs to identify cold and warm fronts over North America. Combinations of fields of 1000-hPa temperature, specific humidity, θ_w, geopotential height, u-wind, v-wind were used as inputs and the CNNs output probabilities of no front, cold front, and warm front at a single grid point. The fields were sourced from the North American Regional Reanalysis (NARR; Mesinger et al., 2006) and laid out on a grid with 32 km spacing. They avoided using data at higher resolutions due to spatial and temporal inconsistencies within the ground truth labels (see their Figure 5). Ground truth front labels were sourced from the NWS' Coded Surface Bulletin (CSB; National Weather Service, 2019), whose spatial

[1]Incorrectly stated as 135 drylines in C15.

and temporal extents are covered by the NARR domain. Grid cells with less than 100 analyzed fronts were not included in the training, validation, or testing datasets. The CNN architecture is shown in their Figure 2 and is summarized as follows. Each CNN contained two or three convolution "blocks" (depending on the image size used for training), with each block containing five layers: two groups of 3×3 convolution and rectified linear unit (ReLU) activation function layers (more on ReLU in Section 2.2), followed by a 2×2 max pooling layer. The features are then collapsed into vectors before the CNN makes its prediction for the target pixel. L19 found that CNNs which ingested larger images performed better than those with smaller images, likely due to more spatial information being given to the model. It was also noted that the two best performing CNNs did not have θ_w included in the list of predictor variables, which is noteworthy considering previous studies have used θ_w in objective algorithms for identifying fronts (e.g., Berry et al., 2011; Catto and Pfahl, 2013; Hewson, 1998). In terms of critical success index (see Section 2.4), the best CNN from L19 significantly outperforms traditional numerical frontal analysis (NFA) methods with 100 and 250 km neighborhoods, showing that CNNs are a viable option for ML architecture when constructing an algorithm to detect frontal boundaries. They also found that the CNNs handled noisy data much better than the traditional NFA methods, particularly with data in lower levels of the troposphere.Lagerquist et al. (2020) developed a new CNN that ingested temperature, specific humidity, u-wind, and v-wind at the surface and 850 hPa. The CNN from Lagerquist et al. (2020) was used to generate a 40-year climatology of cold and warm fronts across North America and outperformed the CNN used in L19.

Matsuoka et al. (2019, hereafter M19) used UNET architectures to detect stationary fronts over and around Japan. Retrieving analysis data from the Japan Meteorological Association's GPV-MSM model, M19 used multiple fields to train the UNETs,

including precipitation, sea level pressure, relative humidity, water vapor content[2], temperature, and wind velocity at 1000, 925, 850, and 700 hPa. Gradients of temperature, θ_e, and wind direction were also calculated at the same pressure levels and used to train the UNETs. They found that temperature and wind velocity at 1000 hPa yielded better results than at other pressure levels. The UNETs did not have significant performance improvements when more variables were added to the inputs, suggesting that either few variables are needed to detect stationary fronts or that more data is needed for the UNET to properly learn relationships between all the variables. M19 also noted that the UNET performance was lower when a typhoon was nearby and attributed this to typhoons disrupting thermodynamic fields used to train the UNETs.

Biard and Kunkel (2019, hereafter BK19) trained a 2D CNN to detect cold, warm, stationary, and occluded fronts. This algorithm will be referred to as DL-FRONT to be consistent with their study. Aside from predicting stationary and occluded fronts in addition to cold and warm fronts, DL-FRONT had a couple characteristics distinct from other model architectures discussed in this chapter. First, DL-FRONT implemented dropout layers after the convolutional layers (with the exception of the final convolutional layer). Dropout randomly omits a user-defined fraction of units within a layer(s) during training to prevent the model from overfitting (Hinton et al., 2012). The units chosen to be omitted change throughout the training process. Overfitting is weakened with dropout as the absence of a fraction of the units encourages the model to learn relationships with less information and prevents large models from learning noise in the input data. BK19 chose to omit 50% of the units in each dropout layer. Second, DL-FRONT did not implement any max pooling like the architecture from L19. The output of DL-FRONT was an image with the same spatial shape as the input image. Using zero-padding in the convolutional layers, the shapes of all layers in the model

[2]M19 does not mention what variable "water vapor" is referring to.

are the same (with the exception of the softmax layer). Inputs to DL-FRONT were fields of sea level pressure, 2m temperature and specific humidity, and 10m u-wind and v-wind sourced from the MERRA-2 dataset (Gelaro et al., 2017). The ground truth labels were sourced from the NWS' CSB. Performing three-fold cross-validation (Refaeilzadeh et al., 2009), BK19 found that DL-FRONT achieved an Area Under the Curve (AUC) score of 0.90. The seasonal front climatologies from the CNNs and the CSB were very similar, showing that the model predictions were representative of the underlying labels used to the train DL-FRONT. It was also found that the algorithm was able to predict many fronts highlighted by Kunkel et al. (2012) as producers of extreme precipitation.

Bochenek et al. (2021) used an RF algorithm to detect fronts over Central Europe. They used over one dozen variables at five different levels from ERA5 reanalysis as the inputs (Hersbach et al., 2018), and the ground truth front labels were taken from the Deutscher Wetterdienst (DWD). Cold, warm, and occluded fronts were all assigned the same label, so the random forest was not predicting the type of front present but rather whether any existed at a particular location. The training datasets consisted of only one to six months of data, however Bochenek et al. (2021) was still able to achieve probability of detection (POD) scores of $\sim$40% with false alarm rates (FAR) under 50%. They also determined that rain water content and total precipitation contributed most to the algorithm's predictions.

Niebler et al. (2022, hereafter N22) trained UNET architectures to detect cold, warm, stationary, and occluded fronts over much of the Northern Hemisphere. Inputs to the UNETs were sourced from ERA5 reanalysis and include temperature, specific humidity, u-wind, v-wind, vertical velocity, at nine pressure levels between the surface and 700 hPa, along with surface pressure and the longitudinal distance between grid points. The UNETs architecture implemented by N22 had a few differences from the

original architecture used in R15. First, the inputs to the models initially undergo convolutions with kernels of size 1×1 rather than using the same kernel size throughout the entirety of the UNET encoder (5×5 in the case of N22). The outputs from these convolutions are passed through a ReLU activation function layer before a batch normalization layer (Ioffe and Szegedy, 2015), which normalizes inputs to the layer such that the inputs have a mean of 0 and a variance of 1. This can provide numerical stability in the model during the training process by preventing exploding or vanishing gradients (Philipp et al., 2018). Second, dropout is also implemented before the main encoder portion of the UNETs with 20% of units randomly omitted with each batch during training. As mentioned earlier, dropout helps to prevent overfitting models to the training data, especially with deep networks containing very large numbers of parameters. Lastly, instead of max pooling, the UNETs implemented average pooling and dropout to down-sample features in the model. 20% of units were omitted with dropout during the down-sampling operations. Whether these characteristics of N22's UNETs improve upon the original architecture from R15 is unclear. They also augmented the training dataset, applying random flips along the longitude and latitude dimensions of the images to increase the number of training samples. Data augmentation is the process of generating modified copies of training data and has been shown to yield better model performance when compared to models trained on unmodified datasets (Yang et al., 2023). N22 also stated that flipping the data might lead to better representation of frontal systems in the Southern Hemisphere. Three models were trained - one with front labels from NWS over North America, another with labels from DWD over Europe, and a third that was trained over both the aforementioned regions. N22 found that the UNETs performed best on the regions they were trained on (e.g., UNET trained on NWS region performed best over NWS region). They also found that the UNET trained over both regions at once nearly exceeded performance

of the other two UNETs over their respective training domains, highlighting the importance of including more training data to improve the a model's ability to generate accurate predictions. Warm fronts were the most difficult front type to detect with the UNETs, likely due to the weaker thermal gradients associated with them, while the UNETs performed best with cold fronts. Overall, the UNETs from N22 showed excellent performance over both NWS and DWD regions and drastically outperformed their baseline methods used as comparisons. The UNETs also had better correlation with the NWS and DWD front climatologies than the objective baseline methods. The downfalls of previous objective methods and improvements in ML architectures motivated us to explore applying modern ML methods to frontal detection.

Method	Strengths	Weaknesses
Clark et al., (2015)	• Only ML-based dryline algorithm in literature (to our best knowledge) • False alarm rate lower than non-ML-based method	• POD lower than non-ML based algorithm
Biard and Kunkel (2019)	• Can discern between cold, warm, stationary, and occluded fronts • Cross-validation prevents overfitting • 97% of predictions within 200 km of fronts	• Inputs only include surface variables • Undercounts warm fronts
Lagerquist et al., (2019)	• Outperforms baseline experiments • Highly optimized pixelwise predictions	• No vertical structures accounted for • Computationally inefficient • Only applied to North America
Matsuoka et al., (2019)	• Includes precipitation as input • Identified seasonal rain-producing fronts	• Only detects stationary fronts • Only applied around Japan
Lagerquist et al., (2020)	• CNNs trained with two vertical levels • Small frontal regions removed to reduce noise	• Computationally inefficient • Only applied to North America
Bochenek et al., (2021)	• Multiple vertical levels in inputs • Discerns between cold, warm, and occluded fronts	• Only applied to Europe • Limited training period
Niebler et al., (2022)	• Trained over multiple domains • Multiple vertical levels in inputs • Can discern between cold, warm, stationary, and occluded fronts	• Struggles to detect weak warm fronts
Justin et al., (2023)	• Good performance outside training domain in Northern Hemisphere • Processes 3-D features and resolves vertical structures with 3-D convolutions • Detects cold, warm, stationary, and occluded fronts	• Computationally inefficient • Cold and warm fronts trained separately from stationary and occluded • High warm and stationary front false alarm rates • Only trained over CONUS

Table 2.2: Strengths and weaknesses of ML-based automated frontal detection methods discussed in this literature review.

Chapter 3

Data and Methods

3.1 UNET3+ Architecture

The UNET3+ is a convolutional neural network (CNN) designed for image segmentation (Huang et al., 2020), improving upon the original UNET from Ronneberger et al. (2015) by adding "full-scale skip connections" and "aggregated feature maps" (more on this later in the section). Image segmentation is the process of labeling various parts of an image as different classes. In this case, the UNET3+ is assigning a class (front type) to every pixel of the model output. The architecture for our new UNET3+ model is shown in Figure 3.1. We will refer to this new architecture as the "five-class model" (abbreviated FCM) throughout this text, with the name stemming from the fact that the model predicts the locations of five types of frontal boundaries.

FCM ingests 4D images as input, with the dimensions representing longitude, latitude, vertical level, and the predictor variables used in the model. In other words, multiple 3D variable maps are input to the model, with the variable maps stacked along the fourth dimension of the input images. Unlike our previous three-model system (J23), the longitude and latitude dimensions of FCM are not fixed, and these dimensions of the input images can be changed under the condition that the lengths of the dimensions are evenly divisible by 8. In each encoder and decoder node of FCM, the images/features pass through two convolution "modules", with each module consisting of a 5×5×5 convolution operation, a batch normalization layer (Ioffe and Szegedy,

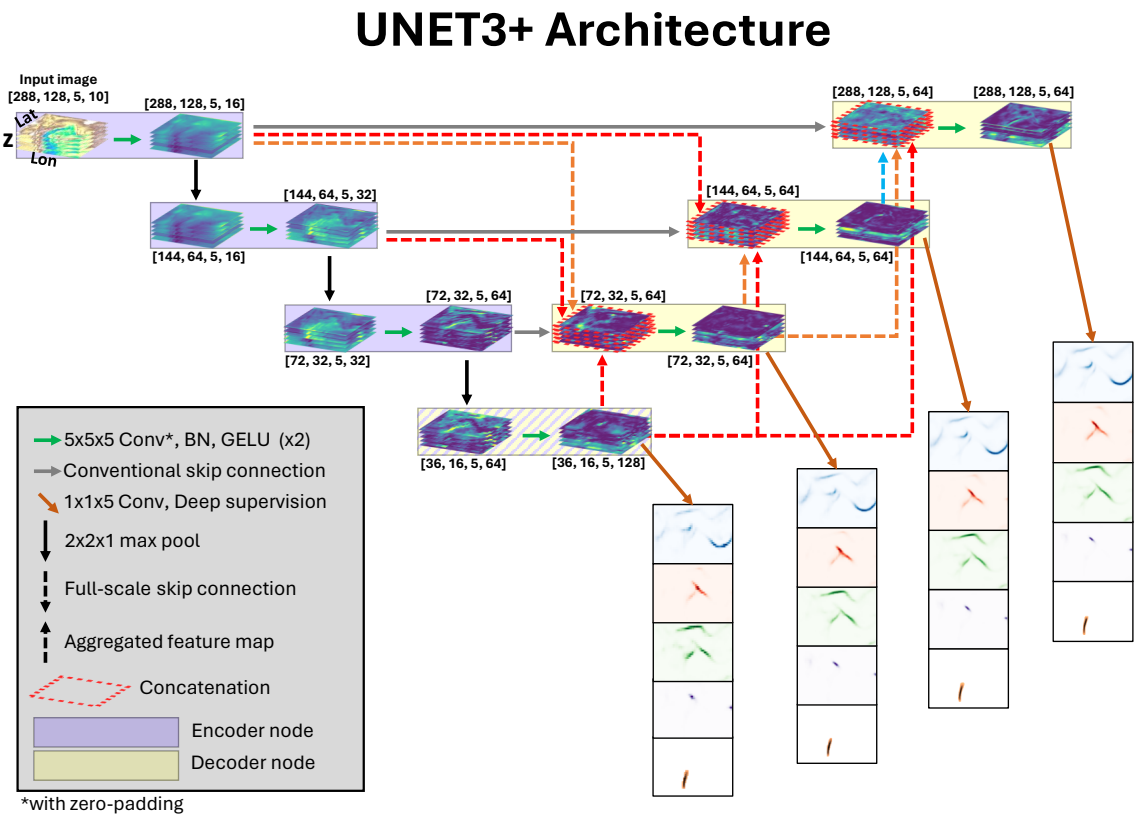


Figure 3.1: Architecture of the UNET3+ model used to predict cold, warm, stationary, and occluded fronts and drylines. This example shows an input size of 288×128×5×10, where the third (vertical) dimension is unmodified until just prior to deep supervision. Note that the bottom node is both an encoder and decoder node.

2015), and a Gaussian Error Linear Unit [GELU; Hendrycks and Gimpel (2023)] activation function layer. All 5×5×5 convolution operations throughout FCM implement zero padding, a process in which layers of zeros are added around an image so that the output of the convolution operation has the same shape as the input (O'Shea and Nash, 2015). Batch normalization layers normalize inputs such that the outputs have a mean of 0 and a variance of 1, which can provide numerical stability to the model (Ioffe and Szegedy, 2015). GELU was chosen as the activation function as it has been

shown to provide superior performance over numerous other activation functions (Lee, 2023). In our previous three-model system, we used the Rectified Linear Unit (ReLU) activation function. ReLU and GELU are defined in Equations (3.1) and (3.2) and plotted in Figure 3.2. ReLU outputs zero for all negative inputs, thus the gradient or derivative for all negative inputs is also zero:

$$\mathrm{ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3.1}$$

If the inputs into a ReLU neuron are repeatedly negative, there is no gradient, rendering the neuron "dead" as its weights cannot be updated. This is known as the dying ReLU problem (Lu, 2020) and is one of the motivations of using GELU in place of ReLU. GELU is a smoother activation function that does not block out all negative inputs and allows gradients to flow smoothly through the model:

$$\mathrm{GELU}(x) = x \cdot \frac{1}{2}[1 + \mathrm{erf}(x/\sqrt{2})] \approx x \cdot \frac{1}{2}(1 + \tanh[\sqrt{2/\pi}(x + 0.044715x^3)]). \tag{3.2}$$

Max pooling operations connect two encoder nodes. As described in Chapter 1, max pooling is a downsampling method that reduces the dimensions of an image using a pre-defined pool size, allowing UNET architectures to convert higher-resolution data into broader features that can be used by the model to further improve its performance. Since the FCM ingests 3D spatial images, the max pooling operations must be performed with 3D pool sizes. However, the vertical dimension of our input images only has a size of 5, so shrinking the size of the vertical dimension with max pooling in our architecture is not feasible. We preserve the size of the vertical dimension while still reducing horizontal dimensions by using a pool size of 2×2×1 in our max pooling operations. We connect encoder and decoder nodes on the same level of the FCM with

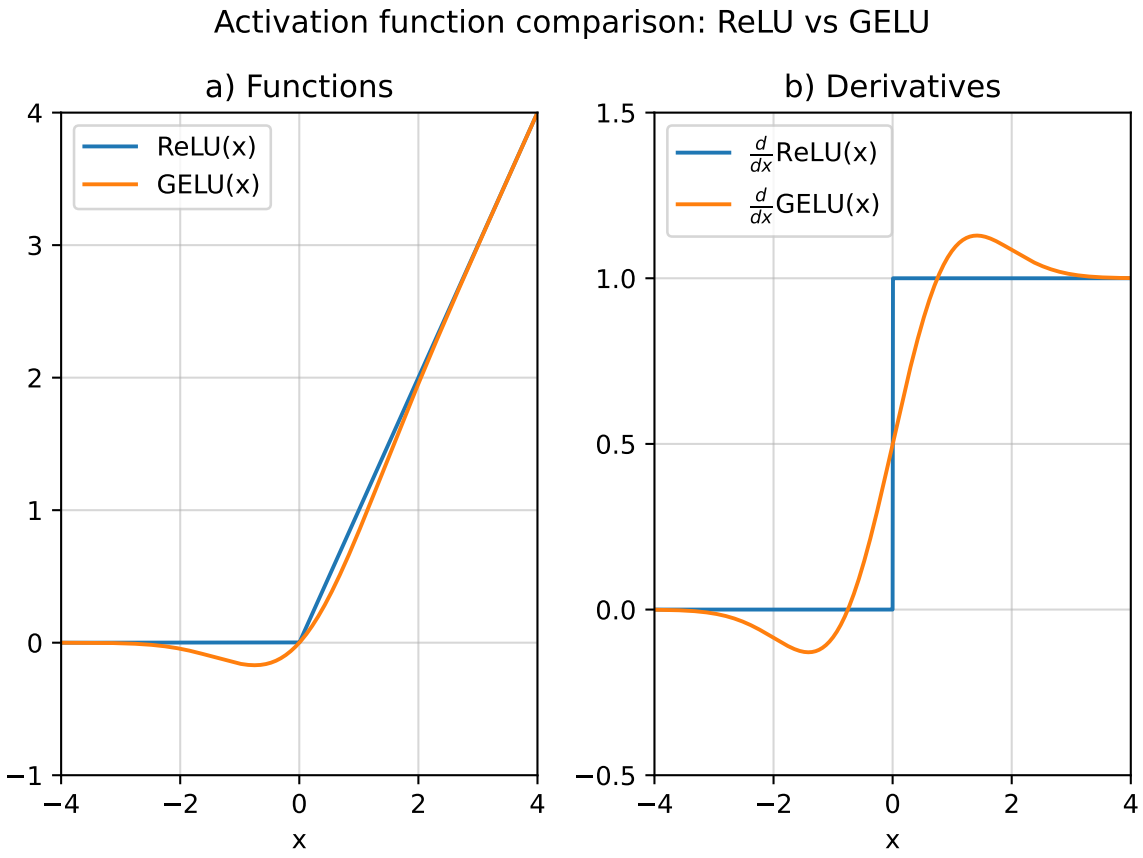


Figure 3.2: a) ReLU and GELU activation functions, and b) derivatives of the ReLU and GELU activation functions.

a conventional skip connection, which takes the output of an encoder node and passes it through one convolution module before connecting it to the decoder node. Full-scale skip connections transport data from a high resolution encoder node to a lower resolution decoder node; images undergo a 3D max pooling operation before passing through a convolution module and being transported to the decoder node. The pool size in each full-scale skip connection is determined by the resolutions of the images that are processed by each of the nodes attached to the connection. For example, if a full-scale skip connection sends images with dimensions 288×128×5 to a decoder node that processes images with dimensions 72×32×5, then the pool size must be 4×4×1. On the

decoder/upsampling side of the UNET3+, nodes are connected by connections containing aggregated feature maps. Aggregated feature maps are low resolution images that have undergone 3D upsampling operations and passed through one convolution module before reaching the target decoder node.

The decoder nodes create concatenated features from all incoming connections before passing the features through a convolution module and performing deep supervision. In the FCM, deep supervision in the decoder nodes (and the bottom encoder/decoder node) starts with features undergoing 5×5×5 convolutions with zero-padding, followed by 1×1×5 convolutions without zero-padding. The absence of zero-padding means that the 1×1×5 convolutions shrink the vertical dimension from size 5 to size 1. We then "squeeze out" the vertical dimension since it has a size of 1, resulting in 2D feature maps with longitude and latitude dimensions. With the exception of the final decoder node, the 2D feature maps are upsampled with pool sizes that result in shapes matching the longitude and latitude dimensions of the original input image of the FCM. After upsampling, the features go through a Softmax function (Bridle, 1989) and turn into probabilities for all the frontal types in the FCM.

3.2 Datasets and Preprocessing

Datasets were made with predictor variables sourced from ERA5 data (Hersbach et al., 2018) and front positions as analyzed by NOAA forecasters at 3-hour intervals (NOAA, 2023) for the period 2008-2020. ERA5 data was chosen given its high spatial resolution (0.25°, 25 km) and global coverage with data at numerous pressure levels available every 3 hours. ERA5 reanalysis has a fixed model state, so the data assimilation process is consistent with time. While the reanalysis data is quality-controlled, it may have spatial biases in boundary positions and be unable to capture some extreme

Variable	Heights/Levels
Dewpoint temperature	2-m AGL, **1000-, 950-, 900-, 850 hPa**
Geopotential height	1000-, 950-, 900-, 850 hPa
Mixing ratio	**2-m AGL, 1000-, 950-, 900-, 850 hPa**
Relative humidity	**10-m AGL, 1000-, 950-, 900-, 850 hPa**
Specific humidity	**2-m AGL**, 1000-, 950-, 900-, 850 hPa
Surface pressure	surface
Temperature	2-m AGL, 1000-, 950-, 900-, 850 hPa
Theta-E	**2-m AGL, 1000-, 950-, 900-, 850 hPa**
U-wind	10-m AGL, 1000-, 950-, 900-, 850 hPa
V-wind	10-m AGL, 1000-, 950-, 900-, 850 hPa
Virtual temperature	**2-m AGL, 1000-, 950-, 900-, 850 hPa**

Table 3.1: Predictor variables used as input to FCM. Variables or levels in **boldface** were derived from variables directly retrieved from ERA5, which are not in boldface.

mesoscale thermodynamic gradients, like those seen with some drylines (Pietrycha and Rasmussen, 2004; Hoch and Markowski, 2005). Ten predictor variables at five vertical levels were used as input (Table 3.1), while the labels were front positions interpolated every one kilometer and transformed to a uniform 0.25° grid to match the ERA5 resolution. This differs from our previous three-model system as wet-bulb temperature and wet-bulb potential temperature (θ_w) are not included in the list of predictors since our permutation studies in J23 suggested that these variables have little influence on the model predictions, and L19 found that front-detecting CNNs with

Dataset (Years)	Image pairs	CF	WF	SF	OF	DL
Training (2008-2017, 2020)	34,766	1.751%	0.523%	1.491%	0.222%	0.0624%
Validation (2018)	3,232	1.954%	0.608%	1.81%	0.193%	0.075%
Testing - CONUS (2019)	2,920	2.027%	0.65%	1.67%	0.238%	0.0478%
Testing - USAD (2019)	1,460	1.009%	0.377%	0.772%	0.291%	0.0056%

Table 3.2: Fraction of pixels containing each front type in the training, validation, and testing datasets. The extents of the CONUS and USAD domains can be found in Figure 3.3. CF = cold front, WF = warm front, SF = stationary front, OF = occluded front, DL = dryline.

θ_w included as input performed worse than those not including θ_w. The front labels were also expanded by one pixel (0.25° or roughly 25 km) in all directions in order to account for positional biases that may exist between the reanalysis and observed boundaries. The data were split into three datasets - training, validation, and testing. The training dataset encompassed years 2008-2017 and 2020, while the validation and testing datasets contained data for 2018 and 2019 respectively. The discontinuity in the years used for the training set is intentional; 2020 had the highest sample of drylines throughout our entire dataset, so it was included in the training data to maximize dryline sample sizes. The training and validation datasets only cover our CONUS domain. Note that the testing dataset for USAD only contains 6-hourly timesteps (0z, 6z, 12z, 18z) as full analyses over USAD are only performed every six hours. The extents of the CONUS and USAD domains are shown in Figure 3.3.

Timesteps used in the training and validation datasets were selected with a two-step process. First, each 3-hourly timestep in the years within each dataset is checked for the presence of at least one of each front type over the CONUS domain. If at least one

of each type is present, the timestep is retained in the dataset. Otherwise, a 50 percent chance determines whether or not the timestep is retained. Filtering is necessary to compensate for the large variation in sample sizes between the different frontal types. Nine pairs of images evenly spaced along the longitude are extracted from each timestep; the shape of the inputs was 128×128×5×10 (longitude × latitude × vertical level × variable), while the labels containing fronts have shape 128×128×6 (longitude × latitude × front type, one of the frontal types is labeled "no front), with each horizontal dimension of the pairs of images having a 25 percent chance of being flipped. In other words, 43.75 percent of all images had at least one horizontal dimension flipped. This was done to increase the number of images in the datasets while maintaining sample size benefits from the timestep selection process, as well as to prevent prevalent horizontal structures from having too great of an influence on FCM's predictions. Rotating the images prevents the model from learning spatial correlations that exist with some frontal types (e.g., the north-south orientation of drylines in the Great Plains with dry air to the west and moist air to the east of the boundary). The process of generating modified copies of images to create a larger dataset is known as data augmentation, which has been shown to improve the performance of deep learning models applied to object recognition tasks (Yang et al., 2023). Note that the data were not augmented based on correlations between the input variables, which could have implications for results stemming from our permutation studies. For example, as moisture content increases, several variables will have their values increased (e.g., relative humidity, specific humidity, mixing ratio, dewpoint temperature, etc.). Correlations between the variables may not assign importance based on their true predictive abilities (e.g., splitting importance up between two closely linked variables, like specific humidity and mixing ratio). Variables were normalized with min-max normalization at each vertical

level to ensure predictor variables are in a common range of [0, 1] and provide numerical stability to the model.

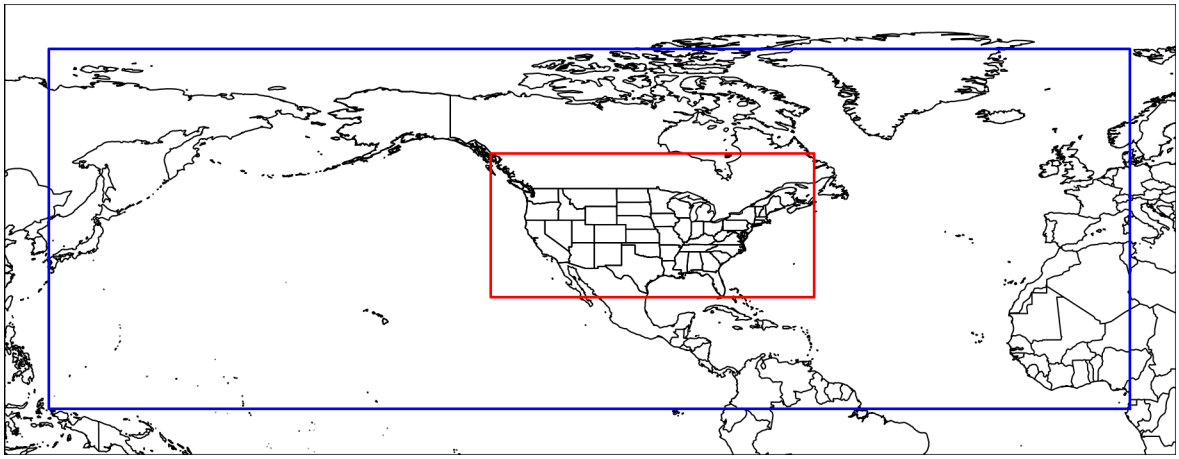


Figure 3.3: CONUS domain (red) and NOAA's USAD (blue). The bounds of the CONUS domain are 132°W, 60.25°W, 25°N, 56.75°N (288×128 pixels on the 0.25° ERA5 grid), while the USAD has bounds of 130°E, 10°E, 0°N, 80°N (960×320 pixels after 1-pixel truncation along each dimension).

3.3 Training and Hyperparameters

FCM was trained in parallel across four NVIDIA A100 GPUs (40 GB variant) over the course of $18\frac{1}{2}$ hours. The hyperparameter choices are summarized in Table 3.3.

The loss function is a key part of the training process and serves as an error metric for the model and optimizes model parameters (Terven et al., 2023). We used a simple manipulation of the fractions skill score [FSS; Roberts (2008)] as the loss function for FCM. FSS is a spatial verification metric that does not penalize predictions that are displaced from the ground truth by a prefined number of pixels, where an FSS of 1 (0) is the best (worst) possible forecast. In the case of FCM, predictions within one pixel (0.25° or roughly 25 km) of a ground truth front are considered hits. The formula for

Hyperparameter	**Value**
Batch size	64
Early stopping	55 epochs
Learning rate	10^{-4}
Loss function	Fractions Skill Score (Roberts, 2008)
Optimizer	Adam (Kingma and Ba, 2014)
Steps per epoch	Training: 10, Validation: 51

Table 3.3: Hyperparameter choices for FCM.

our loss function FSS_{loss} is shown in Equation (3.3) below, with FSS_{loss} equal to 0 (1) for the best (worst) possible forecast.

$$FSS_{loss} = 1 - FSS \tag{3.3}$$

Adam (Kingma and Ba, 2014) was used as the optimizer, with an initial learning rate of 10^{-4} and using the default exponential decay rates from Kingma and Ba (2014). The batch size for both training and validation was 64, performing 10 steps for training and 51 steps for validation in each epoch. In our case, an epoch is defined by one pass over a small subset of the data (10 passes/steps over batches of 64 images). Validation is performed every epoch, and 51 steps ensure that the model sees all images in the validation dataset every time that validation is performed.

Early stopping was used to prevent the model from overfitting to the training dataset. This method of regularization involves stopping the training process early when model error fails to decrease, the exact conditions of which are often subjective and determined prior to model training (see Shen et al. (2022) for more details). Training for FCM was stopped after the validation loss failed to improve for 55 successive

epochs. This rule was set to ensure that the model completed a full iteration over the training dataset without improvements before training was suspended. The model achieved the lowest validation loss at epoch 644, with training being terminated after epoch 699 following 55 consecutive epochs of no improvements in the validation loss.

3.4 Evaluation

To evaluate the performance of FCM over the CONUS and USAD domains, we generated model predictions with images from the respective testing sets for each domain (see Table 3.2) and calculated the critical success index (CSI) using 50, 100, 150, 200, and 250 km neighborhoods (assuming 25 km between grid spaces). Similar to FSS, CSI evaluation using a neighborhood allows for predictions slightly displaced from the ground truth to be counted as hits (e.g., see Clark et al., 2015; Justin et al., 2023; Lagerquist et al., 2019, 2020; Niebler et al., 2022). CSI is calculated at probability thresholds from 0.01 to 1 (0.01, 0.02, 0.03, ..., 1) and is defined in Equation (3.4) below:

$$CSI = \frac{TP}{TP + FP + FN} = \frac{1}{\frac{1}{POD} + \frac{1}{1-FAR} - 1}, \tag{3.4}$$

where TP, FP, and FN are the numbers of true positives, false positives, and false negatives respectively, POD is the probability of detection [Equation (3.5)], and FAR is the false alarm ratio [Equation (3.6)]:

$$POD = \frac{TP}{TP + FN}, \tag{3.5}$$

$$FAR = \frac{FP}{TP + FP}. \tag{3.6}$$

Frequency bias (FB) was also calculated in order to gauge the model's tendency to overpredict or underpredict certain types. FB is defined in Equation (3.7) below.

$$FB = \frac{TP + FP}{TP + FN} \tag{3.7}$$

When FB is greater (less) than 1, the model tends to overpredict (underpredict) events, so a value of 1 is preferred. Confidence intervals at the 95% level were calculated for bulk performance statistics using bootstrapping, iterating over the statistics 1,000 times and retaining statistics for 2,920 (1,460) timesteps each iteration when bootstrapping over the CONUS (USAD) domain.

3.4.1 Permutation Studies

Variable and pressure level importance was determined through single-pass permutations (Breiman, 2001; McGovern et al., 2019). Single-pass permutations involve shuffling values of a particular predictor to see how model performance changes. In our case, we shuffled each variable and/or pressure level in all images within the testing datasets and reevaluated the model across the new datasets containing the shuffled data. 65 permutations were performed for each front type: 50 for single variables at individual pressure levels, 10 for single variables over all pressure levels, and five for all variables at individual pressure levels. The change in POD was used as a metric for importance. The decision to use POD over CSI for importance was made because fronts with exceptionally small sample sizes (e.g., drylines) can see the CSI actually increase during permutation studies due to model probabilities decreasing across the pixels without any targets (in the case of drylines, >99.95% of pixels in the CONUS testing set are empty). This can cause a misleading result whereby parameter importance is understated due to a large drop in the number of false positives. Using POD

prevents this issue and only focuses on pixels containing the frontal types of interest. A decrease in POD indicates that a predictor is important and vice versa for an increase in POD. The change in POD was pixel-based and used the dilated front labels as the targets without any neighborhood approaches like those used in our CSI calculations.

3.4.2 Saliency Maps

Since AI models are effectively black boxes whose predictions cannot be easily explained (McGovern et al., 2019), a method for seeing what parts of an input image a model is focusing on can help better understand how the model makes its predictions. One useful XAI method for understanding model behavior is the generation of saliency maps (Simonyan et al., 2014), which are maps that represent the sensitivity of some unit or neuron in a model with respect to the model's inputs. In our case, we wanted to understand how FCM predictions for the five frontal types could change when input variables were perturbed from their original values. It should be noted that saliency or sensitivity is not necessarily analogous to importance (i.e., high sensitivity $\neq$ high importance) but can still give insight as to how the ambient environment could be modified to improve model predictions. Saliency map generation started with performing one back-propagation pass through FCM, computing a derivative for each variable and front type in each pixel on the 3D input grid (10 variables * 5 frontal types = 50 individual derivatives at each grid point). The maximum derivative across the 10 variables was retained such that each pixel had one saliency value for each front type. These saliency values at each pixel comprise the final saliency maps for each front type at each vertical level. We did not calculate bulk statistics for saliency maps across all timesteps, however we were still able to draw some conclusions from the case study highlighted in Section 3.2.

Chapter 4

Results and Discussion

CONUS				
Front Type	**CSI**	**POD %**	**FAR %**	**FB**
Cold	0.507, 0.627, 0.682	70.9, 80.2, 86.3	36.1, 25.9, 23.6	1.110, 1.083, 1.130
Warm	0.311, 0.426, 0.505	54.8, 67.3, 78.1	58.2, 46.4, 41.2	1.310, 1.255, 1.329
Stationary	0.361, 0.458, 0.510	61.7, 74.0, 82.3	53.5, 45.4, 42.7	1.327, 1.354, 1.435
Occluded	0.386, 0.491, 0.541	61.2, 71.3, 76.2	48.9, 38.7, 34.9	1.199, 1.164, 1.170
Dryline	0.487, 0.624, 0.662	69.1, 78.6, 82.7	37.7, 24.8, 23.1	1.110, 1.046, 1.075
USAD				
Front Type	**CSI**	**POD %**	**FAR %**	**FB**
Cold	0.466, 0.586, 0.645	66.8, 75.9, 80.9	39.3, 28.0, 23.9	1.102, 1.055, 1.062
Warm	0.313, 0.417, 0.480	54.0, 66.4, 75.1	57.3, 47.2, 42.9	1.264, 1.258, 1.315
Stationary	0.263, 0.339, 0.385	50.3, 61.8, 69.7	64.5, 57.1, 53.8	1.417, 1.440, 1.509
Occluded	0.353, 0.469, 0.540	56.0, 68.4, 74.6	51.2, 40.1, 33.8	1.148, 1.141, 1.128
Dryline	0.349, 0.457, 0.525	60.0, 69.9, 78.6	54.5, 43.0, 38.7	1.318, 1.226, 1.282

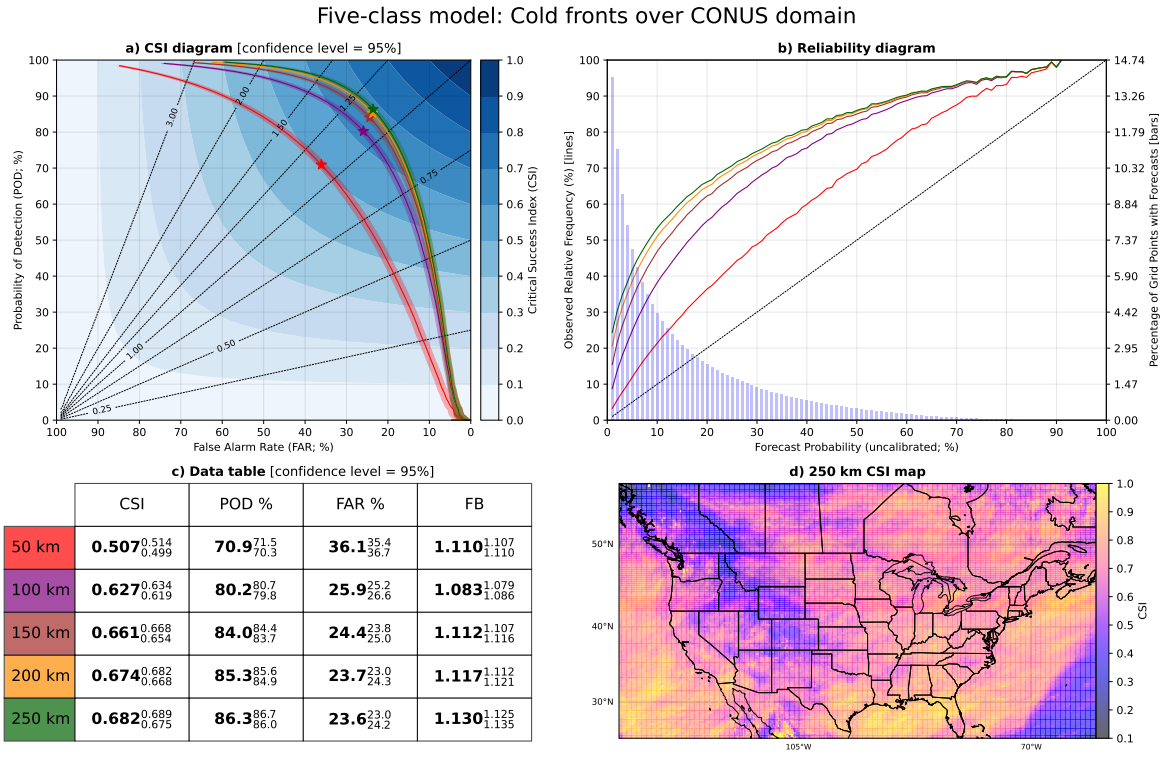
Table 4.1: FCM performance over CONUS (top) and USAD (bottom). The three values in each cell represent scores using 50, 100, and 250 km neighborhoods.

The FCM performance and permutation studies for each of the five frontal types were evaluated with the testing dataset (2019) over CONUS and USAD (Table 4.1

and figs. 4.1 to 4.4, 4.6 to 4.11 and 4.13 to 4.22). The reliability diagrams (panel (b) in the figures previously mentioned) show that the FCM experienced an underforecast bias with all frontal types over both domains, with the exception of stationary fronts over the full domain. Such an underforecast bias did not show up in our previous TMS, however we think the FCM has this bias because of the fact that it predicts five frontal types, and it can be difficult to get the model to be confident in its predictions (i.e., generate high probabilities) without copious amounts of high-quality training data. Reliability will worsen with larger neighborhoods as more true positives (hits) are introduced, which increases the observed relative frequency of the fronts.

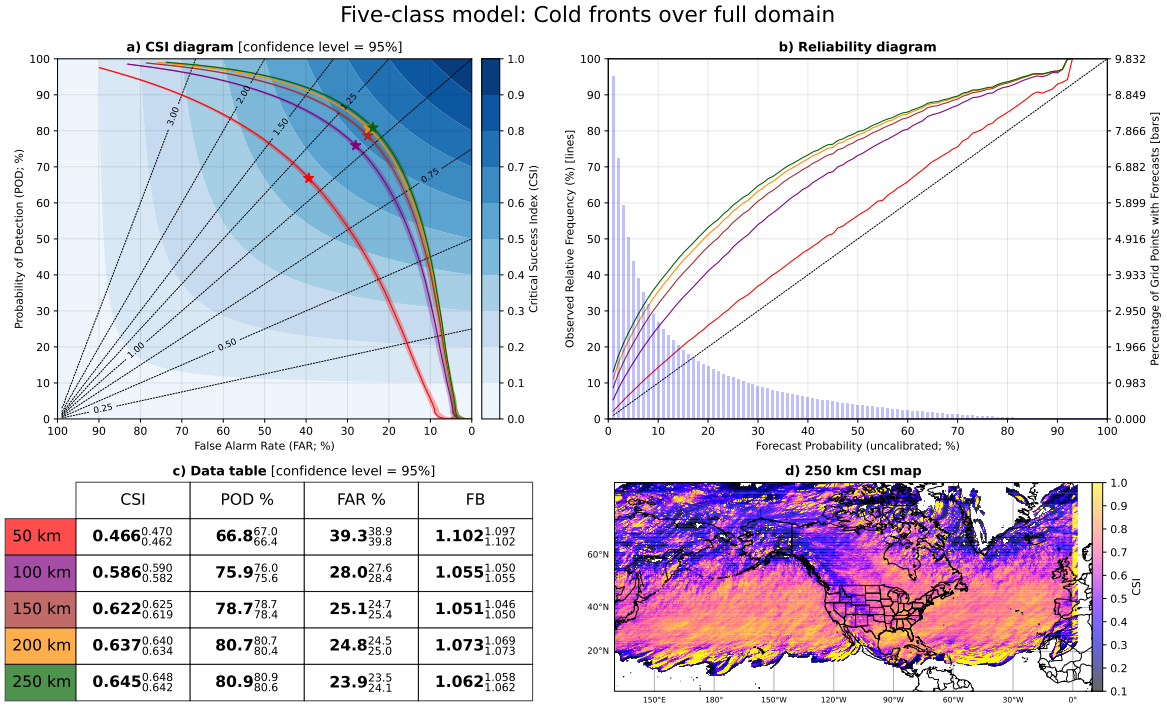
Over CONUS, cold fronts achieved the best performance of all frontal types with a 250 km CSI score of 0.682, successfully hitting 86.3% of pixels containing cold fronts with a 23.6% FAR, outperforming our three-model system from Justin et al. (2023) that achieved a CSI of 0.55 for cold fronts over CONUS. The cold front results from FCM are consistent with findings from J23 and Niebler et al. (2022): both studies found that cold fronts were the best performing boundary type over multiple domains, excluding the binary front type. Cold front performance is weaker over multiple areas, including the Rocky Mountains and areas north of 60°N (Figures 4.1 and 4.2, panel d). The number of cold fronts predicted by the model was slightly higher than the number of ground truth fronts in the test dataset, leading to FB scores between 1.05 and 1.10 over USAD. By comparison, cold front FB scores in J23 over USAD were between 1.10 and 1.27, suggesting that simplification of the model architecture reduced the tendency to overpredict cold fronts. Terrain and the lack of input data at higher altitudes likely complicate cold front detection over the Rocky Mountains. For example, mountains can complicate frontal structures by blocking and diverting flows, and cold air can sink into valleys and clash with existing air masses. The lower frequency of cold fronts at higher latitudes (see J23 and L19) and the fact that distance between longitudes

decreases with northern extent could be an explanation for lower performance over higher latitudes. In the midlatitudes, there is little difference in cold front performance between CONUS and the Atlantic and Pacific Ocean basins, which indicates that FCM is able to generalize oceanic cold fronts despite the training data only covering a very small portion of these basins. The cold front permutation studies shown in Figures 4.3 and 4.4 indicate that v-wind, temperature, and virtual temperature are the most important variables for cold front detection. This is not a surprising result since cold fronts are characterized by temperature contrasts and often wind shifts and moisture gradients (Schultz, 2005). One interesting result was that surface pressure and geopotential heights had much lower importance over USAD versus CONUS. We believe this is related to surface friction; overhanging noses of denser air ahead of surface boundaries have been documented in several studies (e.g., Simpson, 1972; Young and Johnson, 1984; Mitchell and Hovermale, 1977), with Simpson (1972) attributing the overhanging nose to a no-slip lower boundary condition. Since USAD covers much of the Pacific and Atlantic Oceans, we think that the higher importance for surface pressure and geopotential heights over CONUS is the result of overhanging noses caused by friction with land surfaces.



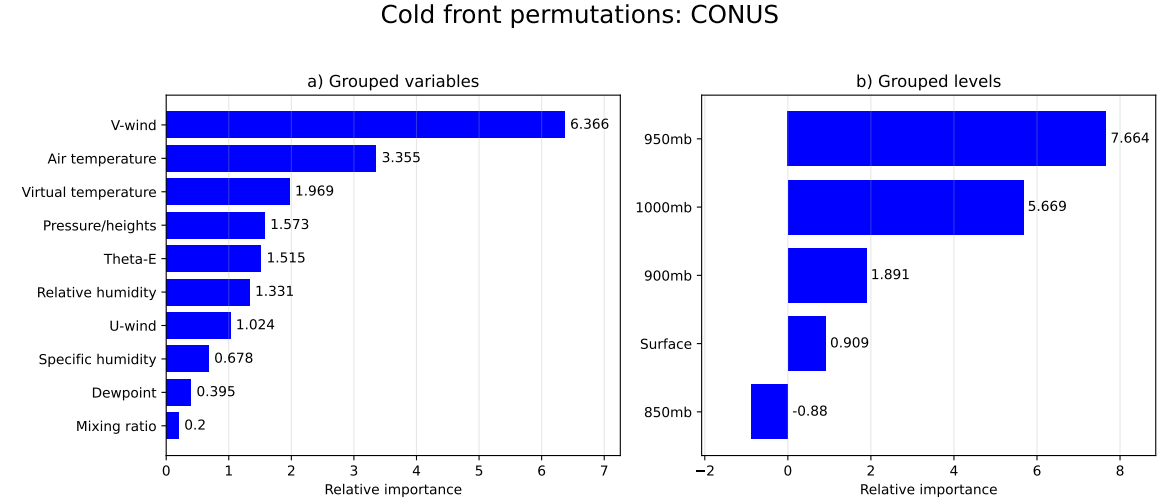
	CSI	POD %	FAR %	FB
50 km	$\mathbf{0.507}^{0.514}_{0.499}$	$\mathbf{70.9}^{71.5}_{70.3}$	$\mathbf{36.1}^{35.4}_{36.7}$	$\mathbf{1.110}^{1.107}_{1.110}$
100 km	$\mathbf{0.627}^{0.634}_{0.619}$	$\mathbf{80.2}^{80.7}_{79.8}$	$\mathbf{25.9}^{25.2}_{26.6}$	$\mathbf{1.083}^{1.079}_{1.086}$
150 km	$\mathbf{0.661}^{0.668}_{0.654}$	$\mathbf{84.0}^{84.4}_{83.7}$	$\mathbf{24.4}^{23.8}_{25.0}$	$\mathbf{1.112}^{1.107}_{1.116}$
200 km	$\mathbf{0.674}^{0.682}_{0.668}$	$\mathbf{85.3}^{85.6}_{84.9}$	$\mathbf{23.7}^{23.0}_{24.3}$	$\mathbf{1.117}^{1.112}_{1.121}$
250 km	$\mathbf{0.682}^{0.689}_{0.675}$	$\mathbf{86.3}^{86.7}_{86.0}$	$\mathbf{23.6}^{23.0}_{24.2}$	$\mathbf{1.130}^{1.125}_{1.135}$

Figure 4.1: Cold front results over CONUS; a) CSI diagram (dashed lines = frequency bias), b) reliability diagram (dashed line = perfect reliability), c) data table with upper and lower performance bounds indicated with superscripts and subscripts, and d) spatial CSI diagram using a 100 km neighborhood. CSI scores lower than 0.1 are not shown on the spatial diagram.



	CSI	POD %	FAR %	FB
50 km	$\mathbf{0.466}^{0.470}_{0.462}$	$\mathbf{66.8}^{67.0}_{66.4}$	$\mathbf{39.3}^{38.9}_{39.8}$	$\mathbf{1.102}^{1.097}_{1.102}$
100 km	$\mathbf{0.586}^{0.590}_{0.582}$	$\mathbf{75.9}^{76.0}_{75.6}$	$\mathbf{28.0}^{27.6}_{28.4}$	$\mathbf{1.055}^{1.050}_{1.055}$
150 km	$\mathbf{0.622}^{0.625}_{0.619}$	$\mathbf{78.7}^{78.7}_{78.4}$	$\mathbf{25.1}^{24.7}_{25.4}$	$\mathbf{1.051}^{1.046}_{1.050}$
200 km	$\mathbf{0.637}^{0.640}_{0.634}$	$\mathbf{80.7}^{80.7}_{80.4}$	$\mathbf{24.8}^{24.5}_{25.0}$	$\mathbf{1.073}^{1.069}_{1.073}$
250 km	$\mathbf{0.645}^{0.648}_{0.642}$	$\mathbf{80.9}^{80.9}_{80.6}$	$\mathbf{23.9}^{23.5}_{24.1}$	$\mathbf{1.062}^{1.058}_{1.062}$

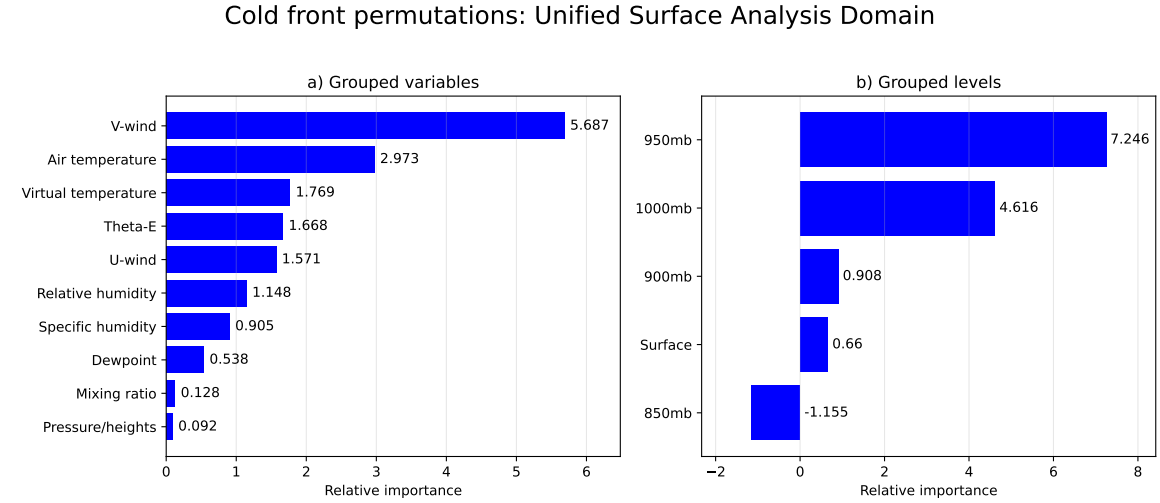
Figure 4.2: Same as Figure 4.1 but for cold fronts over USAD.



c) Variables on single levels

	Surface	1000mb	950mb	900mb	850mb
Air temperature	23 (0.130)	9 (0.439)	4 (0.672)	5 (0.666)	16 (0.232)
Dewpoint	38 (-0.009)	21 (0.158)	39 (-0.042)	47 (-0.211)	35 (0.025)
Virtual temperature	43 (-0.106)	25 (0.119)	12 (0.366)	8 (0.503)	13 (0.348)
U-wind	6 (0.544)	24 (0.120)	34 (0.025)	31 (0.054)	28 (0.082)
V-wind	2 (1.195)	1 (1.518)	3 (0.830)	15 (0.289)	42 (-0.094)
Mixing ratio	40 (-0.088)	41 (-0.092)	45 (-0.144)	46 (-0.181)	48 (-0.238)
Specific humidity	36 (0.020)	29 (0.078)	32 (0.044)	49 (-0.243)	50 (-0.356)
Relative humidity	20 (0.169)	18 (0.215)	19 (0.179)	37 (0.007)	27 (0.086)
Pressure/heights	33 (0.040)	22 (0.131)	7 (0.512)	10 (0.436)	30 (0.057)
Theta-E	44 (-0.113)	26 (0.103)	14 (0.336)	11 (0.388)	17 (0.219)

Figure 4.3: Cold front permutation results over the CONUS domain for a) grouped variables, b) grouped vertical levels, and c) variables on single variables ranked from 1-60 with 1 (60) being the most (least) important variable and level combination.



c) Variables on single levels

	Surface	1000mb	950mb	900mb	850mb
Air temperature	28 (0.147)	9 (0.513)	4 (0.818)	5 (0.696)	29 (0.142)
Dewpoint	34 (0.073)	16 (0.303)	33 (0.109)	44 (-0.059)	27 (0.182)
Virtual temperature	41 (-0.038)	26 (0.184)	11 (0.463)	8 (0.513)	18 (0.260)
U-wind	2 (1.406)	12 (0.453)	49 (-0.368)	48 (-0.353)	22 (0.213)
V-wind	3 (1.184)	1 (1.779)	6 (0.685)	32 (0.112)	38 (-0.010)
Mixing ratio	37 (0.009)	36 (0.013)	43 (-0.056)	46 (-0.164)	45 (-0.125)
Specific humidity	30 (0.134)	19 (0.252)	23 (0.208)	47 (-0.185)	50 (-0.370)
Relative humidity	17 (0.274)	15 (0.313)	24 (0.194)	35 (0.066)	20 (0.251)
Pressure/heights	39 (-0.017)	31 (0.122)	13 (0.428)	21 (0.226)	42 (-0.052)
Theta-E	40 (-0.020)	25 (0.185)	10 (0.494)	7 (0.578)	14 (0.325)

Figure 4.4: Same as Figure 4.3 but for cold fronts over USAD.

FCM performance for warm fronts was similar across the CONUS and USAD domains with 250 km CSI scores of 0.505 and 0.48, respectively, a significant improvement over the three-model system in J23 that achieved CSI scores of 0.36 and 0.37. FCM found 75.1% of the warm fronts within 250 km of the ground truth warm fronts as analyzed by forecasters with a FAR of 42.9% over USAD. FB scores for warm fronts over USAD ranged from 1.26 to 1.32 with FCM and 1.26 to 1.49 with the three-model system in J23, indicating a smaller fraction of warm front false positives from the new algorithm. CSI scores are slightly higher over water, in particular the Atlantic and Pacific ocean basins, than over land (Figures 4.6 and 4.7, panel d). This could be attributed to differences in the intensity of warm fronts over water versus land; Hines and Mechoso (1993) analyzed the evolution of frontal structures during numerical simulations of cyclogenesis and found that simulations with low surface drag exhibited enhanced warm frontogenesis due to the higher wind speeds from low surface drag allowing for more robust warm air advection into the warm frontal zones.

Warm fronts are rarely analyzed over the Rocky Mountains (Figure 4.5b), which implies that sample size is the likely driver of lower CSI scores over the mountainous terrain. The simplified model architecture used in the present study that includes all frontal types likely helped FCM to better differentiate between warm and stationary fronts, both of which experienced high FAR in J23. The warm front permutation studies over USAD shown in Figures 4.8 and 4.9 echoed findings of an observation of a warm front over the Northeast Atlantic by Wakimoto and Bosart (2001, hereafter W01). They found that the observed warm front was more defined aloft with a stronger equivalent potential temperature (θ_E) gradient and intense vertical wind shear. W01 also noted that the warm front at the surface had little θ_E or mixing ratio gradient. Our permutation studies showed that the 850 and 900 hPa levels had greater relative importance than the surface, 1000, and 950 hPa. The results shown in Figure 4.9c show

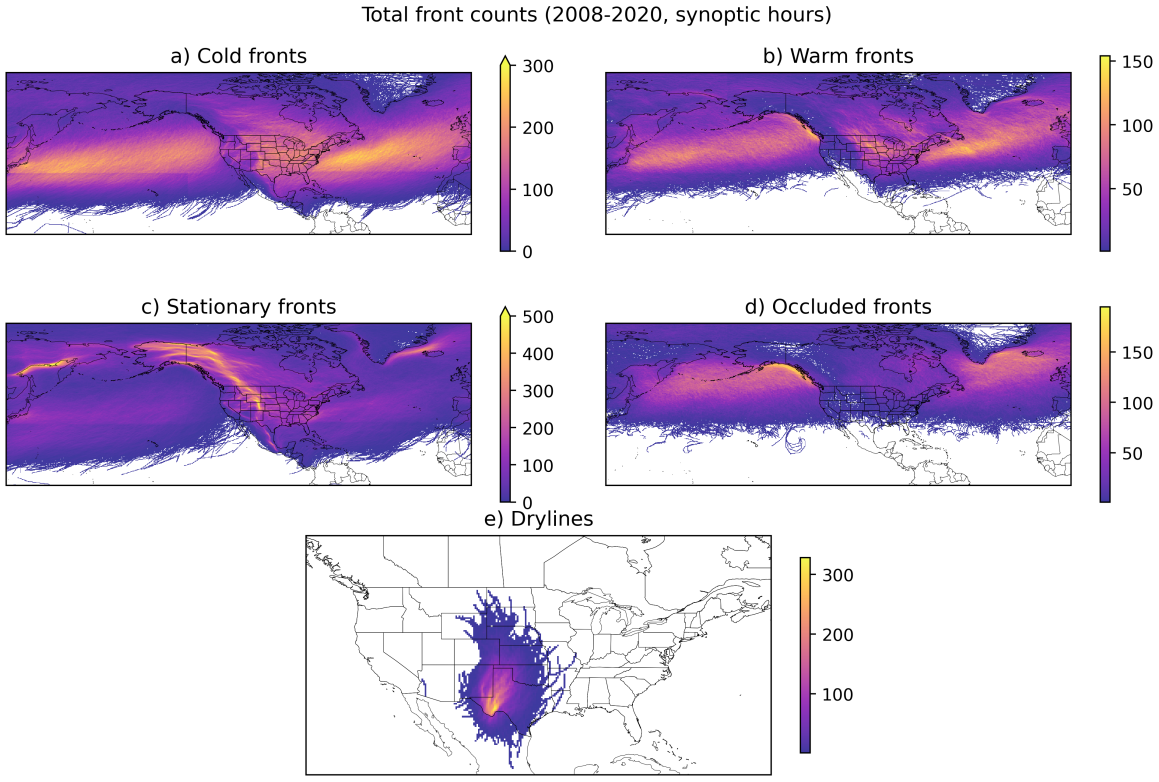
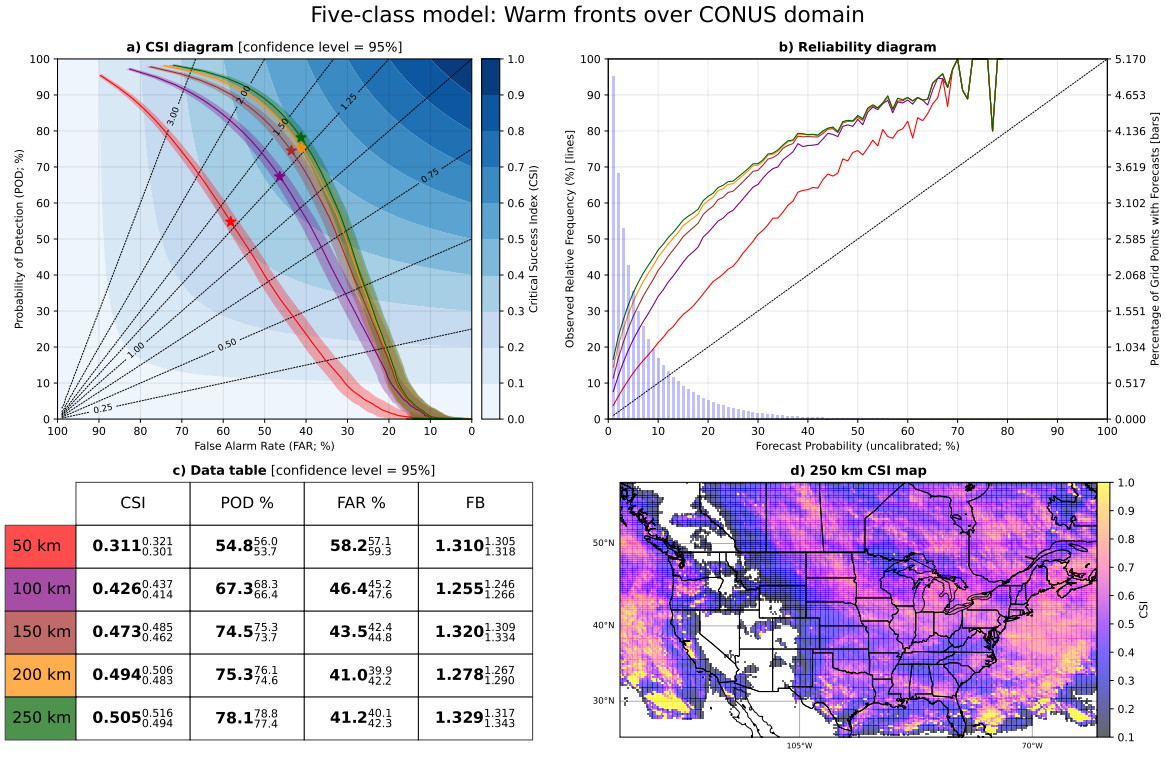


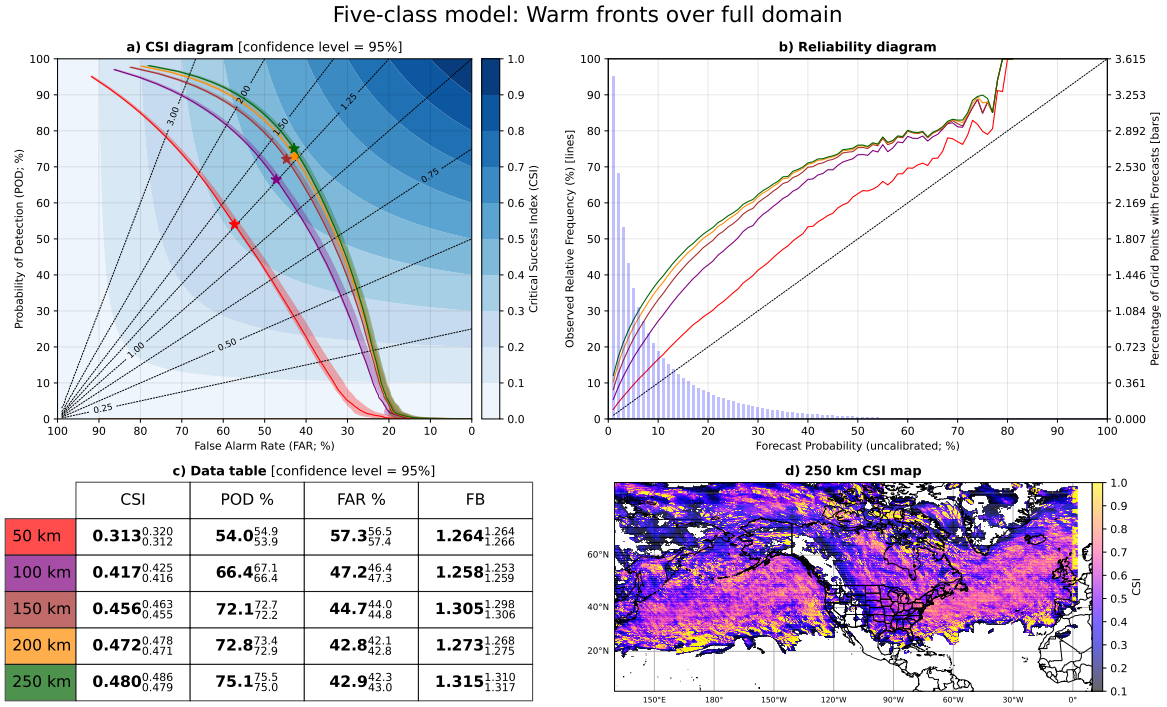
Figure 4.5: Frequency of a) cold fronts, b) warm fronts, c) stationary fronts, d) occluded fronts, and e) drylines drawn by NWS forecasters over USAD for the period 2008-2022 at synoptic hours. Non-synoptic hours are not shown as WPC only draws over North America for these timesteps and frequencies are significantly higher over the WPC domain (see Fig. 2 from Justin et al., 2023)

θ_E has the greatest importance at 950, 900, and 850 hPa, with the surface θ_E having negative relative importance (i.e., hurting warm front predictability). Figure 4.9a show that v-wind and u-wind were the most important variables in detecting warm fronts over USAD, which is supported by the observation of intense vertical wind shear along the oceanic warm front analyzed by W01. The permutation studies suggest that surface pressure is not important to detecting warm fronts, however this may be due to emphasis on geopotential heights of pressure levels above the surface.



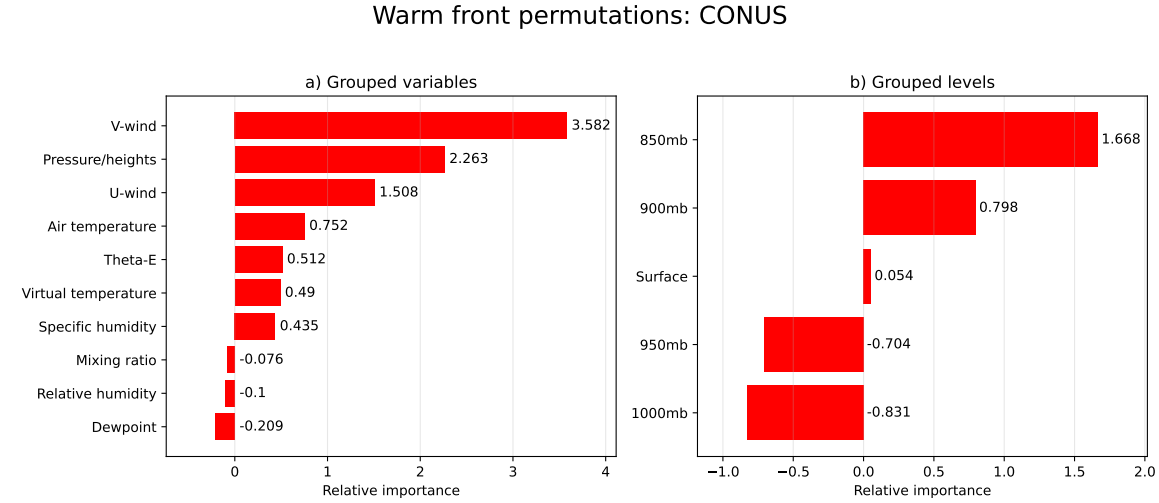
	CSI	POD %	FAR %	FB
50 km	$\mathbf{0.311}^{0.321}_{0.301}$	$\mathbf{54.8}^{56.0}_{53.7}$	$\mathbf{58.2}^{57.1}_{59.3}$	$\mathbf{1.310}^{1.305}_{1.318}$
100 km	$\mathbf{0.426}^{0.437}_{0.414}$	$\mathbf{67.3}^{68.3}_{66.4}$	$\mathbf{46.4}^{45.2}_{47.6}$	$\mathbf{1.255}^{1.246}_{1.266}$
150 km	$\mathbf{0.473}^{0.485}_{0.462}$	$\mathbf{74.5}^{75.3}_{73.7}$	$\mathbf{43.5}^{42.4}_{44.8}$	$\mathbf{1.320}^{1.309}_{1.334}$
200 km	$\mathbf{0.494}^{0.506}_{0.483}$	$\mathbf{75.3}^{76.1}_{74.6}$	$\mathbf{41.0}^{39.9}_{42.2}$	$\mathbf{1.278}^{1.267}_{1.290}$
250 km	$\mathbf{0.505}^{0.516}_{0.494}$	$\mathbf{78.1}^{78.8}_{77.4}$	$\mathbf{41.2}^{40.1}_{42.3}$	$\mathbf{1.329}^{1.317}_{1.343}$

Figure 4.6: Same as Figure 4.1 but for warm fronts over CONUS.



	CSI	POD %	FAR %	FB
50 km	$\mathbf{0.313}^{0.320}_{0.312}$	$\mathbf{54.0}^{54.9}_{53.9}$	$\mathbf{57.3}^{56.5}_{57.4}$	$\mathbf{1.264}^{1.264}_{1.266}$
100 km	$\mathbf{0.417}^{0.425}_{0.416}$	$\mathbf{66.4}^{67.1}_{66.4}$	$\mathbf{47.2}^{46.4}_{47.3}$	$\mathbf{1.258}^{1.253}_{1.259}$
150 km	$\mathbf{0.456}^{0.463}_{0.455}$	$\mathbf{72.1}^{72.7}_{72.2}$	$\mathbf{44.7}^{44.0}_{44.8}$	$\mathbf{1.305}^{1.298}_{1.306}$
200 km	$\mathbf{0.472}^{0.478}_{0.471}$	$\mathbf{72.8}^{73.4}_{72.9}$	$\mathbf{42.8}^{42.1}_{42.8}$	$\mathbf{1.273}^{1.268}_{1.275}$
250 km	$\mathbf{0.480}^{0.486}_{0.479}$	$\mathbf{75.1}^{75.5}_{75.0}$	$\mathbf{42.9}^{42.3}_{43.0}$	$\mathbf{1.315}^{1.310}_{1.317}$

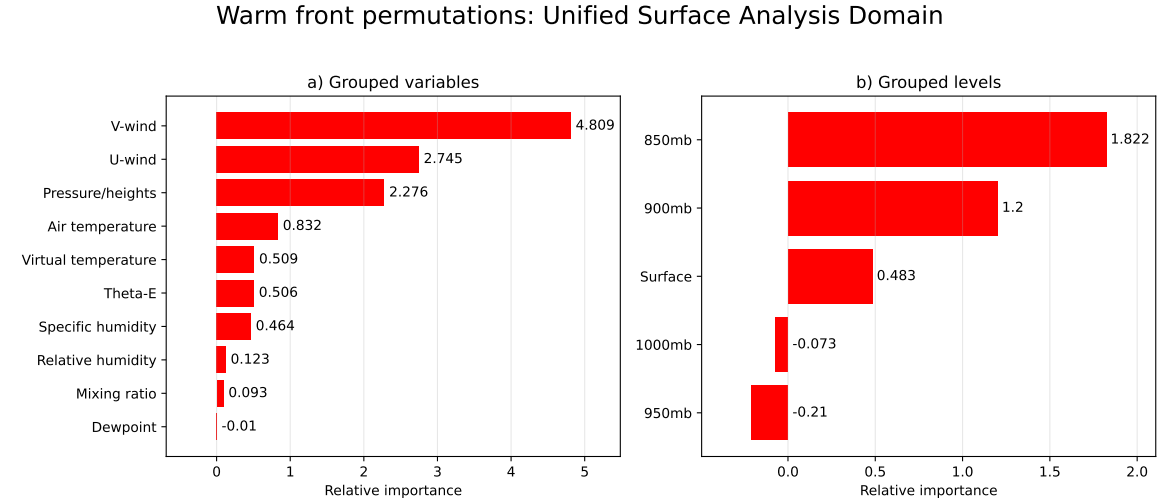
Figure 4.7: Same as Figure 4.2 but for warm fronts over USAD.



c) Variables on single levels

	Surface	1000mb	950mb	900mb	850mb
Air temperature	12 (0.106)	11 (0.154)	20 (0.015)	41 (-0.156)	35 (-0.080)
Dewpoint	29 (-0.056)	47 (-0.235)	50 (-0.287)	31 (-0.068)	28 (-0.054)
Virtual temperature	26 (-0.041)	18 (0.028)	21 (0.006)	36 (-0.084)	37 (-0.089)
U-wind	8 (0.299)	10 (0.161)	16 (0.032)	13 (0.084)	30 (-0.057)
V-wind	5 (0.435)	2 (0.566)	4 (0.480)	1 (0.601)	6 (0.380)
Mixing ratio	42 (-0.165)	32 (-0.071)	39 (-0.138)	49 (-0.267)	45 (-0.207)
Specific humidity	34 (-0.077)	17 (0.029)	27 (-0.045)	23 (-0.034)	14 (0.081)
Relative humidity	33 (-0.071)	40 (-0.140)	48 (-0.242)	43 (-0.190)	22 (-0.009)
Pressure/heights	46 (-0.211)	7 (0.309)	3 (0.516)	9 (0.216)	24 (-0.038)
Theta-E	25 (-0.041)	15 (0.040)	19 (0.023)	38 (-0.121)	44 (-0.199)

Figure 4.8: Same as Figure 4.3 but for warm fronts over CONUS.



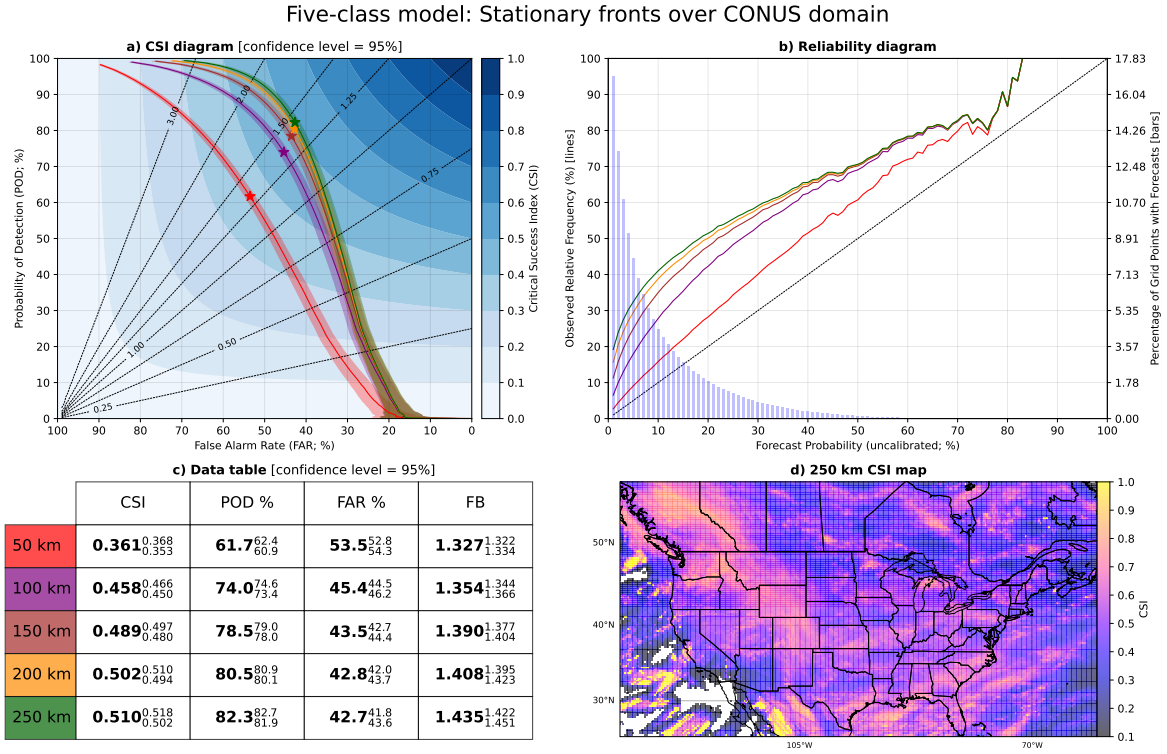
c) Variables on single levels

	Surface	1000mb	950mb	900mb	850mb
Air temperature	43 (-0.044)	26 (0.099)	12 (0.267)	15 (0.178)	23 (0.106)
Dewpoint	35 (0.029)	47 (-0.086)	48 (-0.099)	27 (0.086)	21 (0.115)
Virtual temperature	42 (-0.037)	37 (0.009)	18 (0.162)	16 (0.166)	22 (0.107)
U-wind	8 (0.380)	6 (0.457)	9 (0.329)	13 (0.247)	41 (-0.032)
V-wind	3 (0.635)	1 (1.047)	2 (0.665)	4 (0.621)	7 (0.436)
Mixing ratio	39 (-0.015)	29 (0.061)	30 (0.060)	50 (-0.132)	49 (-0.126)
Specific humidity	32 (0.049)	14 (0.214)	24 (0.101)	38 (-0.010)	33 (0.041)
Relative humidity	25 (0.099)	34 (0.033)	45 (-0.070)	40 (-0.028)	19 (0.157)
Pressure/heights	46 (-0.081)	10 (0.323)	5 (0.497)	11 (0.274)	31 (0.054)
Theta-E	44 (-0.050)	36 (0.019)	17 (0.164)	20 (0.154)	28 (0.070)

Figure 4.9: Same as Figure 4.4 but for warm fronts over USAD.

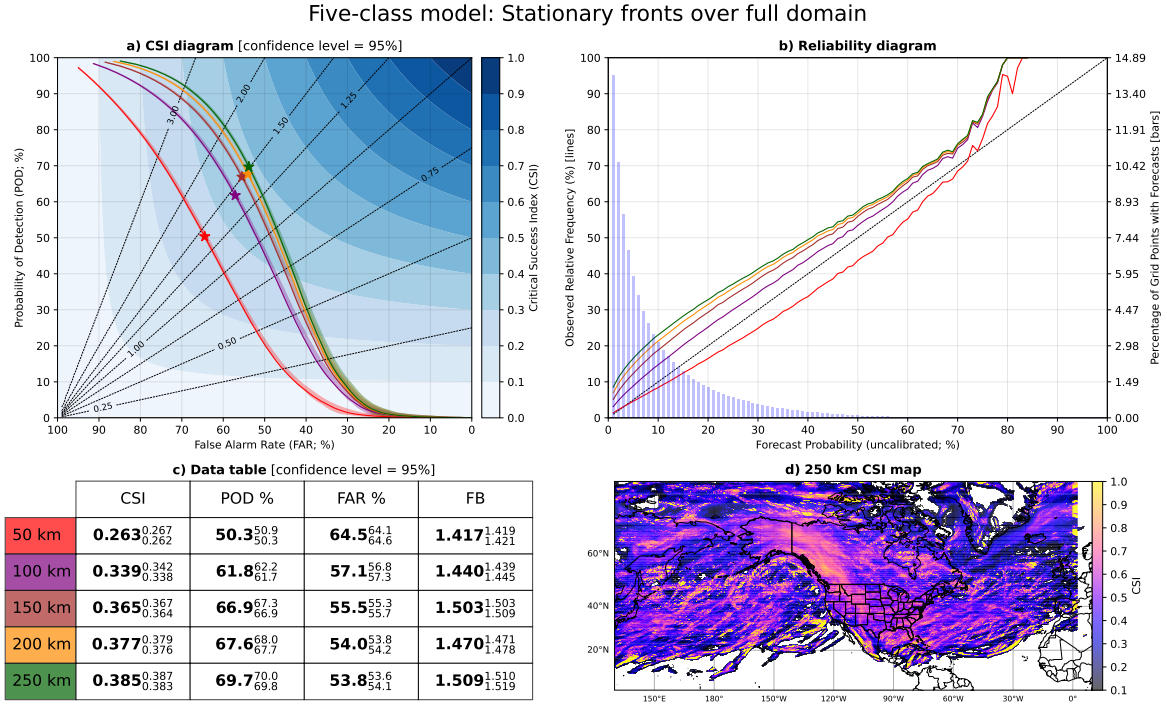
Stationary front performance is among FCM's weaknesses, with the model only achieving CSI scores of 0.51 and 0.385 over CONUS and USAD respectively using a 250 km neighborhood (Table 4.1). However, these scores are a significant improvement over J23, where CSI scores were 0.44 and 0.27. A local maximum in CSI can be noted along the Rocky Mountains in Alberta and British Columbia (Figures 4.10 and 4.11, panel d), coincident with a higher frequency of stationary fronts (Figure 4.5c). While stationary front performance greatly improved over J23, stationary fronts still have the highest FAR of any front type in FCM. We discovered that FCM tends to falsely identify parts of the Intertropical Convergence Zone (ITCZ) as a stationary front, likely due to convergence along the zone (see Figure 4.12) The ITCZ is not included in the training data, so the model is not properly trained to deal with the persistent convergence along the ITCZ. Stationary fronts seem to be one of the more subjective parts of surface analysis and the transition of a cold or warm front to a stationary front is not always definitive. This implies a potential issue with other quasi-permanent boundary features and the approach to training on individual timesteps, both of which could be sources of the high FAR for stationary fronts. The model only ingests data for one timestep, so the model has no information on previous positions of fronts and is unable to base predictions on the movement of the fronts. The stationary front permutation studies Figures 4.13 and 4.14 indicate that air temperature is the most important variable for stationary front detection. V-wind has a strong negative relationship with FCM's stationary front performance, however the reason for this is not clear. U-wind has stronger importance over the CONUS domain compared to USAD, which we suspect to be linked to the predominantly North-South orientation of stationary fronts frequently observed along and over the Rocky Mountains (Figure 4.5c). We also expected greater importance over CONUS for pressure levels at greater elevation due

to the elevated and complex terrain over the Rocky Mountains, however only the 950 hPa level had notably higher importance over CONUS.



	CSI	POD %	FAR %	FB
50 km	$\mathbf{0.361}^{0.368}_{0.353}$	$\mathbf{61.7}^{62.4}_{60.9}$	$\mathbf{53.5}^{52.8}_{54.3}$	$\mathbf{1.327}^{1.322}_{1.334}$
100 km	$\mathbf{0.458}^{0.466}_{0.450}$	$\mathbf{74.0}^{74.6}_{73.4}$	$\mathbf{45.4}^{44.5}_{46.2}$	$\mathbf{1.354}^{1.344}_{1.366}$
150 km	$\mathbf{0.489}^{0.497}_{0.480}$	$\mathbf{78.5}^{79.0}_{78.0}$	$\mathbf{43.5}^{42.7}_{44.4}$	$\mathbf{1.390}^{1.377}_{1.404}$
200 km	$\mathbf{0.502}^{0.510}_{0.494}$	$\mathbf{80.5}^{80.9}_{80.1}$	$\mathbf{42.8}^{42.0}_{43.7}$	$\mathbf{1.408}^{1.395}_{1.423}$
250 km	$\mathbf{0.510}^{0.518}_{0.502}$	$\mathbf{82.3}^{82.7}_{81.9}$	$\mathbf{42.7}^{41.8}_{43.6}$	$\mathbf{1.435}^{1.422}_{1.451}$

Figure 4.10: Same as Figure 4.1 but for stationary fronts over CONUS.



	CSI	POD %	FAR %	FB
50 km	$0.263^{0.267}_{0.262}$	$50.3^{50.9}_{50.3}$	$64.5^{64.1}_{64.6}$	$1.417^{1.419}_{1.421}$
100 km	$0.339^{0.342}_{0.338}$	$61.8^{62.2}_{61.7}$	$57.1^{56.8}_{57.3}$	$1.440^{1.439}_{1.445}$
150 km	$0.365^{0.367}_{0.364}$	$66.9^{67.3}_{66.9}$	$55.5^{55.3}_{55.7}$	$1.503^{1.503}_{1.509}$
200 km	$0.377^{0.379}_{0.376}$	$67.6^{68.0}_{67.7}$	$54.0^{53.8}_{54.2}$	$1.470^{1.471}_{1.478}$
250 km	$0.385^{0.387}_{0.383}$	$69.7^{70.0}_{69.8}$	$53.8^{53.6}_{54.1}$	$1.509^{1.510}_{1.519}$

Figure 4.11: Same as Figure 4.2 but for stationary fronts over USAD.

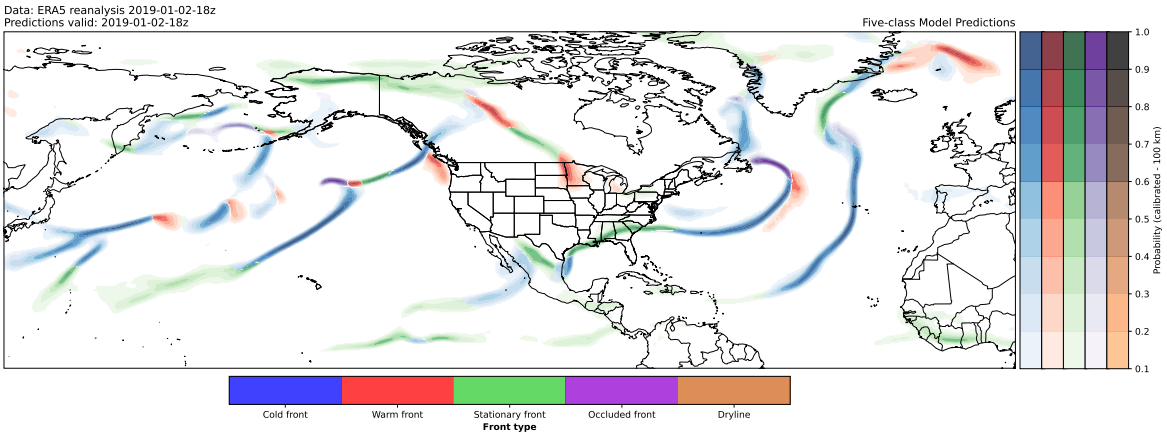
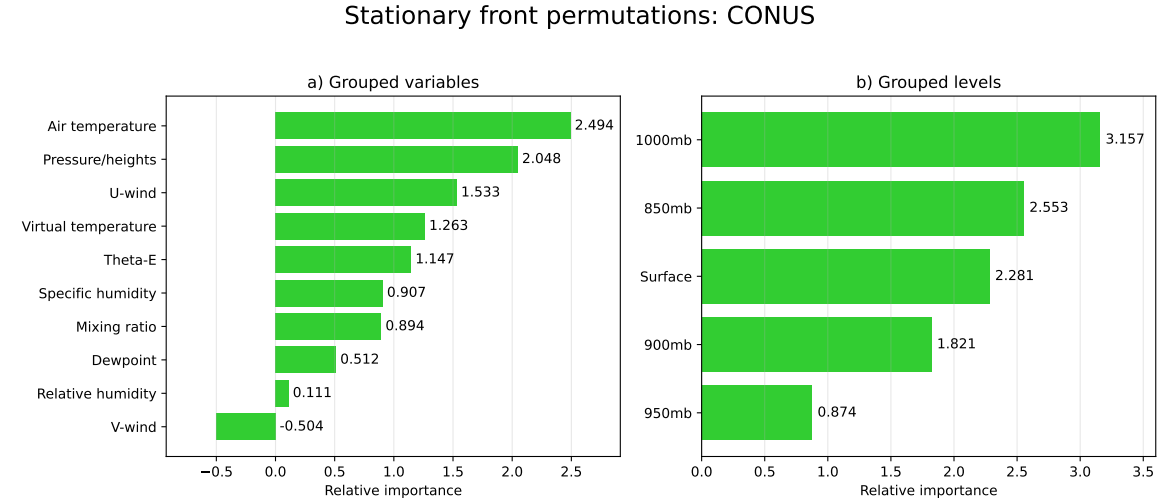


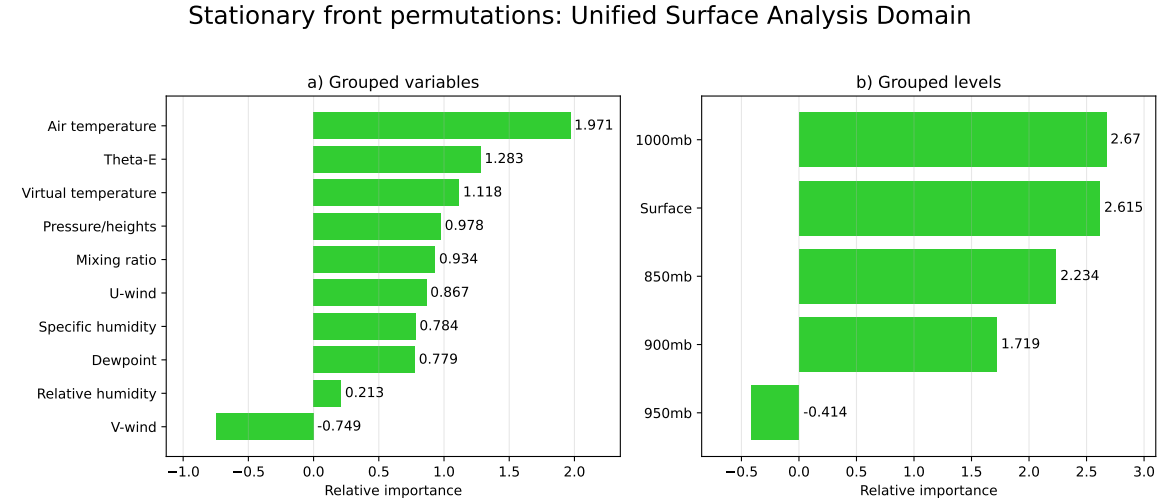
Figure 4.12: FCM predictions for 1800 UTC Jan 2 2019 over USAD.



c) Variables on single levels

	Surface	1000mb	950mb	900mb	850mb
Air temperature	11 (0.194)	2 (0.661)	3 (0.591)	4 (0.335)	16 (0.167)
Dewpoint	28 (0.059)	34 (-0.001)	30 (0.021)	23 (0.102)	27 (0.064)
Virtual temperature	29 (0.046)	6 (0.294)	5 (0.316)	8 (0.272)	17 (0.156)
U-wind	15 (0.177)	12 (0.188)	26 (0.089)	44 (-0.116)	43 (-0.105)
V-wind	46 (-0.224)	48 (-0.360)	50 (-0.469)	49 (-0.436)	47 (-0.250)
Mixing ratio	33 (0.008)	18 (0.153)	13 (0.184)	20 (0.126)	21 (0.116)
Specific humidity	19 (0.139)	9 (0.256)	25 (0.090)	31 (0.021)	38 (-0.072)
Relative humidity	40 (-0.078)	37 (-0.065)	39 (-0.077)	41 (-0.090)	36 (-0.025)
Pressure/heights	1 (0.839)	22 (0.108)	42 (-0.097)	45 (-0.129)	35 (-0.014)
Theta-E	32 (0.017)	10 (0.209)	7 (0.283)	14 (0.178)	24 (0.095)

Figure 4.13: Same as Figure 4.3 but for stationary fronts over CONUS.



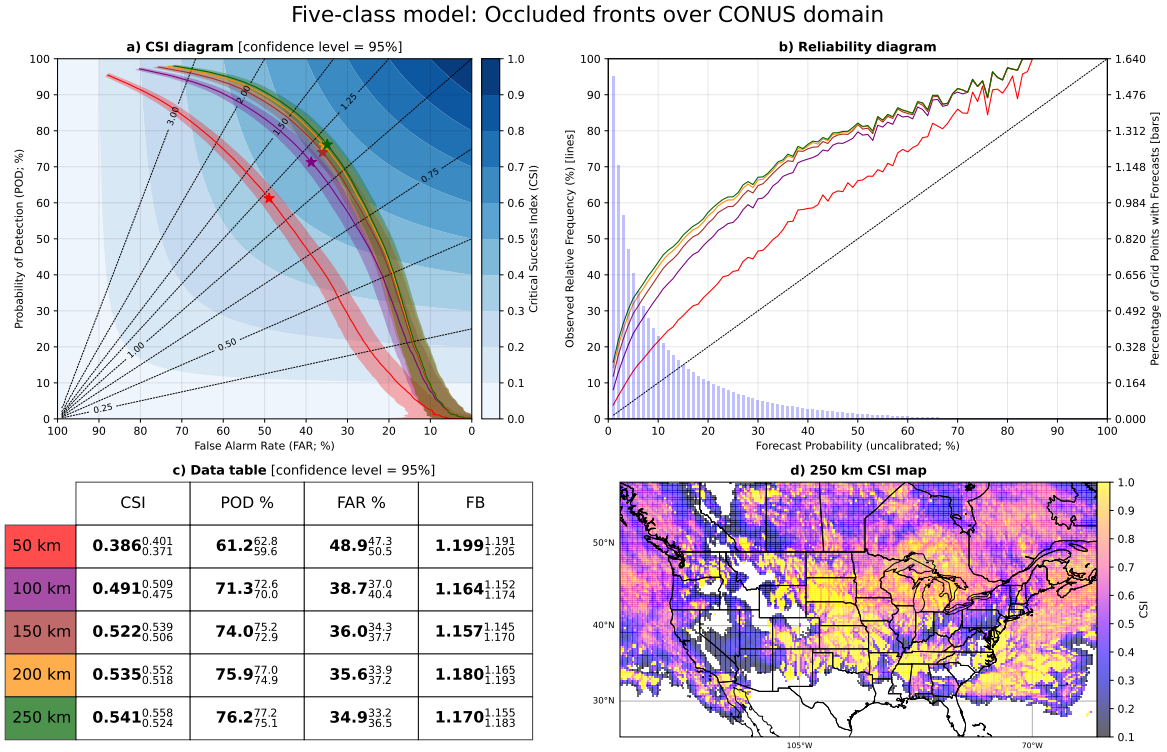
c) Variables on single levels

	Surface	1000mb	950mb	900mb	850mb
Air temperature	12 (0.297)	1 (0.736)	2 (0.517)	16 (0.222)	22 (0.166)
Dewpoint	11 (0.307)	28 (0.139)	31 (0.113)	17 (0.212)	26 (0.150)
Virtual temperature	15 (0.227)	4 (0.461)	7 (0.346)	18 (0.212)	23 (0.165)
U-wind	40 (0.011)	35 (0.064)	41 (-0.013)	44 (-0.084)	43 (-0.057)
V-wind	46 (-0.145)	50 (-0.425)	49 (-0.409)	48 (-0.300)	45 (-0.131)
Mixing ratio	25 (0.152)	10 (0.312)	13 (0.288)	21 (0.192)	20 (0.194)
Specific humidity	8 (0.326)	9 (0.324)	27 (0.141)	38 (0.034)	39 (0.020)
Relative humidity	34 (0.064)	32 (0.081)	37 (0.037)	36 (0.041)	30 (0.114)
Pressure/heights	3 (0.474)	33 (0.080)	47 (-0.150)	42 (-0.031)	19 (0.199)
Theta-E	24 (0.153)	6 (0.383)	5 (0.441)	14 (0.252)	29 (0.128)

Figure 4.14: Same as Figure 4.4 but for stationary fronts over USAD.

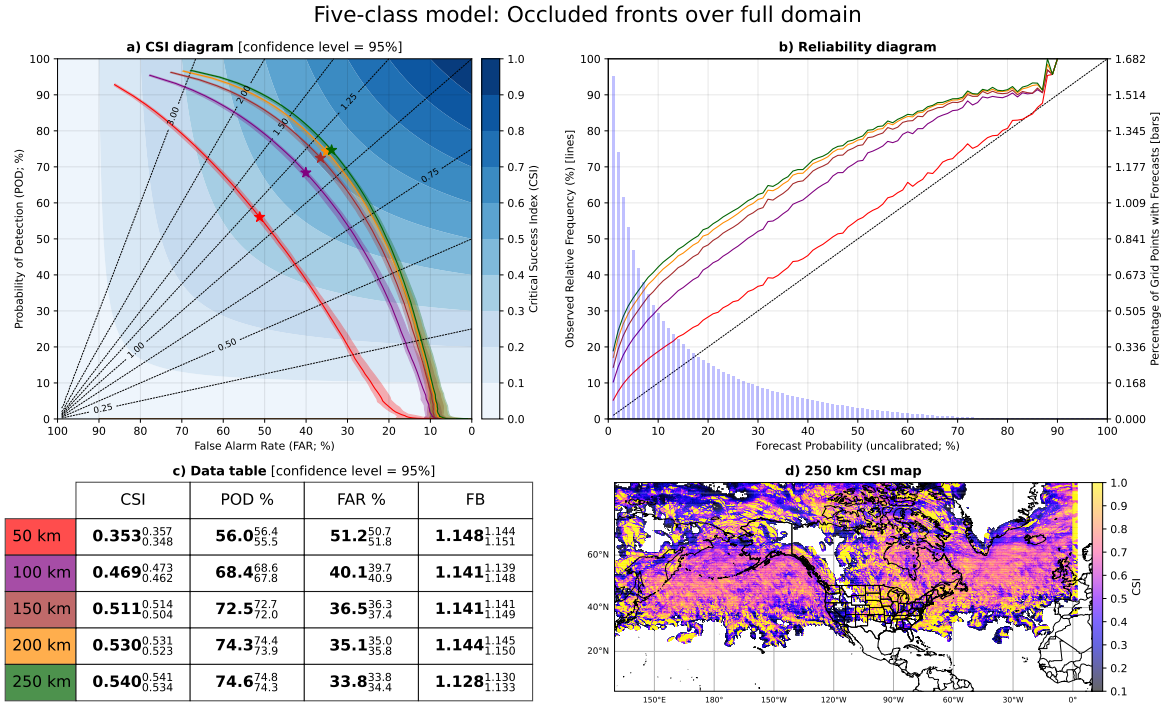
Occluded front performance was nearly the same across CONUS and USAD; 250 km CSI scores were about 0.54 over both domains. The spatial CSI plots in Figures 4.15 and 4.16 (panel d) show a CSI maximum over the Central US east of the Rocky Mountains, with broader areas of high performance over much of the Pacific and Atlantic Ocean basins north of 30°N. The model appears to resolve the occlusion process for Norwegian model-like cyclones (Bjerknes and Solberg, 1922), with occluded fronts wrapping around cyclones during the occlusion process (see the Case Study section for an example), in contrast to the Shapiro-Keyser cyclone model where an occluded front does not occur and the "wrap-up" process is associated with the bent-back warm front (Shapiro and Keyser, 1990). However, the intersections of the cold, warm, and occluded fronts (commonly referred to as "triple points") predicted by the model are not always colocated with the triple points shown in NOAA's surface analyses. This could be due to the fact that most occlusions are warm-type occlusions, which occur when air behind former cold fronts is warmer than air ahead of the warm fronts (Schultz and Mass, 1993; Schultz et al., 2014) and perhaps FCM interprets portions of the warm-type occlusions as warm fronts. The occluded front permutation studies in Figures 4.17 and 4.18 show that wind and pressure variables were most important for occluded front detection, similar to our findings in J23 with the three-model system. This is supported by multiple studies showing wind shifts coincident with occluded fronts (e.g., Market and Moore, 1998; Steenburgh et al., 1997; Schultz and Mass, 1993; Shafer et al., 2006). In addition, the wind components are important in the "wrap-up" process, where an occluded front becomes elongated and wraps around the low pressure center cyclonically. This wrap-up is influenced by deformation of the thermal ridge that is created after the cold and warm fronts collide (Schultz and Vaughan, 2011; Martin, 1999), however it is unclear how much this contributed to the relative

importance assigned to u-wind and v-wind as FCM appears to struggle with identifying wrapped occlusions (discussed more in the Case Study section). Similar to cold fronts, occluded fronts are often associated with pressure troughs and thus sharp rises in pressure can be observed upon their passage (e.g., Shafer et al., 2006). The unique nature of occluded fronts being in close proximity to low pressure centers also likely contributes to greater relative importance for pressure variables. The most important vertical levels included the surface, 1000, and 950 hPa, which are the lowest vertical levels in our datasets and the first levels at which an occluded front will form following the collision of a cold front with a warm front. Since the features at 950 hPa are aloft, analysts may use different data sources when drawing occluded fronts (e.g., satellite imagery), while we are simply training the model on thermodynamic fields.



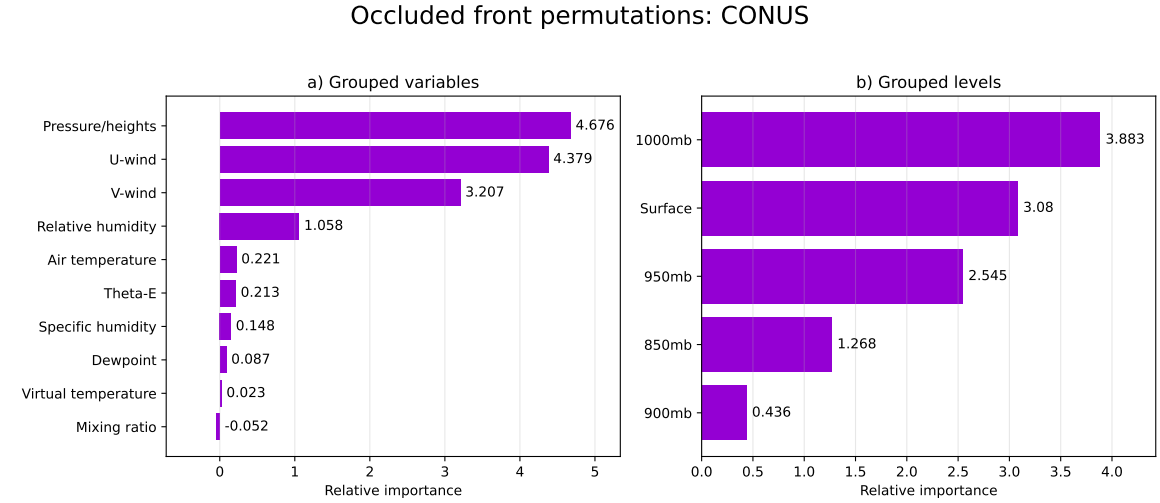
	CSI	POD %	FAR %	FB
50 km	$0.386^{0.401}_{0.371}$	$61.2^{62.8}_{59.6}$	$48.9^{47.3}_{50.5}$	$1.199^{1.191}_{1.205}$
100 km	$0.491^{0.509}_{0.475}$	$71.3^{72.6}_{70.0}$	$38.7^{37.0}_{40.4}$	$1.164^{1.152}_{1.174}$
150 km	$0.522^{0.539}_{0.506}$	$74.0^{75.2}_{72.9}$	$36.0^{34.3}_{37.7}$	$1.157^{1.145}_{1.170}$
200 km	$0.535^{0.552}_{0.518}$	$75.9^{77.0}_{74.9}$	$35.6^{33.9}_{37.2}$	$1.180^{1.165}_{1.193}$
250 km	$0.541^{0.558}_{0.524}$	$76.2^{77.2}_{75.1}$	$34.9^{33.2}_{36.5}$	$1.170^{1.155}_{1.183}$

Figure 4.15: Same as Figure 4.1 but for occluded fronts over CONUS.



	CSI	POD %	FAR %	FB
50 km	$\mathbf{0.353}^{0.357}_{0.348}$	$\mathbf{56.0}^{56.4}_{55.5}$	$\mathbf{51.2}^{50.7}_{51.8}$	$\mathbf{1.148}^{1.144}_{1.151}$
100 km	$\mathbf{0.469}^{0.473}_{0.462}$	$\mathbf{68.4}^{68.6}_{67.8}$	$\mathbf{40.1}^{39.7}_{40.9}$	$\mathbf{1.141}^{1.139}_{1.148}$
150 km	$\mathbf{0.511}^{0.514}_{0.504}$	$\mathbf{72.5}^{72.7}_{72.0}$	$\mathbf{36.5}^{36.3}_{37.4}$	$\mathbf{1.141}^{1.141}_{1.149}$
200 km	$\mathbf{0.530}^{0.531}_{0.523}$	$\mathbf{74.3}^{74.4}_{73.9}$	$\mathbf{35.1}^{35.0}_{35.8}$	$\mathbf{1.144}^{1.145}_{1.150}$
250 km	$\mathbf{0.540}^{0.541}_{0.534}$	$\mathbf{74.6}^{74.8}_{74.3}$	$\mathbf{33.8}^{33.8}_{34.4}$	$\mathbf{1.128}^{1.130}_{1.133}$

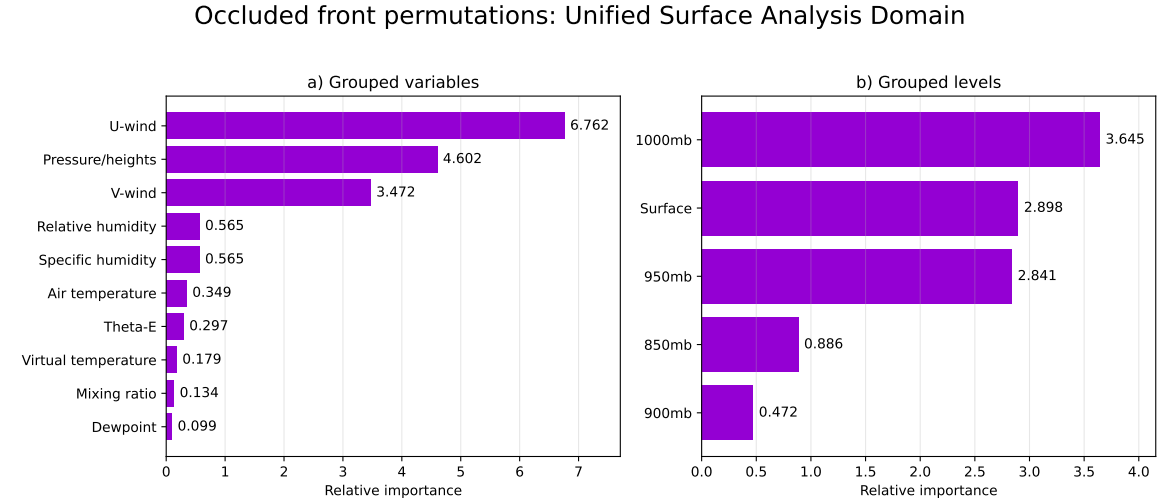
Figure 4.16: Same as Figure 4.2 but for occluded fronts over USAD.



c) Variables on single levels

	Surface	1000mb	950mb	900mb	850mb
Air temperature	44 (-0.118)	46 (-0.133)	22 (0.031)	33 (-0.035)	50 (-0.269)
Dewpoint	34 (-0.045)	20 (0.065)	42 (-0.109)	41 (-0.097)	19 (0.068)
Virtual temperature	49 (-0.207)	48 (-0.160)	23 (0.024)	29 (-0.021)	47 (-0.135)
U-wind	2 (1.449)	3 (1.045)	8 (0.403)	17 (0.099)	15 (0.207)
V-wind	11 (0.354)	7 (0.414)	9 (0.385)	6 (0.533)	10 (0.373)
Mixing ratio	35 (-0.047)	43 (-0.109)	45 (-0.124)	37 (-0.063)	26 (0.002)
Specific humidity	27 (-0.006)	36 (-0.056)	25 (0.015)	31 (-0.033)	28 (-0.016)
Relative humidity	13 (0.211)	14 (0.210)	16 (0.125)	18 (0.080)	12 (0.271)
Pressure/heights	38 (-0.063)	5 (0.730)	1 (1.473)	4 (0.846)	21 (0.039)
Theta-E	40 (-0.077)	39 (-0.070)	32 (-0.035)	24 (0.019)	30 (-0.025)

Figure 4.17: Same as Figure 4.3 but for occluded fronts over CONUS.



c) Variables on single levels

	Surface	1000mb	950mb	900mb	850mb
Air temperature	48 (-0.068)	41 (0.015)	14 (0.220)	30 (0.087)	50 (-0.219)
Dewpoint	43 (0.008)	25 (0.125)	44 (0.004)	38 (0.021)	18 (0.154)
Virtual temperature	47 (-0.066)	46 (-0.051)	22 (0.135)	31 (0.081)	49 (-0.143)
U-wind	2 (1.572)	1 (1.762)	5 (0.842)	39 (0.020)	19 (0.153)
V-wind	8 (0.622)	7 (0.636)	9 (0.353)	11 (0.235)	13 (0.221)
Mixing ratio	23 (0.125)	36 (0.061)	33 (0.067)	37 (0.035)	34 (0.066)
Specific humidity	20 (0.139)	21 (0.136)	10 (0.311)	24 (0.125)	32 (0.078)
Relative humidity	15 (0.193)	16 (0.187)	27 (0.114)	28 (0.101)	12 (0.225)
Pressure/heights	40 (0.018)	4 (0.890)	3 (1.549)	6 (0.765)	29 (0.087)
Theta-E	42 (0.011)	45 (-0.000)	35 (0.065)	17 (0.174)	26 (0.119)

Figure 4.18: Same as Figure 4.4 but for occluded fronts over USAD.

Drylines over the Great Plains of the United States were detected well by the FCM. Using a 100 km neighborhood, 78.6% of drylines were located with a 24.8% FAR, yielding a 100 km CSI of 0.624 over the CONUS domain (Figure 4.19). Performance was lower over USAD as the FCM often predicts drylines in areas of Mexico where drylines are not frequently drawn, leading to a higher FAR (Figure 4.20). Recent evidence has suggested that some of the false alarms from the FCM may be due to changes in forecasters' drawing habits and that drylines were previously under-drawn by analysts (Hosek, 2024). Our dryline permutation studies (Figures 4.21 and 4.22) show that mixing ratio, v-wind, and specific humidity were the most important variables for dryline detection. Drylines were the only front type predicted by the FCM to have a moisture variable as its most important predictor (mixing ratio), which is physically consistent with the fact that drylines are defined by sloped terrain and their often extreme moisture gradients (Schaefer, 1986; Pietrycha and Rasmussen, 2004; Hoch and Markowski, 2005). Specific humidity also had high importance, likely due to its similar conservation properties compared to those of mixing ratio (both are conserved following dry adiabatic motion and unaffected by changes in pressure or temperature). However, given that specific humidity and mixing ratio have very similar properties, it seems plausible that the model could predict drylines just as well with only one of these variables, making mixing ratio or specific humidity far more important than the permutation studies suggest. The high importance assigned to v-wind was a surprising find, however we believe this can be tied to the nocturnal low-level jet's southerly component that accelerates on the moist side of the dryline as the dryline begins its westward regression during the nocturnal hours (Parsons et al., 1991, 2000). Surface pressure also had elevated importance; observations by Parsons et al. (1991); Ziegler and Hane (1993) showed that high (low) pressure exists on the dry (moist) side of the dryline. 900 and 850 hPa were determined to be the most important pressure levels for

dryline detection, followed by the surface level. While we are unsure as to why higher pressure levels (in terms of elevation) were preferred over lower levels, we think this may be a result of diurnal heating effects that control the progression and regression of the dryline (Schaefer, 1973) and the nocturnal low-level jet having a presence above the 900 hPa level as found by Parsons et al. (1991).

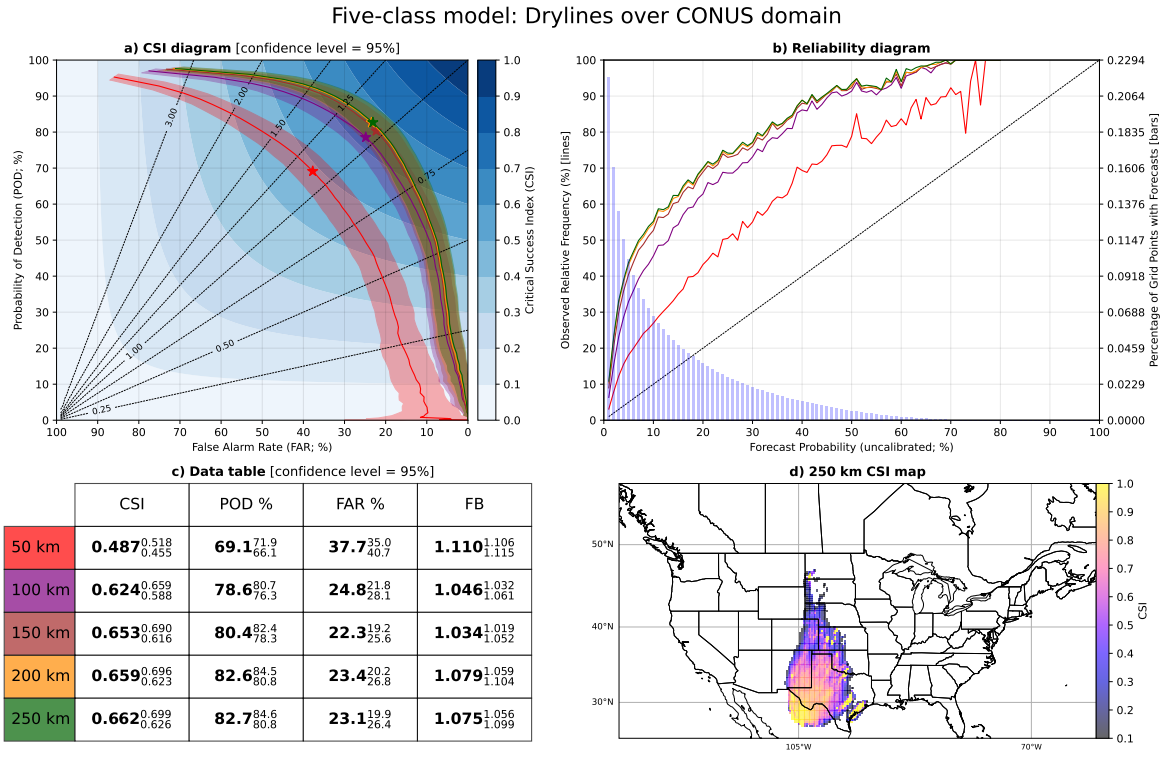
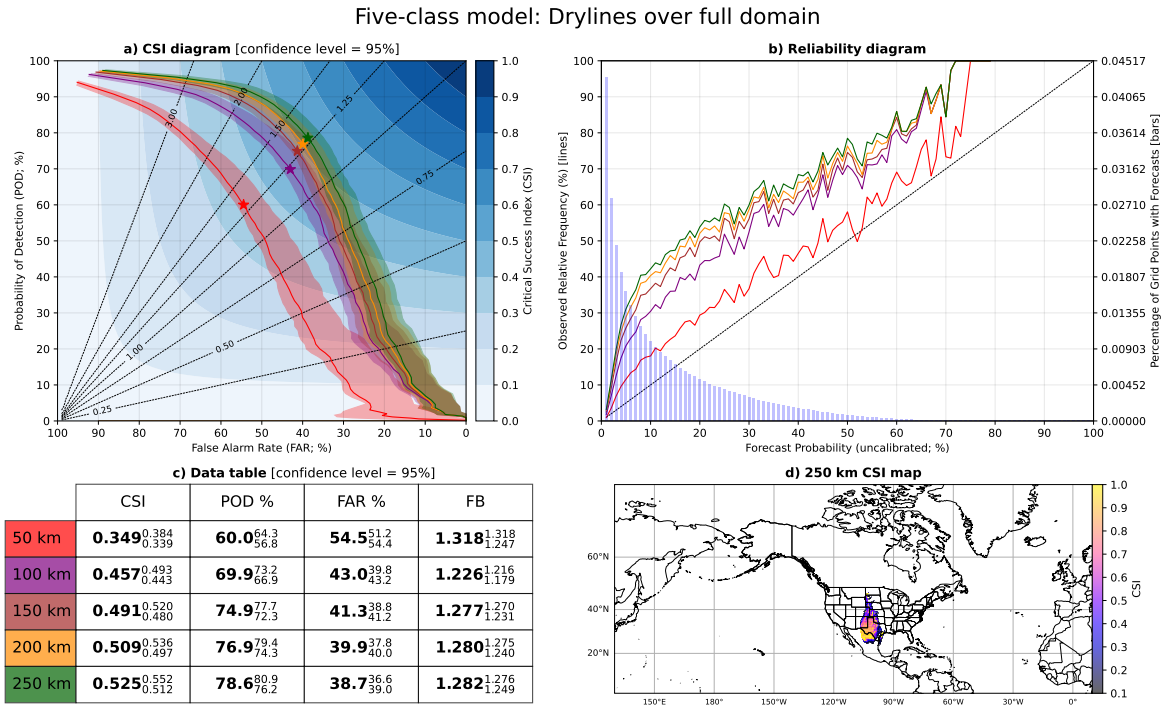
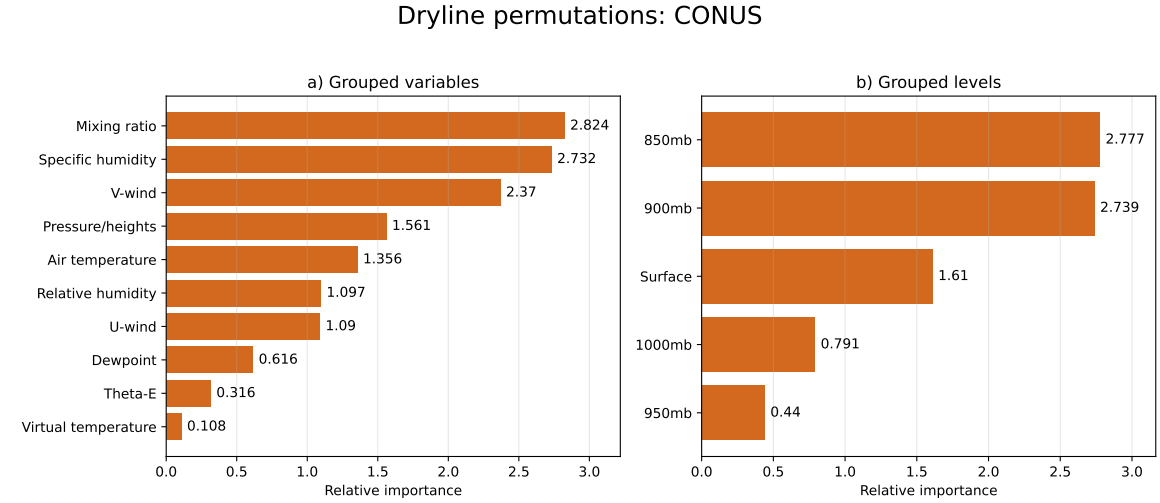


Figure 4.19: Same as Figure 4.1 but for drylines over CONUS.



	CSI	POD %	FAR %	FB
50 km	**0.349**$^{0.384}_{0.339}$	**60.0**$^{64.3}_{56.8}$	**54.5**$^{51.2}_{54.4}$	**1.318**$^{1.318}_{1.247}$
100 km	**0.457**$^{0.493}_{0.443}$	**69.9**$^{73.2}_{66.9}$	**43.0**$^{39.8}_{43.2}$	**1.226**$^{1.216}_{1.179}$
150 km	**0.491**$^{0.520}_{0.480}$	**74.9**$^{77.7}_{72.3}$	**41.3**$^{38.8}_{41.2}$	**1.277**$^{1.270}_{1.231}$
200 km	**0.509**$^{0.536}_{0.497}$	**76.9**$^{79.4}_{74.3}$	**39.9**$^{37.8}_{40.0}$	**1.280**$^{1.275}_{1.240}$
250 km	**0.525**$^{0.552}_{0.512}$	**78.6**$^{80.9}_{76.2}$	**38.7**$^{36.6}_{39.0}$	**1.282**$^{1.276}_{1.249}$

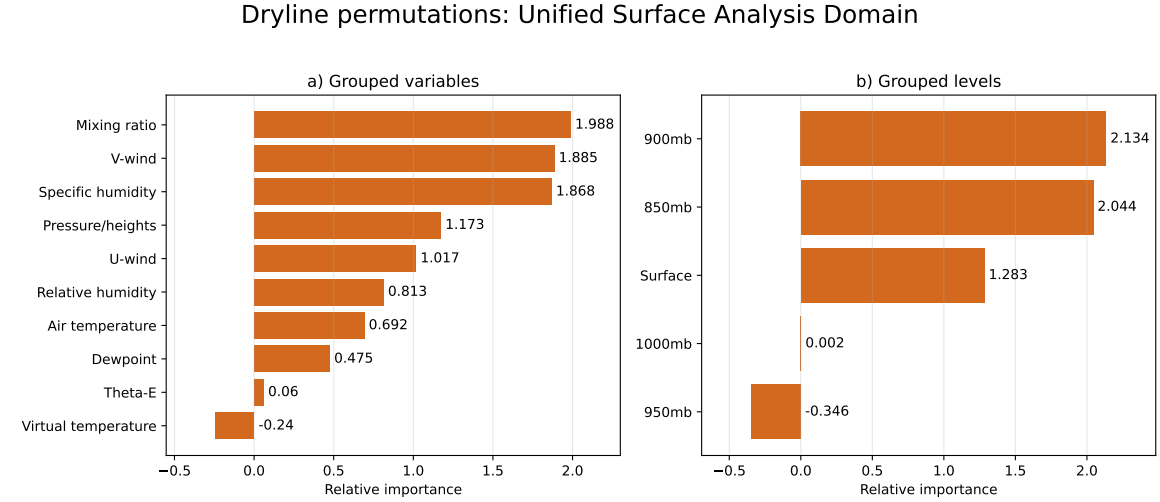
Figure 4.20: Same as Figure 4.2 but for drylines over USAD.



c) Variables on single levels

	Surface	1000mb	950mb	900mb	850mb
Air temperature	26 (0.095)	24 (0.101)	34 (-0.020)	17 (0.204)	11 (0.380)
Dewpoint	16 (0.230)	44 (-0.097)	46 (-0.109)	25 (0.099)	31 (0.025)
Virtual temperature	28 (0.051)	42 (-0.084)	50 (-0.293)	33 (-0.013)	13 (0.299)
U-wind	9 (0.391)	14 (0.280)	48 (-0.144)	49 (-0.261)	47 (-0.132)
V-wind	38 (-0.051)	6 (0.470)	12 (0.376)	10 (0.388)	21 (0.132)
Mixing ratio	20 (0.136)	5 (0.485)	1 (0.773)	3 (0.699)	15 (0.278)
Specific humidity	18 (0.195)	2 (0.742)	4 (0.600)	7 (0.454)	35 (-0.033)
Relative humidity	32 (0.001)	23 (0.114)	19 (0.165)	22 (0.117)	37 (-0.047)
Pressure/heights	8 (0.445)	30 (0.045)	27 (0.094)	45 (-0.107)	43 (-0.091)
Theta-E	41 (-0.082)	36 (-0.046)	39 (-0.066)	40 (-0.068)	29 (0.045)

Figure 4.21: Same as Figure 4.3 but for drylines over CONUS.



c) Variables on single levels

	Surface	1000mb	950mb	900mb	850mb
Air temperature	31 (0.078)	44 (-0.028)	49 (-0.114)	14 (0.281)	6 (0.491)
Dewpoint	15 (0.276)	40 (0.019)	43 (-0.027)	26 (0.133)	32 (0.074)
Virtual temperature	22 (0.167)	47 (-0.047)	50 (-0.314)	34 (0.043)	9 (0.411)
U-wind	12 (0.366)	13 (0.325)	37 (0.032)	46 (-0.042)	42 (-0.011)
V-wind	27 (0.126)	4 (0.540)	7 (0.423)	11 (0.381)	21 (0.176)
Mixing ratio	23 (0.156)	10 (0.401)	1 (0.623)	2 (0.592)	16 (0.274)
Specific humidity	20 (0.184)	3 (0.548)	5 (0.503)	8 (0.419)	29 (0.097)
Relative humidity	41 (0.015)	28 (0.118)	18 (0.233)	19 (0.189)	36 (0.034)
Pressure/heights	17 (0.265)	25 (0.143)	24 (0.149)	39 (0.031)	33 (0.069)
Theta-E	35 (0.036)	38 (0.032)	45 (-0.028)	48 (-0.051)	30 (0.084)

Figure 4.22: Same as Figure 4.4 but for drylines over USAD.

4.1 Case Study: 2023 Christmas Winter Storm

With the goal of better understanding FCM's strengths and weaknesses, we analyzed FCM model output (Figure 4.23), mid-level water vapor satellite imagery (Figure 4.24), WPC surface analyses (Figure 4.25), and observations at the surface and 850 hPa (Figure 4.26, Figure 4.27) for four timesteps - 0000 and 1200 UTC on Dec 26 and Dec 27, 2023. From Dec 24-27, 2023, a cyclone situated over the Midwest region of the United States brought upwards of 12 inches (30.5 cm) of snow and 1.5 inches (3.8 cm) of ice across areas in South Dakota and Nebraska[1]. This precipitation was driven by an occluded front that was wrapped around its parent surface cyclone, with the precipitation occurring in the northwest quadrant of the cyclone. This is commonplace for winter precipitation as the northwest quadrants of surface cyclones are coincident with rising motion in frontogenetical circulations (Ganetis et al., 2018).

Starting with 0000 UTC Dec 26 2023, WPC surface analysis indicated an occluding cyclone over the Upper Midwest with a nearby cold frontal zone extending from Southern Minnesota to Quebec (Figure 4.25a). The occluded front extends north of a triple point where a cold front is overtaking a warm front to the east as suggested by the WPC analysis and FCM predictions. This occlusion was well-defined with sharp temperature gradients at the surface and 850 hPa and a notable wind shift and pressure trough at the surface. The FCM had no issue finding the occluded front as calibrated probabilities exceeded 80% along most of the frontal zone (Figure 4.23). The frontal zone that extended from Southern Minnesota to Quebec was analyzed by WPC as a stationary front in Minnesota and a cold front along the rest of its extent. Notably, the FCM was able to highlight the transition from the stationary portion to the cold portion of the frontal zone, though the transition is displaced from that indicated in

[1]https://www.weather.gov/abr/2023ChristmasStorm

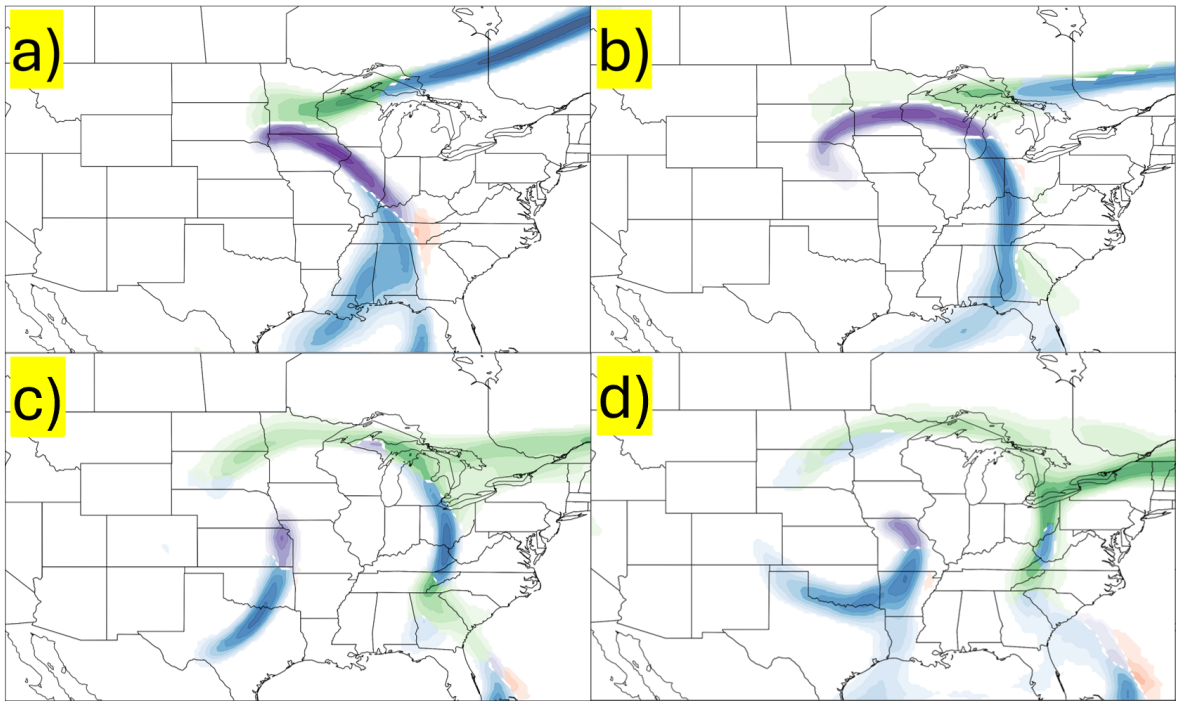


Figure 4.23: FCM predictions over CONUS using GFS model data: a) 0000 UTC Dec 26 2023, b) 1200 UTC Dec 26 2023, c) 0000 UTC Dec 27 2023, d) 1200 UTC Dec 27 2023. Note that all predictions use forecast hour 0 for the respective initialization times and are calibrated to a 100 km neighborhood with filled contours at 10% intervals. (blue = cold front, red = warm front, green = stationary front, purple = occluded front)

WPC's surface analysis. FCM had lower probabilities along the stationary portion of the zone closest to the center of the cyclone. FCM also identified a broad cold frontal region in the Southeastern US attached to the southern end of the occluded front. The model was likely less confident in this region due to a more diffuse temperature gradient, weaker wind shift, and the lack of a prominent pressure trough. In fact, this cold front has almost no presence at 850 hPa, though a weak moisture gradient can be noted where WPC analyzed a cold front. Satellite imagery shows the cyclone in the Upper Midwest with an attendant high pressure system off the New England coast.

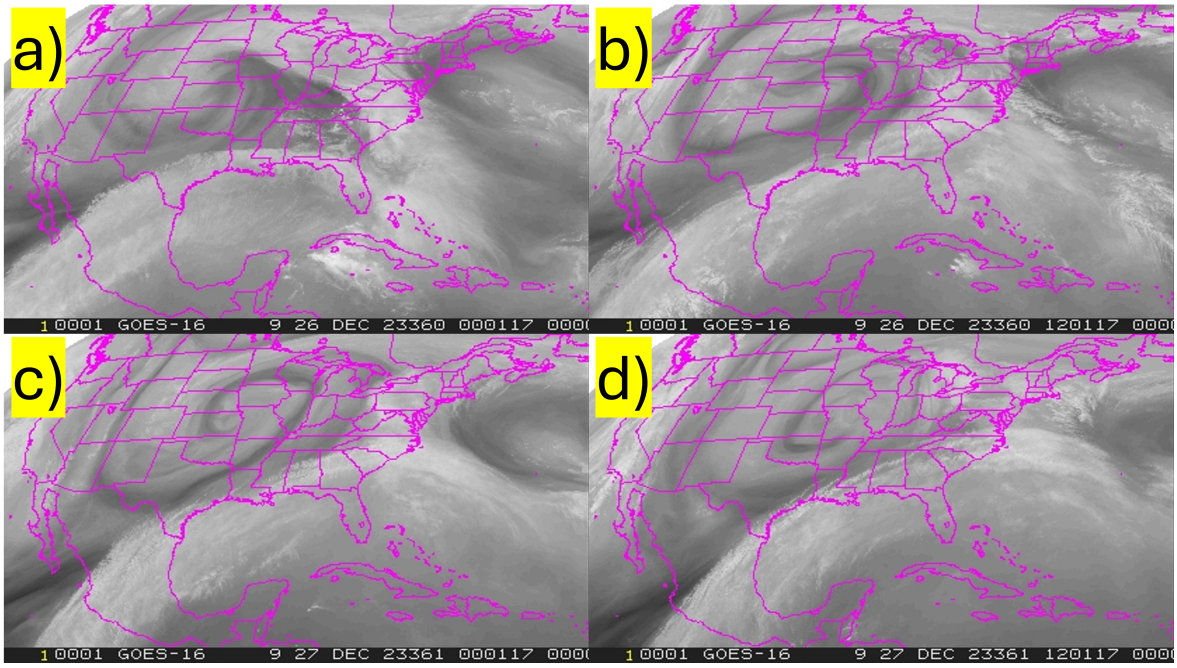


Figure 4.24: GOES-16 mid-level water vapor imagery (band 9) showing a cyclone over CONUS: a) 0001 UTC Dec 26 2023, b) 1201 UTC Dec 26 2023, c) 0001 UTC Dec 27 2023, d) 1201 UTC Dec 27 2023.

Higher mid-level water vapor levels can be noted along and ahead of the occluded front, where precipitation is ongoing.

At 0300 UTC Dec 26 (not shown), WPC analyzed the previous occluded front as a cold front, likely due to air behind the front being much colder than air out ahead of it. Since this air was north of an analyzed warm front, this suggests that the occluded front analyzed three hours prior may have been a rare instance of a cold-type occlusion, which occurs when air behind the former cold front undercuts warmer air ahead of the warm front (Schultz et al., 2014). Confirming this hypothesis will require an in-depth analysis of vertical cross sections along the occlusion, which is beyond the scope of this thesis. This front (now analyzed as a cold front) collided with a stationary boundary sitting in Southern Minnesota, at which point a new occlusion was noted in the 0600 UTC Dec 26

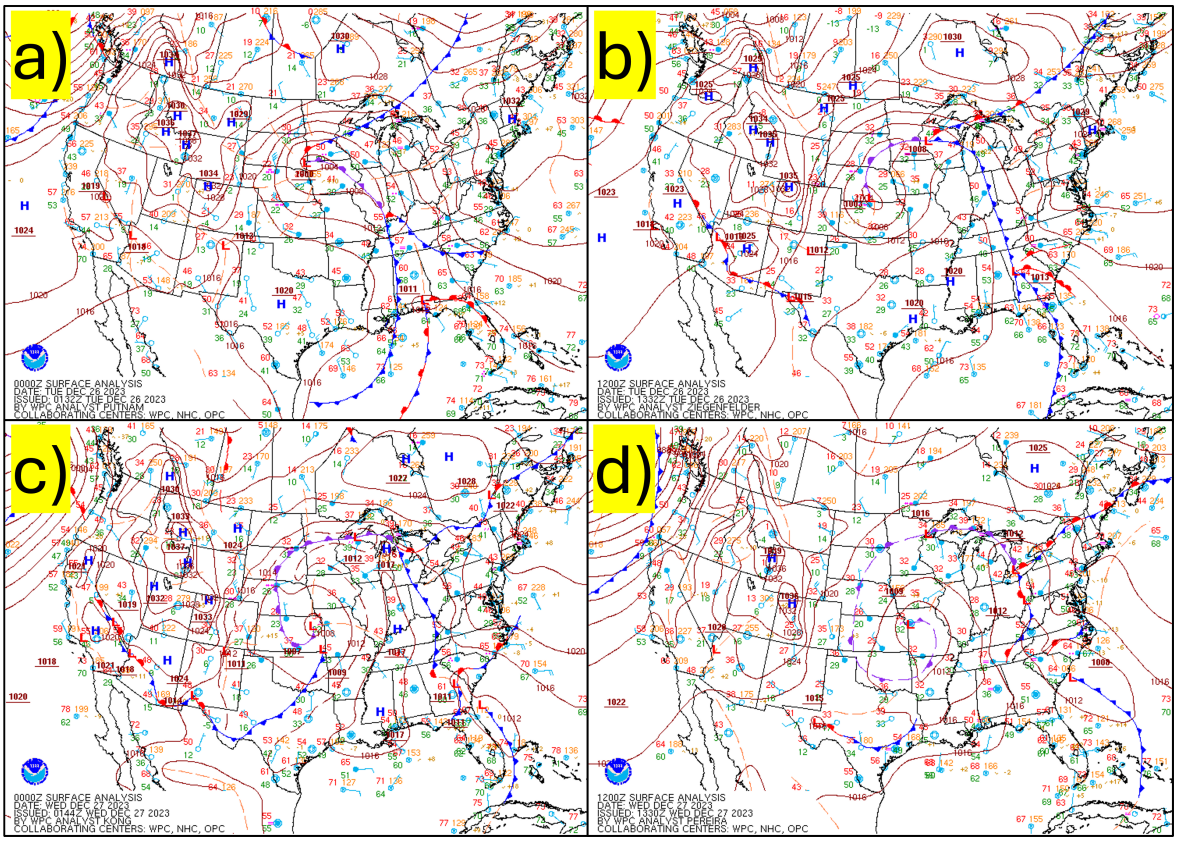


Figure 4.25: WPC surface analyses over CONUS: a) 0000 UTC Dec 26 2023, b) 1200 UTC Dec 26 2023, c) 0000 UTC Dec 27 2023, d) 1200 UTC Dec 27 2023.

WPC surface analysis (not shown). At 1200 UTC on Dec 26, WPC's surface analysis shows a triple point in Western Wisconsin, with the a stationary front extending to the ENE, a cold front extending down to the Florida panhandle, and an occluded front that has started to wrap cyclonically around the low pressure center. The extension of the occluded front around the cyclone is likely due to deformation of the thermal ridge as described in Schultz and Vaughan (2011); Martin (1999). The occluded front is highly visible on satellite in Minnesota and South Dakota with precipitation north and west of the front, highlighted by the mid-level water vapor imagery. FCM was able to resolve the beginning of the wrap-up process, placing calibrated probabilities exceeding 70% around the northwest side of the low and identifying the occluded front responsible for

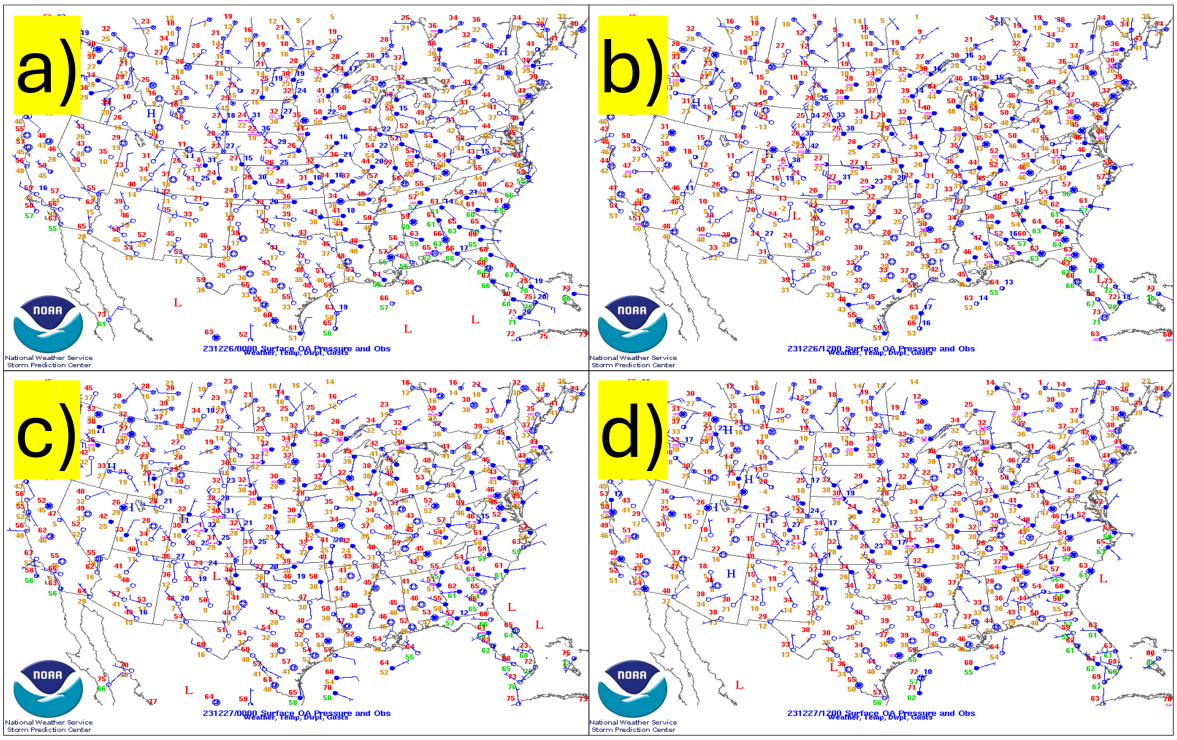


Figure 4.26: Objectively analyzed surface maps from the Storm Prediction Center: a) 0000 UTC Dec 26 2023, b) 1200 UTC Dec 26 2023, c) 0000 UTC Dec 27 2023, d) 1200 UTC Dec 27 2023.

heavy winter precipitation in the Upper Midwest. Observations at the surface and 850 hPa show a wind shift and sharp temperature gradient along the occluded front, and a pressure trough is noted in the WPC surface analysis. The cold front across Ontario and Quebec had lower probabilities from FCM, likely due to the front transitioning into a quasi-stationary state as suggested by the 1200 UTC surface analysis from WPC. This frontal zone was accompanied by a temperature gradient and strong convergence of v-wind, consistent with our cold front permutation studies (see Figures 4.3 and 4.4). The cold front extending down to the Florida panhandle also exhibited a temperature gradient and wind shift and was detected by FCM with probabilities exceeding 80% in areas along the front. The model was less confident in its predictions closer to the Gulf

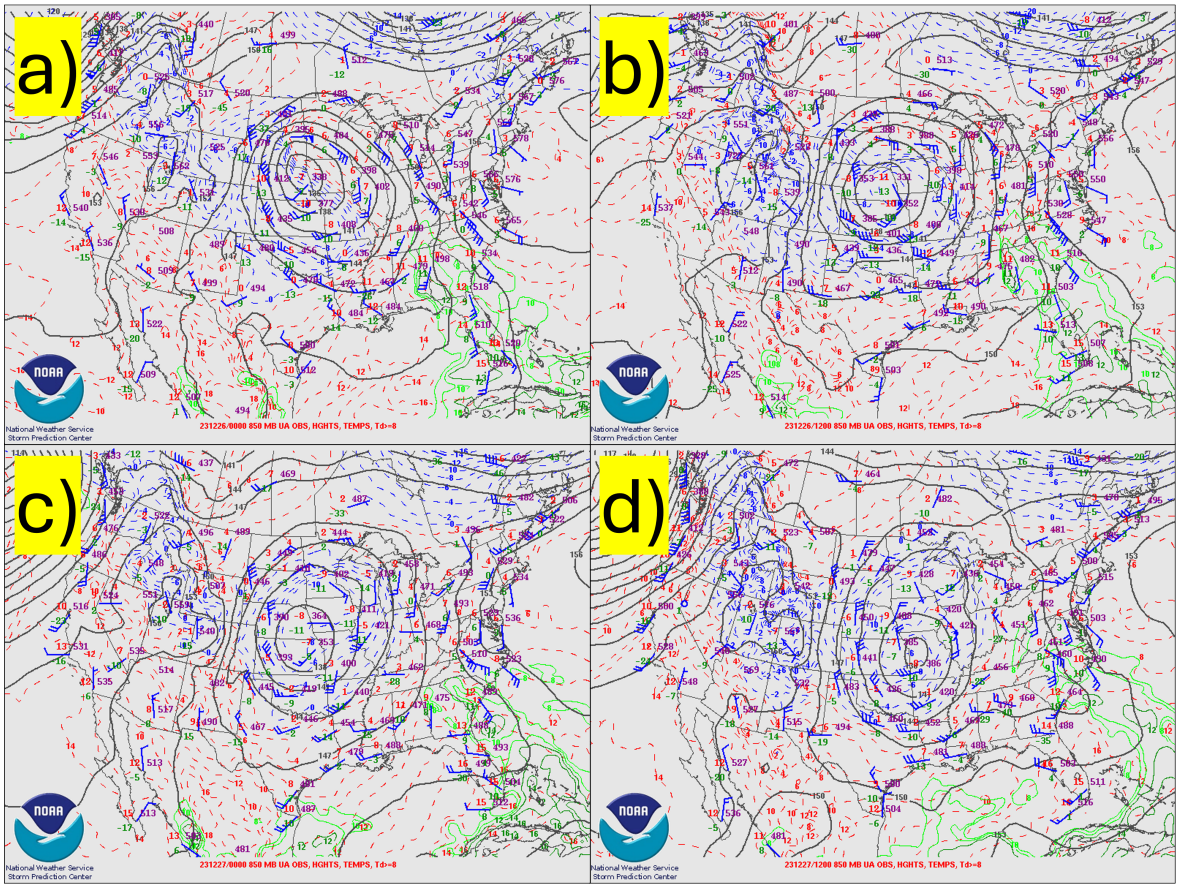


Figure 4.27: Objectively analyzed 850 hPa maps from the Storm Prediction Center: a) 0000 UTC Dec 26 2023, b) 1200 UTC Dec 26 2023, c) 0000 UTC Dec 27 2023, d) 1200 UTC Dec 27 2023.

of Mexico as a secondary surface low with multiple fronts had developed in the state of Georgia. The warm front analyzed by WPC extending east of this secondary low was not detected, even though we believe the surface observations corroborate WPC's analysis of the warm front. We believe that a combination of the weak temperature gradient along the warm front and the presence of other boundaries in close proximity limited FCM's ability to detect this warm front.

By 0000 UTC Dec 27, the occluded front was wrapped three-quarters of the way around the cyclone, as evident by mid-level water vapor imagery. The dark corridor in

the mid-level water vapor imagery behind the occluded front is representative of the dry conveyor belt, also known as the dry intrusion, which is a region of air that descends from the upper troposphere (Browning, 1997). WPC analyzed this front along the transition between the dry conveyor belt and high mid-level water vapor concentrations coincident with winter precipitation. The occluded front had almost completely disappeared from FCM model output, with the exceptions of small corridors over the Upper Peninsula of Michigan and the Midwest, with the corridor in the Midwest being immediately south of the surface low and attached to a newly-developed cold front in the Southern Plains. Surface and 850 hPa analyses show that the temperature gradient along the occluded front has weakened significantly, though some wind convergence is still present along the occluded front in the Great Lakes region. FCM had a mix of stationary, occluded, and cold front probabilities along the northernmost section of the analyzed occluded front, likely due to the wind convergence being accompanied by weak thermal gradients. The highest occluded front probabilities along this section of the analyzed occluded front were located near the triple point over Michigan, but FCM still determined much of this frontal zone to be of the stationary type. FCM output did not exhibit any notable changes between 0000 UTC and 1200 UTC on Dec 27, however the occluded front probabilities near the triple point disappeared. At 1200 UTC Dec 27 the occluded front is still visible on mid-level water vapor imagery and completely wrapped around the low pressure center. While a subtle temperature gradient still exists at the surface and 850 hPa, there is no wind convergence that clearly demarcates the weakening occlusion. We believe that including satellite imagery as input to a future UNET3+ architecture will help with occluded front detection. Satellite imagery could also help with the detection of other frontal types, as we learned that forecasters often use satellite imagery to locate fronts where surface observations are few and far between (e.g., oceans).

4.2 Case Study: Saliency Maps

Saliency maps for our chosen case study of 0000 UTC Feb 4 2019 are shown in Figures 4.28 to 4.32, while the WPC fronts for this timestep are visualized in Figure 4.33. Cold front predictions in Figure 4.28a show high probabilities for a cold front in the Central US, with multiple cold fronts elsewhere across the domain. The Central US cold front probabilities were nearly maxed out with very little saliency at all vertical levels. This suggests that FCM is confident in its predictions of strong cold fronts, meaning that small changes to the input variables would not impact the model's interpretation of the data. Our cold front permutation studies (Figures 4.1 and 4.4) show that while v-wind and air temperature are the most important variables for cold front detection, several other input variables are contributing to cold front detection. We believe this is why strong cold fronts appear to have lower saliency than weaker cold fronts. Another cold front was located off the coast of Southern California with lower probabilities than the Central US cold front. This cold front was associated with the highest saliency values seen across the whole domain at the time. Saliency was very small along the axis of the cold front, suggesting that the model was confident in the central location of the front, however the saliency jumps outside of the highest probabilities. The model appears to be uncertain about the full extent of the cold front, which is something we have noticed with low-probability cold fronts. Saliency surrounding this front was highest at 1000 and 950 hPa, the two most important levels for cold front detection according to our permutation studies.

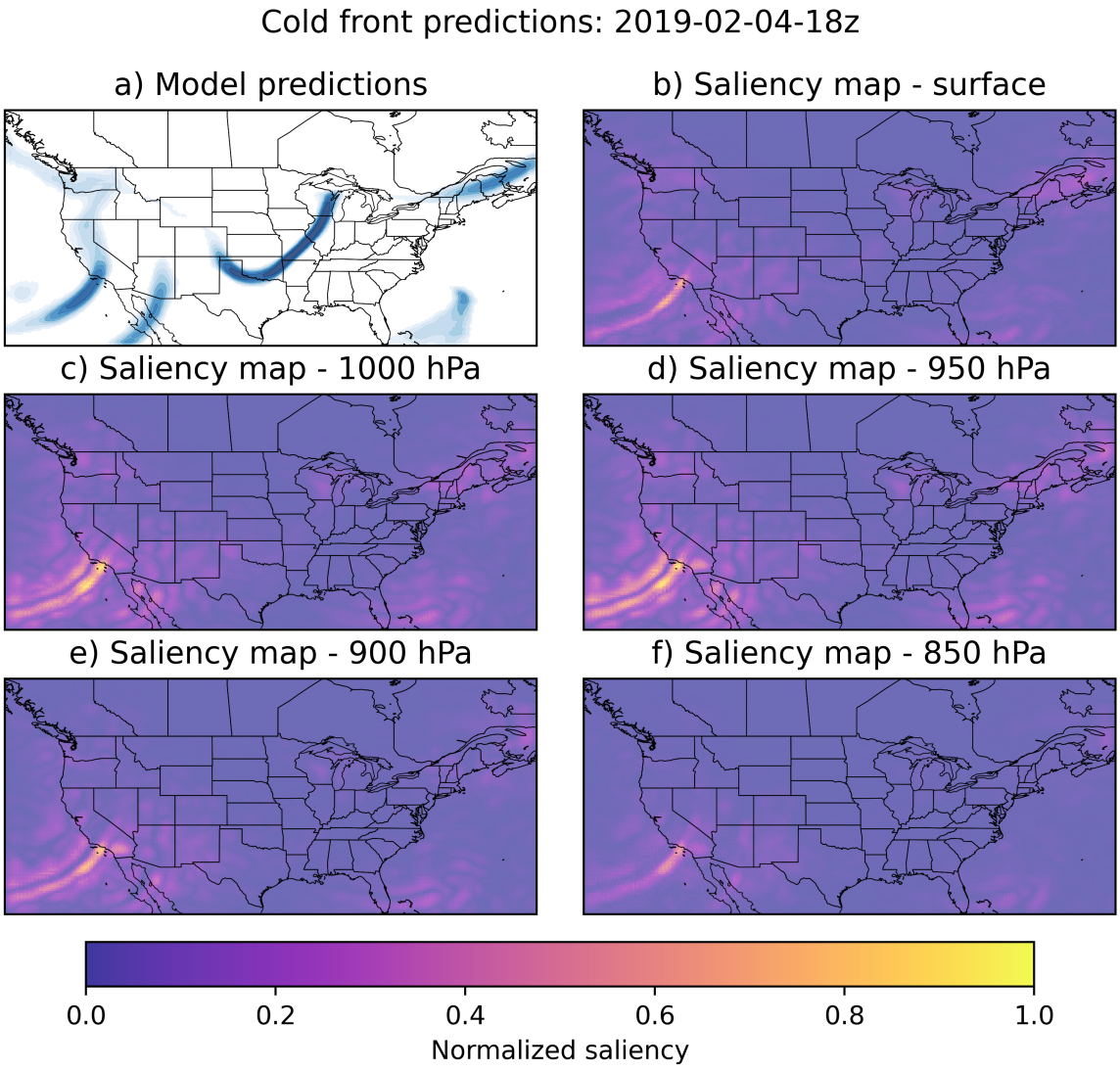


Figure 4.28: FCM cold front predictions over CONUS contoured at 10% intervals (a) and saliency maps at b) the surface, c) 1000 hPa, d) 950 hPa, e) 900 hPa, and f) 850 hPa at 1800 UTC Feb 2 2019.

Warm front predictions and saliency maps can be found in Figure 4.29. The model identified a zonally-oriented warm front over the Great Lakes region, closely matching the WPC surface analysis (Figure 4.33). The highest saliency values were at 1000 hPa and north of the warm front. We believe this is because the model is uncertain about how far north to extend the warm front due to weaker thermal gradients and wind shifts often associated with warm fronts, especially over land. As mentioned earlier, FCM warm front performance was higher over water than land, which we attributed to less surface friction over water. This could be an explanation for why the warm front in the Western Atlantic shown in Figure 4.29 has lower saliency than the warm front over the Great Lakes; FCM could be more certain on its placement of oceanic warm fronts since they tend to be stronger and more defined when surface friction is weak (Hines and Mechoso, 1993). Like cold fronts, local minimums in saliency were observed along the axes of oceanic warm fronts, and saliency was greater near the fronts where probabilities were lower.

Stationary fronts exhibited some interesting characteristics with saliency. The predictions and saliency maps are shown in Figure 4.30. Stationary fronts with high probabilities, especially those over the Rocky Mountains, were associated with low saliency. This behavior was also observed with strong cold fronts and oceanic warm fronts. The low saliency is likely a result of multiple variables working together. In other words, a change to a particular variable may not substantially influence the model's interpretation of the data because there are several other variables being used to generate high probabilities. FCM identified a possible stationary front in the Great Lakes region, covering an area also highlighted as possibly containing a warm front (see Figure 4.29a). High saliency was observed at various points along this region where FCM produced conflicting probabilities. We found that saliency increased dramatically in regions where stationary fronts were transitioning to cold or warm fronts. For

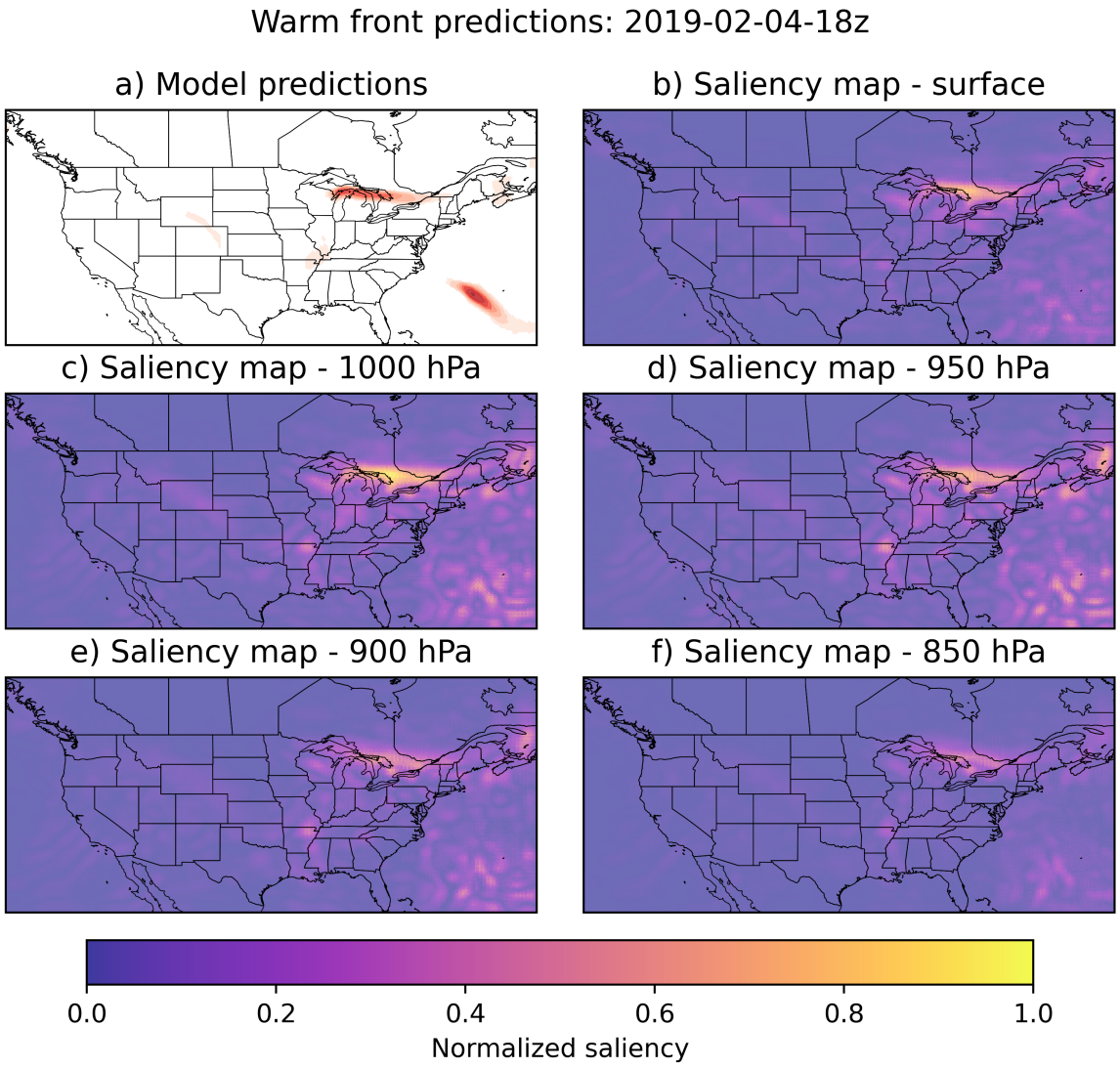


Figure 4.29: Same as Figure 4.28 but for warm fronts.

example, as stationary front probabilities decrease with eastern extent in Maine, the saliency increases along with the cold front probabilities shown in Figure 4.28. This suggests there are features that greatly influence FCM's interpretation of transitions between frontal zones, and small changes to the surrounding environment could result in the model favoring one front type over the others. In the Pacific Northwest, high saliency can be noted where stationary front probabilities are lower than those to the

northwest and southeast. This is another transition zone where FCM suggests a stationary front is turning into a cold front, which is a likely source for the high saliency. The Pacific Northwest is a region where FCM has relatively poor cold front performance when compared to the rest of the CONUS domain, likely due to oceanic cold fronts moving onto land in the region, which could also make the model more sensitive to environmental changes in this region.

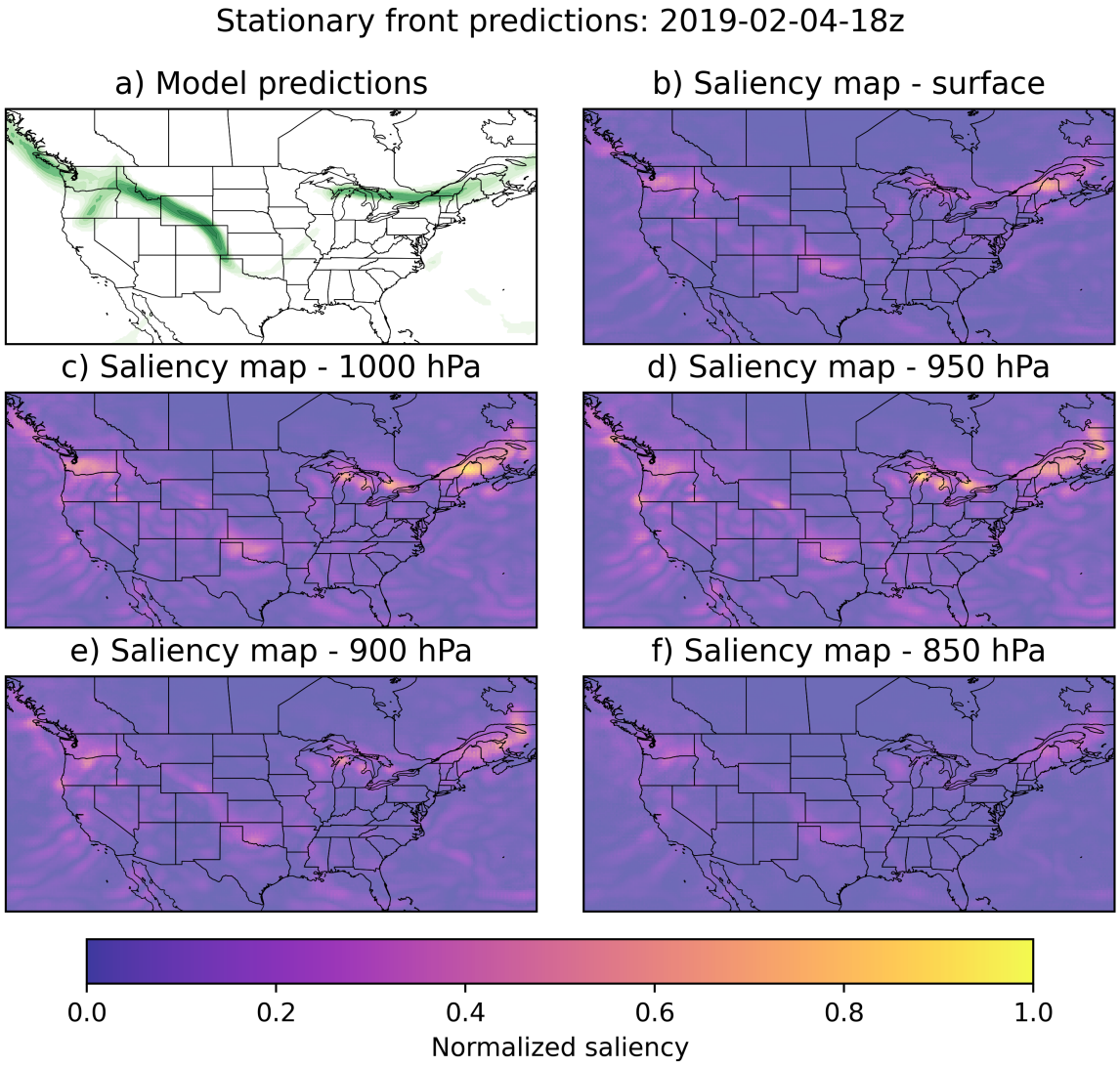


Figure 4.30: Same as Figure 4.28 but for stationary fronts.

Occluded front probabilities and saliency maps for the case study are visualized in Figure 4.31. This case study only has one occluded front, however through analyses of other case studies not shown in this thesis we were still able to come across some intriguing results. A feature we have seen in many occluded fronts we have analyzed is a maximum in saliency centered right over the triple point - the intersection between the cold, warm, and occluded fronts associated with midlatitude cyclones. This "hot spot" in occluded front saliency is where FCM cold, warm, and occluded front probabilities all decrease close to the triple point. While this is not a surprising find, it does show that FCM is behaving in a way that can be related to real processes observed in the atmosphere. The saliency hot spot on the triple point is also where the cold front is actively overrunning the warm front, extending the occluded front further south with time. In other words, a small change to the environment near the triple point, such as letting the cold front overrun a greater portion of the warm front, will likely result in a sharp increase in occluded front probabilities where saliency at the triple point is greatest. The vertical level with the highest saliency on the triple point was 950 hPa, where geopotential height had the greatest importance across all 50 variable and level combinations. This makes sense given that warm occlusions, the most common type of occlusion, result in the occluded front being forced above the surface (Schultz and Mass, 1993; Schultz et al., 2014).

Dryline predictions and saliency maps can be found in Figure 4.32. FCM identified no drylines in this case study, however WPC analyzed a dryline in Texas (Figure 4.33). Along the axis of the WPC dryline, saliency was relatively low, while saliency was higher surrounding the dryline. As described above, this behavior was also noted with cold fronts identified by FCM, however there was no dryline predicted by FCM. We are unsure as to why saliency is low along the WPC dryline. The WPC analysis for this case study (not shown) shows convergence along the dryline near its intersection

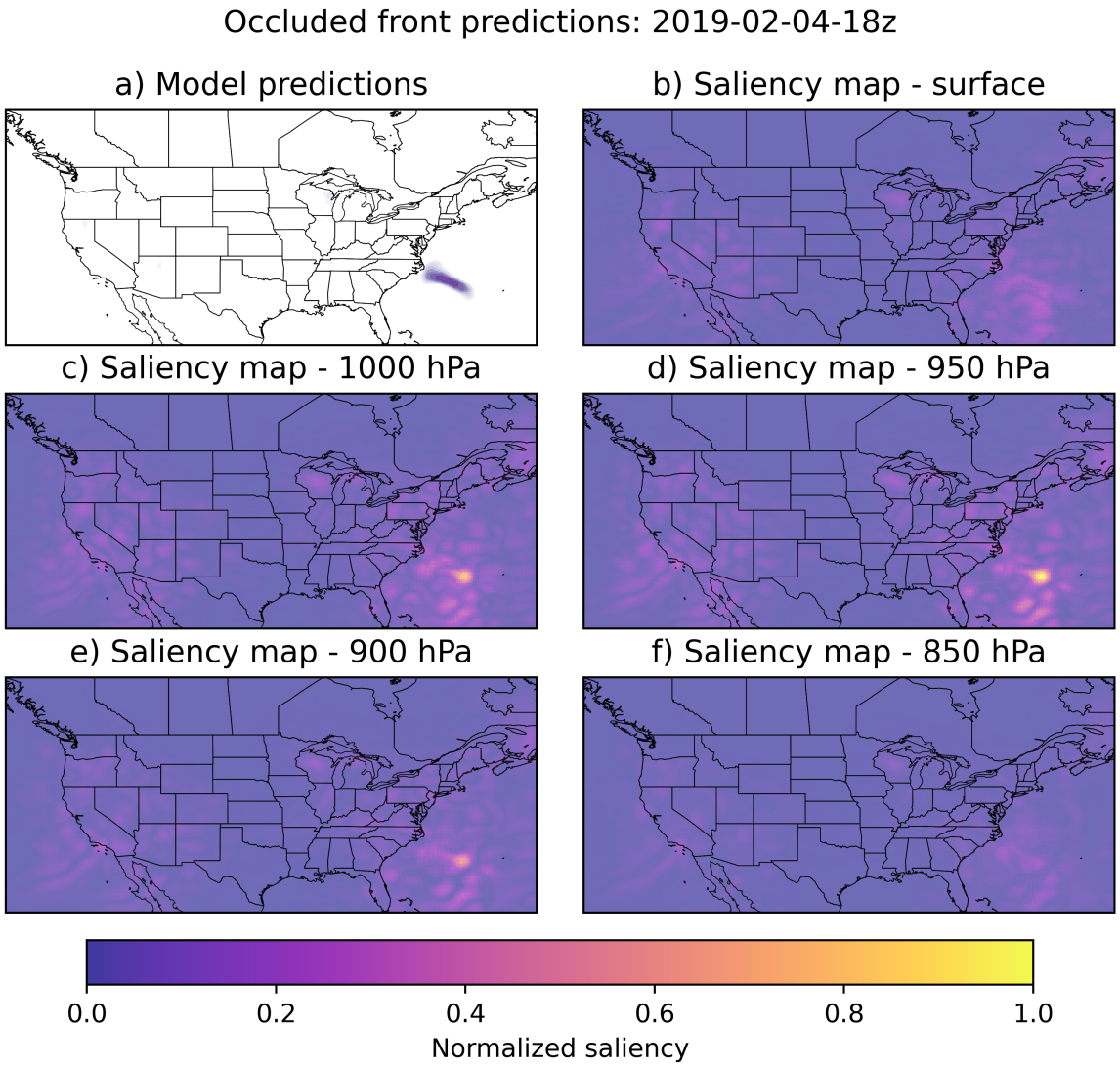


Figure 4.31: Same as Figure 4.28 but for occluded fronts.

with the cold front in Northern Texas and Southern Oklahoma, however the moisture gradients appear relatively weak, with surface dewpoints only changing by ∼10°F across the dryline analyzed by WPC. No prominent temperature gradients were present. We believe that this result may reflect the ∼25 km resolution of the ERA5 data not being sufficient to capture the convergence along the small portion of the dryline near the Red River. We think that the high saliency surrounding the dryline is likely related

to the weak moisture gradients - perhaps increasing the magnitude of the moisture gradient along the dryline could increase FCM dryline probabilities in the region.

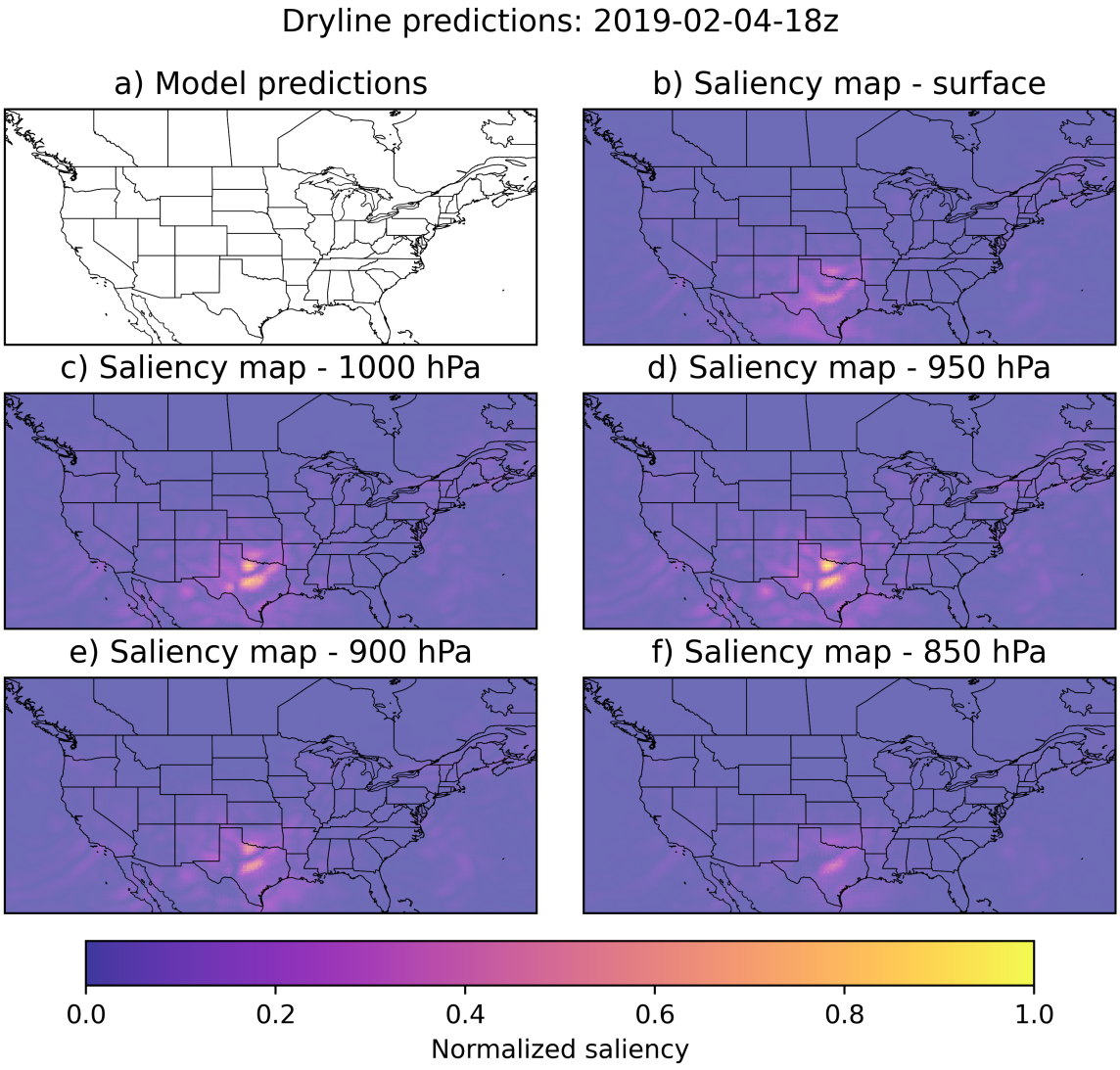


Figure 4.32: Same as Figure 4.28 but for drylines.

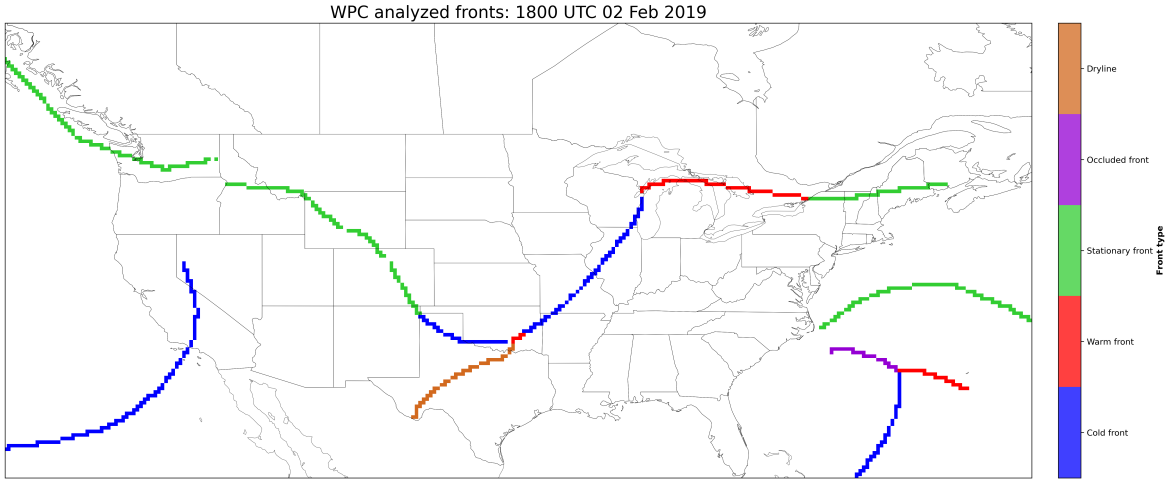


Figure 4.33: Cold, warm, stationary, and occluded front and dryline positions in the WPC analysis for 1800 UTC Feb 4 2019.

Chapter 5

Conclusions and Future Work

Results within this thesis have shown that this new "five-class" model (FCM) can effectively detect cold, warm, stationary, and occluded fronts and drylines over CONUS and NOAA's Unified Surface Analysis domain. This model is a significant improvement over our three-model system (TMS) highlighted in J23, with greater fractions of fronts detected and lower false alarm rates. FCM outperforms TMS with cold, warm, stationary, and occluded fronts at the five neighborhoods assessed (50, 100, 150, 200, 250 km), and all performance improvements were statistically significant at the 95% confidence level. In terms of trainable parameters, FCM is 25 times smaller than each of the models in TMS, which shows that more parameters do not always result in better performance. Combining all frontal types into a single model and the addition of drylines is likely a contributing factor to the performance improvements; FCM is able to learn relationships and discern between all the frontal types, whereas TMS was unable to directly discern between certain frontal types because cold and warm fronts were separated from stationary and occluded fronts into their own models. Unfortunately, stationary front false alarm rates are still very high and the Intertropical Convergence Zone (ITCZ) is still being identified by FCM as a stationary front, which we may be able to resolve by including the ITCZ inside the training domain. Wet-bulb temperature and θ_w were not included in the list of predictors for FCM as permutation studies for TMS showed that these variables contributed very little to model performance, however since FCM is just one model it is unclear if the exclusion of these variables

had any impact on FCM's performance. Assessing the true effect of wet-bulb temperature and θ_w on FCM's ability to detect fronts will require the training of a separate UNET3+ architecture. Permutation studies performed on FCM were consistent with those performed on TMS and showed that wind, temperature, and pressure and moisture variables contributed to FCM's performance, however some differences were noted, particularly with stationary fronts. FCM permutation studies indicated that v-wind negatively impacted stationary front performance, whereas TMS permutation studies from the model predicting stationary and occluded fronts suggested that v-wind was an important predictor for stationary fronts. The cause(s) behind this discrepancy are not known, however we believe that results from permutation studies conducted on FCM are more representative of variables' true contributions to front detection since FCM is able to directly learn relationships between the five frontal types it was trained to predict. We think that accounting for correlations between the input variables through greater data augmentation could improve the performance of future model architectures. Analysis of FCM predictions with saliency maps shows that model behavior can be directly related to physical processes observed in the real atmosphere. While the saliency maps cannot explain all model behaviors, combining them with permutation studies gave some insight as to what regions of the domain could greatly affect FCM's interpretation of the underlying frontal types.

We achieved our goal of operationalizing FCM as it is now used operationally at WPC. Over the coming months we hope to get constructive feedback on how effectively forecasters are able to use our new algorithm to improve identification and classification of frontal boundaries. This feedback will help us make important decisions about how we structure and train future algorithms for identifying fronts. We are also collaborating with forecasters at The Weather Company (TWC) to assist them in drawing fronts globally out to 120 hours lead-time, a labor-intensive process that motivated

TWC to adopt our FCM algorithm and use real-time Global Forecast System (GFS) and European Centre for Medium-Range Weather Forecasts (ECMWF) model data as input to FCM to highlight frontal zones across the globe. While no performance statistics have been calculated globally due to the lack of available ground truth front labels, initial impressions of performance in the Southern Hemisphere are poor, suggesting that we need a larger domain of training data in order to resolve fronts outside of USAD. We hope to eventually develop an algorithm that is able to resolve fronts across the globe with spatially-uniform performance; such could enable us to perform long-term climatologies of different types of frontal boundaries across the globe.

Our future work will include training new model architectures on high-resolution model data such as the North American Mesoscale Model (NAM), High Resolution Rapid Refresh (HRRR) model, and the Rapid Refresh Forecast System (RRFS) model that is currently under development. HRRR and RRFS are convective-allowing models on 3-km grids, so small-scale convective feedbacks will likely complicate the training of future algorithms with HRRR and RRFS data as input. However, a front detection algorithm trained on high-resolution data could have benefits over ERA5 reanalysis, such as being able to precisely locate boundaries that can influence storm-scale processes like tornadogenesis (e.g., Maddox et al., 1980; Markowski et al., 1998; Sills et al., 2004; Childs and Schumacher, 2019). With the exponential growth of machine learning in meteorological research (Chase et al., 2022), future high-resolution deep learning models could also be used to train other model architectures to predict fronts as many recently developed machine learning algorithms have been able to outperform traditional Numerical Weather Prediction (NWP) at forecasting precipitation and variables such as temperature, specific humidity, wind speed, and geopotential (e.g., Sønderby et al., 2020; Bi et al., 2022; Andrychowicz et al., 2023; Zhang et al., 2023). Permutation studies like those performed in this thesis could be used to increase our understanding

of fronts and improve existing NWP models, eventually providing the basis for better reanalysis datasets that can be used to train future machine learning models.

ERA5 data on single and pressure levels was 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.

Reference List

Andrychowicz, M., L. Espeholt, D. Li, S. Merchant, A. Merose, F. Zyda, S. Agrawal, and N. Kalchbrenner, 2023: Deep learning for day forecasts from sparse observations. arXiv, https://doi.org/10.48550/arXiv.2306.06079.

Berry, G., M. J. Reeder, and C. Jakob, 2011: A global climatology of atmospheric fronts. *Geophysical Research Letters*, **38**.

Bi, K., L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, 2022: Pangu-Weather: A 3D high-resolution model for fast and accurate global weather forecast. arXiv, https://doi.org/10.48550/arXiv.2211.02556.

Biard, J. C., and K. E. Kunkel, 2019: Automated detection of weather fronts using a deep learning neural network. *Advances in Statistical Climatology, Meteorology and Oceanography*, **5 (2)**, 147–160.

Bjerknes, J., 1919: On the structure of moving cyclones. *Monthly Weather Review*, **47**, 95–99.

Bjerknes, J., and H. Solberg, 1922: Life cycle of cyclones and the polar front theory of atmospheric circulation. 1-61 pp.

Bochenek, B., Z. Ustrnul, A. Wypych, and D. Kubacka, 2021: Machine learning-based front detection in Central Europe. *Atmosphere*, **12 (10)**.

Breiman, L., 2001: Random forests. *Machine learning*, **45**, 5–32.

Bridle, J., 1989: Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. *Advances in Neural Information Processing Systems*, D. Touretzky, Ed., Morgan-Kaufmann, Vol. 2.

Browning, K. A., 1997: The dry intrusion perspective of extra-tropical cyclone development. *Meteorological Applications*, **4 (4)**, 317–324.

Catto, J. L., and S. Pfahl, 2013: The importance of fronts for extreme precipitation. *Journal of Geophysical Research: Atmospheres*, **118**, 10,791–10,801.

Chase, R. J., D. R. Harrison, A. Burke, G. M. Lackmann, and A. McGovern, 2022: A machine learning tutorial for operational meteorology. Part I: Traditional machine learning. *Weather and Forecasting*, **37 (8)**, 1509–1529.

Childs, S. J., and R. S. Schumacher, 2019: An updated severe hail and tornado climatology for Eastern Colorado. *Journal of Applied Meteorology and Climatology*, **58**, 2273–2293.

Clark, A. J., A. MacKenzie, A. McGovern, V. Lakshmanan, and R. A. Brown, 2015: An automated, multiparameter dryline identification algorithm. *Weather and Forecasting*, **30 (6)**, 1781–1794.

Clarke, L. C., and R. J. Renard, 1966: The U.S. Navy numerical frontal analysis scheme: Further development and a limited evaluation. *Journal of Applied Meteorology and Climatology*, **5 (6)**, 764–777.

Dagon, K., J. Truesdale, J. C. Biard, K. E. Kunkel, G. A. Meehl, and M. J. Molina, 2022: Machine learning-based detection of weather fronts and associated extreme precipitation in historical and future climates. *J. Geophys. Res.*, **127 (21)**.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137 (656)**, 553–597.

Denker, J., and C. Burges, 1995: *The Mathematics Of Generalization*, chap. Image Segmentation and Recognition. CRC Press.

Ganetis, S. A., B. A. Colle, S. E. Yuter, and N. P. Hoban, 2018: Environmental conditions associated with observed snowband structures within Northeast U.S. winter storms. *Monthly Weather Review*, **146 (11)**, 3675–3690.

Gelaro, R., and Coauthors, 2017: The modern-era retrospective analysis for research and applications, Version 2 (MERRA-2). *Journal of Climate*, **30 (14)**, 5419–5454.

Hendrycks, D., and K. Gimpel, 2023: Gaussian Error Linear Units (GELUs). arXiv, https://doi.org/10.48550/arXiv.1606.08415.

Hersbach, H., and Coauthors, 2018: ERA5 hourly data on pressure levels from 1959 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), (Accessed on 15-MAR-2021).

Hewson, T. D., 1998: Objective fronts. *Meteorological Applications*, **5**, 37–65.

Hines, K. M., and C. R. Mechoso, 1993: Influence of surface drag on the evolution of fronts. *Monthly Weather Review*, **121 (4)**, 1152–1176.

Hinton, G. E., N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, 2012: Improving neural networks by preventing co-adaptation of feature detectors. arXiv, https://doi.org/10.48550/arXiv.1207.0580.

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. *Weather and Forecasting*, **5**, 613–626, https://doi.org/10.1175/1520-0434(1990)005¡0613:CFAATF¿2.0.CO;2.

Hoch, J., and P. Markowski, 2005: A climatology of springtime dryline position in the U.S. great plains region. *J. Clim.*, **18 (12)**, 2132–2137.

Hope, P., and Coauthors, 2014: A comparison of automated methods of front recognition for climate studies: A case study in Southwest Western Australia. *Monthly Weather Review*, **142**, 343–363.

Hosek, M., 2024: A 15-year climatology of WPC-analyzed drylines, OU Weather and Climate Seminar, Spring 2024.

Huang, H., and Coauthors, 2020: UNet 3+: A full-scale connected UNet for medical image segmentation. arXiv, https://doi.org/10.48550/arxiv.2004.08790.

Ioffe, S., and C. Szegedy, 2015: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv, https://doi.org/10.48550/arXiv.1502.03167.

Janiesch, C., P. Zschech, and K. Heinrich, 2021: Machine learning and deep learning. *Electronic Markets*, **31 (3)**, 685–695.

Justin, A. D., C. Willingham, A. McGovern, and J. T. Allen, 2023: Toward operational real-time identification of frontal boundaries using machine learning. *Artificial Intelligence for the Earth Systems*, **2**, e220 052.

Kalnay, E., and Coauthors, 2018: The NCEP/NCAR 40-year reanalysis project. *Renewable energy*, Vol. 1, Routledge, 146–194.

Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. arXiv, https://doi.org/10.48550/arxiv.1412.6980.

Kohonen, T., 1982: Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, **43 (1)**, 59–69.

Krause, J., B. Sapp, A. Howard, H. Zhou, A. Toshev, T. Duerig, J. Philbin, and L. Fei-Fei, 2016: The unreasonable effectiveness of noisy data for fine-grained recognition. *Computer Vision – ECCV 2016*, Springer International Publishing, 301–320.

Kunkel, K. E., D. R. Easterling, D. A. R. Kristovich, B. Gleason, L. Stoecker, and R. Smith, 2012: Meteorological causes of the secular variations in observed extreme precipitation events for the conterminous United States. *Journal of Hydrometeorology*, **13 (3)**, 1131–1141.

Lagerquist, R., J. T. Allen, and A. McGovern, 2020: Climatology and variability of warm and cold fronts over North America from 1979 to 2018. *Journal of Climate*, **33**, 6531–6554.

Lagerquist, R., A. McGovern, and D. J. Gagne II, 2019: Deep learning for spatially explicit prediction of synoptic-scale fronts. *Weather and Forecasting*, **34**, 1137–1160.

LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, 1989: Backpropagation applied to handwritten zip code recognition. *Neural Computation*, **1 (4)**, 541–551.

Lederer, J., 2021: Activation functions in artificial neural networks: A systematic overview. arXiv, https://doi.org/10.48550/arXiv.2101.09957.

Lee, M., 2023: Gelu activation function in deep learning: A comprehensive mathematical analysis and performance. arXiv, https://doi.org/10.48550/arXiv.2305.12073.

Lu, L., 2020: Dying ReLU and initialization: Theory and numerical examples. *Communications in Computational Physics*, **28 (5)**, 1671–1706.

Maddox, R. A., L. R. Hoxit, and C. F. Chappell, 1980: A study of tornadic thunderstorm interactions with thermal boundaries. *Monthly Weather Review*, **108**, 322–336.

Market, P. S., and J. T. Moore, 1998: Mesoscale evolution of a continental occluded cyclone. *Monthly Weather Review*, **126 (7)**, 1793–1811.

Markowski, P. M., E. N. Rasmussen, and J. M. Straka, 1998: The occurrence of tornadoes in supercells interacting with boundaries during VORTEX-95. *Weather and Forecasting*, **13 (3)**, 852–859.

Martin, J. E., 1999: The separate roles of geostrophic vorticity and deformation in the midlatitude occlusion process. *Monthly Weather Review*, **127 (10)**, 2404–2418.

Matsuoka, D., S. Sugimoto, Y. Nakagawa, S. Kawahara, F. Araki, Y. Onoue, M. Iiyama, and K. Koyamada, 2019: Automatic detection of stationary fronts around Japan using a deep convolutional neural network. *Scientific Online Letters of the Atmosphere*, **15**, 154–159.

McGovern, A., R. Lagerquist, D. J. Gagne, 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. *Bulletin of the American Meteorological Society*, **100**, 2175–2199.

Mesinger, F., and Coauthors, 2006: North American regional reanalysis. *Bulletin of the American Meteorological Society*, **87 (3)**, 343–360.

Mitchell, K. E., and J. B. Hovermale, 1977: A numerical investigation of the severe thunderstorm gust front. *Monthly Weather Review*, **105 (5)**, 657–675.

National Weather Service, 2013: Unified surface analysis manual. National Weather Service Analysis Centers, https://www.wpc.ncep.noaa.gov/sfc/UASfcManualVersion1.pdf, 33 pp.

National Weather Service, 2019: National Weather Service coded surface bulletins, 2003-. Zenodo, 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. *Weather and Climate Dynamics*, **3**, 113–137.

NOAA, 2023: NOAA unified surface analysis fronts. Zenodo, https://doi.org/10.5281/zenodo.7505022.

O'Shea, K., and R. Nash, 2015: An introduction to convolutional neural networks. arXiv, https://doi.org/10.48550/arxiv.1511.08458.

Parsons, D. B., M. A. Shapiro, R. M. Hardesty, R. J. Zamora, and J. M. Intrieri, 1991: The finescale structure of a West Texas dryline. *Monthly Weather Review*, **119 (5)**, 1242–1258.

Parsons, D. B., M. A. Shapiro, and E. Miller, 2000: The mesoscale structure of a nocturnal dryline and of a frontal–dryline merger. *Monthly Weather Review*, **128 (11)**, 3824–3838.

Philipp, G., D. Song, and J. G. Carbonell, 2018: The exploding gradient problem demystified - definition, prevalence, impact, origin, tradeoffs, and solutions. arXiv, https://doi.org/10.48550/arXiv.1712.05577.

Pietrycha, A. E., and E. N. Rasmussen, 2004: Finescale surface observations of the dryline: A mobile mesonet perspective. *Weather Forecast.*, **19 (6)**, 1075–1088.

Refaeilzadeh, P., L. Tang, and H. Liu, 2009: Cross-Validation. *Encyclopedia of Database Systems*, L. Liu, and M. T. Özsu, Eds., Springer US, Boston, MA, 532–538.

Renard, R. J., and L. C. Clarke, 1965: Experiments in numerical objective front analysis. *American Meteorological Society*, **93**, 547–556.

Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteorological Applications*, **15**, 163–169.

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Springer International Publishing, Cham, 234–241.

Schaefer, J. T., 1973: The motion and morphology of the dryline. NOAA Tech. Memo. ERL NSSL-66, 81 pp.

Schaefer, J. T., 1986: *The Dryline*, 549–572. American Meteorological Society, Boston, MA, https://doi.org/10.1007/978-1-935704-20-1.

Schemm, S., I. Rudeva, and I. Simmonds, 2015: Extratropical fronts in the lower troposphere–global perspectives obtained from two automated methods. *Quarterly Journal of the Royal Meteorological Society*, **141**, 1686–1698.

Schultz, D. M., 2005: A review of cold fronts with prefrontal troughs and wind shifts. *Monthly Weather Review*, **133**, 2449–2472.

Schultz, D. M., B. Antonescu, and A. Chiariello, 2014: Searching for the elusive cold-type occluded front. *Monthly Weather Review*, **142 (8)**, 2565–2570.

Schultz, D. M., and C. F. Mass, 1993: The occlusion process in a midlatitude cyclone over land. *Monthly Weather Review*, **121 (4)**, 918–940.

Schultz, D. M., and G. Vaughan, 2011: Occluded fronts and the occlusion process: A fresh look at conventional wisdom. *Bulletin of the American Meteorological Society*, **92**, 443–466.

Shafer, J. C., W. J. Steenburgh, J. A. W. Cox, and J. P. Monteverdi, 2006: Terrain influences on synoptic storm structure and mesoscale precipitation distribution during IPEX IOP3. *Monthly Weather Review*, **134 (2)**, 478–497.

Shapiro, M. A., and D. Keyser, 1990: *Extratropical Cyclones, The Erik Palmén Memorial Volume*. Amer. Meteor. Soc., 167-191 pp.

Shen, R., L. Gao, and Y.-A. Ma, 2022: On optimal early stopping: Over-informative versus under-informative parametrization. arXiv, https://doi.org/10.48550/arXiv.2202.09885.

Sills, D. M. L., J. W. Wilson, P. I. Joe, D. W. Burgess, R. M. Webb, and N. I. Fox, 2004: The 3 November tornadic event during Sydney 2000: Storm evolution and the role of low-level boundaries. *Weather and Forecasting*, **19 (1)**, 22–42.

Simmonds, I., K. Keay, and J. A. T. Bye, 2012: Identification and climatology of Southern Hemisphere mobile fronts in a modern reanalysis. *Journal of Climate*, **25**, 1945–1962.

Simonyan, K., A. Vedaldi, and A. Zisserman, 2014: Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv, https://doi.org/10.48550/arXiv.1312.6034.

Simpson, J. E., 1972: Effects of the lower boundary on the head of a gravity current. *Journal of Fluid Mechanics*, **53 (4)**, 759–768.

Smith, R., and M. Reeder, 1988: On the movement and low-level structure of cold fronts. *Monthly Weather Review*, **116**, 1927–1944.

Steenburgh, W. J., C. F. Mass, and S. A. Ferguson, 1997: The influence of terrain-induced circulations on wintertime temperature and snow level in the washington cascades. *Weather and Forecasting*, **12 (2)**, 208–227.

Sønderby, C. K., and Coauthors, 2020: MetNet: A neural weather model for precipitation forecasting. arXiv, https://doi.org/10.48550/arXiv.2003.12140.

Terven, J., D. M. Cordova-Esparza, A. Ramirez-Pedraza, and E. A. Chavez-Urbiola, 2023: Loss functions and metrics in deep learning. arXiv, https://doi.org/10.48550/arXiv.2307.02694.

Thomas, C. M., and D. M. Schultz, 2019: Global climatologies of fronts, airmass boundaries, and airstream boundaries: Why the definition of "front" matters. *Monthly Weather Review*, **147**, 691–717.

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. *Bulletin of the American Meteorological Society*, **73**, 459–472.

Uppala, S. M., and Coauthors, 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131 (612)**, 2961–3012.

Wakimoto, R. M., and B. L. Bosart, 2001: Airborne radar observations of a warm front during FASTEX. *Monthly Weather Review*, **129**, 254–274.

Yang, S., W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, 2023: Image data augmentation for deep learning: A survey. arXiv, https://doi.org/10.48550/arXiv.2204.08610.

Young, G. S., and R. H. Johnson, 1984: Meso- and microscale features of a Colorado cold front. *Journal of Applied Meteorology and Climatology*, **23 (9)**, 1315–1325.

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. *Monthly Weather Review*, **127**, 1725–1744, https://doi.org/10.1175/1520-0493(1999)127¡1725:AFOFCO¿2.0.CO;2.

Zhang, Y., M. Long, K. Chen, L. Xing, R. Jin, M. I. Jordan, and J. Wang, 2023: Skillful nowcasting of extreme precipitation with NowcastNet. *Nature*, **619 (7970)**, 526–532.

Ziegler, C. L., and C. E. Hane, 1993: An observational study of the dryline. *Monthly Weather Review*, **121 (4)**, 1134–1151.