THESIS

CSU-MLP GEFS DAY-1 "FIRST-GUESS" EXCESSIVE RAINFALL FORECASTS: AGGREGATE EVALUATION AND SYNOPTIC REGIMES OF BEST- AND WORST-PERFORMING FORECASTS

Submitted by:

Jacob A. Escobedo

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2022

Master's Committee:

Advisor: Russ Schumacher

Susan van den Heever
Daniel Cooley

Copyright by Jacob A. Escobedo 2022

All Rights Reserved

ABSTRACT

CSU-MLP GEFS DAY-1 "FIRST-GUESS" EXCESSIVE RAINFALL FORECASTS: AGGREGATE EVALUATION AND SYNOPTIC REGIMES OF BEST- AND WORST-PERFORMING FORECASTS

Forecasting excessive rainfall, particularly flash flood-producing rainfall, is an important problem that remains difficult due to the small spatial scales and varying temporal scales at which they occur. One important operational product that highlights areas for potential excessive rainfall and flash flood occurrences is the Excessive Rainfall Outlook (ERO) issued by the NOAA Weather Prediction Center (WPC), which provides outlooks for lead times of 1-3 days. To address the need for additional tools for WPC forecasters while forming a given ERO, the Colorado State University Machine Learning Probabilities (CSU-MLP) system, a probabilistic forecast system for excessive rainfall (and other convective hazards), was developed to produce forecasts to be used as a "first-guess" ERO. CSU-MLP employs the use of a random forest (RF) algorithm trained using NOAA's Second-Generation Global Ensemble Forecast System Reforecast (GEFS/R) and precipitation observations, while using the operational GEFS with the trained model to produce real-time forecasts. Initially developed as a medium range guidance (2-3 day lead time), CSU-MLP has produced day-1 forecasts that have been evaluated favorably during the 4-week Flash Flood and Intense Rainfall Experiment (FFaIR) in the summer of 2020. However, CSU-MLP day-1 forecasts have been observed to have daily forecast skill that can vary widely between days. This work will include an aggregate evaluation of CSU-MLP day-1 forecasts over a longer period of study (3 March 2019 – 15 October 2020) than that of FFaIR, and an identification of synoptic regimes for which these forecasts tend to perform at their best and worst.

Results show that CSU-MLP day-1 forecasts are reliable, provide adequate discrimination of excessive rainfall events (AuROC =0.819), and have comparable performance, evaluated by use of the

Brier skill score (BSS), to that of the ERO (CSU-MLP BSS = 0.081; ERO BSS = 0.085). However, CSU-MLP forecasts have a higher frequency of categorical probabilities (≥ 0.05) which results in larger variations of daily BSS. Synoptic regimes of best-performing daily forecasts reveal a tendency for these regimes to be characterized by moderate to strong large-scale forcing and relatively high low-level and column moisture. This would include warm-season regimes with moderate amplitude upper-level troughs, tropical cyclones, cut-off lows, and cool-season regimes where strong forcing is co-located near an abundant moisture source. Forecasts tend to perform worst when there is strong large-scale forcing and low-level and column moisture is relatively low, such as cool-season regimes with large amplitude troughs and surface cyclones but higher levels of atmospheric moisture are not present nor as widespread. This work has implications for WPC forecasters as they use the "first-guess" forecasts while developing the ERO for a given day, as well as implications for future CSU-MLP system model iterations and/or designs.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Excessive rainfall and flash flooding events can result in a loss of life, injury, and damage to property, infrastructure, and agriculture. From 1991-2020, flooding resulted in 85 fatalities per year in the United States, more than from any other convective weather hazard (NOAA 2022b). Over this same time period, the top 28 costliest flooding events resulted in an estimated $144.4 billion in costs (NOAA 2022a). These statistics emphasize the importance of improving forecasting for flash floods to help mitigate the costly effects of these events.

One of the first studies that identified general characteristics of flash flood-producing systems was that of Maddox et al. (1979). While distinguishing systems by different synoptic and mesoscale characteristics, they found several commonalities between the cases they analyzed:

- Heavy rains were produced by convective storms
- Surface dewpoint temperatures were very high (> 60° F)
- Large moisture contents were present through a deep tropospheric layer (less than roughly 5°F difference between temperature and dew point temperature at different layers)
- Vertical wind shear was weak to moderate through the cloud depth (15-25 knots)

Building upon these findings, Doswell et al. (1996) developed an "ingredients-based" approach for forecasting excessive precipitation using the following simple equation:

$$P = \bar{R}D \tag{1.1}$$

where P is total precipitation, $\bar{R}$ is the average rain rate, and D is the duration of the precipitation. $\bar{R}$ can be equated as such:

$$\bar{R} = Ewq \tag{1.2}$$

where E is precipitation efficiency, w is vertical velocity, and q is water vapor mixing ratio. This involves identifying areas for potentially high rain rates, and thus, ingredients for deep moist convection: high low-level moisture, steep mid-level lapse rates, lifting mechanisms for convective initiation. Increased moisture through the depth of the troposphere will also support greater E and higher $\bar{R}$, as less evaporation occurs due to both entrainment of environmental air at edges of the cloud and sub-cloud evaporation of precipitation. This means areas of high relative humidity throughout the troposphere is another ingredient for excessive rainfall events. Lastly, a long-lasting rainfall event will increase P, where D is function of the size, shape, orientation, and movement of a precipitation system (Doswell 1996; Schumacher and Johnson 2005). This not only involves identifying areas with long duration low-level forcing and horizontal moisture transport for initiation and maintenance of convection, but also identifying areas where convection remains nearly stationary or repeatedly traverses a location.

With ingredients for excessive rainfall in mind, forecasters must be able to recongnize the synoptic pattens that result in extreme rainfall and flash flooding. Maddox et al. (1979) identified some of the typical synoptic and mesoscale patterns that were associated with the flash floods in the CONUS. Among those are the "synoptic" and "frontal" patterns. "Synoptic" events (Figure 1.1; from Schumacher 2017) are characterized by a slow moving upper-level trough with an associated quasi-stationary surface front. Ahead of the trough, there is a southerly low-level jet (LLJ) that results in northward advection of

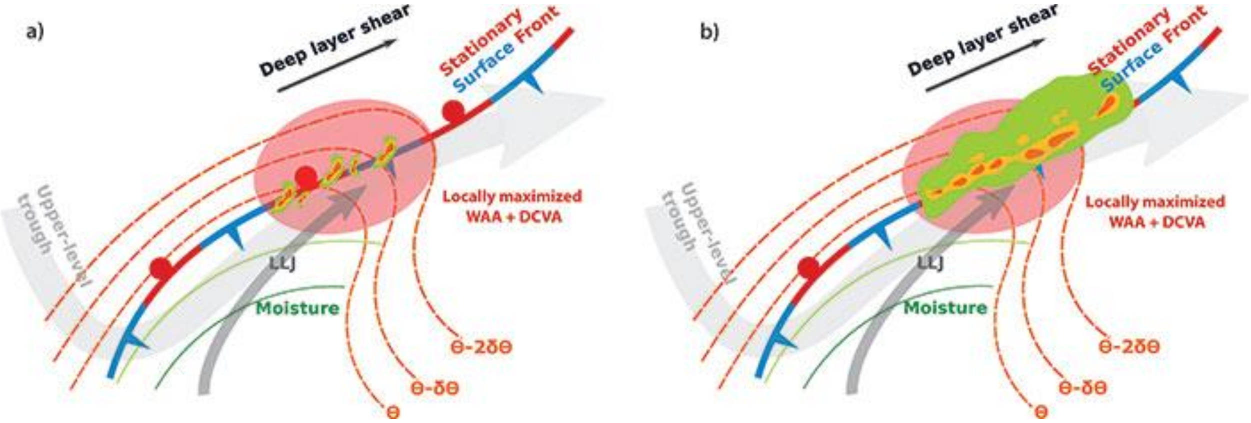


Figure 1.1. Illustration of Maddox et al. (1979) synoptic flash flood events from Figure 1 of Schumacher (2017), where a) depicts the environmental conditions during convective initiation and b) depicts conditions at a later time when convection as evolved into an MCS.

moisture and warmer temperatures roughly parallel to the surface front. Shaded in red are the areas where forcing for ascent is maximized due to differential cyclonic vorticity advection (DCVA) and low-level warm air advection (WAA). It is at this location that convection is typically initiated and grows upscale into a mesoscale convective system (MCS) within the warm sector. Due to the typical slow evolution of the synoptic environnment, these events can last multiple days and affect a broad region. "Frontal" events (Figure 1.2; from Schumacher 2017) are characterized by a frontal boundary (usually an east-to-west orientated warm front or stationary front) where a southerly low-level jet (LLJ) flows over the front with locally maximized forcing for ascent (shaded in red in Figure 1.2a,b) via WAA and frontogenesis. This results in convective initation on the cool side of the front. Often there is an upper-level ridge axis near convective inititation, with a weak mid-level shortwave upstream of maximum forcing and mid- to upper-level winds that flow nearly parallel to the surface boundary, resulting in a system where new convection is continuosly initiated upstream of mature convection.

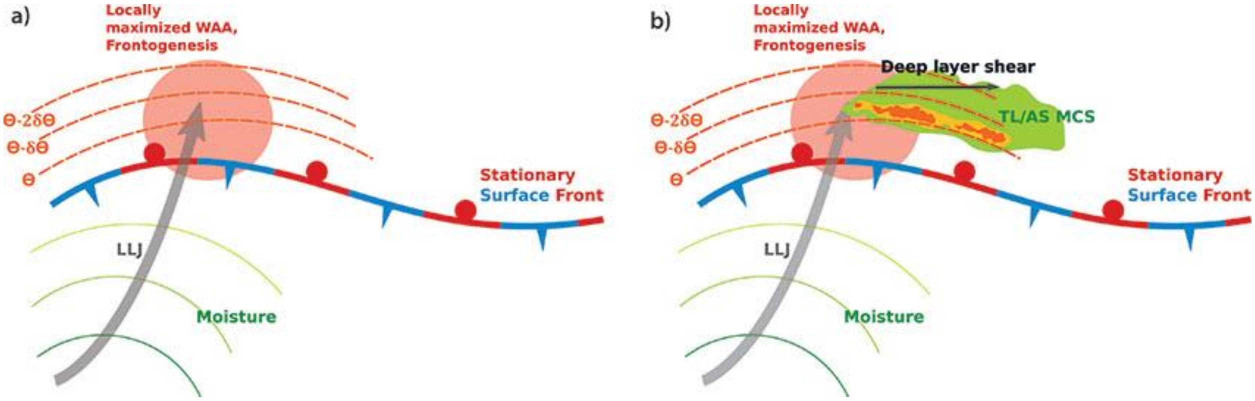


Figure 1.2. Illustration of Maddox et al. (1979) frontal flash flood events from Figure 2 of Schumacher (2017), where a) depicts environmental conditions preceding convection and b) depicts conditions at later time where an MCS has formed on the cool side of the front.

Additional studies have investigated and categorized the synoptic regimes for excessive rainfall and/or flash floods within the CONUS, often looking at various regional domains (or sub-regions within) such as the southeast (Konrad 1997; Moore et al. 2015), the northeast (LaPenta et al. 1995; Marquardt

Collow et al. 2016; Agel et al. 2019), the central CONUS (Bradley and Smith 1994; Lavers and Villarini 2013; Mullens 2021), east of the Rockies (Schumacher and Johnson 2005; Peters and Schumacher 2014), the west CONUS (Maddox et al. 1980; Ely et al. 1994; Warner et al. 2012; Neimann et al. 2013), and the entire CONUS (Kunkel et al. 2012; Barlow et al. 2019). Despite the various definitions of excessive rainfall used to identify event days, the different regions of the CONUS examined, and the different pattern identification techniques used, most identify synoptic regimes east of the Rockies that resemble those of the "synoptic" events and "frontal" events identified by Maddox et al. (1979. Many of these studies also categorize additional synoptic regimes that result in extreme rainfall. For example, extratropical cyclones (ETCs) are a common regime category east of the Rockies (e.g. Konrad 1997; LaPenta et al. 1997; Kunkel et al. 2012; Mullens 2021), where excessive rainfall is observed in close proximity to surface cyclones. Mullens (2021) also distinguishes regimes by considering forcing by fronts that are well removed from the parent surface cyclone. Tropical cyclones (TCs) and their inland remnants are another flood-producing identified regime (e.g. Schumacher and Johnson 2005; Kunkel et al. 2012; Mullens 2021), as their high moisture contents, strong dynamics, large size, and occasionally slow speed provide the necessary ingredients for extreme rainfall. Across the western CONUS, extreme rainfall producing synoptic regimes are dominated by the interaction of landfalling (and sometimes inland penetrating) atmospheric rivers (AR) and orography in the cool-season (e.g. Ely et al. 1994; Neiman et al. 2012; Rivera et al. 2014; Warner et al. 2012), and warm-season monsoon regimes associated with anomalous moisture and shortwaves occurring around the large-scale upper-level ridge typical across the region (e.g. Adams and Comrie 1997; Maddox et al. 1980).

While ingredients for excessive rainfall are brought together by the synoptic conditions, extreme rainfall that produces flash floods typically occurs from MCSs which form and evolve as direct result of cold pool interactions with the surrounding environment. Maddox et al. (1979) identified these events as “mesohigh” events. MCSs can take on several different archetypes (Parker and Johnson 2000), however, only a few are commonly responsible for flash flood-producing rainfall (Schumacher and Johnson 2005, Keene and Schumacher 2013). This means that forecasters must be able to anticipate the formation of MCSs and their structural evolution, in addition to synoptic regime recognition, to better identify regions most at risk for excessive rainfall for a given day. To add to this complexity, flash floods also depend on the relevant hydrology (e.g. Viterbo et. al 2020). This includes consideration of antecedent soil conditions, basin size, distribution of rainfall within a basin, soil type, land use, and terrain slope, to name a few.

Although the overall environments and ingredients for excessive rainfall are generally well understood and can sometimes be identified days in advance, forecasting these events with precision remains a challenging problem due to the relatively small spatial and varying temporal scales at which excessive rainfall occurs. Forecasters must rely on numerical weather prediction (NWP) model output (especially at longer lead times) to construct quantitative precipitation forecasts (QPFs). However, NWP has limited predictability of convective systems (e.g. Melhauser and Zhang, 2012; Nielsen and Schumacher 2016) as well as observed biases for excessive rainfall (Herman and Schumacher 2016). As a result, accurate QPF for heavy rainfall in the warm season has been challenging (Fritsch and Carbone 2004; Sukovich et al. 2014). Nonetheless, forecaster QPF has consistently improved and shown more skill compared to NWP (Novak et al. 2014). In recent years, the advent of convection-allowing models (CAMs) as well as statistical postprocessing methods (e.g. Hamill et al. 2015; Hamill and Scheuerer 2018; Whan and Schmeits 2018; Loken et al. 2019) have resulted in a plethora of additional tools to be used to enhance QPF forecasting.

One important operational product that highlights areas for potential excessive rainfall and flash flood occurrences within the CONUS is the Excessive Rainfall Outlook (ERO; Erickson et al. 2021) issued by the NOAA Weather Prediction Center (WPC). These outlooks provide forecasts, at days 1-3 lead times, for the probabilities that rainfall will exceed Flash Flood Guidance (FFG; Schmidt et al. 2007) within 40 kilometers of a point of interest. FFG is created by the River Forecast Centers and is an estimate of rainfall over a given time duration that may cause small streams to flood when the stream is at bankfull. Probabilities are displayed through four categories of risk: Marginal (MRGL; 5-10%), Slight (SLGT; 10-20%), Moderate (MDT, 20-50%), and High (HIGH; >50%). WPC issues three EROs around 0900 UTC every day (with periodic updates), each valid 1200UTC – 1200 UTC for a given day 1-3 forecast period (corresponding to 3-27-, 27-51-, and 51-75-hour forecast times). Among some of the challenges of QPF forecasting mentioned in the previous paragraph, another challenge is that NWP and other postprocessing methods do not provide direct information about whether QPF is excessive for a given location. To address these challenges, a forecast system – the Colorado State University Machine Learning Probabilities (CSU-MLP) system- based on NWP model reforecasts, historical observations of excessive rainfall, and machine-learning algorithms- was developed to provide WPC forecasters with guidance to be used as a "first-guess" when developing their ERO (Herman and Schumacher 2018c). CSU-MLP uses the random forest algorithm (RF; Breimann 2001), which is composed of a set of decision trees that individually make a unique classification prediction based on inputs to the tree. RFs are used because of their capability to potentially correct for model forecast biases, such as QPF being systematically too high or too low, spatial biases in extreme QPF, and/or temporal biases in the initiation and evolution of extreme precipitation features (e.g. Herman and Schumacher 2016).

As described in Schumacher et al. (2021), multiple versions of CSU-MLP have been iteratively improved upon. Initially developed in 2017 (CSU-MLP versions denoted by the year of model development) and designed to forecast for days 2-3 lead times (Herman and Schumacher 2018a,c), alterations to the CSU-MLP system have come from changes to the definition of excessive rainfall. Specifically, this involved the inclusion of flash flood local storm reports (LSRs), the use of the Climatology-Calibrated Precipitation Analysis (CCPA; Hou 2013) instead of Stage IV data, and the use of different 24-hr average recurrence interval (ARI) exceedances for different regions of the CONUS

(Schumacher et al. 2021). Another key addition to the CSU-MLP system was the production of day-1 forecasts from the 2019 and 2020 versions, of which the most recent (2020 version) was objectively and subjectively evaluated at the Flash Flood and Intense Rainfall Experiment (FFaIR; Trojiniak et al. 2020) in the summer of 2020 and has since been incorporated as the most recent CSU-MLP version (along with day-2 and day-3 forecasts) into WPC operations (Schumacher et al. 2021).

Although the CSU-MLP 2020 version (herein, CSU-MLP for brevity) forecasts have aggregate skill comparable to that of the ERO (Schumacher et al. 2021), individual forecast performance has been observed to range widely. It is hypothesized that, similar to NWP forecasts for extreme precipitation, CSU-MLP forecasts perform best on days with strong synoptic forcing associated with large-scale troughs and cyclones and worst on warm-season days where mesoscale and storm-scale interactions largely influence the locations of excessive rainfall and flash flood events. While previous work from Herman and Schumacher (2018a) investigated the spatial and temporal predictors with the most influence on the RF architecture (i.e. feature importance), this work will investigate those synoptic regimes for which CSU-MLP forecasts- specifically day-1 forecasts- perform best and worst.

This thesis is organized as follows: Chapter 2 will provide some additional details about CSU-MLP model training and forecasts. Chapter 3 provides aggregate CSU-MLP forecast evaluation in comparison to the ERO for the period of study, which supplements findings from Schumacher et al. (2021). Chapter 4 will investigate the characteristics of daily forecast skill across different regions and seasons. Chapter 5 shows the different synoptic regimes of best- and worst-performing CSU-MLP day-1 forecasts. Lastly, discussion and conclusions from this work are in Chapter 6.

CHAPTER 2: CSU-MLP DAY-1 GEFS/R RANDOM FOREST MODEL

As introduced in Chapter 1, CSU-MLP is a forecast system that postprocesses NWP forecasts using a RF model to generate excessive rainfall outlooks designed to resemble the WPC ERO. The predictor and predictand datasets used for training the RF model are described in this chapter, along with the generation of real-time forecasts.

2.1 Predictors: GEFS/R Forecasts

Forecasts from NOAA's Second-Generation Global Ensemble Forecast System Reforecast (GEFS/R; Hamill 2013) are used to generate the predictor dataset to train the CSU-MLP Day-1 RF model (herein, RF model for brevity). The GEFS/R is a global, convection-parameterized 11-member ensemble with T254L42 resolution, corresponding to an effective horizontal grid spacing of ~55 km at 40° latitude. GEFS/R forecasts are initialized daily at 0000 UTC dating back to December 1984, with forecast fields every three hours out to 72 hours past initialization and every six hours after. The GEFS/R dataset was chosen for RF model training as the ensemble system used to generate these reforecasts is nearly static throughout its 30+ year period of coverage with only updates to the operational data assimilation system. This benefits RF model training by providing a larger forecast dataset (i.e. more data) as opposed to using other NWP models where more frequent updates limit the amount of forecasts during a static model system.

The predictor dataset is assembled by using the day-1 1200 UTC – 1200 UTC forecast period, corresponding to forecast hours 12-36, which includes nine timesteps of forecast fields to capture temporal variations of these fields. In addition, spatial variations are included in the RF model. This is done by assembling forecast fields in a forecast gridpoint-relative sense, where forecast fields within a radius of three gridpoints are used to train the RF model at a specific gridpoint on the GEFS/R grid. This is fundamentally the same predictor assembly process used in Herman and Schumacher (2018a,c) and Hill and Schumacher (2021). The RF model is trained to learn relationships between these predictors and occurrences of excessive rainfall for each gridpoint of the GEFS/R domain within the CONUS. Specific forecast fields used for the predictor dataset include the ensemble median of QPF, precipitable water, mean sea level pressure, CAPE, CIN, 2-m temperature and relative humidity, 10-m wind components, and vertical wind shear (e.g. Herman and Schumacher 2018a,c; Schumacher et al. 2021). Daily initializations from January 2003 through August 2013 are employed in the training dataset.

2.2 Predictands: Excessive Rainfall from CCPA and Flash Flood Reports

Excessive rainfall is defined, within the framework of RF model, using quantitative precipitation estimate (QPE) exceedances of average recurrence intervals (ARIs), which correspond to the expected duration (given the local climatology) between exceedances of rainfall of a given threshold, and local storm reports of flash floods (LSRs). QPE is used from CCPA which is a combination of Stage IV precipitation analysis (Lin and Mitchell 2005) and the NOAA Climate Prediction Center unified global daily gauge analysis (CPC; Xie et al. 2010). These two precipitation datasets are combined to create a daily 24-hour precipitation analysis that maintains the spatial patterns of Stage IV but statistically adjusted to have a long-term average and climate probability distribution close to that of CPC, which is generally considered more accurate due to rigorous quality control (Hou et al. 2014). Flash flood LSRs are recorded by NOAA National Weather Service Warning Forecast Offices, however, criteria for what to consider a flash flood vary between offices (Clark et al. 2014).

Occurrences of excessive rainfall are defined as CCPA 24-hour precipitation from 1200 UTC – 1200 UTC exceeding 1-yr or 2-yr (ARI), or the occurrence of a flash flood report in the same period. Either of these occurrences that fall within 40km of a GEFS/R gridpoint is considered an excessive rainfall event for a given day and location. These excessive rainfall events and non-events are aligned with their corresponding spatial and temporal day-1 GEFS/R predictors to produce a complete training dataset for each gridpoint on the GEFS/R grid. Each training gridpoint is additionally segregated into eight different training regions (e.g. Herman and Schumacher 2018a,c; Hill and Schumacher 2021; Schumacher et al.

2021), as shown in Figure 2.1. Due to some of the uncertainties in the CCPA/Stage IV datasets, and the varying correspondence of flash floods with different ARIs across different regions (e.g Herman and Schumacher 2018b; Hill and Schumacher 2021), the use of 1-yr or 2-yr ARI exceedances in the predictand dataset varies. Specifically, 1-yr ARI exceedances are used for the NE, NGP, SE, and SGP and 2-yr ARI exceedances are used for MDWST, PCST, SW, and ROCK regions (Schumacher et al. 2021).

2.3 Generating Real-time Forecasts

Real-time forecasts, generated by the GEFS/R-based RF model (herein CSU-MLP), are produced using the forecasts from the operational GEFS ensemble median for each of the predictors described previously (QPF, precipitable water, mean sea level pressure, CAPE, CIN, 2-m temperature and relative humidity, 10-m wind components, and vertical wind shear). Probabilistic forecasts for excessive rainfall are made at each model gridpoint using the relative frequency of excessive rainfall predicted by all decision-trees in the RF. CSU-MLP forecasts from each region are merged into a single grid of probabilities across the CONUS, with smoothing applied across the regional boundaries to avoid sharp gradients in probabilities. Forecasts are then displayed graphically to appear similar to the ERO (Figure 2.2), where probabilities are classified within each corresponding categorical risk for excessive rainfall. In

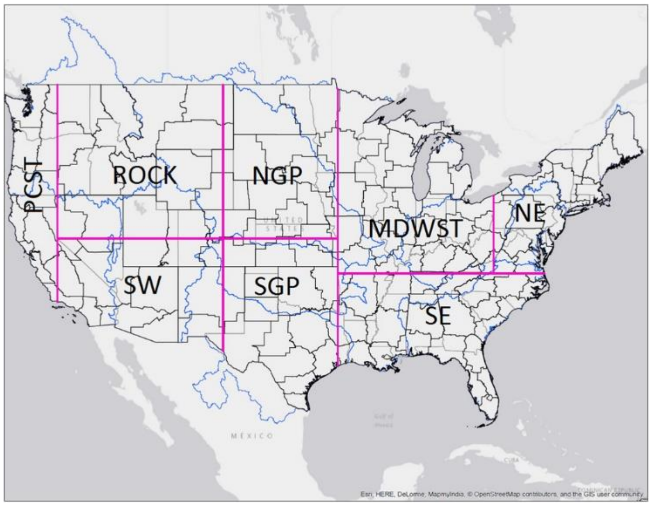


Figure 2.1. Map depicting the eight training regions and the labels describing each.

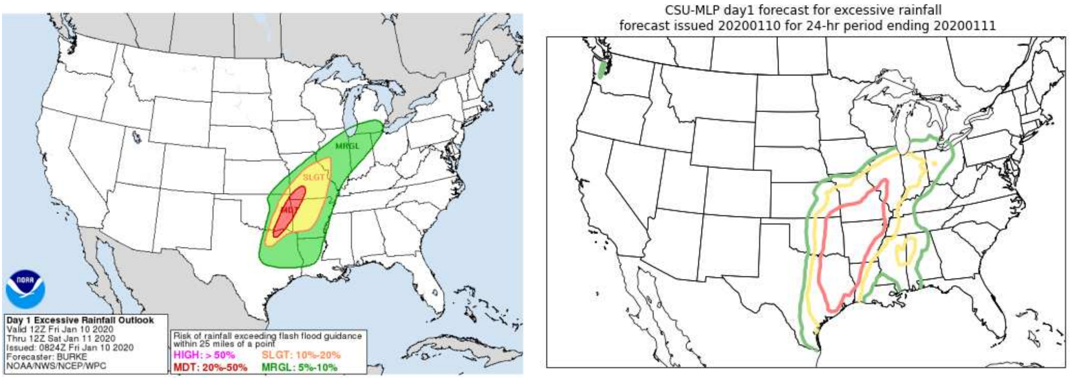


Figure 2.2. Example of (a) WPC day-1 ERO and (b) CSU-MLP day-1 "first-guess" forecasts valid the 24-hr period from 1200 UTC 10 January 2020 through 1200 UTC 11 January 2020.

total, daily forecasts used in this subsequent analysis include or consider 589 daily forecasts between 3 March 2019 and 15 October 2020. Some forecasts were retrospectively generated, with a limit to availability of the operational GEFS to generate forecasts prior to 3 March 2019.

CHAPTER 3: AGGREGATE CSU-MLP EVALUATION

This chapter will focus on evaluation of CSU-MLP frequency, reliability, discrimination, and skill using the same period of study from Schumacher et al. (2021) that includes 589 aggregated forecasts generated between 3 March 2019 through 15 October 2020. These characteristics will also be compared to those of the ERO.

3.1 UFVS Verification Dataset and Climatology

Verification is performed by use of the Unified Flood Verification System (UFVS; Erickson 2019, 2021), developed by WPC. The UFVS includes flash flood observations that occur within a given 1200 UTC – 1200 UTC timeframe, specifically, flash flood LSRs and U.S. Geological Survey (USGS) river gauge observations. Due to the spatial discontinuities of LSRs and the limited observational coverage of river gauges, UFVS also includes Stage IV precipitation exceeding FFG and Stage IV precipitation exceeding 5-yr ARI thresholds. These are used as flash flood proxies that are intended to capture flash flood occurrences missed by the observational datasets. UFVS flash flood observations and proxies include a 40-km neighborhood radius as to be consistent with the definition of the ERO (and CSU-MLP). Forecast grids from CSU-MLP are interpolated to the UFVS and ERO grid for a gridpoint-based verification. Figure

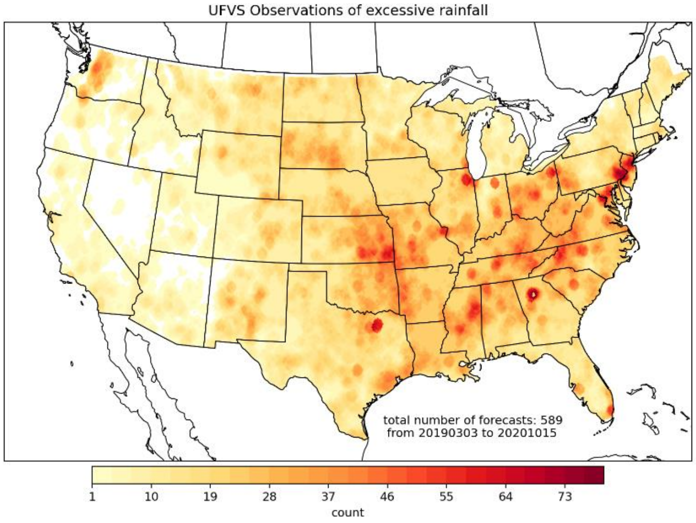


Figure 3.1. Observations of excessive rainfall from UFVS from 3 March 2019 – 10 October 2020.

3.1 shows the total number of UFVS observations across the CONUS within the period of study. UFVS observations occurred at a higher frequency in the eastern two-thirds of the CONUS in contrast to the western third, where there were no UFVS observations in some locations.

The UFVS is also used to construct a daily climatological forecast of excessive rainfall to be used as a baseline to calculate skill scores for CSU-MLP and the ERO forecasts, respectively. The daily climatological forecast is calculated using the frequency of UFVS observations from 1 October 2016 – 30 September 2020. Due to limitations in the temporal extent of the UFVS observational dataset, this results in a relatively short climatology where there are many small areas that have a relatively high or low frequency of UFVS observations that may not be representative of the true climatology for excessive rainfall. Specifically, frequencies are categorized by day of the year for each gridpoint within the CONUS domain and subsequently smoothed by calculating a rolling 31-day mean centered about the day of interest and smoothed by a spatial Gaussian filter across the domain. Figure 3.2 (a) and (b) help to visualize the effect of the spatial filter for all UFVS observations during the period of study.

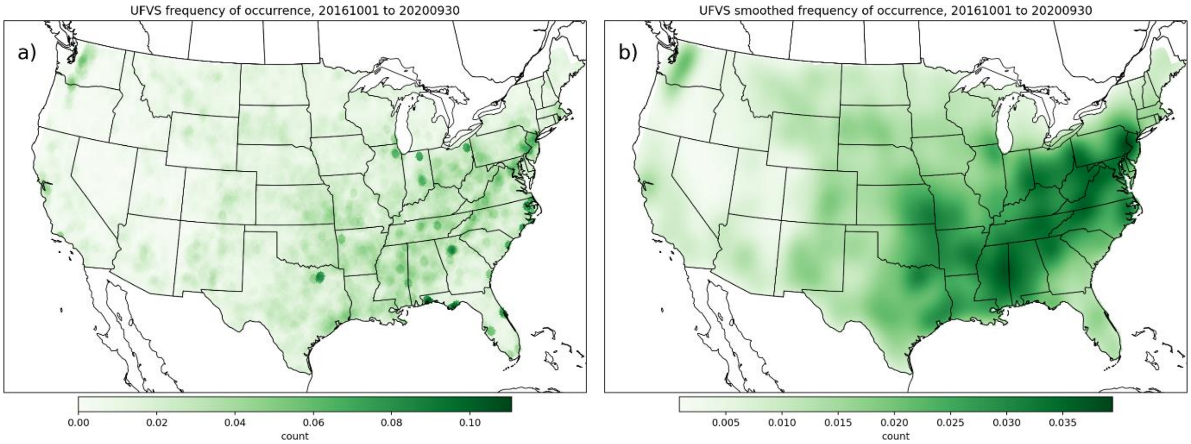


Figure 3.2. Observations of excessive rainfall from UFVS for the climatological period of 1 October 2016 – 30 September 2020 for (a) unfiltered gridded observations and (b) spatially smoothed gridded observations.

3.2 Forecast Frequency

Figure 3.3 shows CSU-MLP (left column) and ERO (right column) forecast fractional frequencies aggregated temporally and displayed spatially across the CONUS for forecast probabilities greater than each categorical threshold (5% MRGL, 10% SLGT, 20% MDT, 50% HIGH). Both CSU-MLP and the ERO show high frequency of categorical forecasts across the central and eastern CONUS in contrast to the western CONUS. This not surprising given the higher frequency of UFVS observations during the 2016-2020 period across these areas. Both forecasts also show decreasing frequency of forecast probabilities for increasing categorical thresholds. MRGL excessive rainfall probabilities are relatively common, particularly in the SE and MDWST border regions where roughly 1 in 5 days were forecasted as such. However, frequency of higher categorical thresholds decreases substantially, particularly for MDT and HIGH thresholds. Comparison between CSU-MLP and the ERO shows that CSU-MLP forecasts a higher frequency of forecasts for MRGL, SLGT, and MDT thresholds, particularly, SLGT and MDT frequencies are higher across nearly all regions of the central and eastern CONUS for CSU-MLP as well as in the SW region. Notable regions of highest forecast frequency differences are in the central CONUS, as well as in the southern portion of the NE region.

3.3 Forecast Reliability

An important characteristic of a probabilistic forecast is the reliability of the forecast, which measures the observed frequency of an event relative to a probabilistic forecast threshold. A probabilistic forecast is considered reliable when the observed relative frequency matches the probabilistic forecast of that event. For CSU-MLP and the ERO, probabilities of excessive rainfall are given by the four categorical risk contours, corresponding to a range of probabilities within each category. Reliability is assessed in a spatial sense, where given a certain gridpoint coverage of each category, the percentage of excessive rainfall gridpoint observations that overlap the forecasted category is calculated (Figure 3.4). For CSU-MLP, forecasted probabilities are reliable within the range of probabilities corresponding to each categorical risk, denoted by the green and red horizontal lines. Percent coverage area for CSU-MLP is lower across all four thresholds compared to the ERO, which is consistent with the higher frequency of forecasts at the MRGL, SLGT, and MDT categories forecasted by

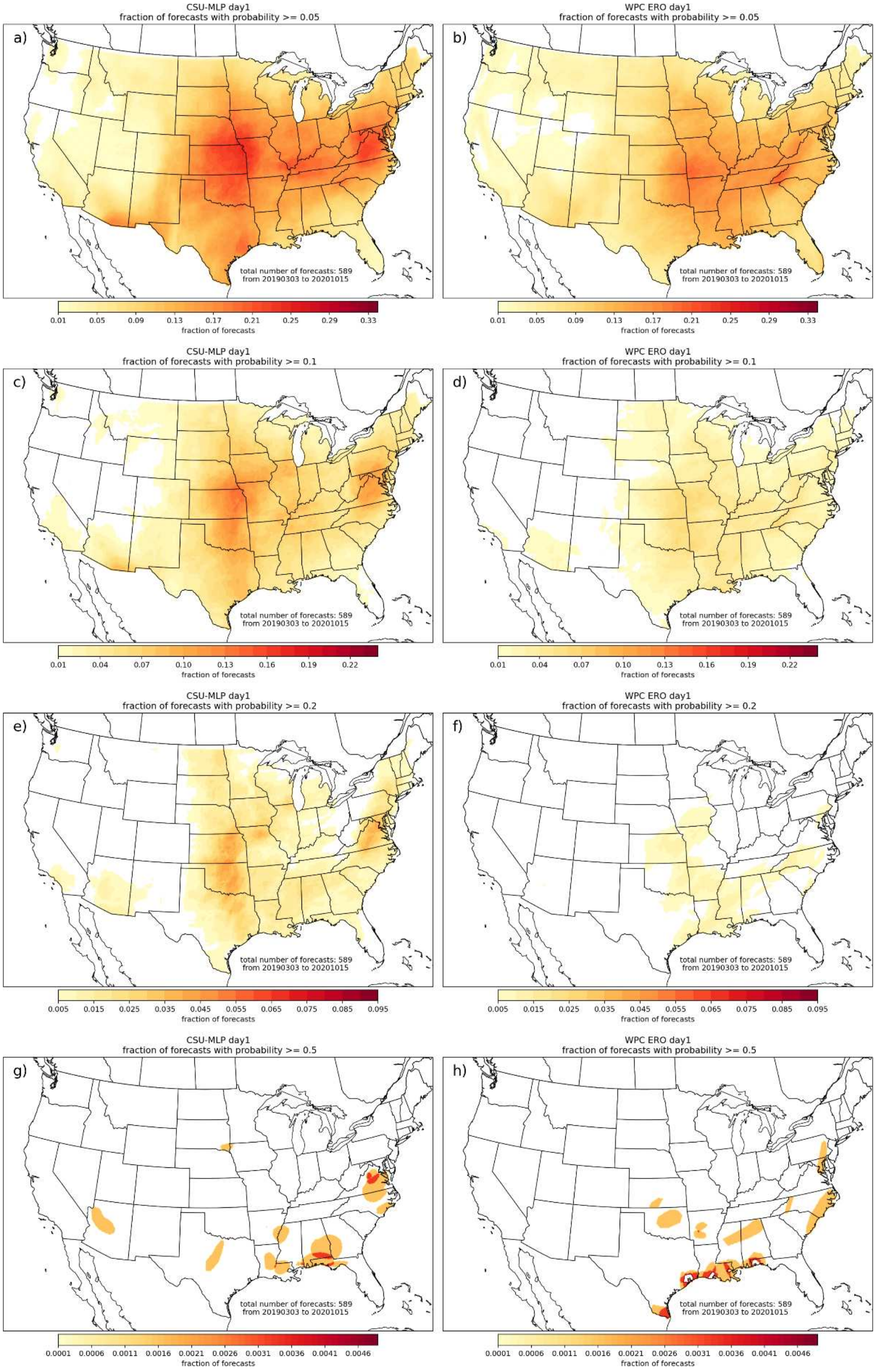


Figure 3.3. Fraction of forecasts days for (left) CSU-MLP and (right) ERO with probability greater than (a),(b) 5%, (c), (d) 10%, (e), (f) 20%, and (g), (h) 50%.

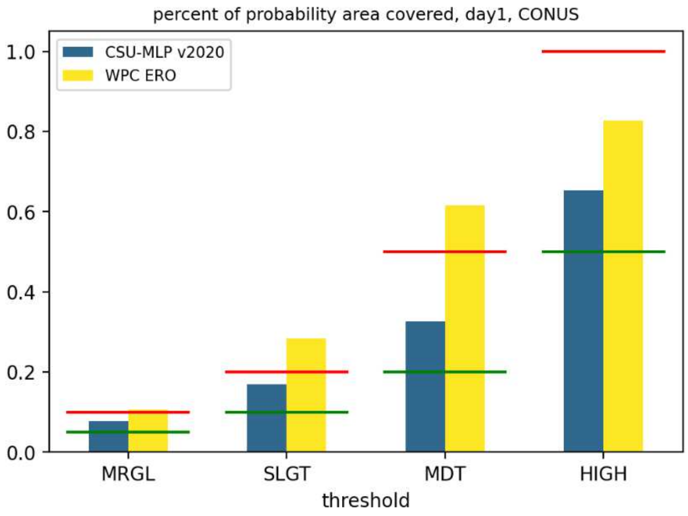


Figure 3.4. Percent of probability area covered by UFVS observations for CSU-MLP and ERO forecast. Green (red) lines indicate the lowest (highest) percentage for which aggregated categorical forecasts are reliable.

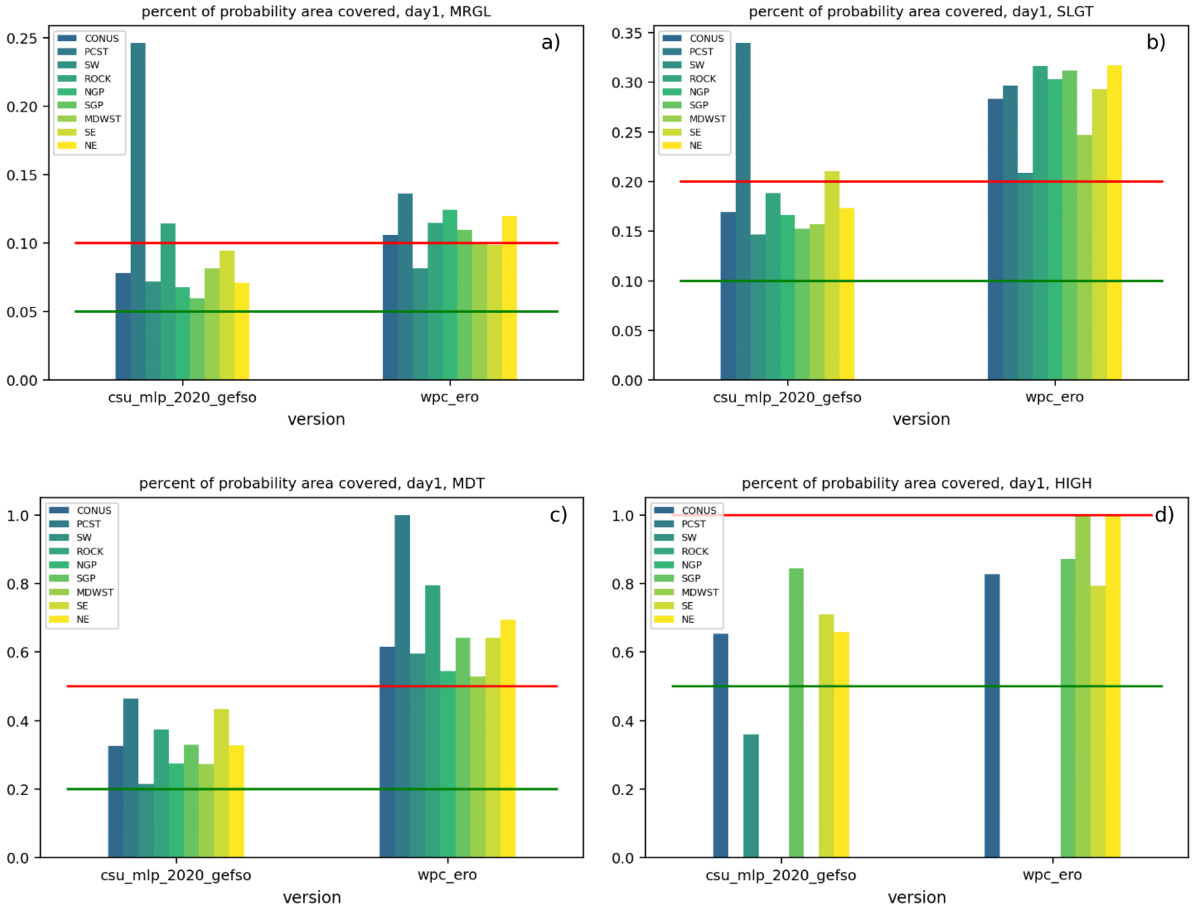


Figure 3.5. Percent of probability are covered by UFVS observations for forecasts in each region of CONUS and for (a) MRGL, (b) SLGT, (c) MDT, and (d) HIGH categorical risks. Green (red) bars indicate the lowest (highest) percentage for which aggregated categorical forecasts are reliable.

CSU-MLP (Figure 3.3). The ERO underpredicts the risk for excessive rainfall in its categorical forecast areas for MRGL, SLGT, and MDT thresholds, and is reliable within the threshold range for the HIGH category. Figure 3.5 shows the percent coverage area for forecasts subdivided by each of the eight training regions. CSU-MLP is reliable for most categorical thresholds in most regions, however, CSU-MLP underpredicts the coverage of excessive rainfall for MRGL forecasts in the PCST and ROCK regions, for SLGHT forecasts in the PCST and SE regions, and overpredicts coverage for HIGH forecasts in the SW.

3.4 Forecast Resolution/Discrimination

Forecasts are also characterized by their ability to discriminate between excessive rainfall events and non-events. This is measured by probability of detection (POD; also known as sensitivity/true positive rate) and probability of false detection (POFD; also known as false alarm ratio which is equivalent to 1 – specificity). POD measures how often an event is correctly forecasted to occur given the total number of events that occurred. This can be written as Equation 3.1:

$$POD = \frac{TP}{TP + FN} = \frac{TP}{\#\, Observed\ Events} \tag{3.1}$$

where TP are true positives (event is correctly forecasted) and FN are false negatives (event is incorrectly not forecasted). POFD measures how often a non-event is incorrectly forecasted to occur (false positives, FP) given the total number of non-events. This can be written as Equation 3.2:

$$POFD = \frac{FP}{FP + TN} = \frac{FP}{\#\, Observed\ Non - Events} \tag{3.2}$$

A perfectly discriminating forecast system would be able to identify all observed events (POD = 1) while avoiding instances of false detection (POFD = 0). For CSU-MLP and the ERO, the midpoint probability of each categorical threshold is used as the discriminating probability for events (forecasted probability above this threshold) and non-events (forecasted probability below this threshold). POD and POFD statistic pairs at each threshold for CONUS-wide forecasts are plotted graphically (Figure 3.6) in what is known as the Relative Operating Characteristic, or ROC curve. ROC curves for the daily climatological forecast and that of no discriminating skill (dashed line) are also plotted. The area under the ROC (AuROC) is a measure of the total discrimination, where a perfectly discriminating forecast has an AuROC equal to 1 and a non-discriminating forecast has an AuROC equal to 0.5. AuROC values for each region for CSU-MLP and the ERO are shown in Figure 3.7. CSU-MLP and the ERO both display discriminating capabilities for CONUS and all eight regions within, particularly for the CONUS, SW, NGP, SGP, MDWST, NE, and SE where AuROC are greater than 0.74. Both PCST and ROCK regions have lower POD for lower probability thresholds where CSU-MLP AuROC scores are 0.666 and 0.716, respectively,

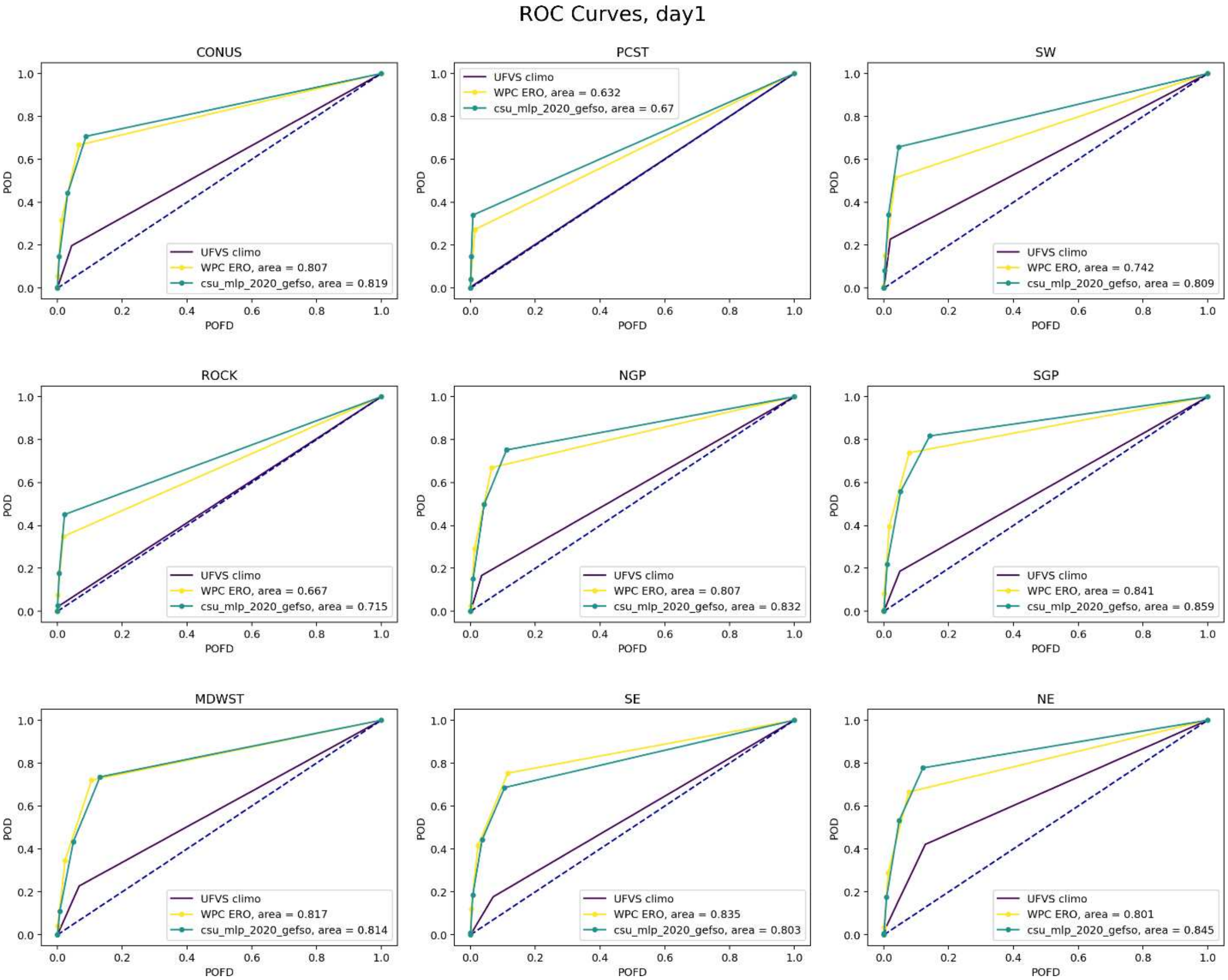


Figure 3.6. ROC curves for CONUS and regional forecasts for CSU-MLP and ERO forecasts using midpoint probabilities to calculate POD and POFD.

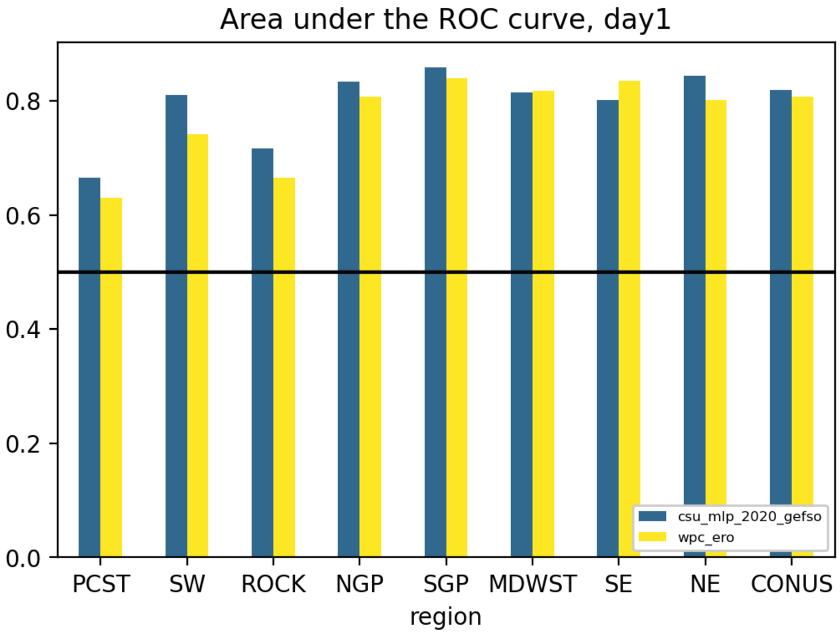


Figure 3.7. AuROC for CSU-MLP and ERO forecasts for CONUS and each region

and ERO AuROC scores are 0.630 and 0.666, respectively. This is due to the underconfident MRGL forecast areas (Fig. 3.5) where excessive rainfall events are missed due to smaller MRGL forecast areas. ERO AuROCs are smaller than all CSU-MLP AuROCs (except for SE region) owing to the ERO's underconfident SLGT and MDT probabilities which leads to a much lower POD compared to that of CSU-MLP.

3.5 Forecast Skill

CSU-MLP and ERO forecast skills are evaluated against the daily climatological forecast for excessive rainfall using the Brier skill score (BSS), which is a measure of the forecast Brier score (BS) to that of the climatological BS. BS measures the average squared differences between a probabilistic forecast value, P_i , and the observed, O_i (1 for an excessive rainfall event or 0 for non-event):

$$BS = \frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2 \qquad (3.3)$$

where n is the total number of probability-observation pair samples. BSS is formulated as:

$$BSS = \frac{BS_{climo} - BS_{forecast}}{BS_{climo}} = 1 - \frac{BS_{forecast}}{BS_{climo}} \tag{3.4}$$

A forecast that performs well relative to climatology must have a $BS_{forecast}$ that is low relative to BS_{climo}. This would mean the forecast has probabilities close to 1 or 0 for corresponding excessive rainfall events and non-events across an aggregate of forecast-observation pairs and/or climatology has probabilities far from 1 or 0 for corresponding excessive rainfall events and non-events. BSS is calculated using midpoint probabilities for each categorical forecast as well as for each daily climatological forecast. Daily climatology is below 5% across most locations through much of the year, so BSS is sensitive to probabilistic forecasts from CSU-MLP and the ERO. However, the daily climatological forecast is higher than 5% (sometimes 10%) in many locations across summer months when excessive rainfall events are more common. In these situations, BSS is sensitive to both the forecasted probabilities and climatology.

BSS is calculated by aggregating samples in time for across the CONUS domain to evaluate spatial patterns in skill, as displayed in Figure 3.8. Skills for both (a) CSU-MLP and (b) ERO forecasts are greater than that of climatology across much of the central and eastern regions of the CONUS. Western regions show much more variability, especially for CSU-MLP, as a small number of forecasts and the occurrence or non-occurrence of individual UFVS observations greatly influence local BSS. Figure 3.9 shows that BSS aggregated in time and space, where ERO forecasts show slightly greater aggregate skill across the CONUS as well as four of the eight regions. This contrasts with BSS for day-2 and day-3 forecasts, where CSU-MLP demonstrates slightly higher skill than the ERO (Schumacher et al. 2021).

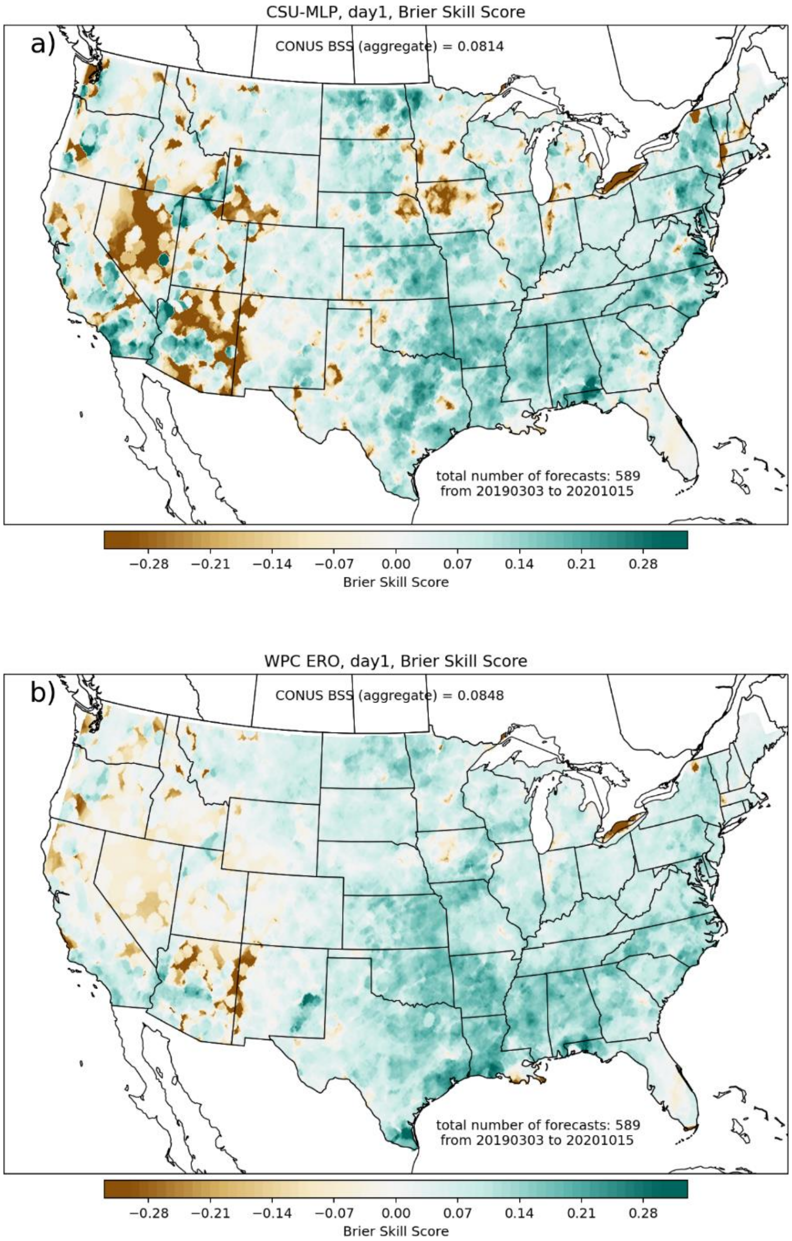


Figure 3.8. Temporally aggregated BSS for (a) CSU-MLP and (b) ERO calculated at each gridpoint.

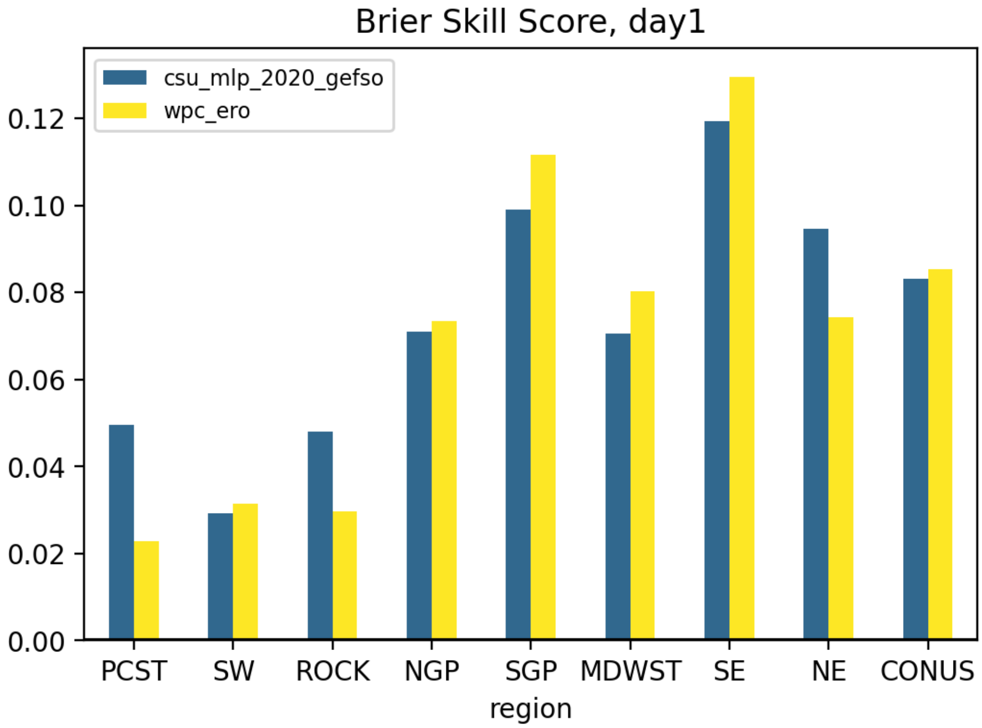


Figure 3.9. Aggregate BSS for CSU-MLP and the ERO for each region and all forecasts.

CHAPTER 4: DAILY SKILL CHARACTERISTICS OF CSU-MLP AND THE ERO AT DAY-1

In the previous chapter, CSU-MLP (day-1) forecasts were shown to be reliable, with discrimination and skill comparable to that of the WPC ERO. However, CSU-MLP skill relative to climatology can vary between individual daily forecasts where days with widespread UFVS observational coverage result in higher BSS and days with minimal coverage result in lower BSS (Figure 4.1). Also, CSU-MLP's higher forecast frequency for MRGL, SLGT, and MDT categorical thresholds compared to the ERO can be both beneficial and detrimental to daily skill. For example, Figure 4.1 shows two forecast examples where CSU-MLP forecasted both higher probabilities and a larger area of categorical risk compared to the ERO. In the top row, both CSU-MLP and the ERO performed well as UFVS observations were widespread across risk areas, but CSU-MLP had a much higher BSS (0.34) relative to the ERO (0.18). In the bottom row, UFVS observations are less numerous and displaced from highest probabilities which gives CSU-MLP a much lower BSS (-0.54) compared to the ERO (-0.01). Daily CSU-MLP forecasts in comparison with daily ERO forecasts will be briefly investigated. This will lead into the methodology for the identification of forecast cases of "interest" to be used when selecting best- and worst-performing forecast days for identifying synoptic regimes.

4.1 Distributions of Daily UFVS, CSU-MLP, and ERO Coverage

Daily observations of excessive rainfall from UFVS (Figure 4.2) takes the form of an exponential distribution. Days with zero UFVS observations are the most frequent for CONUS and all regions, although, these occurrences across the CONUS are much less common than for individual regions. Figure 4.3 shows daily forecast coverage of any categorical risk forecasted by CSU-MLP and the ERO, and, not surprisingly, the smallest coverage bin (0-400 gridpoints) has the highest frequency. Figure 4.4 shows the same distributions as Figure 4.2 but excluding days with no UFVS observations. For each region, days with excessive rainfall observations tend to stay below 1000 gridpoints of coverage, while

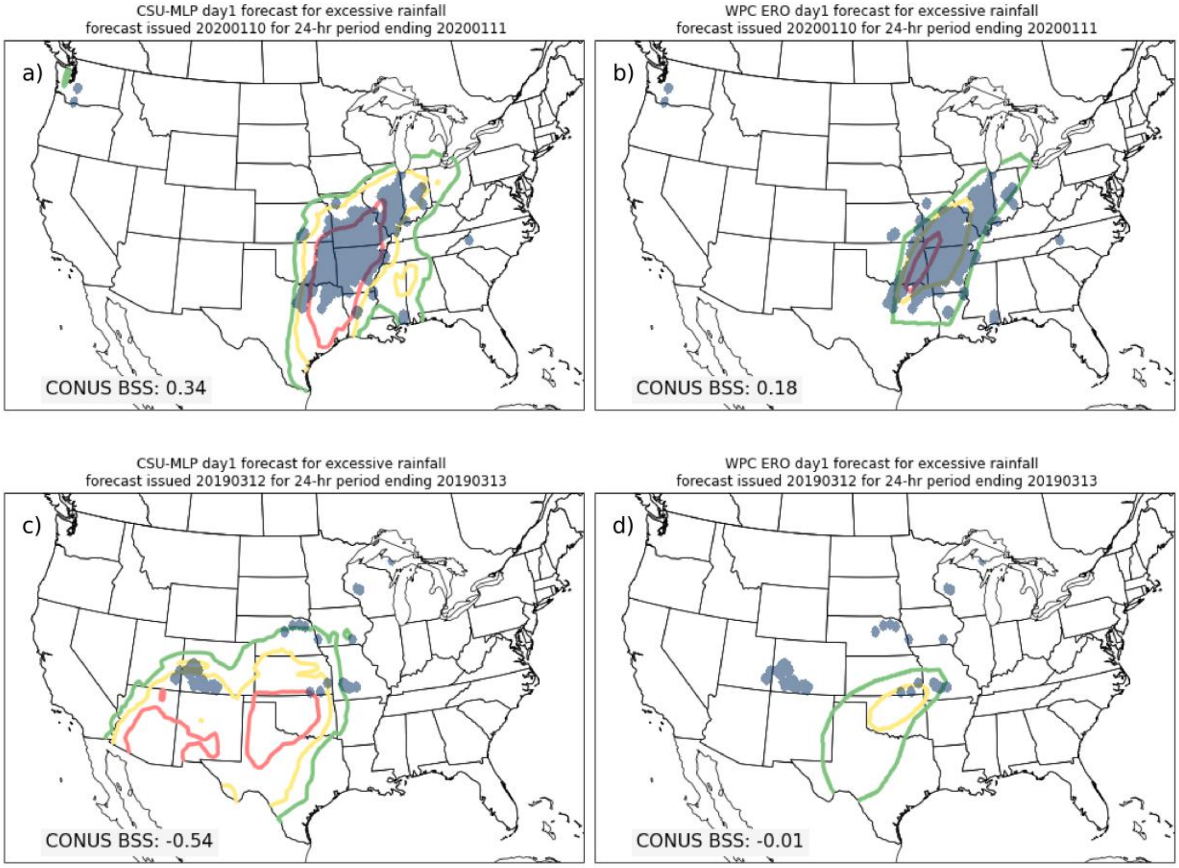


Figure 4.1. Example forecasts days where CSU-MLP has (a) high and (c) low BSS along with ERO forecasts for those same days. Categorical risks for excessive rainfall are in contours (same color scale as Figure 2.2) and UFVS observations are color-filled.

for the entire CONUS, observations tend to stay below 4000 gridpoints. Likewise, Figure 4.5 shows distributions of CSU-MLP and ERO gridpoint coverage excluding days without any categorical risk. For all regions except NE and NGP, CSU-MLP forecasts have a higher frequency of forecasts in the lowest bin compared to the ERO, which contrasts with the ERO having higher frequency in the lowest bin for most regions considering all daily forecasts (Figure 4.3). This would suggest the tendency for CSU-MLP to forecast relatively small categorical risk areas when the ERO would have no coverage.

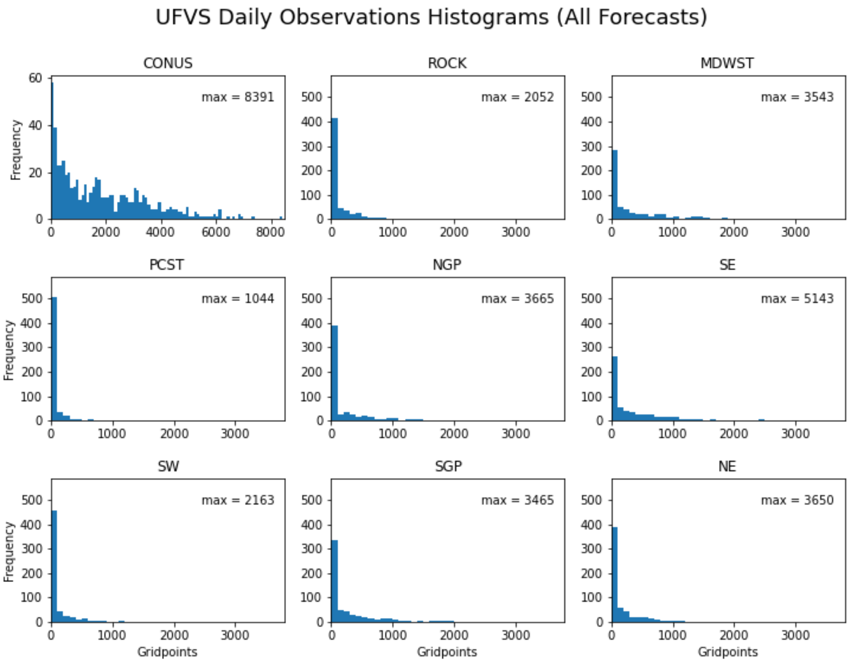


Figure 4.2. Frequency distributions of daily UFVS excessive rainfall gridpoint coverage for CONUS and all regions through the 589-day period of study, with bin widths of 100 gridpoints.

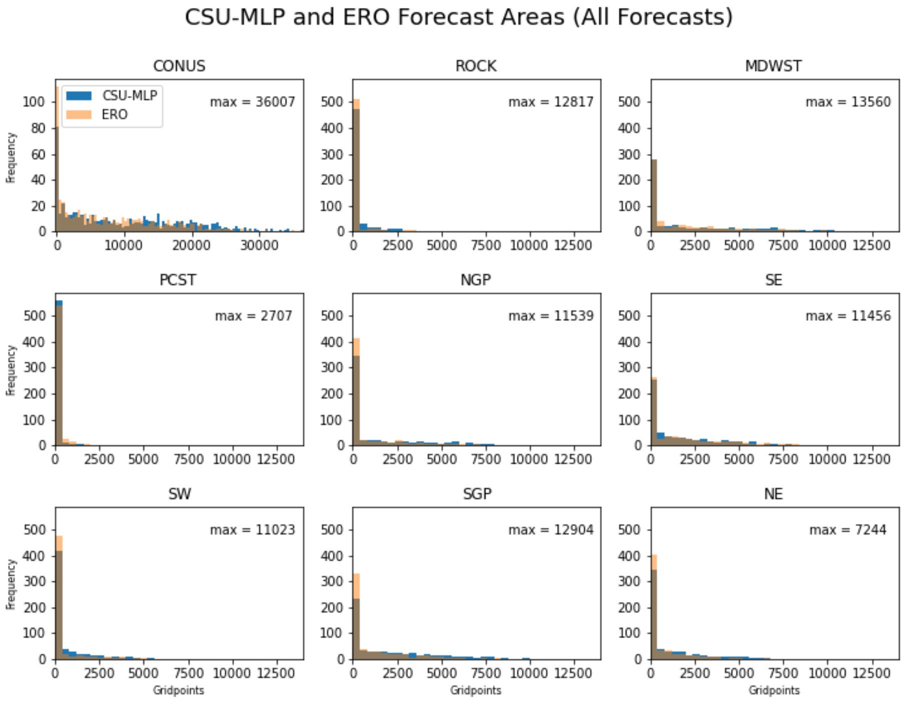


Figure 4.3. Frequency distribution of daily CSU-MLP (blue) and ERO (orange) gridpoint coverage for CONUS and all regions through the 589-day period of study, with bin widths of 400 gridpoints.

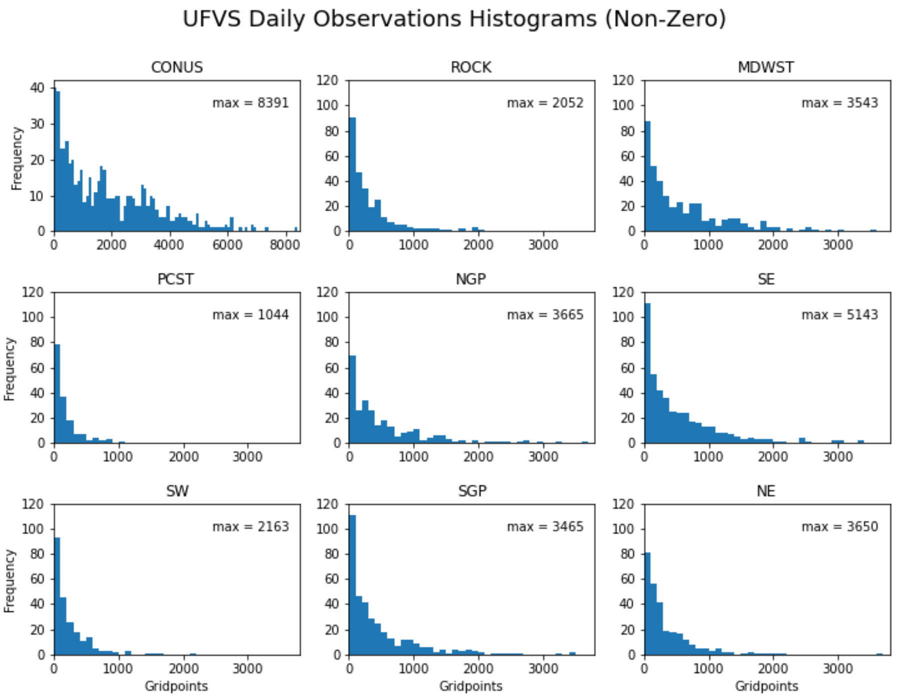


Figure 4.4. Same as Figure 4.2 but excluding days with zero excessive rainfall observations for the given regional domain.

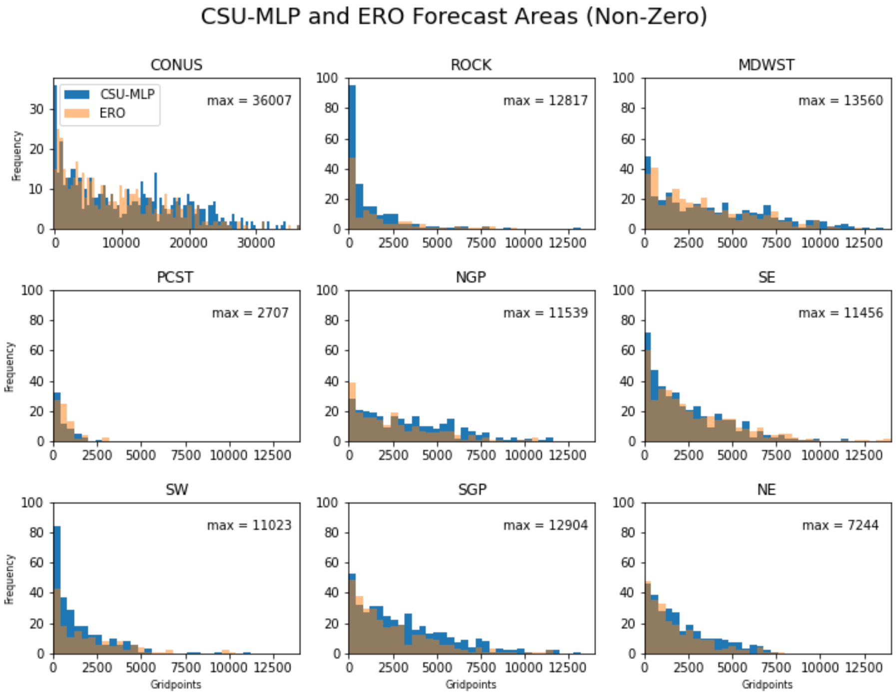


Figure 4.5. Same as Figure 4.3 but excluding days with zero coverage of CSU-MLP and ERO categorical risk, respectively, for the given regional domain.

4.2 CSU-MLP and ERO Daily Forecast Skill Comparison

To assess daily skill, a daily BSS is calculated for both CSU-MLP and ERO forecasts, where forecast-observation pairs are aggregated within the CONUS and each of the regional domains for a given forecast day. Figure 4.6 shows daily BSS distributions for CSU-MLP (blue) and ERO (orange) for all forecasts. Across the regional domains, most forecast days have equal skill relative to the climatological forecast (BSS close to zero) as these are predominantly days where the climatological forecast and CSU-MLP/ERO forecasts do not have categorical risk within the region and no excessive rainfall observations occur. For CONUS and all regions, both CSU-MLP and the ERO have distributions centered slightly positive of zero considering only forecasts in which at least one UFVS observations occurred (Figure 4.7). This indicates that both forecasts have greater skill compared to the climatological forecast on these days. However, CSU-MLP has a larger tail towards negative BSS values (except for PCST) indicating the increased tendency for CSU-MLP to have poor individual forecast days compared to the ERO. Figure 4.8

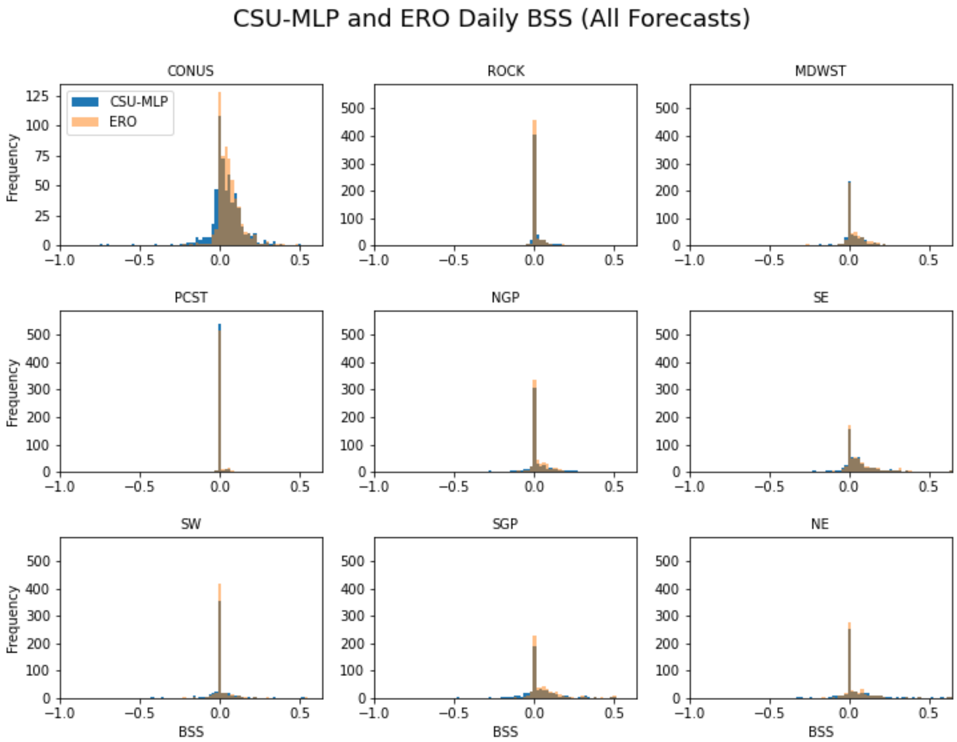


Figure 4.6. Daily BSS histograms by region for CSU-MLP (blue) and ERO (orange) day-1 forecasts for the period of study. Brown indicates overlap between CSU-MLP and the ERO.

directly compares daily BSS between CSU-MLP and the ERO by taking the BSS difference for each day, but only for days where a UFVS observation occurred in a given region. Although median differences are near zero for most regions, some regions have a noticeable negative skewness (NGP, SGP, MDWST, NE) indicating a relatively higher frequency of daily forecasts where CSU-MLP performs worse than the ERO. Looking at Figures 4.7 and 4.8 in conjuction, these negative BSS difference forecasts seem to be from days in which CSU-MLP has a large negative BSS relative to the ERO as opposed to days where the ERO has large positive BSS relative to CSU-MLP (such as the examples provided in Figure 4.1). Figure 4.8 also highlights the larger variance in skill performance within the MDWST, NE, and SGP regions, indicating that daily BSS from CSU-MLP and ERO forecasts are observed to be more dissimilar- mainly due to the larger negative tail.

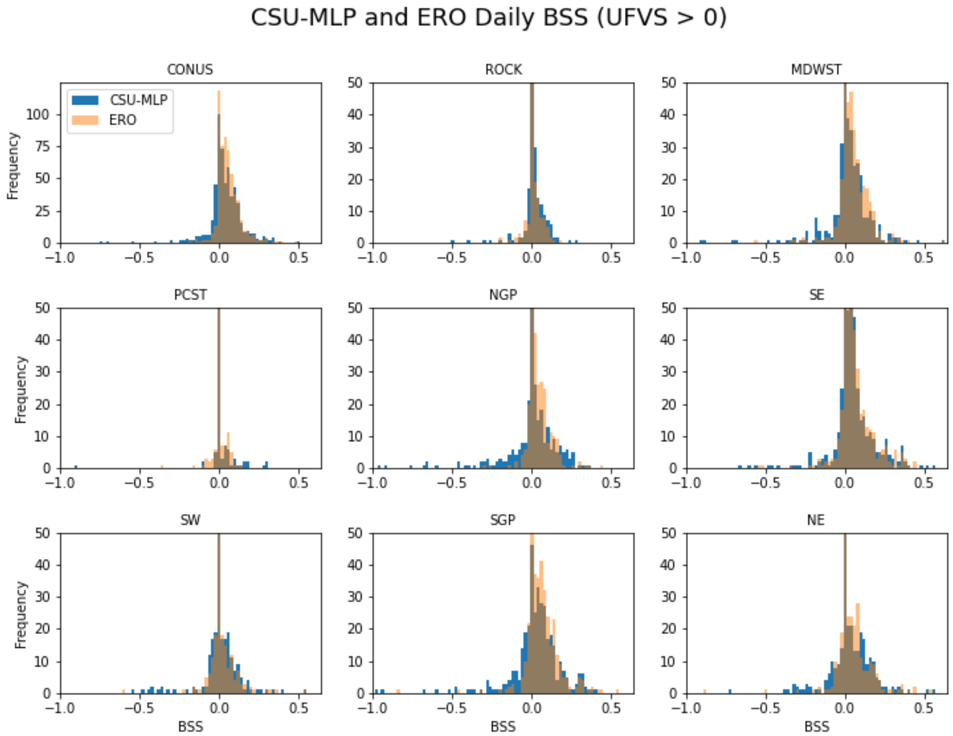


Figure 4.7. Same as Figure 4.6 but for forecast days where there was a UFVS observation within the given regional domain.

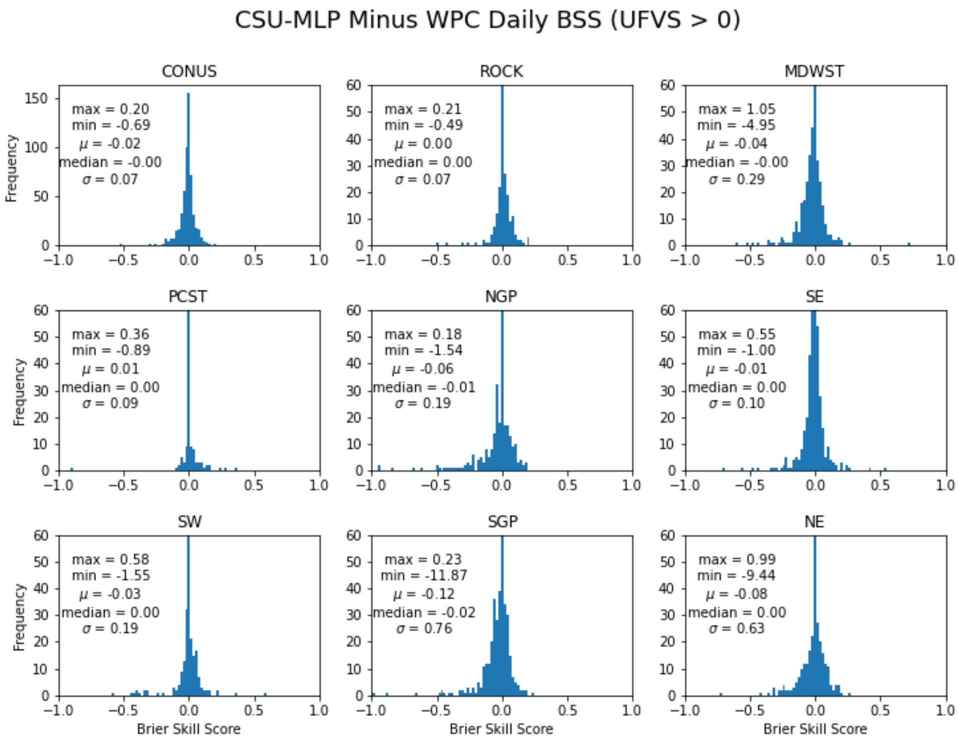


Figure 4.8. Daily BSS difference by region between CSU-MLP and ERO day-1 forecasts, only considering days with a UFVS observation within the given regional domain

These plots show that CSU-MLP daily skill can vary substantially from day-to-day, and there is reason to investigate potential causes not only to improve upon future iterations of models within the CSU-MLP system but also to provide WPC forecasters additional guidance for when daily forecasts tend to perform best/worst. This will be done by selecting high/low daily BSS CSU-MLP forecasts, for CONUS and regional domains, and identifying the synoptic regimes among these forecasts. Daily forecasts to be considered in this analysis will be described in the next section.

4.3 Selecting Cases

While it is important for CSU-MLP to be able to correctly forecast near-zero probability for excessive rainfall on quiescent days, these are not necessarily days where WPC forecasters would potentially utilize the CSU-MLP forecast, especially since CSU-MLP tends to forecast categorical probabilities more frequently than the ERO (Figure 3.4). This would lead one to only look at forecasts where CSU-MLP has any categorical probabilities within a given domain to discard of quiescent forecasts from the distribution. Of course, CSU-MLP forecasts do not detect all excessive rainfall events either (see ROC curves from Figure 3.7) and case selection of this method could miss potential low daily BSSs, but these forecast days tend to be summer days where BSS > 0 due to a relatively low number of undetected UFVS observations and widespread high climatology probabilities (≥ 0.05) which result in a high BS_{climo}. However, this selection method also includes additional undesirable CSU-MLP forecasts, specifically, low BSS forecasts that are sensitive to small MRGL categorical areas that do not verify with UFVS observations. These small areas are a combination of small categorical objects (i.e. a closed contour) within a domain, or, for regional forecasts, small areas along the domain border that are a part of a larger contiguous object that extends through adjacent regions. To eliminate these kinds of forecasts, CSU-MLP must forecast at least a MRGL (≥ 0.05) categorical area of at least 400 gridpoints within a given domain to be used in the synoptic analysis. This 400-gridpoint size is designed to approximately resemble the smallest ERO categorical object that WPC might issue in an ERO (Figure 4.9). This also ensures that regional cases only consider categorical areas that extend far enough into a region to be a forecast case of "interest". The CSU-MLP 400-gridpoint requirement still retains some undesirable forecasts in the distribution, specifically, quiescent forecasts where the daily climatology forecast has high values (≥ 0.05) across a given domain. This is similar to the quiescent summertime

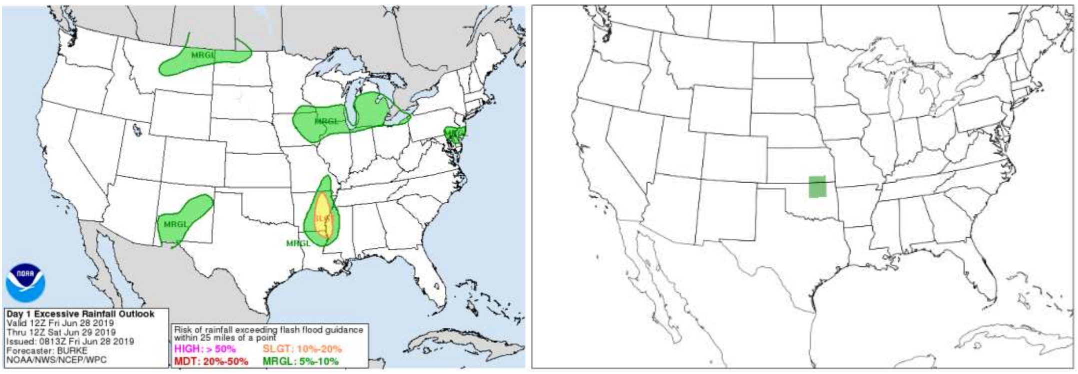


Figure 4.9. (Left) An example day-1 ERO with multiple categorical objects and (right) a map showing the size of 400 gridpoints.

forecasts eliminated by the 400-gridpoint threshold, however, in these forecasts, high climatology is widespread enough so that BS_{climo} is still greater than $BS_{forecast}$ which results in a positive (sometimes largely) BSS. To discard these undesirable cases, forecasts with high climatological forecasts (>0.05) within the domain must also have UFVS observational coverage of at least 5% percent the high climatological forecast gridpoint coverage. This eliminates summertime cases that are "quiescent" relative to the daily climatology. The remaining forecasts compose the distribution of forecasts cases of "interest" to be used to classify worst- and best-performing forecasts.

Figure 4.10 shows box-and-whisker plots of daily BSS for all forecasts, with selected forecasts plots in Figure 4.11. Daily BSS through for all forecasts shows the same distribution as the histogram

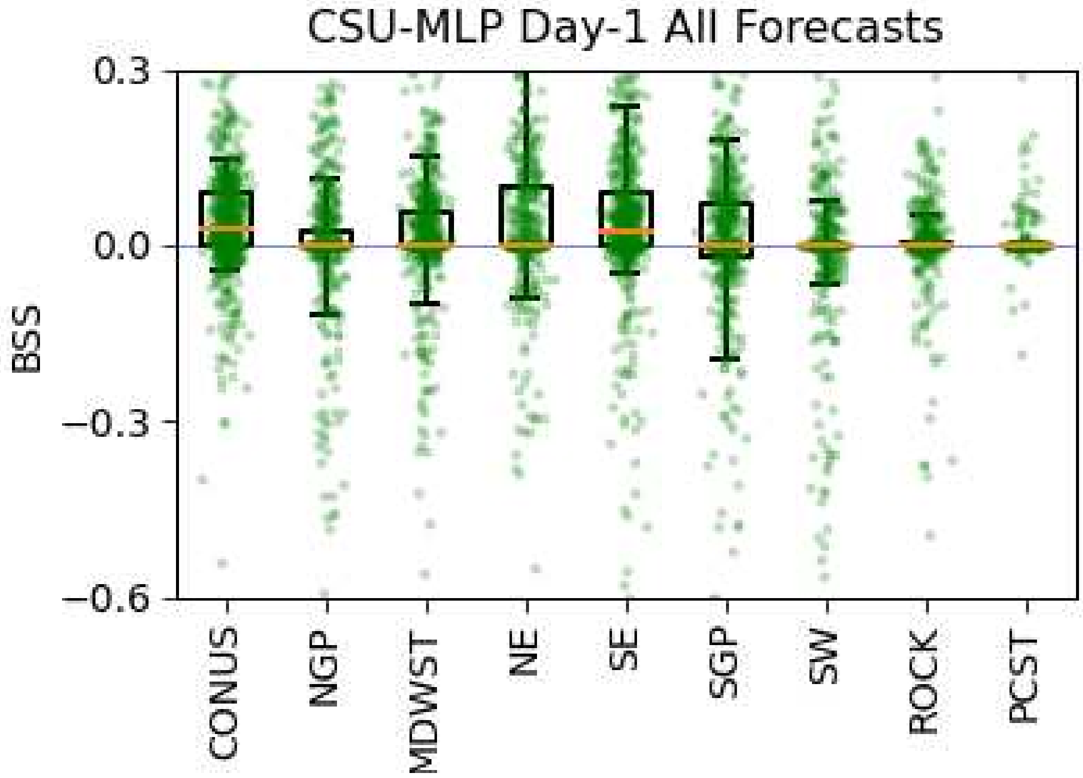


Figure 4.10. Box-and-whisker plots of daily BSS for CSU-MLP forecasts for all 589 forecasts during the period of study along with individual BSS shown with green markers. Boxes denote the 25th and 75th percentile BSS, with whiskers denoting the 10th and 90th percentile BSS. Orange lines denote the median daily BSS for each region.

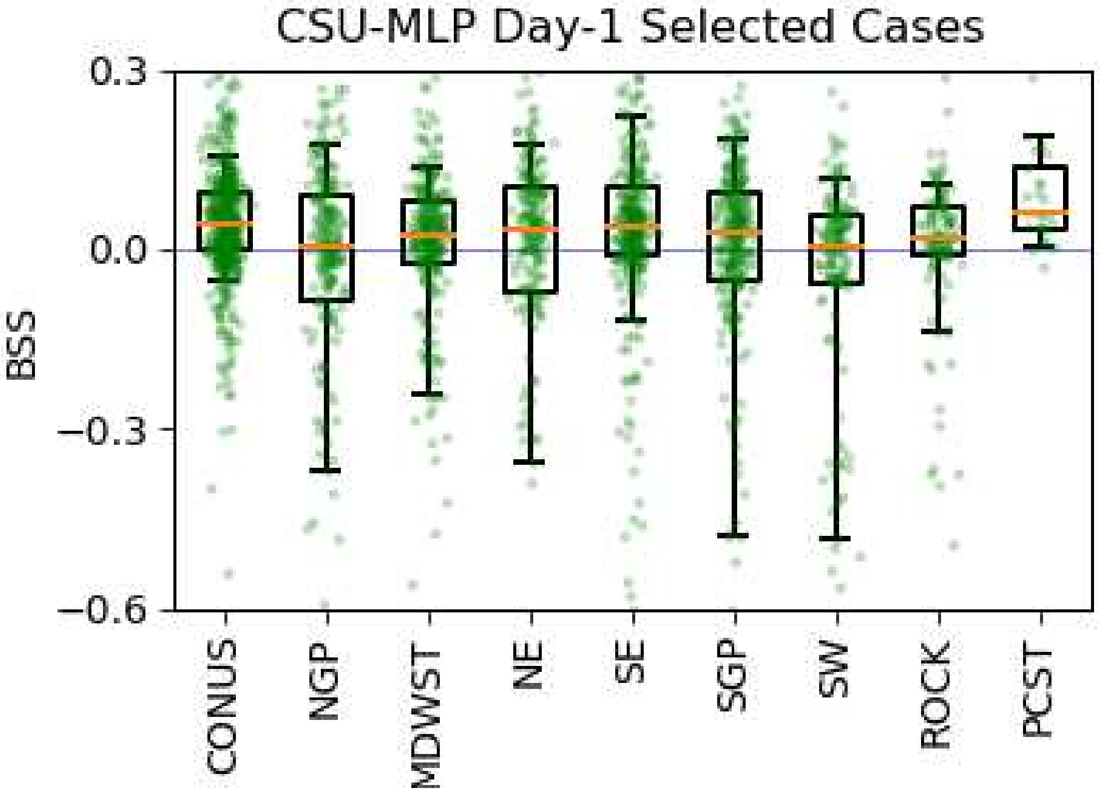


Figure 4.11. Same as Figure 4.10 but for selected forecast cases of "interest".

from Figure 4.6, but median and quantile thresholds are shown explicitly. All regions show median daily BSS close to zero as this is influenced by the large number of forecasts days with quiescent conditions. However, selected forecasts have quiescent days excluded and, as a result, median values shifted to slightly higher BSS for most of the regions. In addition, the width of the interquartile range (between 25th and 75th percentile) increases, indicating that the distribution for forecasts of "interest" is relatively wide for a given region compared to daily skill aggregated across the CONUS. Five of the eight regions show large decreases in the 10th percentile threshold, without much change to the 90th percentile threshold. This is consistent with the types of forecasts that were discarded, as the large chunk of quiescent days removed would naturally broaden the distribution and removal of quiescent days with high climatology would reduce the number of high daily BSS forecasts. The best performing region is the PCST region, where the 25th and 75th percentile BSS are highest among all the regions (although with a small number of forecasts of interest). Selected forecast daily BSS distributions are also displayed by meteorological season (Figure 4.12). For the CONUS, median values do not change between seasons, however, the width of the 10th to 90th threshold is wider for DJF and SON. Regional distributions have much broader interquartile ranges compared to CONUS. Many regions have extremely low 10th percentile thresholds for DJF and SON, while maintaining similar 90th percentile thresholds across seasons. While the larger spread in these seasons could be due to smaller samples of forecasts of "interest", the consistency of this spread across multiple regions would suggest that CSU-MLP forecasts are most vulnerable to low skill days in these seasons.

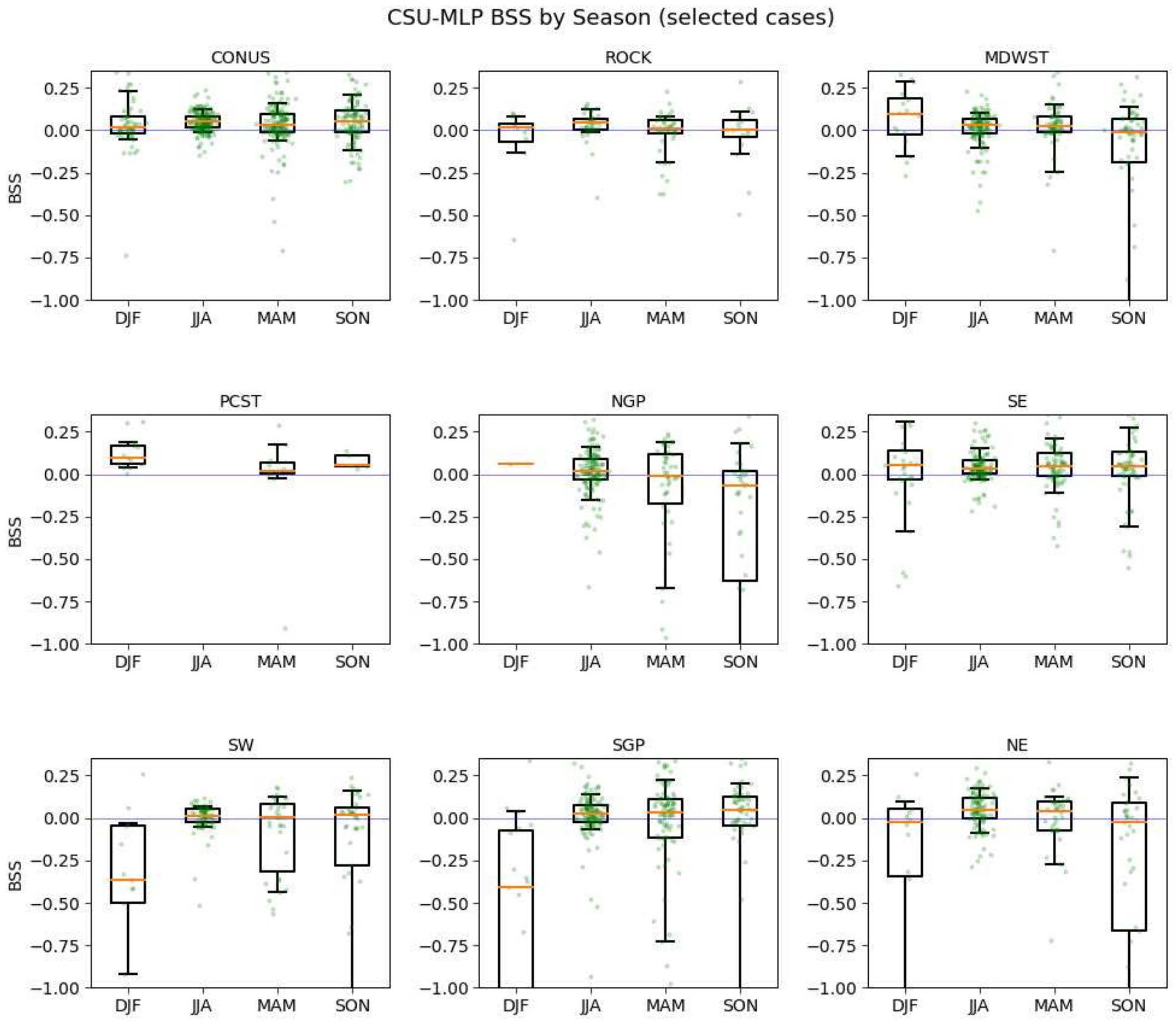


Figure 4.12. Same as Figure 4.11 but with distributions of daily BSS among each meteorological season for each region.

4.4 Defining Best- and Worst-Performing Forecasts

Best- and worst-performing forecasts are defined as forecasts with daily BSS greater than or equal to 75th percentile thresholds and less than or equal to 25th percentile thresholds, respectively. This results in a total of 125 cases selected for each best- and worst-performing CONUS-wide forecasts (Table 4.1) with BSS thresholds of 0.09 and 0.00, respectively. Among the regional forecasts, the greatest number of cases selected were for the SE (58) and the least number were for the PCST (7). Best-performing BSS thresholds ranged between 0.05 (SW) and 0.14 (PCST), whereas worst-performing BSS thresholds ranged between 0.03 (PCST) and -0.09 (NE). These selected cases are used for synoptic regime identification in the following chapter.

Table 4.1. Daily BSS thresholds for best-performing forecasts (> 75th percentile) and worst-performing forecasts (< 25th percentile) by region along with number of cases selected to each performance group, respectively.

Region	No. Cases for 75th/25th	75th BSS Threshold	25th BSS Threshold
CONUS	125	0.09	0.00
NGP	31	0.09	-0.09
MDWST	51	0.08	-0.02
NE	27	0.10	-0.07
SE	58	0.11	-0.01
SGP	49	0.09	-0.05
SW	18	0.06	-0.06
ROCK	19	0.07	-0.01
PCST	7	0.14	0.03

CHAPTER 5: SYNOPTIC REGIMES OF BEST- AND WORST-PERFORMING CSU-MLP FORECASTS

In this chapter, synoptic regimes are identified among the best- and worst-performing forecasts of "interest" from Chapter 4. These forecasts of "interest" are forecasts in a given domain where CSU-MLP has a categorical risk area of at least 400-gridpoints and, if the climatological frequency is high (>0.05), the UFVS observational gridpoint coverage is at least 5% of the high climatological gridpoint coverage. In the following sections, the methodology for synoptic regime classification is explained along with the analysis of those identified synoptic regimes.

5.1 Synoptic Regime Classification

For best- and worst-performing forecasts of interest, respectively, forecasts are subjectively categorized into groups of similar synoptic regimes using ERA5 reanalysis (Hersbach et al. 2020). This analysis is done by examination of 12, 18, 00, and 06 UTC synoptic charts within the day-1 timeframe corresponding to each forecast. These charts include fields of mean sea-level pressure (MSLP), 2-m dew point temperature, 500-hPa geopotential heights and winds, 250-hPa geopotential heights and winds, precipitable water (PWAT), 850-hPa dew point temperature and winds, standardized anomaly 500-hPa geopotential heights and PWAT. Subjective categorization attempted to consider not only the locations and evolutions of large-scale troughs, surface cyclones, and moisture fields, but also diagnosing the primary synoptic forcing mechanisms (if any) that lead to convective initiation using the synoptic charts (i.e. cyclones, fronts, orography) in conjunction with analysis of composite radar. Regime type is also categorized using the two event types described in Chapter 1 from Maddox et al. (1979), along with additional categories such as extratropical cyclone (ETC), tropical cyclone (TC), cold front/warm front (not to be confused with "frontal" from Maddox et. al 1979), atmospheric river (AR), and cut-off low (COL) that are discriminated in similar manners to Kunkel et al. (2012) and Mullens (2021). These events are categorized when the UFVS observations and/or maximum CSU-MLP categorical risk contours are within close proximity (roughly within 200 km) to the synoptic feature of interest (i.e. surface cyclone, frontal boundary). ETC events are categorized as such when a developed surface cyclone (i.e. closed MSLP contour) moves closely (within roughly 200 km) over an area of excessive rainfall risk. This is separate from cold/warm front categories, where a swath of UFVS observations and/or CSU-MLP categorical risk area is located along a frontal boundary well removed (roughly 500km) from a parent surface cyclone. Cold front and warm front regimes, while also having some overlap with "synoptic" and "frontal" regime types, are distinguished by the total movement of the frontal boundary throughout the forecast period, where these boundaries move on the order of 1000 km through a forecast period. In contrast, "synoptic" and "frontal" events have boundaries that are relatively stationary (moving on the order of 100km through the forecast period). TC regimes are categorized as such when UFVS observations and/or CSU-MLP categorical risk are associated with rainfall from landfalling or inland moving TCs (as identified by the National Hurricane Center). COL regimes are defined as regimes where there is a closed 500-hPa geopotential height contour near the area of excessive rainfall and/or forecasted risk. Lastly, AR regimes are identified where there is an elongated local maximum (horizontal length of roughly 1000 km and horizontal width of roughly 100 km) of PWAT near the area of excessive rainfall and/or forecasted risk area.

Composites charts are then created by averaging of each of the individual 00 UTC reanalysis fields within the regime, but only considering regimes with a minimum of four representative days. As a result, not all best- and worst-performing forecasts of "interest" were used in composite analysis as some did not belong in a regime containing at least three additional similar cases. Figure 5.1 and Figure 5.2 provide an example group of forecasts days that were categorized under the same synoptic regime, where subplots of two different chart types are shown using reanalysis valid 00 UTC for each forecast period. In this example, most cases have a 500-hPa trough axis (Figure 5.2) that extends through the central and southern CONUS that results in large-scale ascent across the southeast along with a surface

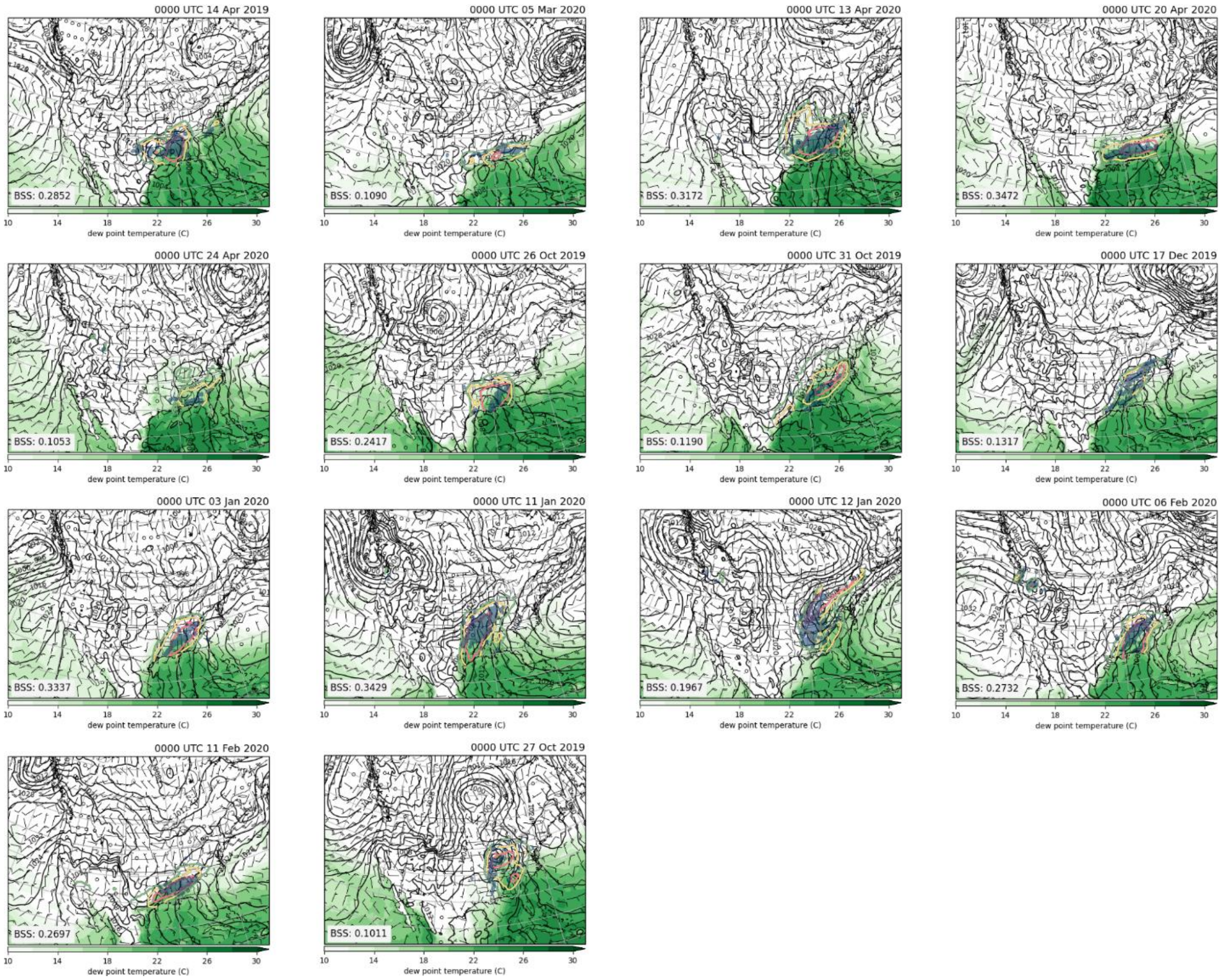


Figure 5.1. ERA5 reanalysis at 00 UTC (middle of the forecast period) for each case of regime 7 of the best-performing CONUS regimes with contours of MSLP, 10-m wind barbs, 2-m surface dew point in shading, colored contours of categorical risk from the CSU-MLP forecast, and UFVS observations (dark gray shading). Daily BSS is shown in the bottom-left corner of each plot.

cyclone (Figure 5.1). The surface cyclone has a moist sector to the southeast that quickly moves northward as a warm front, and a north-south oriented cold front to the south of the cyclone that moves eastward. A broad swath of UFVS observations and higher CSU-MLP probabilities occur near the path of the surface cyclone as these areas received sustained rainfall from lift of a conditionally unstable low-level airmass along and north of the surface warm-front as well as rainfall from a cold front passage. The regions of high CSU-MLP probabilities closely align with UFVS observations which results in a set of forecasts with relatively high BSS.

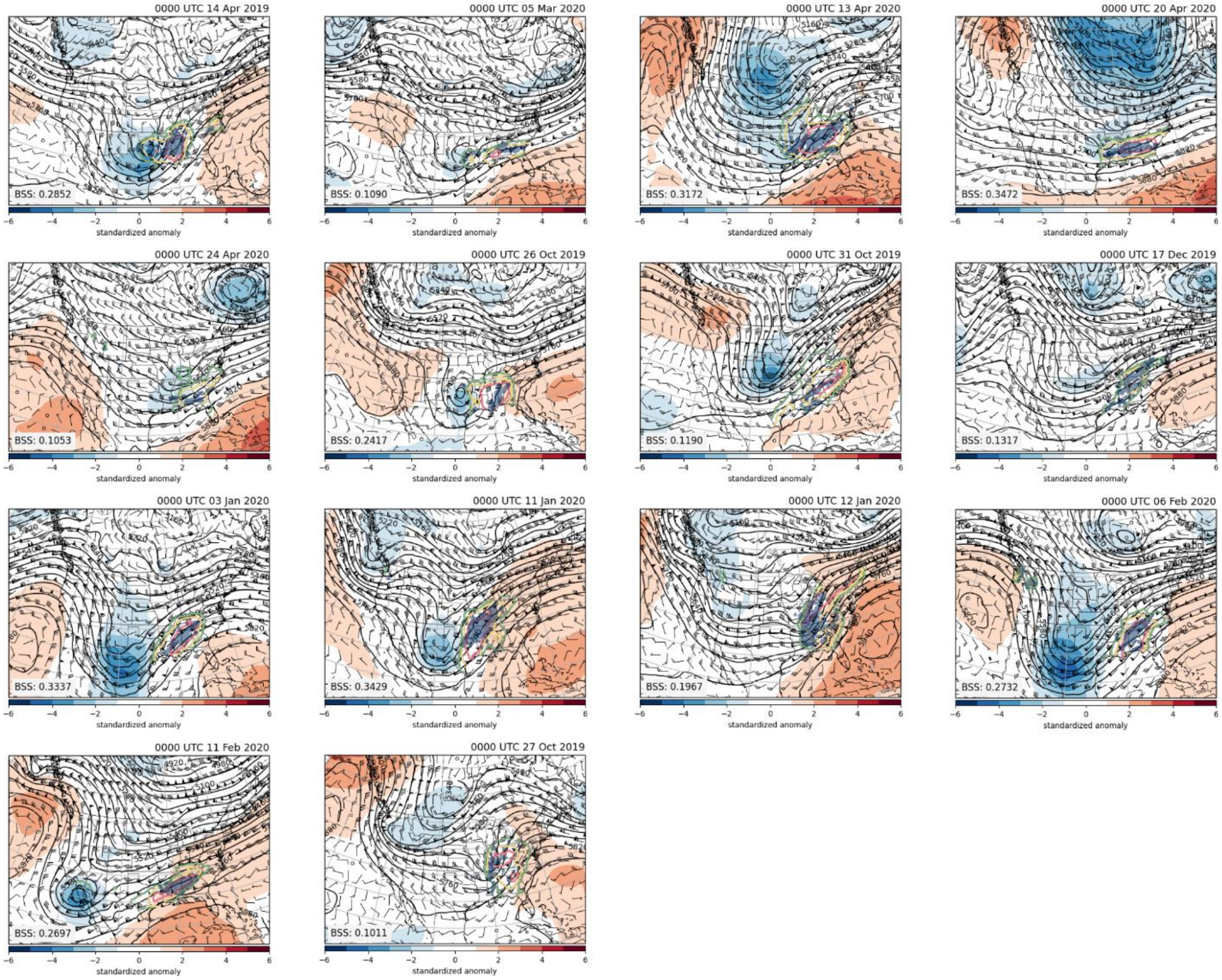


Figure 5.2. ERA5 500-hPa reanalysis at 00 UTC (middle of the forecast period) for each case of regime 7 of the best-performing CONUS regimes with contours of geopotential height, wind barbs, standardized height anomaly in shading, colored contours of categorical risk from the CSU-MLP forecast, and UFVS observations (dark gray shading). Daily BSS is shown in the left corner of each plot

Regime groups are identified using best- and worst-performing forecasts of CONUS-wide daily skill and regional daily skill. The following sections will examine the worst-performing (11) and best-performing (9) regimes identified from CONUS-wide forecasts. Many of the of regimes identified from regional forecasts are analogous to one or more the CONUS regimes, so these will not be shown explicitly. Within the text, regimes will be referred to by their regime number (which is arbitrarily assigned) and percentile threshold (i.e. R1-25th corresponds to regime 1 of the worst-performing forecasts).

5.2 CONUS Worst-Performing Synoptic Regimes

Table 5.1 shows average daily BSS statistics for worst-performing CSU-MLP forecasts evaluated for the CONUS, along with the ERO average for those given cases and the difference between CSU-MLP and the ERO. All regimes have CSU-MLP forecasts that performed worse than the daily climatological forecasts. Half of the regimes had average ERO BSS less than zero, indicating that WPC forecasters also struggled with these regimes. However, CSU-MLP skill was worse on average than that of the ERO, which is consistent with CSU-MLP forecasts having a larger negative tail in daily BSS distributions (see Figure 4.7). Listed in Table 5.1 are additional regions where the same synoptic regime was identified for worst-performing forecasts in those regions. Synoptic groups with multiple additional regions, like R9-25th, provides a sense of the larger areal extent these synoptic regimes had for excessive rainfall threat. The season with the most frequent occurrence of cases within the regime are shown in the last column.

Table 5.1. Characteristics of each identified worst-performing synoptic regime using CONUS-wide daily BSS. Number of cases comprising a regime is provided, along with the average CSU-MLP daily BSS among those cases (BSS Avg.), the average ERO daily BSS (ERO BSS Avg.), and the difference between CSU-MLP and ERO average daily BSS. Other regions indicate regions in which analogous regimes were identified using regional daily BSS (as defined in Chapter 4). The season with the most frequent occurrence of cases within the regime are shown in the last column

Regime No.	No. cases	BSS Avg.	ERO BSS Avg.	BSS Diff.	Other Regions	Regime Type	Season
1	4	-0.03	0.02	-0.05	NE, SW	cold front	JJA
2	4	-0.03	0.00	-0.03	MDWST, NE	mesohigh	JJA
3	8	-0.05	0.05	-0.10	MDWST, NGP	frontal	JJA
4	5	-0.07	0.04	-0.11	NGP, MDWST, ROCK, SW	Synoptic, frontal	JJA
5	7	-0.12	-0.03	-0.09	SE	TC	SON
6	17	-0.11	-0.04	-0.07	SE, SGP	Cold front	MAM
7	8	-0.11	-0.01	-0.10	SE, SGP	Warm front	MAM
8	9	-0.16	-0.01	-0.15	MDWST, NGP, ROCK, SGP	ETC	MAM
9	6	-0.15	-0.01	-0.13	NE	ETC	MAM
10	8	-0.06	0.01	-0.06	SE	Cold front	SON
11	7	-0.24	0.00	-0.24	ROCK, SW	Cold front	DJF

5.2.1. Warm-Season Regimes

Figures 5.3 and 5.5 show composite reanalysis fields for R1-25th and R2-25th, respectively, showing (a) composite heights and winds at 250-hPa, (b) composite 850-hPa winds and PWAT, (c) composite 500-hPa heights and winds, and (d) composite MSLP and 2-m dew point temperature. Additionally, the relative frequency of CSU-MLP forecasting at least a MRGL (≥ 0.05) categorical risk among the forecast days of the regime is displayed in (c) to highlight the area of risk for excessive rainfall. Both regimes feature a mid- and upper-level trough across the eastern CONUS, with PWAT between 30-40 mm in the areas of risk. Daytime convection increases within or near risk areas and

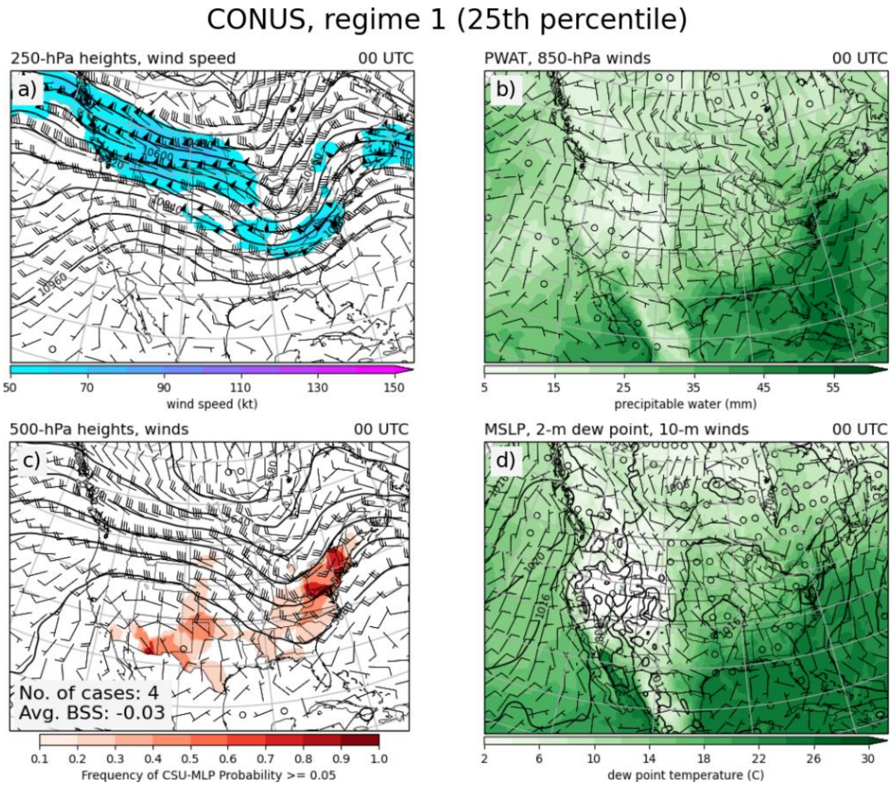


Figure 5.3. ERA5 composite reanalysis for R1-25th showing (a) composite 250-hPa winds, isotachs, and geopotential heights, (b) composite 850-hPa winds and PWAT, (c) composite 500-hPa winds, geopotential heights, and frequency of CSU-MLP forecasting at least 0.05 probability at a given location among the regime cases, and (d) composite MSLP, 2-m winds, and 10-m dewpoint temperatures. All composites use the 00 UTC reanalysis during the forecast period of each case.

composite radar indicates that convective evolution is predominately controlled by mesoscale and storm scale interaction (see radar images of example cases from Figure 5.4 and 5.6). However, R1-25th has a higher amplitude trough-ridge pattern across the CONUS with a more distinct eastward moving

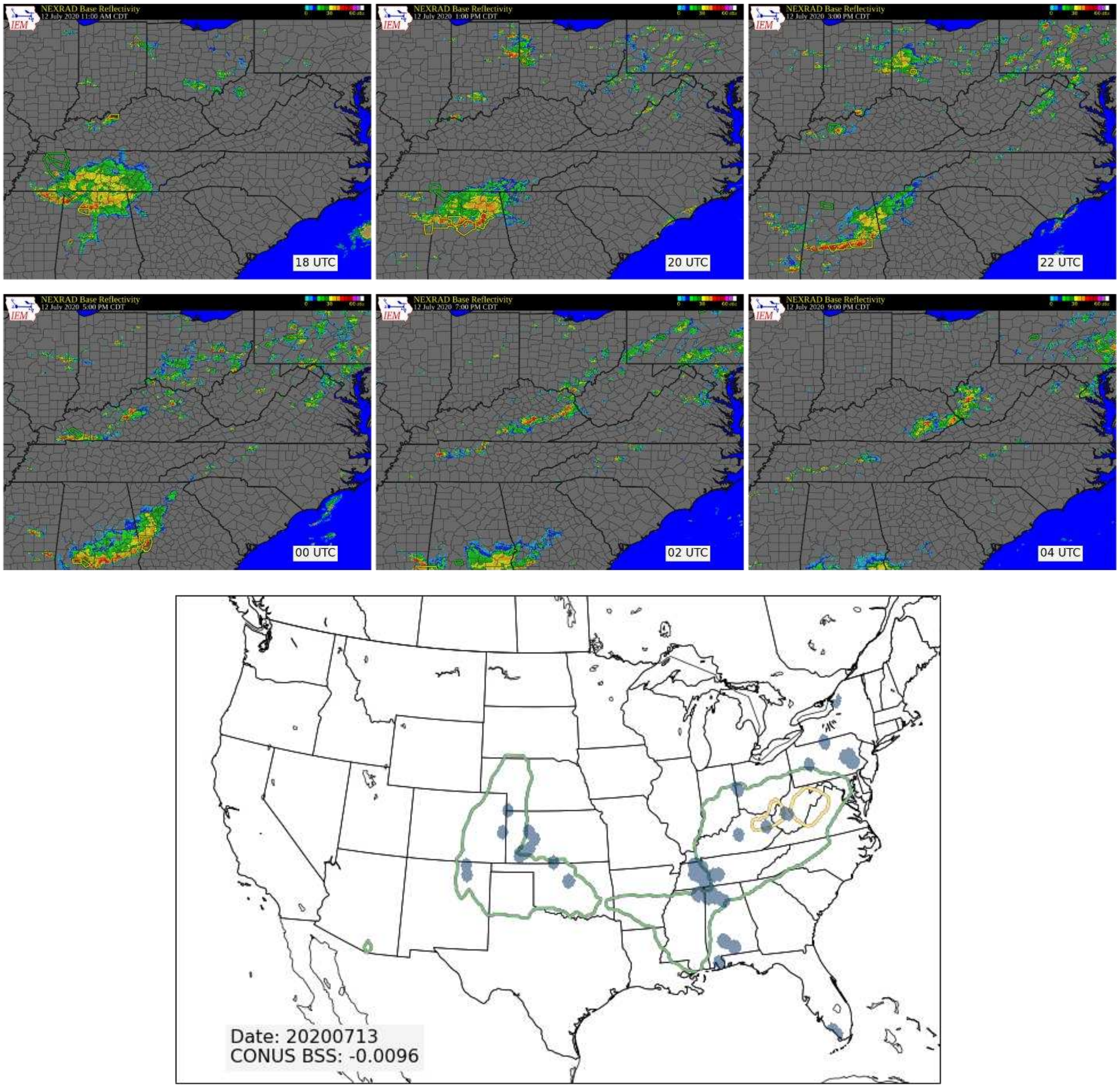


Figure 5.4. Composite radar imagery with flash flood warning in green polygons (top two rows) and CSU-MLP forecast contours and UFVS observations (bottom) for the 24-hr period forecast period spanning 12 UTC 12 July 2020 – 12 UTC 13 July 2020. This forecast case was categorized to R1-25th. Radar images are from 18, 20, 22, 00, 02, and 04 UTC within the forecast time period (images obtained from the Iowa Environmental Meosnet at http://mesonet.agron.iastate.edu/GIS/radmap_api.phtml).

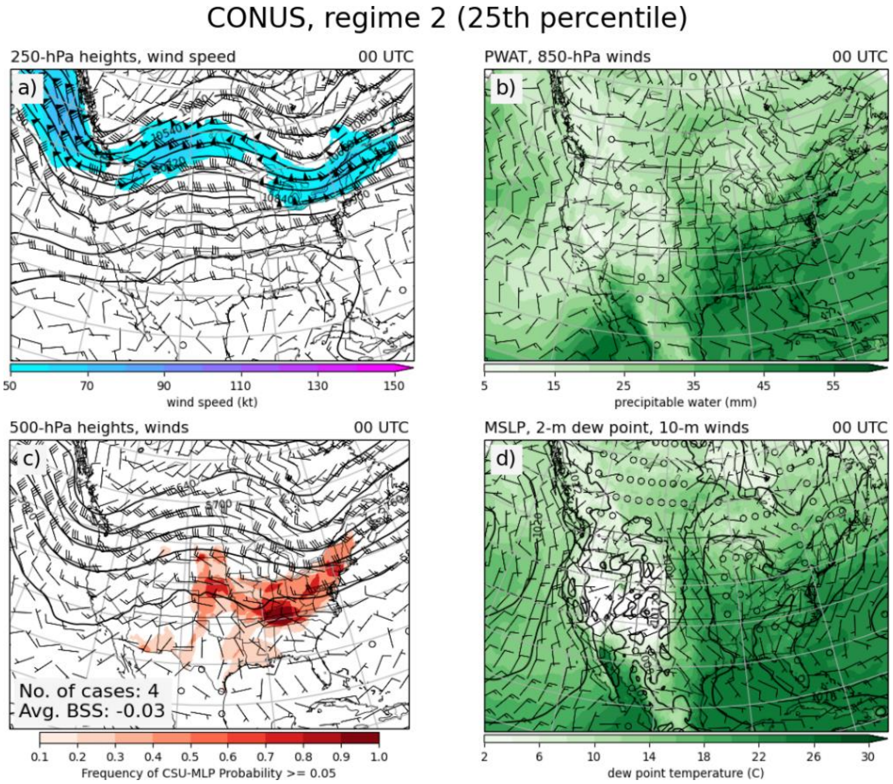


Figure 5.5. Same as Figure 5.3 but for R2-25th.

surface boundary (in contrast to R2-25th) oriented north-south near the coast, denoted by the east-west moisture gradient. In most of these cases, this boundary was a cold front with a weak parent surface low well north of the risk area. This led to R1-25th primary forcing for convection to be categorized as "cold front", much like one of the warm-season patterns identified and described in Agel et al. (2019). Additionally, R1-25th hints at a potential monsoon pattern in the southwestern CONUS associated with weak mid- and upper-level winds and increased moisture in southern Arizona, although excessive rainfall observations were variable between cases. R2-25th is different from R1-25th in that the gradual moisture gradient is oriented across a weak east-west boundary, and mid- and upper-level winds are generally weak in this area with little low-level synoptic forcing. In many cases, there was not a distinguishable surface boundary across the risk area, but composite radar indicated that convection was more organized than R1-25th, as convective complexes grew upscale with a wider coverage of precipitation. Figure 5.6 shows radar imagery of such an evolution from an R2-25th case, where there is greater coverage of rainfall by larger thunderstorm complexes near and south of the categorical risk area in the lower Midwest. This is in contrast to that of the R1-25th case (Figure 5.4), where afternoon and evening convection associated with the frontal boundary is less widespread in the lower Midwest region. Thus, the primary forcing is categorized as "mesohigh" (Maddox et al. 1979) for R2-25th as convective organization is predominantly influenced by cold pool evolutions.

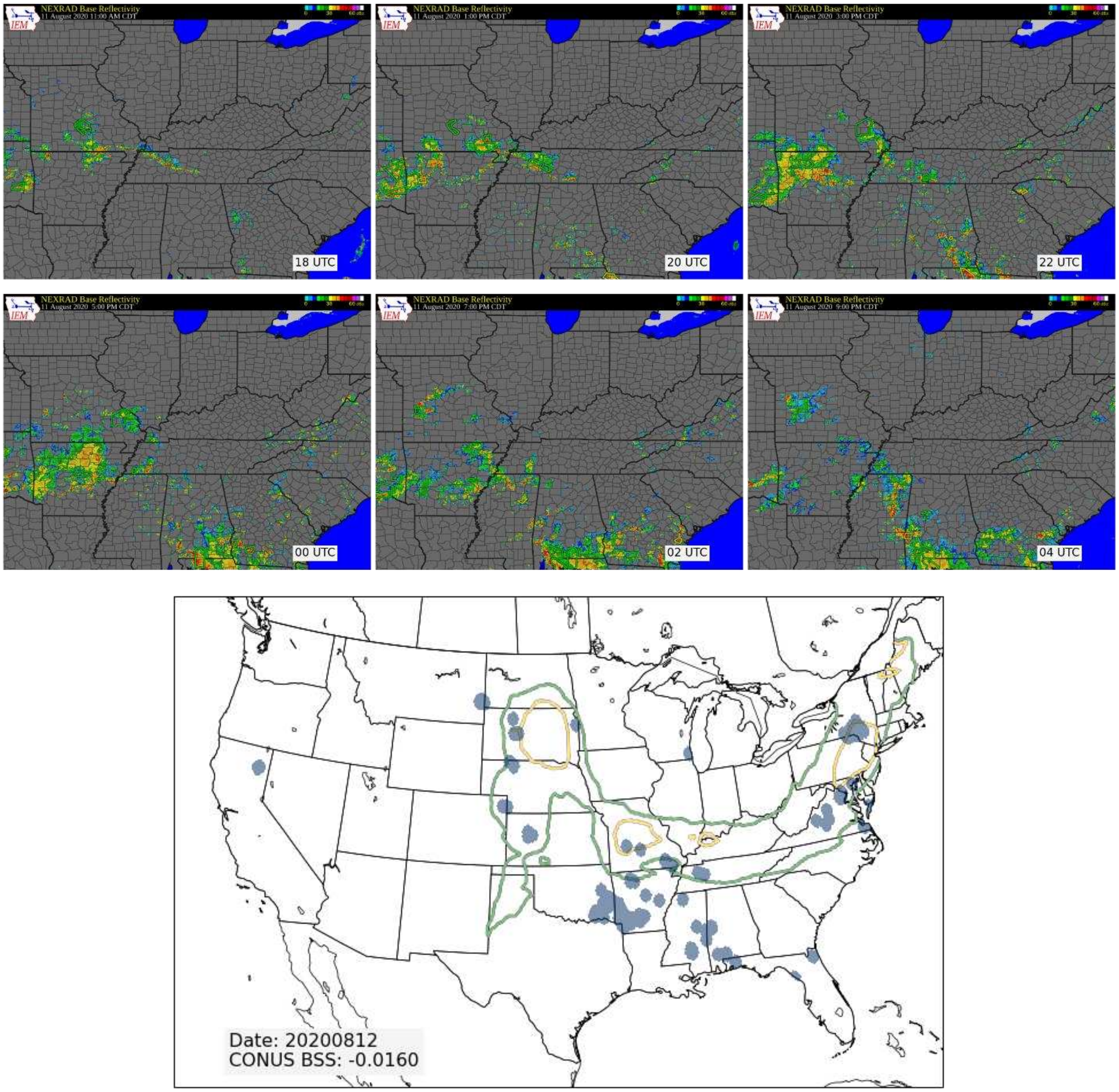


Figure 5.6. Same as Figure 5.4 but for the 24-hr forecast period spanning 12 UTC 11 August 2020 – 12 UTC 12 August 2020. This forecast case was categorized to R2-25th.

Figure 5.7 and 5.9 shows composite reanalysis for R3-25th and R4-25th, respectively, where CSU-MLP identifies risk areas in the northern plains and the upper Midwest. R3-25th is characterized by zonal mid- and upper-level flow across the CONUS, a weak east-west surface boundary across the upper Midwest associated with an occluded surface low pressure in southern Canada, a southerly to southwesterly low-level jet in the central CONUS, and 30-40 mm of PWAT. In most cases, convection forms at night on the cool-side of the weak surface boundary, where training of convective cells appears to occur at different locations. This is a typical pattern for nocturnal MCSs and "frontal" flash flood events (Maddox et al. 1979) resulting in a classification of "frontal" (Figure 5.8 provides an example radar evolution of a case of this regime). R4-25th (Figure 5.9) is characterized by a slow-moving large-scale trough across the western CONUS and a ridge axis centered over the central CONUS. Associated with the southwesterly mid-level flow across the Rockies is relatively low surface pressure in the lee of the Rockies from Montana down through New Mexico. At low-levels, southeasterly flow results in a

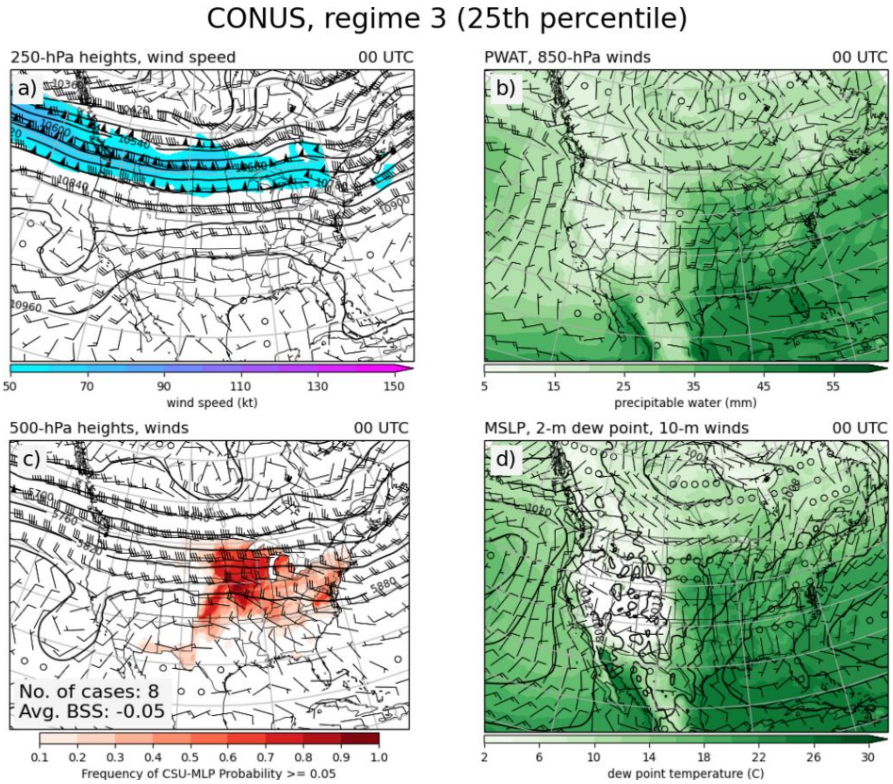


Figure 5.7. Same as Figure 5.3 but for R2-25th.

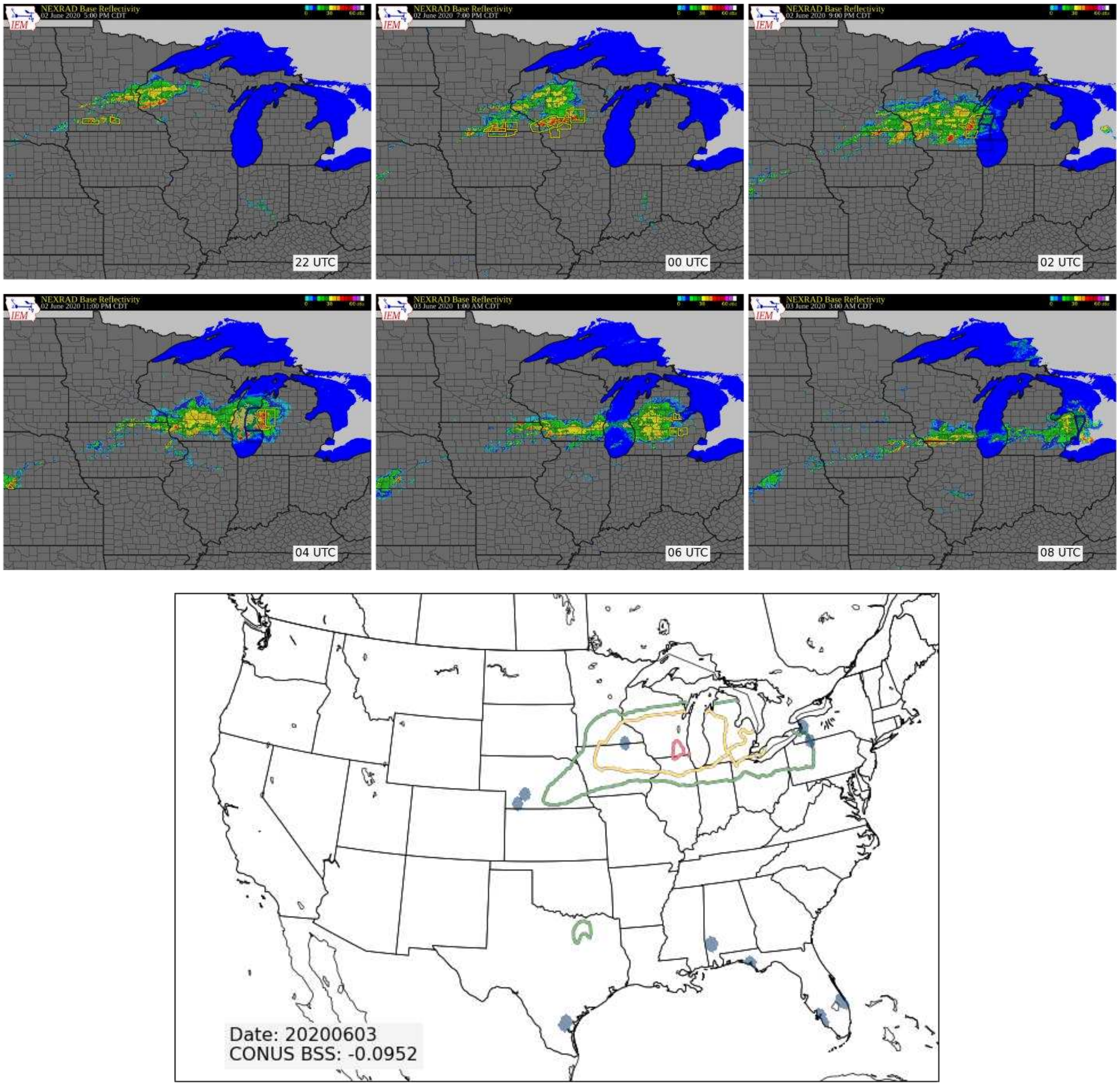


Figure 5.8. Same as Figure 5.4 but for the 24-hr forecast period spanning 12 UTC 2 June 2020 – 12 UTC 3 June 2020. This forecast case was categorized to R3-25th. Radar images are from 22, 00, 02, 04, 06, and 08 UTC within the forecast time period.

narrow corridor of moisture through the central and northern plains, with a northwest-southeast warm front positioned in the upper Midwest. Although CSU-MLP forecasts categorical risk across a large swath, highest frequencies and probabilities are located along this warm front. In these cases, convection is often initiated along the cool-side of the warm front, as there is a slight perpendicular component to 850-hPa winds that would result in forcing associated with the "frontal" event type. There

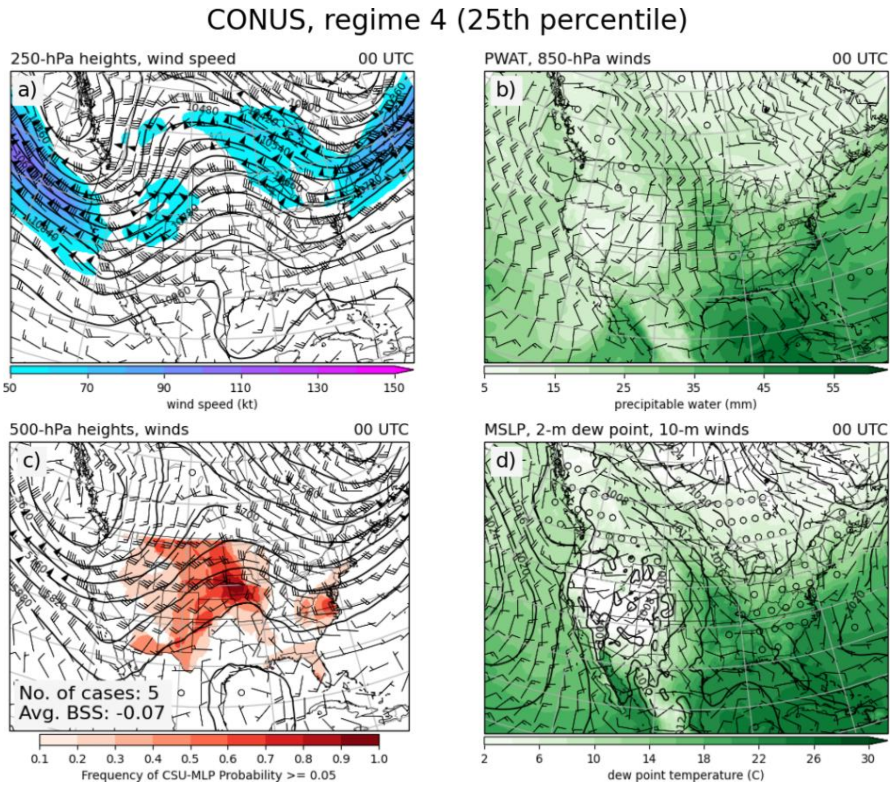


Figure 5.9. Same as Figure 5.3 but for R4-25th.

is also strong synoptic forcing associated with the large-scale trough, which provides an environment for convection initiation across the Rockies and upscale growth to MCSs across the warm sector in the plains. This slow-moving pattern also results in consecutive days of excessive rainfall potential, much like that of the "synoptic" events from Maddox et al. (1979). Due to the various forcing mechanisms for potentially excessive rainfall, this regime's forcing was classified as a hybrid between "synoptic" and "frontal".

The last synoptic regime among the worst-performing warm-season forecasts is that of offshore TCs in the northern Gulf of Mexico (R5-25th; Figure 5.10). TCs in these regimes tend to be slow-moving and/or approaching landfall during the forecast period, with widespread PWAT values greater than 40 mm and highest values around 55 mm to the east of the TC but offshore. A key feature of these events is the limited northward extent of inland moisture. This is in large part due to a relatively strong surface anticyclone across the eastern CONUS, with dry northeasterly low-level winds across the southeast.

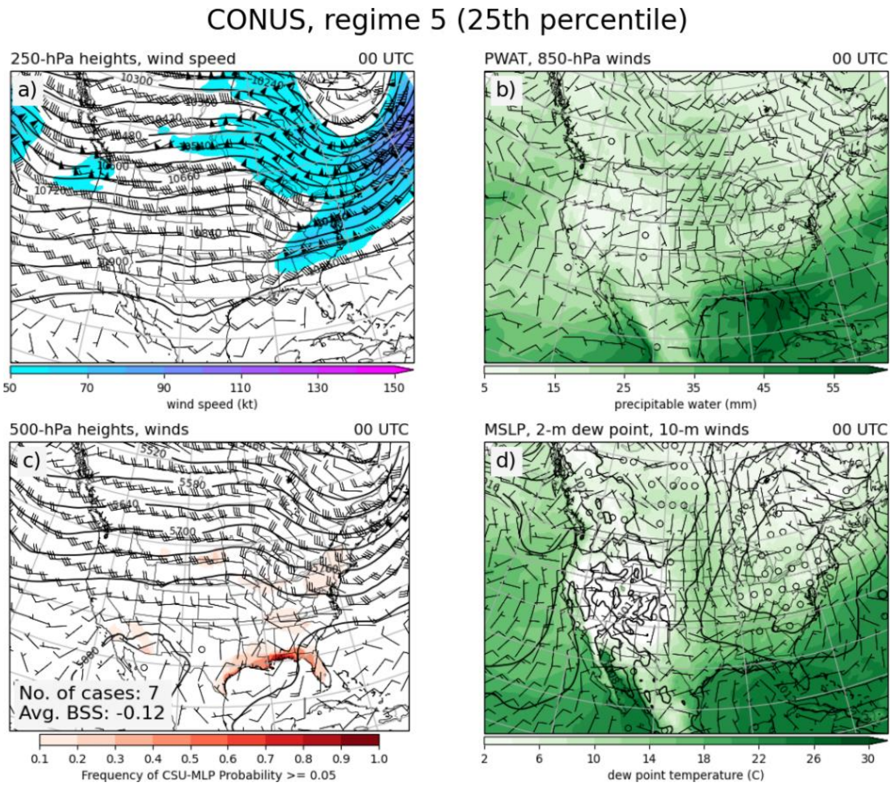


Figure 5.10. Same as Figure 5.3 but for R5-25th

Although CSU-MLP categorical probabilities are confined along the coastline, average forecast skill is the worst among the warm-season regimes.

Among all five of these regimes categorized, three of them have relatively few cases (R1-25th, R2-25th, and R4-25th). This might suggest that these synoptic regimes rarely occur for poor CSU-MLP forecasts, however, this is only when considering CONUS-wide forecast skill. When considering skill among the different training region domains, these regimes occur more often. Table 5.2 compares average daily BSS for CSU-MLP and the ERO for the five warm-season worst-performing CONUS regimes, and the different regions in which the same regime was identified for either best- or worst-performance at the regional level. R1-25th, R2-25th, and R4-25th all show increase in the number of cases identified for each regime at the regional level, and performance among these regions can vary. This suggests that CSU-MLP forecasts in these regimes do not perform poorly across multiple regions simultaneously for a given day. Instead, CSU-MLP seems to identify a broad area of risk with higher categorical probabilities with extreme rainfall usually occurring within a smaller area of the forecasted categorical risk. Except for R4-25th, this seems to be due to the lack of strong large-scale forcing. While Table 5.2 does not show R3-25th and R5-25th to have inter-regional variability, they do have analogous best-performing regimes (R3-75th and R4-75th, respectively) which suggests variable performance for those regimes as well.

Table 5.2. Regional daily BSS average statistics for regimes analogous to worst-performing CONUS warm-season regimes.

CONUS Regime No.	Region, Percentile	No. cases	BSS Avg.	ERO BSS Avg.	BSS Diff.
1	CONUS 25th	4	-0.03	0.02	-0.05
	NE 25th	9	-0.17	0.01	-0.19
	SW 25th	5	-0.80	-0.01	-0.78
	SW 75th	5	0.07	0.02	0.05
2	CONUS 25th	4	-0.03	0.00	-0.03
	NE 25th	9	-0.14	-0.07	-0.07
	NE 75th	12	0.16	0.08	0.08
	MDWST 25th	13	-0.30	-0.10	-0.19
3	CONUS 25th	8	-0.05	0.05	-0.10
	NGP 25th	4	-0.18	-0.01	-0.18
	MDWST 25th	4	-0.26	-0.05	-0.20
	MDWST 75th	6	0.11	0.13	-0.02
4	CONUS 25th	5	-0.07	0.04	-0.11
	NGP 25th	9	-0.34	0.02	-0.36
	NGP 75th	8	0.14	0.11	0.03
	MDWST 75th	4	0.13	0.17	-0.04
	ROCK 75th	10	0.13	0.07	0.06
	SW 25th	8	-1.01	-0.31	-0.70
5	CONUS 25th	7	-0.12	-0.03	-0.09
	SE 25th	9	-0.38	-0.20	-0.19

5.2.2 Spring/Fall Regimes

Five regimes were identified that occurred during the transition seasons of spring and fall. Figure 5.11 shows reanalysis composites for R6-25th that resulted in poor CSU-MLP forecasts across the southern CONUS. It is characterized by a large-scale trough axis that extends through the southwestern CONUS and into northwest Mexico, PWAT ranging from 20-30 mm across the risk area, and a relatively strong 250-hPa jet stream. There is also the presence of a positive tilt 500-hPa trough across the central CONUS. Associated with this trough is an anticyclone centered over the central CONUS with a cold front in southern Texas intersected with a PWAT maximum in southeast Texas. Among these cases, there was a cold front that was either moving southeastward or stationary, with southerly 850-hPa flow and rainfall to the north of the surface boundary. Typical cases of this regime included a broad area of moderate stratiform precipitation across the southern plains, with isolated instances of embedded convection. However, the progressive nature of the cold front and/or the relatively modest PWAT values (25-30 mm) limit the occurrence of excessive rainfall. Convection in these cases is clearly identifiable with the cold front with the parent surface low well removed from the area, thus, forcing is determined to be "cold front". A total of 17 cases were identified in R6-25th- this was the most numerous among worst-performing regimes. This regime will be revisited in the final section of this chapter where it will be contrasted to the most numerous best-performing regime.

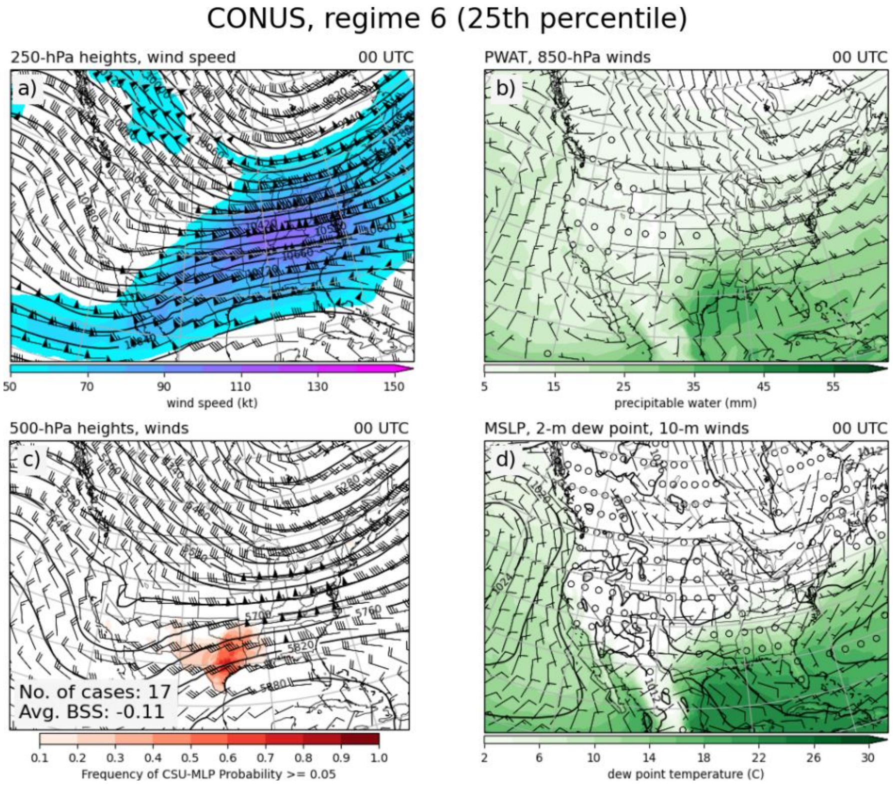


Figure 5.11. Same as Figure 5.3 but for R6-25th

R7-25th (Figure 5.12) is similar to R6-25th, except southerly low-level flow has strengthened and extended down to the surface due to lowering surface pressures in lee of the Rockies as the southwest mid-level trough is positioned slightly eastward. A warm front is often moving quickly to the north in these cases, and this is usually where convection forms (if any), thus, the forcing mechanism is determined to be "warm front". Dew point temperatures remain at or below 10°C around the area of risk, and PWAT is around 20-30 mm. Once again, it seems the modest moisture parameters are the primary inhibitors for excessive rainfall, as some cases have very little precipitation within and around CSU-MLP risk areas, let alone extreme precipitation.

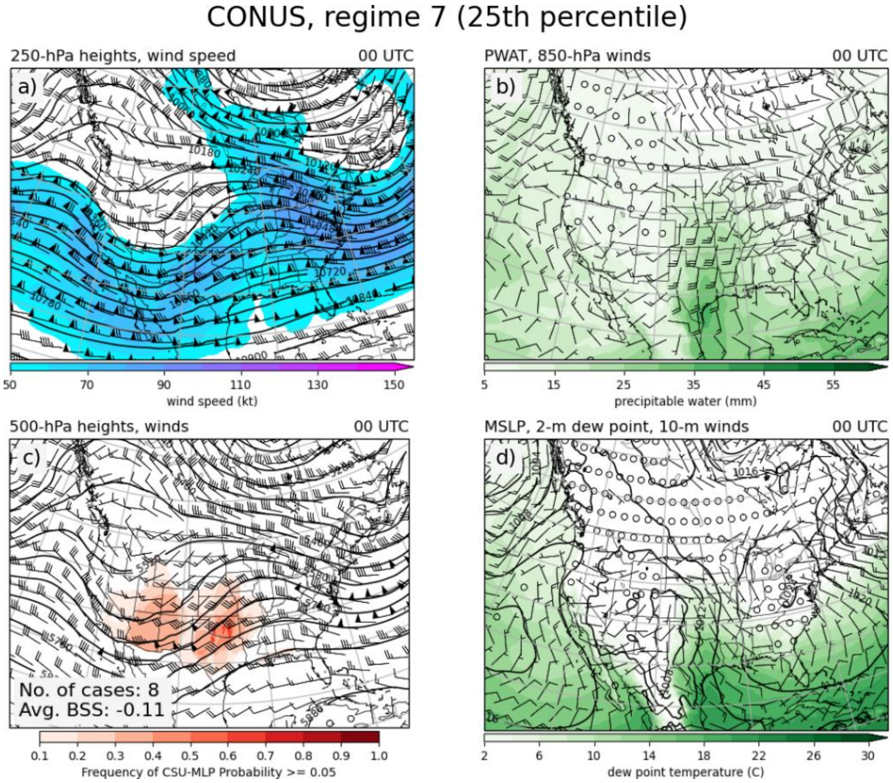


Figure 5.12. Same as Figure 5.3 but for R7-25th

Figure 5.13 shows the reanalysis composite for R8-25th, which is characterized by a strong large-scale trough similar to that of R6-25th and R7-25th, but with a stronger downstream ridge. There is also a surface cyclone across the central plains, situated beneath the left exit region and right entrance region of two separate jetstreaks, with an eastward extending warm front positioned across the Midwest. The strong cyclone induces a composite 20-30 kt southerly 850-hPa flow within a broad moist sector characterized by modest dew point temperatures (10-16 C) and PWAT (20-30 mm). This southerly flow impinges upon the warm front and turns westward, with the higher PWAT values north of the surface warm front and in the "wrap-around" region of the cyclone. In these cases, convection often formed along the warm front near the eastward moving surface cyclone, as this area was also the most frequent risk of excessive rainfall from CSU-MLP. Thus, forcing was categorized as "ETC". However, despite the very strong forcing associated with these ETCs of this regime, it seems the rather modest moisture fields played a role in the limited coverage of excessive rainfall observations.

Figures 5.14 and 5.15 show reanalysis composite of R9-25th and R10-25th, respectively, which both highlight areas of risk for excessive rainfall along the east coast. R9-25th is characterized by a trough axis through the northeast, with a jetstreak located at the base of the trough. In correspondence with

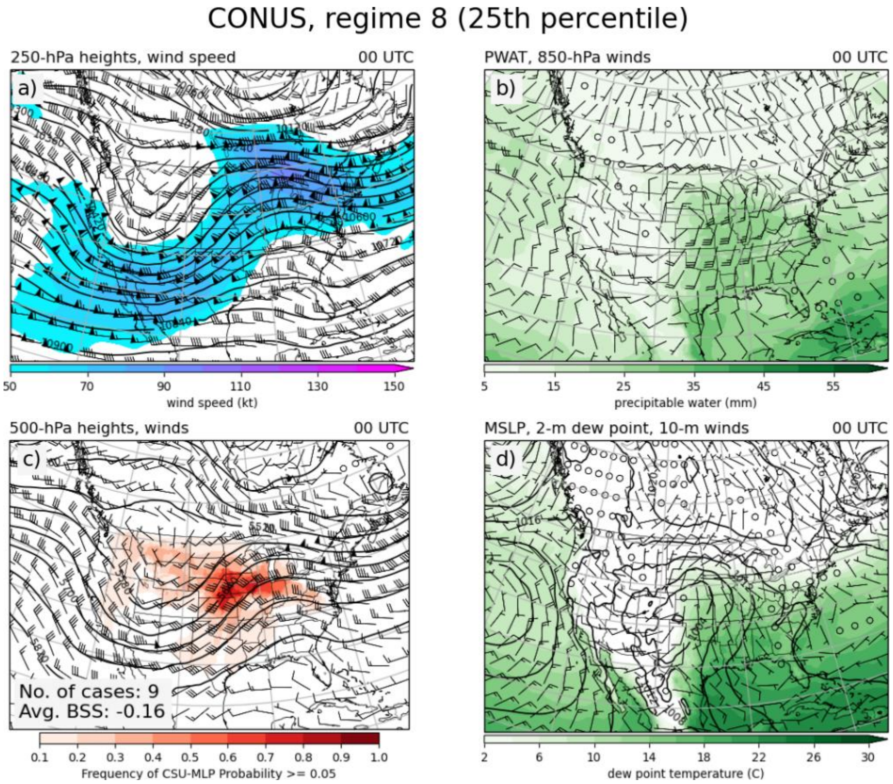


Figure 5.13. Same as Figure 5.3 but for R8-25th

the left exit region of the jetstreak is a surface cyclone off the coast. Higher dew point temperatures are removed from land, but southeasterly 850-hPa winds that wrap around the cyclone show elevated PWAT values (15-20 mm) that makes its way inland. Forcing for this regime is classified as "ETC", however, convective precipitation rarely occurred over land. R10-25th has a highly amplified pattern across the CONUS, with a trough axis across the east and an occluded surface cyclone in eastern Canada. Highest dew point temperatures and PWAT are located offshore, with PWAT less than 20 mm adjacent to the coast. A strong, eastward moving cold front is located across the risk area, which is well removed from the parent surface low. Thus, forcing is categorized as "cold front". Again, the lack of a broad area of higher low-level and column moisture despite strong forcing from large-scale troughs seems to be an inhibiting factor for excessive rainfall.

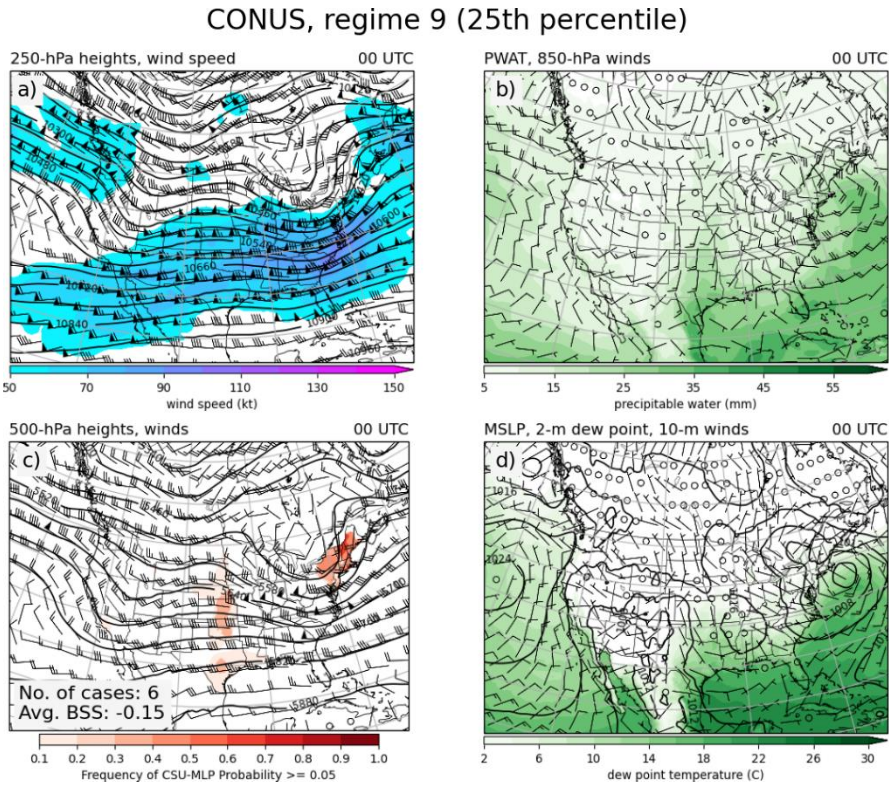


Figure 5.14. Same as Figure 5.3 but for R9-25th

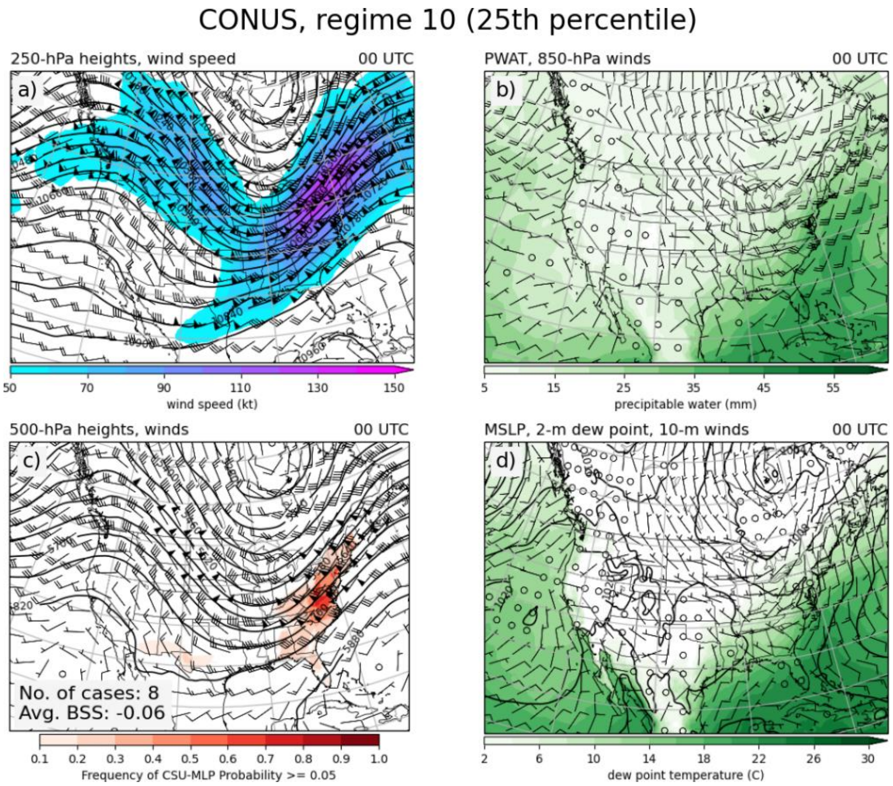


Figure 5.15. Same as Figure 5.3 but for R10-25th

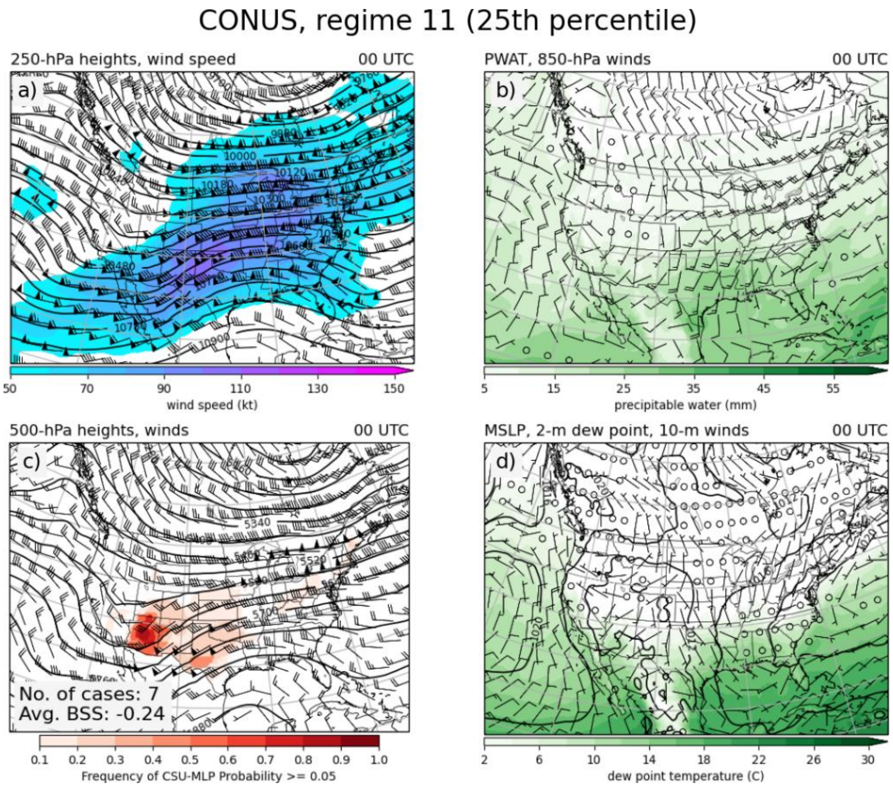
In contrast to some of the worst-performing warm-season regimes, all the spring/fall regional regimes corresponding to the CONUS-based regimes show poor average skill (Table 5.3). Even compared to the ERO, which does not perform well for each of these regimes, skill from CSU-MLP is much lower on average. Despite the strong synoptic forcing associated with all these events, modest low-level and column moisture appears to be the prohibiting factor for excessive rainfall.

5.2.3 Winter Regimes

Among the worst-performing CONUS regimes, there was only one regime where most occurrences were in the winter. This regime is characteristic of an inland penetrating cold front with southwesterly moisture transport that affects the southwestern CONUS (e.g. Rivera et al. 2012). Figure 5.16 shows this regime, with a large-scale trough axis centered across the western CONUS and southwesterly mid-level flow across the southwest CONUS. This composite was formed using the 12 UTC reanalysis among the cases, as the excessive rainfall threat was for rainfall that occurred during the first

Table 5.3. Regional daily BSS average statistics for regimes analogous to worst-performing CONUS spring/fall regimes.

CONUS Regime No.	Region, Percentile	No. cases	BSS Avg.	ERO BSS Avg.	BSS Diff.
6	CONUS 25th	17	-0.11	-0.04	-0.07
	SGP 25th	24	-1.81	-0.29	-1.52
	SE 25th	9	-0.63	-0.77	0.15
7	CONUS 25th	8	-0.11	-0.01	-0.10
	SGP 25th	16	-1.6	-0.18	-1.42
	SE 25th	9	-0.32	-0.26	-0.06
8	CONUS 25th	9	-0.16	-0.01	-0.15
	NGP 25th	11	-0.8	-0.07	-0.73
	MDWST 25th	8	-1.35	-0.33	-1.02
	SGP 25th	4	-1.13	0.00	-1.13
9	CONUS 25th	6	-0.15	-0.01	-0.13
	NE 25th	6	-1.77	-0.04	-1.73
10	CONUS 25th	8	-0.06	0.01	-0.06
	SE 25th	5	-0.03	0.01	-0.04


CONUS, regime 11 (25th percentile)

a) 250-hPa heights, wind speed 00 UTC

b) PWAT, 850-hPa winds 00 UTC

c) 500-hPa heights, winds 00 UTC

d) MSLP, 2-m dew point, 10-m winds 00 UTC

Figure 5.16. Same as Figure 5.3 but for R11-25th

half of the day-1 time period. Composite PWAT of 20-25 mm is located ahead of a mid-level front to the southwest of the Baja peninsula, and this moisture penetrates southern Arizona and New Mexico. However, inland PWAT values are below 20 mm which would appear to limit excessive rainfall in the area despite the strong trough and orographic ascent along the terrain in central Arizona. Table 5.4 shows the skill statistics for this regime when identified looking at only regional forecasts, with CSU-MLP having much lower skill compared to ERO. There was a large enough sample of cases (4) of this regime where forecasts performed well, but these are largely outnumbered by worst-performing cases (15).

Table 5.4. Regional daily BSS average statistics for regimes analogous to worst-performing CONUS winter regimes.

CONUS Regime No.	Region, Percentile	No. cases	BSS Avg.	ERO BSS Avg.	BSS Diff.
11	CONUS 25th	7	-0.24	0.00	-0.24
	SW 25th	15	-0.52	-0.18	-0.34
	SW 75th	4	0.11	0.04	0.07
	ROCK 25th	7	-0.21	0.00	-0.21

5.3 CONUS Best-Performing Synoptic Regimes

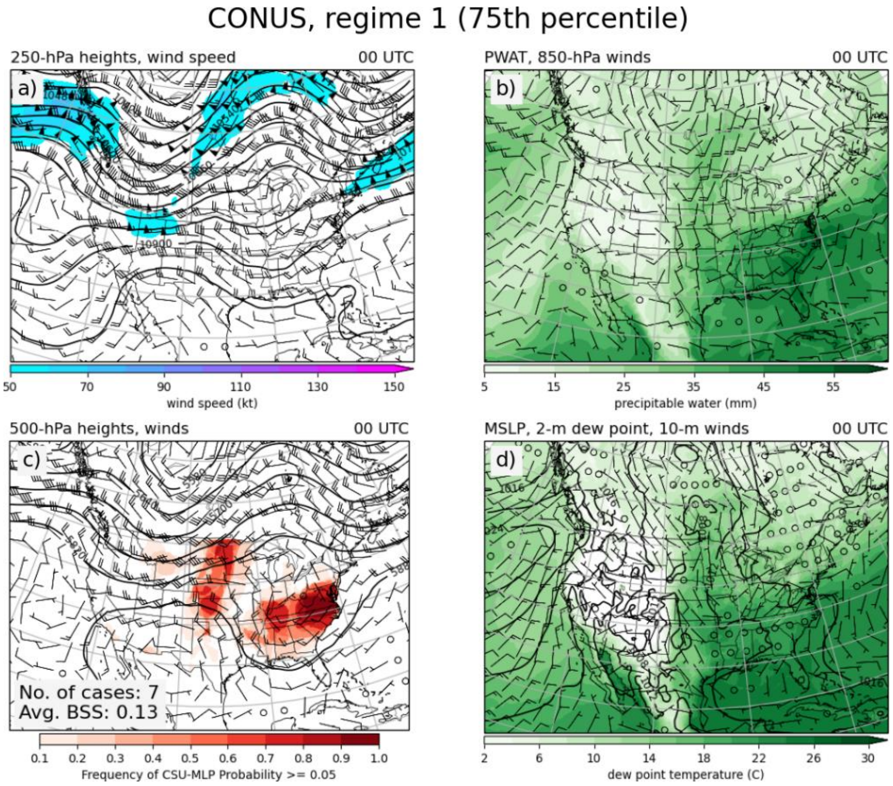
Table 5.5 shows average daily BSS statistics for best-performing CSU-MLP forecasts evaluated for the CONUS, along with the ERO daily BSS for those given cases. CSU-MLP forecasts have either equal or slightly greater skill than the ERO for each of the regimes. However, based on the daily BSS difference distributions from Figure 4.8, these regimes are not necessarily forecast regimes where CSU-MLP skill is greatest compared to the ERO (except for R7-75th where average daily BSS difference is 0.11).

5.3.1 Warm-Season Regimes

Figure 5.17 shows reanalysis composites for R1-75th, characterized by a slow-moving trough across the northern Rockies and a COL in the eastern CONUS. Although there are two areas that CSU-MLP frequently forecasts for, this regime was composited to focus on the COL in the eastern CONUS, as

Table 5.5. Same as Table 5.1 but for best-performing synoptic regimes using CONUS-wide daily BSS.

Regime No.	No. cases	BSS Avg.	ERO BSS Avg.	BSS Diff.	Other Regions	Regime Type	Season
1	7	0.13	0.10	0.03	NE, SE	COL	JJA
2	11	0.17	0.14	0.02	NGP	synoptic	JJA
3	9	0.12	0.12	0.00	MDWST	frontal	JJA
4	8	0.25	0.23	0.02	SE, SGP	TC	JJA
5	5	0.17	0.16	0.01	SE	TC	SON
6	31	0.16	0.15	0.01	SE, SGP, MDWST, NGP	synoptic	MAM
7	7	0.18	0.11	0.08	NE	ETC	MAM
8	6	0.19	0.08	0.11	PCST	AR, orography	DJF
9	14	0.23	0.19	0.04	SE, SGP, MDWST, NE	ETC	DJF


CONUS, regime 1 (75th percentile)

250-hPa heights, wind speed 00 UTC

PWAT, 850-hPa winds 00 UTC

500-hPa heights, winds 00 UTC

MSLP, 2-m dew point, 10-m winds 00 UTC

Figure 5.17. Same as Figure 5.3 but for R1-75th

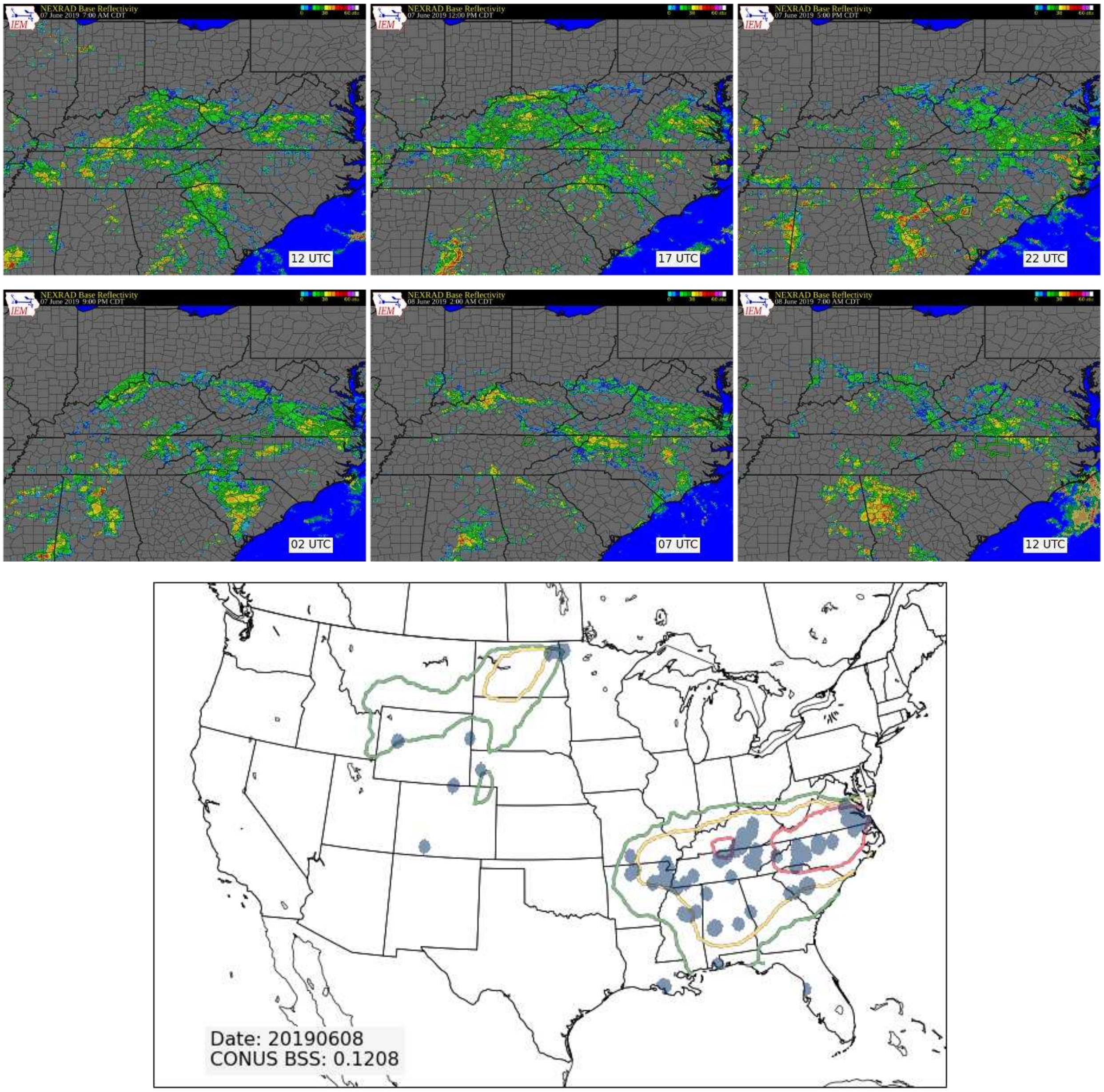


Figure 5.18. Same as Figure 5.4 but for the 24-hr forecast period spanning 12 UTC 7 June 2019 – 12 UTC 8 June 2019. This forecast case was categorized under R1-75th. Radar images are from 12, 17, 22, 02, 07, and 12 UTC within the forecast time period.

performance in the northern plains was variable for these cases. Associated with the COL, an inverted surface trough axis extends through North Carolina, with easterly upslope flow impinging upon the Appalachians. There is abundant low-level moisture adjacent to the surface trough, with widespread dew point temperatures greater than 18 C and PWAT ranging from 35-45 mm. The COL also provides instability for widespread convection, as often happens with COL in the summer as they move over a moist low-level air mass (Nieto et al. 2008; Shepherd et al. 2011). Daytime convection increases around the risk area, with cold-pool interactions playing a large role in the evolution of convection (Figure 5.18 shows radar evolution of an example case of this regime). Thus, forcing is categorized as "COL ". R2-75th (Figure 5.19) is nearly identical to R1-75th, but cases are selected such that high skill forecasts occur on all days across the central and northern plains. This region is influenced by a slow-moving large-scale trough across the northern Rockies. Southerly low-level flow transports a corridor of 25-35 mm PWAT through the central and northern plains. In these cases, there is often a weak surface cyclone that develops in the northern plains, with a slow-moving front serving as focus for convective initiation. Multiple MCSs develop and propagate into the warm sector at various times of the day. This regime has close resemblance to the "synoptic" regime from Maddox et al. (1979) and is categorized as such.

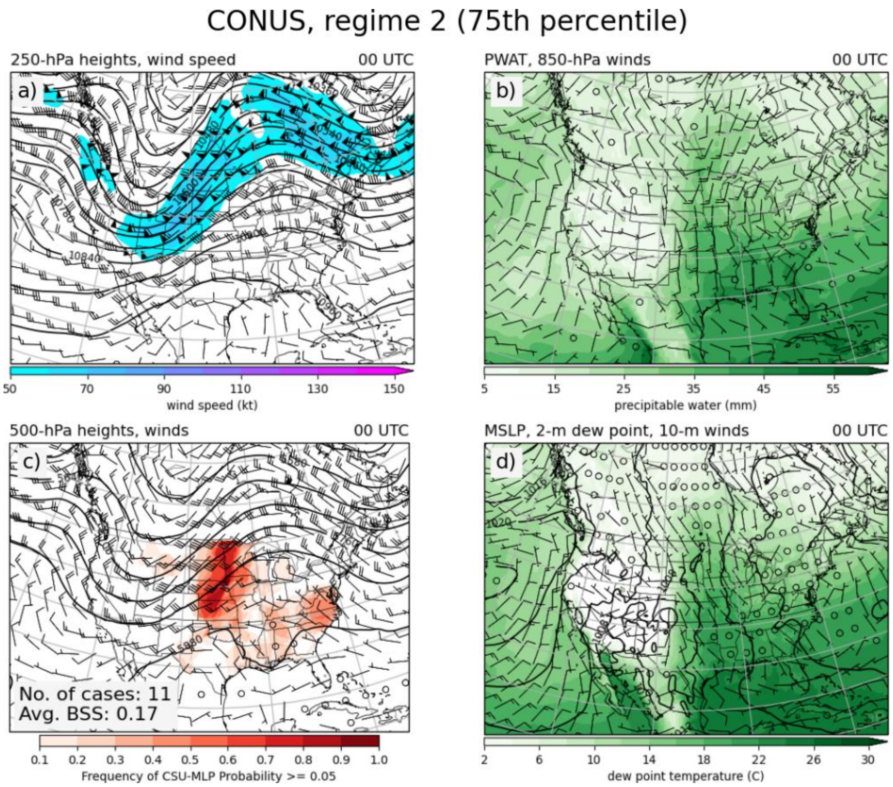


Figure 5.19. Same as Figure 5.3 but for R1-75th

Figure 5.20 shows R3-75th, which is analogous to R3-25th from the worst-performing cases (Figure 5.7). This is characterized by zonal mid- and upper-level flow across the CONUS, a weak boundary across the upper Midwest, a southerly to southwesterly low-level jet in the central CONUS, and PWAT approaching 40 mm. Nocturnal convection is often initiated along the cool side of the surface front, with numerous excessive rainfall observations are across the forecasted risk area (see radar case example in Figure 5.21). Differences in composite PWAT between the two show that the best-performing regime has higher PWAT across the central plains and upper Midwest (Figure 5.22). Geopotential heights are also higher across southern Canada for the best-performing regime, with a ridge in height difference implying a weaker zonal jet in the northern plains.

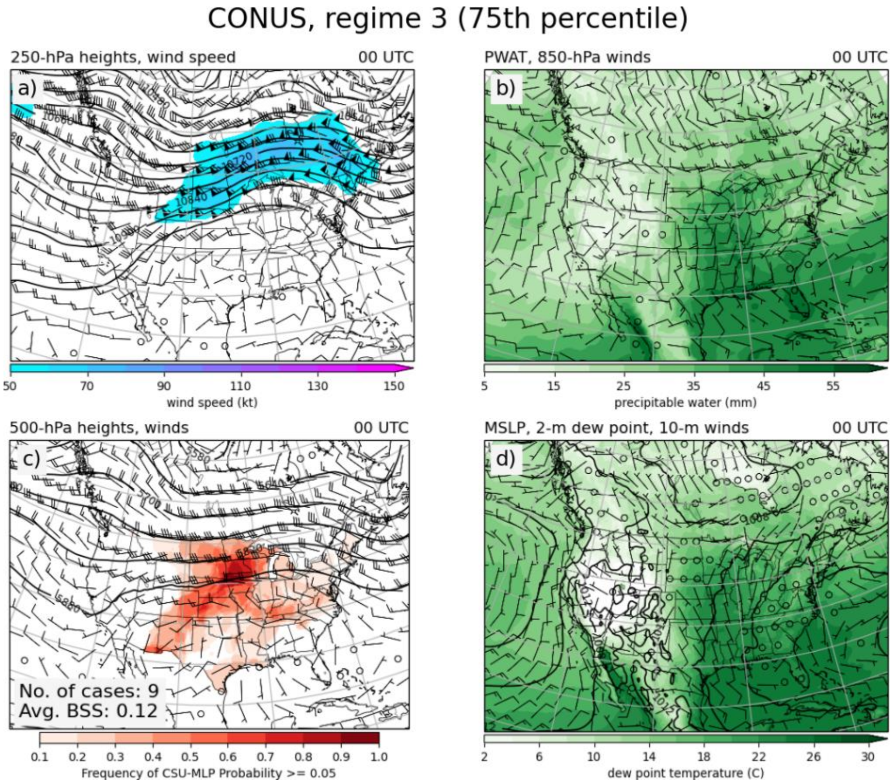


Figure 5.20. Same as Figure 5.3 but for R1-75th

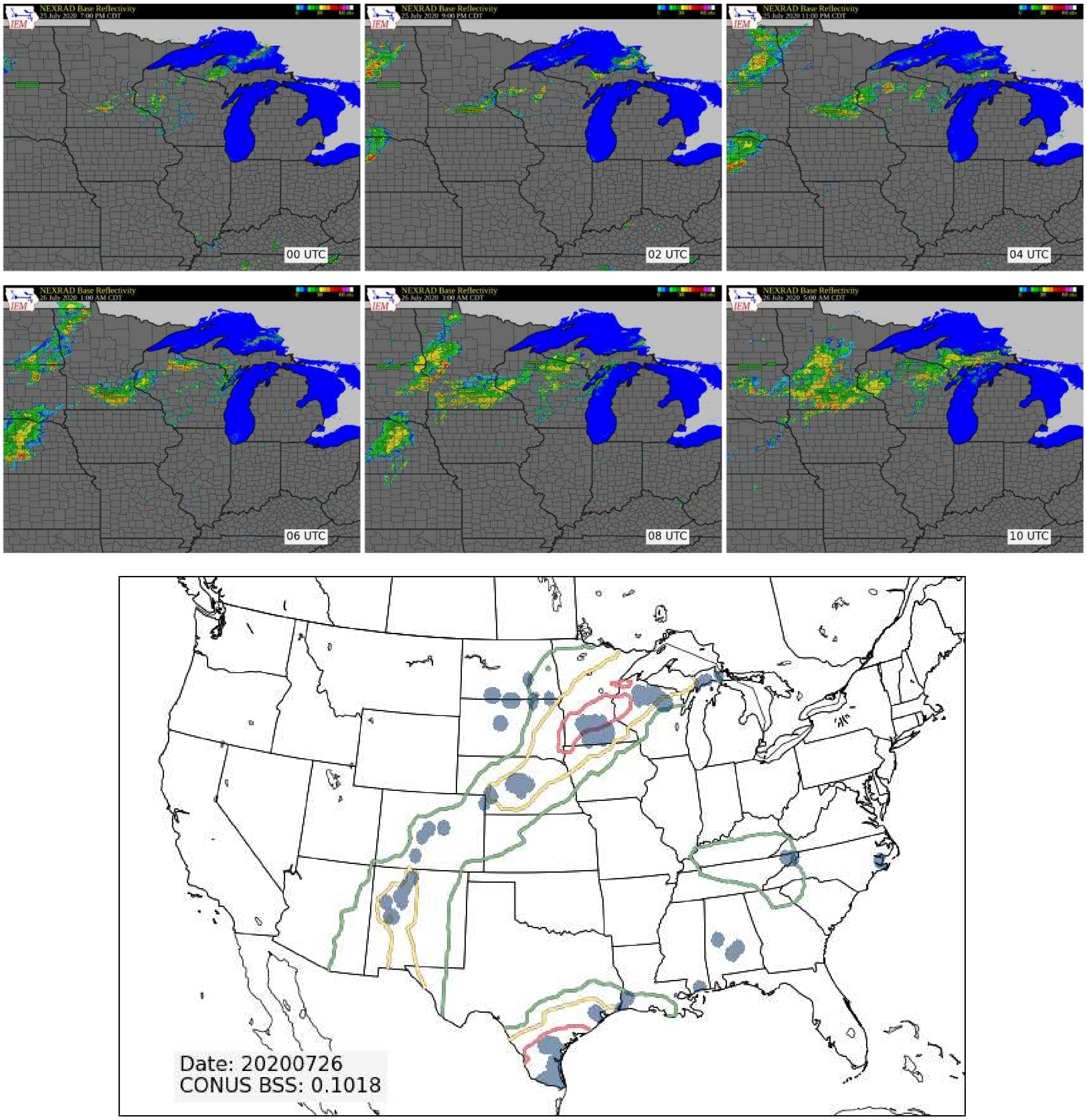


Figure 5.21. Same as Figure 5.4 but for the 24-hr forecast period spanning 12 UTC 25 July 2020 – 12 UTC 26 July 2020. This forecast case was categorized under R3-75th.

Figures 5.23 and 5.25 show landfalling TC regimes along the gulf coast (R4-75th) and inland moving TC's in the southeast (R5-75th), respectively. These regimes are separated amongst each other by whether landfall occurred during the forecast time period. The landfalling TCs are characterized by widespread PWAT above 40 mm, with a maximum over 50 mm in Louisiana. Similar to R5-25th from the

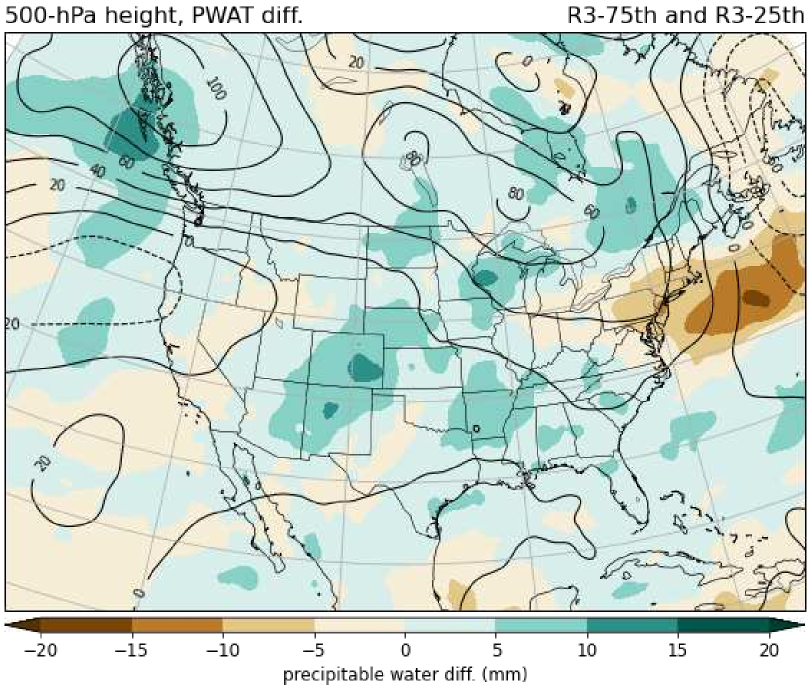


Figure 5.22. Difference plot of ERA5 reanalysis composite 500-hPa geopotential heights (contour) and composite PWAT (color fill) between R3-75th and R3-25th.

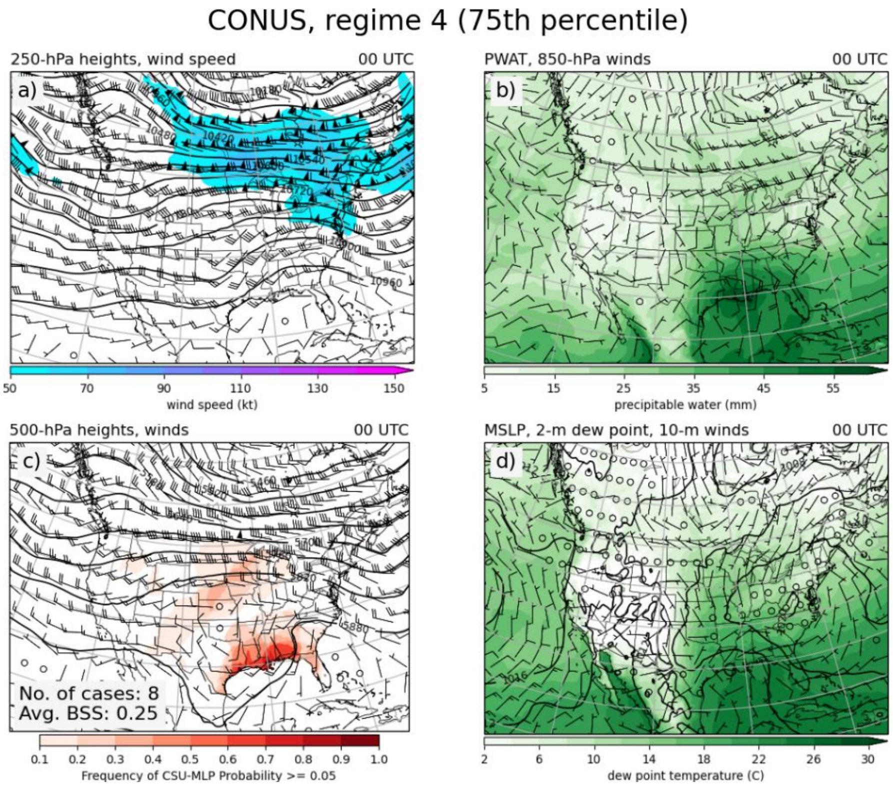


Figure 5.23. Same as Figure 5.3 but for R4-75th

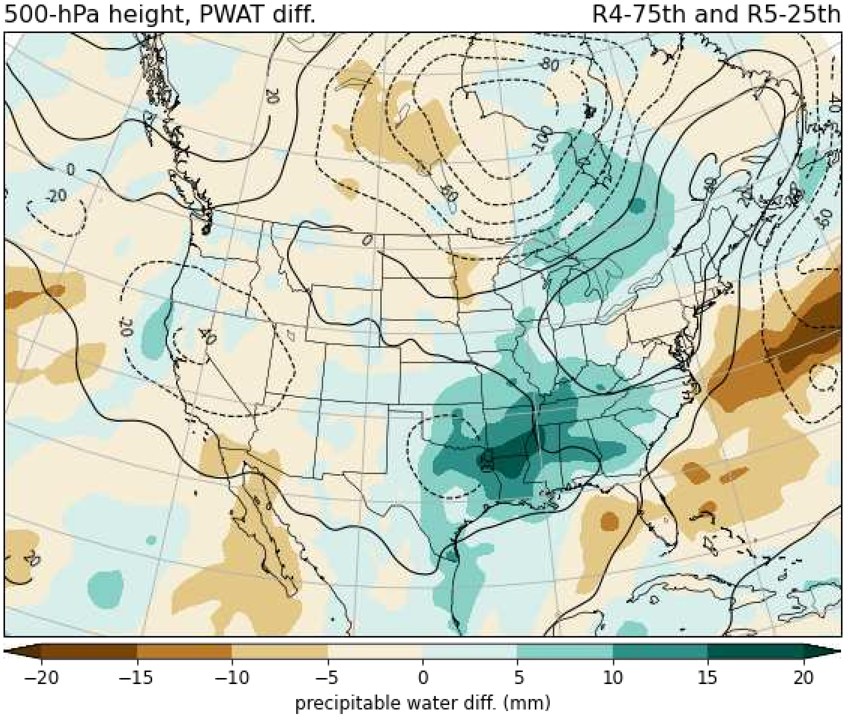


Figure 5.24. Same as Figure 5.16 but between R4-75th and R5-25th.

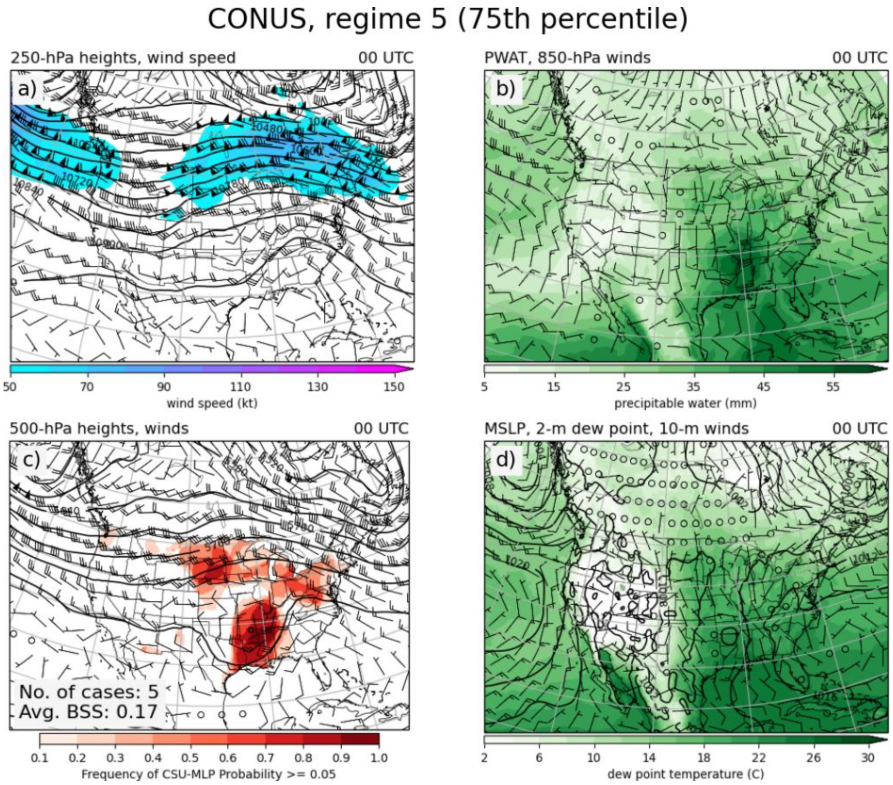


Figure 5.25. Same as Figure 5.3 but for R4-75th

worst-performing forecasts, there is a strong surface anticyclone centered over the northeast CONUS. However, there is greater low-level moisture and PWAT to the north surrounding the TC (Figure 5.24) which results in a larger coverage of CSU-MLP risk that yields a high skill forecast. Inland moving TC's also have high skill as the associated large core of high PWAT is well inland which results in a broad area of excessive rainfall risk.

Among all five of these regimes, CSU-MLP forecasts performed well for analogous regimes identified at the regional level (Table 5.6). R2-25th was the only one with a worst-performing group at the regional level (NGP), however, there were more than three times as many best-performing cases (16) than worst-performing cases (5). The ERO performed well in all the regional forecasts, especially for the R4-75th and R5-75th where the ERO outperforms CSU-MLP in the SGP and SE regions. In contrast to the worst-performing warm-season regimes, these regimes had either relatively stronger forcing among similar moisture (R1-75th, R2-75th, R4-75th, R5-75th vs R1-25th, R2-25th) or higher moisture among weaker forcing (R3-75th vs R3-25th).

Table 5.6. Regional daily BSS average statistics for regimes analogous to best-performing CONUS warm-season regimes.

CONUS Regime No.	Region, Percentile	No. cases	BSS Avg.	ERO BSS Avg.	BSS Diff.
1	CONUS 75th	7	0.13	0.10	0.03
	NE 75th	7	0.19	0.15	0.04
	SE 75th	7	0.17	0.15	0.02
2	CONUS 75th	11	0.17	0.14	0.02
	NGP 75th	16	0.19	0.11	0.08
	NGP 25th	5	-0.18	0.00	-0.18
3	CONUS 75th	9	0.12	0.12	0.00
	MDWST 75th	6	0.11	0.13	-0.02
4	CONUS 75th	8	0.25	0.23	0.02
	SE 75th	8	0.34	0.28	0.06
	SGP 75th	7	0.25	0.35	-0.11
5	CONUS 75th	5	0.17	0.16	0.01
	SE 75th	10	0.2	0.26	-0.06

5.3.2 Spring/Fall Regimes

Figure 5.26 shows the reanalysis composite for R6-75th, which is a characterized by a slow-moving large-scale trough axis over the central CONUS. Lower surface pressures are present across the central CONUS just ahead of the mid- and upper-level trough, denoting the average location of quasi-stationary fronts among the cases in this regime. A corridor of PWAT ranging from 30-40 mm is in the warm sector ahead of the front, along with a southwesterly LLJ parallel to the surface boundary from east Texas up through the Great Lakes. In these cases, there is often a shortwave that travels northeast downstream of the trough axis that serves as a focus for convective initiation along and near the quasi-stationary surface boundary. Convection forms into a variety of MCSs that traverse out into the warm sector to the south and/or east, with repeated episodes of heavy rainfall due to the slow-moving large-

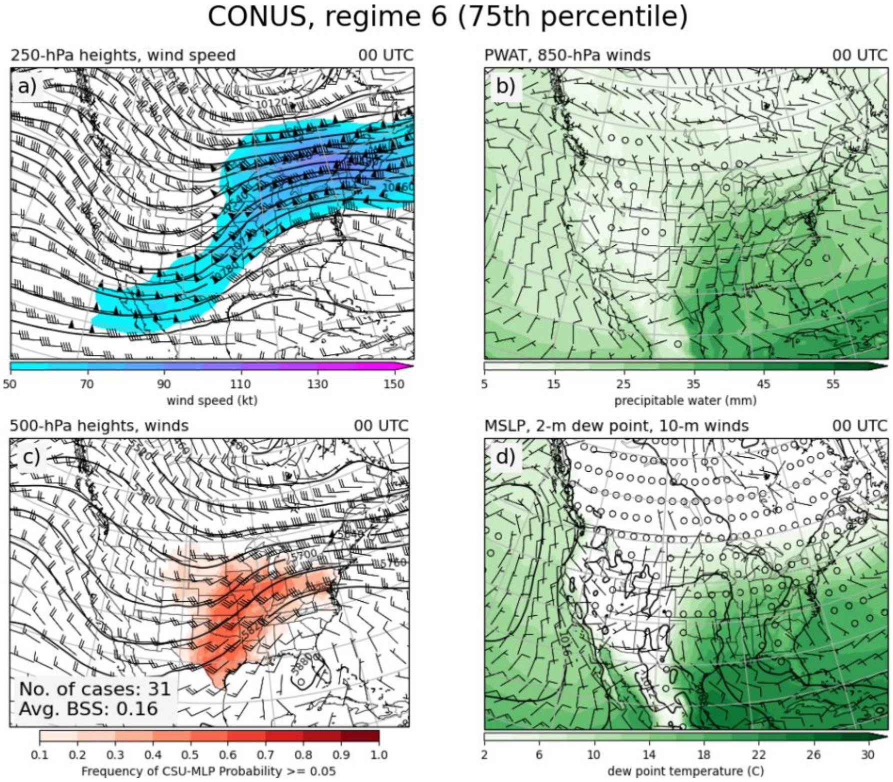


Figure 5.26. Same as Figure 5.3 but for R4-75th

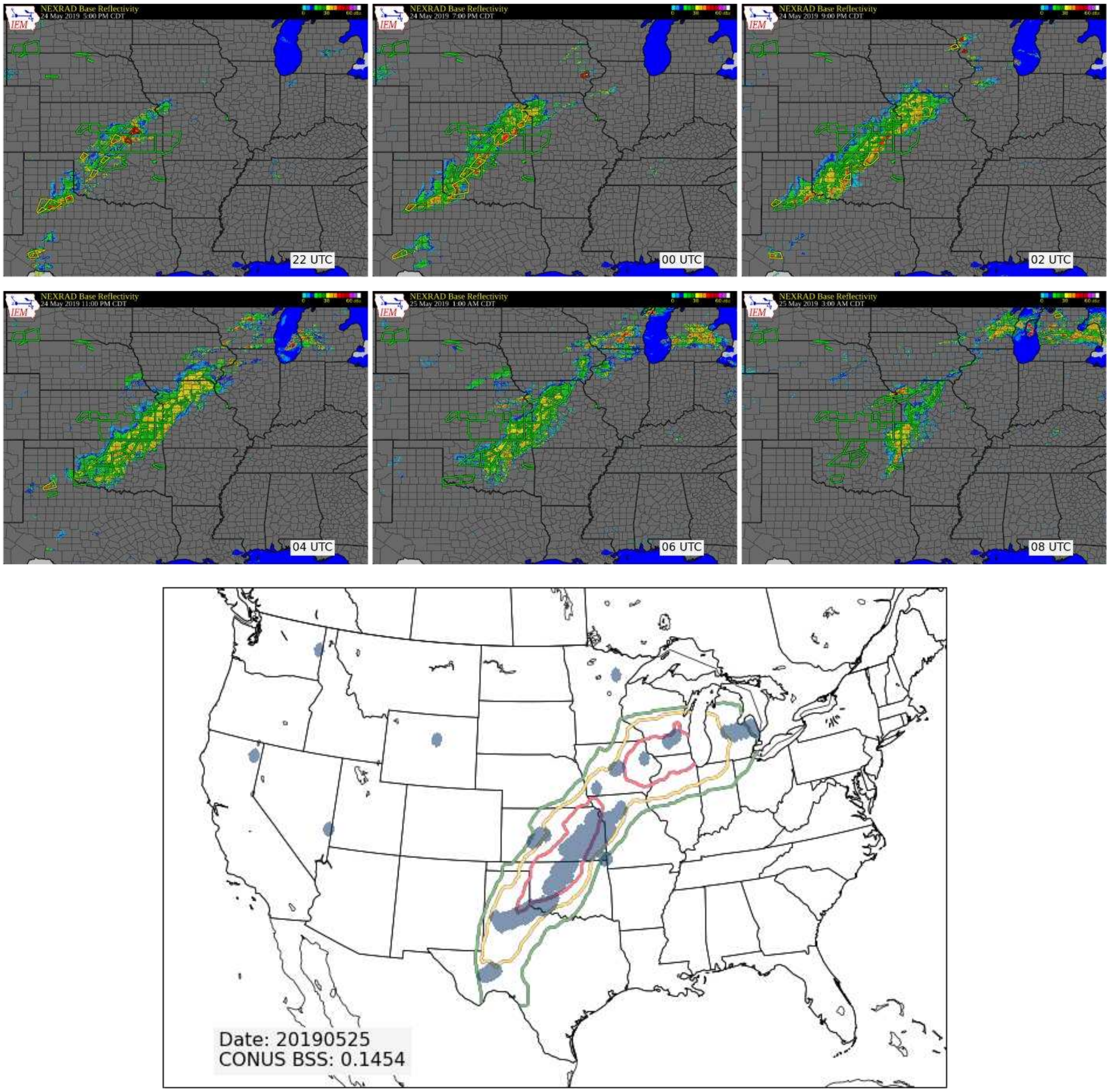


Figure 5.27. Same as Figure 5.4 but for the 24-hr forecast period spanning 12 UTC 24 May 2019 – 12 UTC 25 May 2020. This forecast case was categorized under R6-75th.

scale pattern. Figure 5.27 and 5.28 show case examples in which MCSs from on consecutive days across the central CONUS, resulting in widespread flash flood warnings. This pattern resembles nearly exactly the "synoptic" type flash flood events from Maddox et al. (1979) and is categorized as such. This regime had the greatest number of cases (31) for any identified regime and represents a typical best-performing

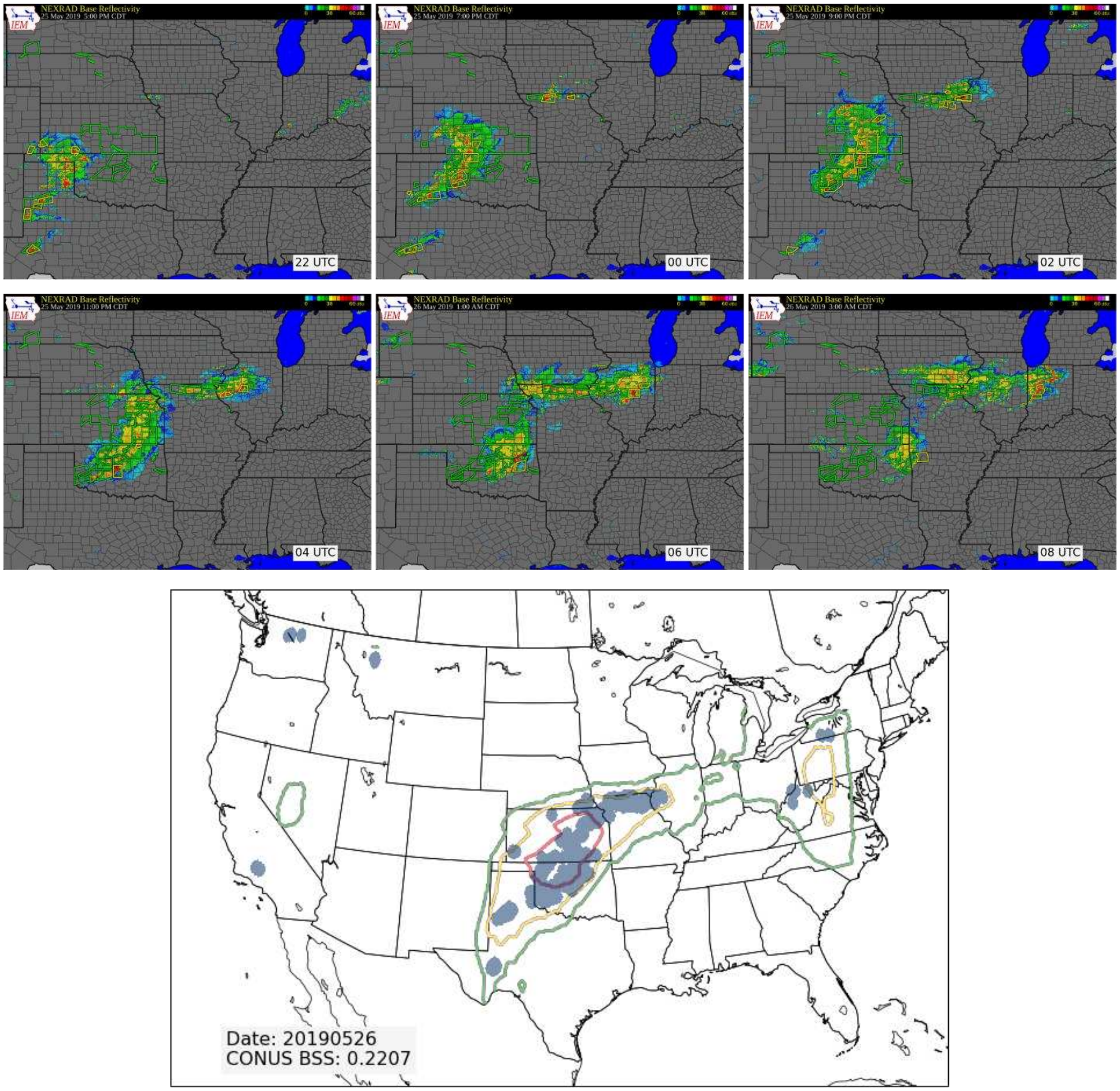


Figure 5.28. Same as Figure 5.4 but for the 24-hr forecast period spanning 12 UTC 25 May 2019 – 12 UTC 26 May 2020. This forecast case was categorized under R6-75th.

forecast scenario with moderate forcing and relatively higher moisture events (this regime will be compared to R6-25th at the end of the chapter).

The other spring/fall regime of the best-performing forecasts is R7-75th (Figure 5.29), which is characterized by a large-scale trough axis across the eastern CONUS and a strong surface cyclone moving northeast across the northeast CONUS. Strong 850-hPa southwesterly flow transports a swath of

25-35 mm PWAT east surface cold front. This somewhat analogous to R10-25th from the worst-performing forecasts. However, this regime has the cyclone center much further west such that the attending warm and cold front are across land along with the moist sector. In addition, the mid- and upper-level trough-ridge pattern is more amplified, with the primary jetstreak residing across the top of the ridge axis. Forcing is categorized as "ETC" and is analogous to pattern "C2" from Agel et al. (2019).

Performance for analogous spring/fall regimes categorized at the regional level all have slightly higher skill than the ERO (Table 5.7). There are numerous regional cases for R6-75th, particularly for SGP forecasts with 35. CSU-MLP average daily BSS is slightly higher compared to the ERO which is consistent with warm-season regimes. These spring/fall regimes also have more instances of excessive rainfall than worst-performing regimes seemingly due to the higher low-level moisture and PWAT across the forecasted risk areas despite comparable or greater synoptic forcing.

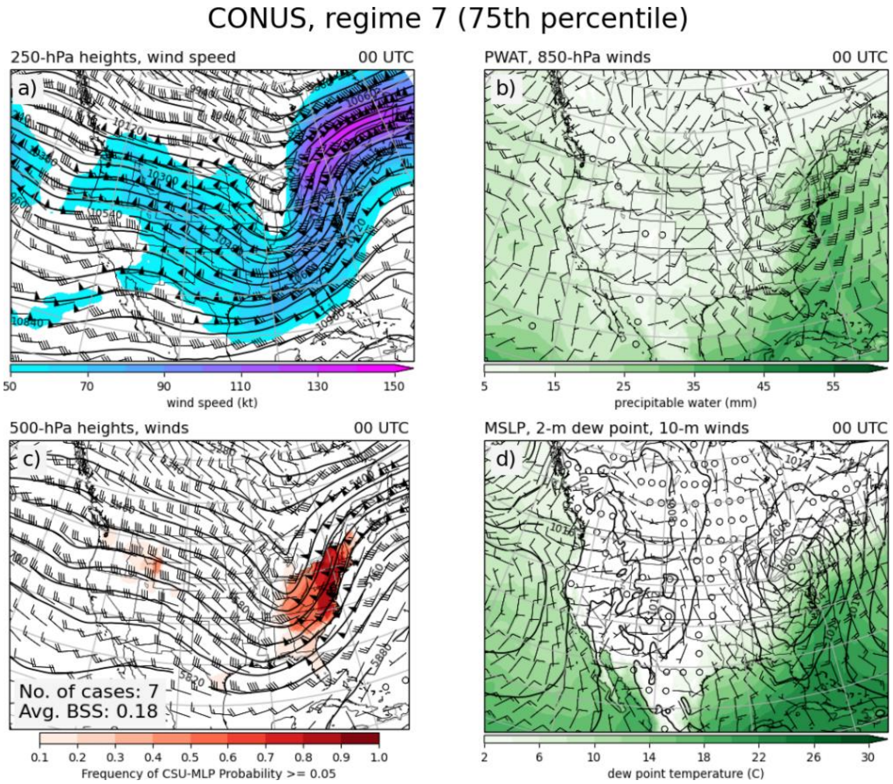


Figure 5.29. Same as Figure 5.3 but for R4-75th

Table 5.7. Regional daily BSS average statistics for regimes analogous to best-performing CONUS spring/fall regimes.

CONUS Regime No.	Region, Percentile	No. cases	BSS Avg.	ERO BSS Avg.	BSS Diff.
6	CONUS 75th	31	0.16	0.15	0.01
	SGP 75th	35	0.21	0.18	0.03
	SE 75th	16	0.21	0.20	0.01
	MDWST 75th	10	0.22	0.22	0.00
	NGP 75th	10	0.19	0.19	0.00
7	CONUS 75th	7	0.18	0.11	0.08
	NE 75th	8	0.21	0.13	0.08

5.3.3 Winter Regimes

R8-75th and R9-75th (Figure 5.30 and 5.31, respectively) are the two regimes with most frequent case occurrence in DJF, although, these regimes were common in spring/fall months. R9-75th is that of landfalling AR in the northwestern CONUS (e.g. Warner et al. 2012). A large-scale ridge axis is anchored across the western CONUS, with a jet streak at the top of the ridge axis. A surface ridge is co-located

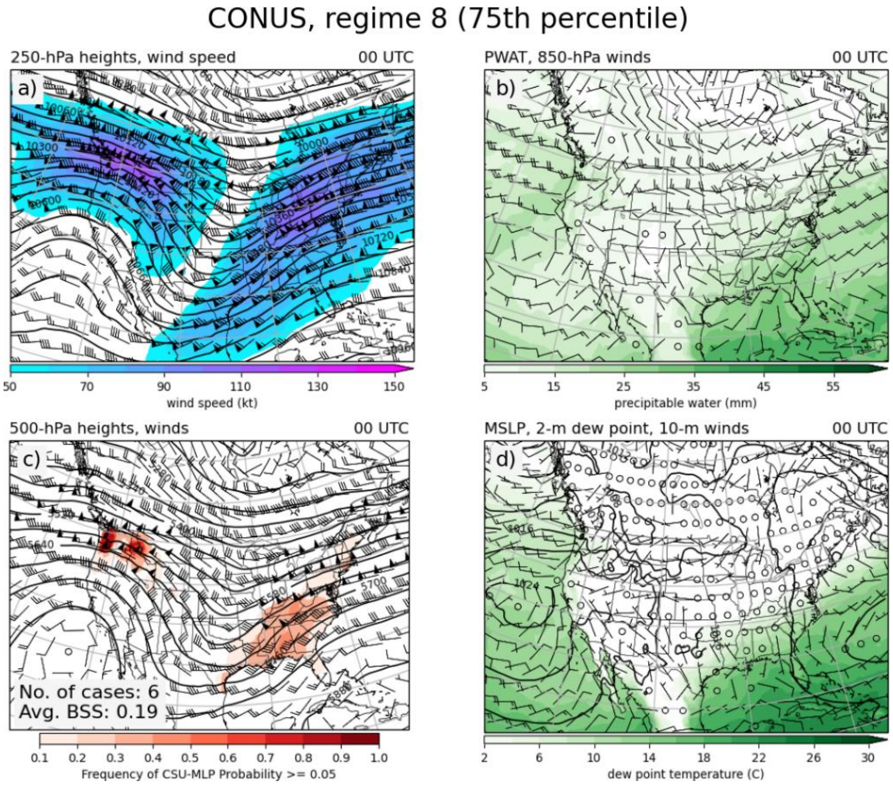


Figure 5.30. Same as Figure 5.3 but for R8-75th

with the upper-level ridge, with westerly low-level flow and a relatively moist air mass impinging upon the terrain of the Washington. PWAT just offshore of the risk area is 20-25 mm, which is not particularly high for these events. However, this is largely due to slightly different orientations of the narrow ARs which cause the locally high PWAT values to be smoothed when composited.

R10-75th is characterized by a strong large-scale trough across the central CONUS, with a low surface pressure across the southeast and northward moving warm front to the east of the low. Widespread PWAT of 30-40 mm are present across the southeast, with the 850-hPa southwesterly jet extending well north of the warm front position. In these cases, there is usually a surface cyclone that develops at various locations within the composite low-pressure region, and persistent convective initiation along the warm front with a broad shield of stratiform precipitation with embedded convection extending northward. In addition, some cases feature a squall line forced by the trailing cold

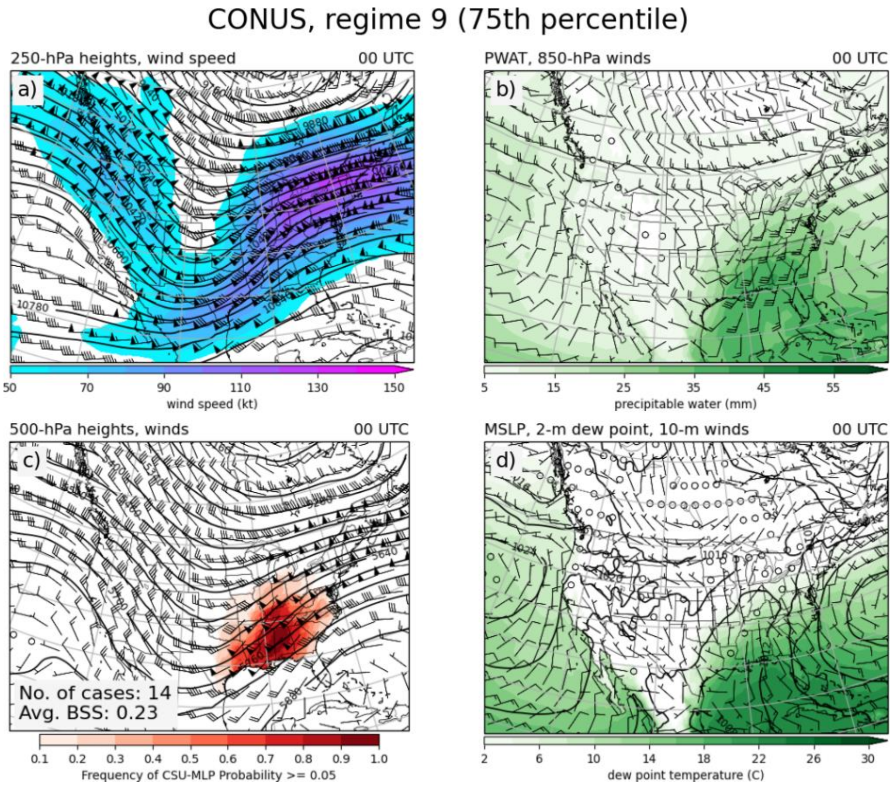


Figure 5.31. Same as Figure 5.3 but for R8-75th

front that traverses some of the same regions that experienced precipitation from the warm front. Due to the strong forcing provided by the large-scale trough and surface cyclone and fronts, this regime is categorized as ETC. This regime resembles that of the "Strong IVT" regime identified by Moore et al. (2015).

R8-75th does not have any analogous regimes identified at the regional level (Table 5.7). While the six cases identified at the CONUS-level showed a clear pattern for CSU-MLP to have high performance for these AR events, there was not a large enough sample of best-performing forecasts in the PCST (only 7 selected cases among the period of study) to create a group of at least four cases, as locations of the ARs varied latitudinally across the region. R9-75th shows a different regional pattern among some of the other large-scale regimes, where regional skill varies among the different regions affected. However, this seems to be largely due to the location of the system relative to the defined regional boundaries, as poor regional forecasts had small areas of risk near the SE, MDWST, and NE that did not observe excessive rainfall. ERO forecasts for these worst-performing regional forecasts do not perform well either, but they have greater skill than CSU-MLP which is largely from CSU-MLP forecasting greater categorical risk area into these adjacent regions.

Table 5.8. Regional daily BSS average statistics for regimes analogous to best-performing CONUS spring/fall regimes

CONUS Regime No.	Region, Percentile	No. cases	BSS Avg.	ERO BSS Avg.	BSS Diff.
8	CONUS 75th	6	0.19	0.08	0.11
9	CONUS 75th	14	0.23	0.19	0.04
	SE 75th	17	0.25	0.19	0.07
	SGP 25th	5	-0.75	-0.04	-0.71
	MDWST 75th	13	0.17	0.12	0.05
	MDWST 25th	10	-1.05	-0.37	-0.68
	NE 25th	4	-2.89	-0.18	-2.71

5.4 Contrasting Typical Best- and Worst-Performing Forecasts

A general pattern emerges when investigating best- and worst-performing CSU-MLP forecasts, particularly for regimes outside of the warm-season: both tend to be from strong synoptic forcing events, however, moisture levels and its areal extent seem to play a role in forecast performance. Two examples of these types of events are R6-75th (Figure 5.21) and R6-25th (Figure 5.8) with each regime containing the most numerous cases out of all the regimes (31 and 17, respectively). R6-75th was classified as a "synoptic" regime due to the widespread convective episodes across a broad area of the southern and central CONUS, and R6-25th was classified as a "cold front" regime due to strong forcing for precipitation along and north of a cold/stationary front in the southern plains. Figure 5.25 shows difference plots using the composite reanalysis from these two regimes. The position of the 250-hPa

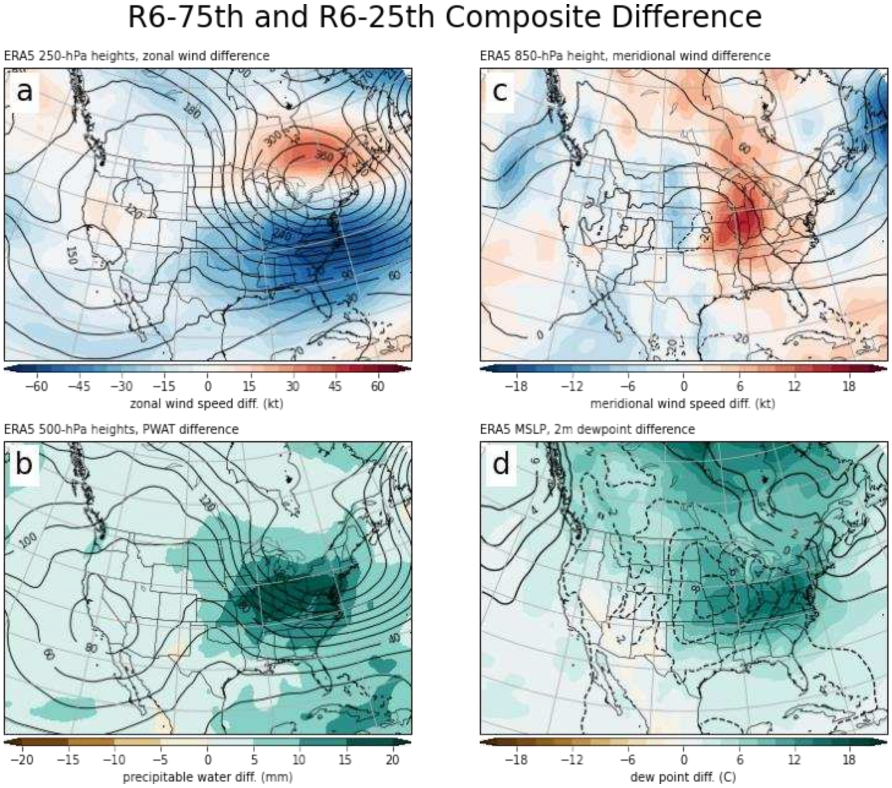


Figure 5.25. Difference plots of ERA5 reanalysis composite for (a) 250-hPa geopotential heights (contour) and zonal wind speed (color fill), (b) 500-hPa geopotential heights (contour) and PWAT (color fill), (c) 850-hPa geopotential heights (contour) and meridional wind speed (color fill), and (d) MSLP (contour) and 2-m dew point temperature (color fill) between R6-75th and R6-25th.

jetstream is much further north in R6-75th (as indicated by the positive zonal wind contour fill (a) across southeast Canada) due to a strong ridge across the eastern CONUS that forced the jet stream northeastward across the central CONUS. This places the central CONUS in the right entrance regions of the jet stream which favors upper-level divergence and low surface pressure. In contrast, the jet stream across the southeast CONUS places the central CONUS in the left entrance region which favors upper-level convergence and high surface pressures. The difference in MSLP is evident in (d), where low MSLP in R6-75th are co-located with high MSLP in R6-25th, resulting in a broad warm/moist sector with southerly low-level moisture transport, as composite PWAT (b), 2-m dew point temperatures (d) and meridional component of 850-hPa winds (c) are much higher for R6-75th. CSU-MLP risk for excessive rainfall is confined to the southern plains for R6-25th, where PWAT is marginally less (0-5 mm) than R6-75th with a similar 850-hPa meridional wind speed. However, the lower 2-m dew point temperatures and northerly winds for R6-25th (Figure 5.8) resulted in limited or non-existent instability for surface-based convection among these cases.

The contrast between these two most common regimes is apparent. The stronger ridging in the eastern CONUS in advance of a trough in R6-75th helps establish a broad warm sector and threat for excessive rainfall ahead of the trough which typically yield favorable forecasts for excessive rainfall. These would be conditions that occur during spring or fall where there is a common overlap between relatively high atmospheric moisture and larger amplitude troughs (e.g. R7-75th, R9-75th) or for large amplitude troughs in the summer that affect the northern regions of the CONUS (e.g. R1-75th, R2-75th). Whereas the jet stream positioned further south along with the strong surface anticyclone in the central CONUS in R6-25th confines moisture to the southern CONUS where excessive rainfall probabilities typically verify with minimal UFVS observations. These would be much like conditions during cooler months where forcing for ascent is prevalent but higher atmospheric moisture conditions are limited (e.g. R6-25th, R7-25th, R8-25th, R9-25th, R10-25th, R11-25th).

CHAPTER 6: DISCUSSION AND CONCLUSIONS

Aggregate results show that CSU-MLP day-1 forecasts are reliable, provide adequate discrimination of excessive rainfall events and non-events (AuROC =0.819), and have comparable skill to that of the ERO (CSU-MLP BSS = 0.081; ERO BSS = 0.085). However, CSU-MLP tends to forecast larger areas and higher probabilities of categorical risk compared to the ERO. This tendency results in daily BSS from CSU-MLP in having a larger variance compared to the ERO, where days with widespread UFVS observations rewards skill but day with little UFVS punishes skill. Given best- and worst-performing forecasts, consistent patterns were observed among different synoptic regimes during different times of the year (Table 6.1). For warm-season regimes, synoptic forcing for convection is generally weakest but instability and moisture for convection is widespread. Thus, extreme precipitation is influenced largely by MCS and cold-pool evolutions. Areas of CSU-MLP categorical risk were typically in the NGP, MDWST, and NE regions. Among these, forecasts that consistently perform best were the ones with strong synoptic forcing (e.g. R1-75th, R2-75th), whereas forecasts with weak synoptic forcing (e.g. R2-25th, R3-25th, R3-75th) had inconsistency in their performance, particularly for large areas of forecasted risk where there was inter-regional variability. TC regimes were also among the warm-season regimes that performed well (R4-75th, R5-75th), but when the TC remained offshore with dry air inland (R5-25th), forecasts were among the worst-performing. For the rest of the year, strong synoptic forcing (e.g. ETCs and fronts) is omnipresent among the regimes but with more variable column moisture. Forecasts that consistently perform best have composite PWAT values that range from 25-45 mm across a large swath of CSU-MLP risk areas (e.g. R6-75th, R7-75th, R9-75th), whereas forecasts that perform worst have composite PWAT values ranging from 15-35 mm (e.g. R6-25th, R7-25th, R8-25th, R9-25th, R10-25th, R11-25th) with limited extent of CSU-MLP risk area. These regimes occurred most frequently during spring and fall months (although winter occurrences were common) and across various regions east of the

Table 6.1. Regime identifier along with each regime type's synoptic forcing and strength classification, composite PWAT across the CSU-MLP excessive rainfall risk area, regions with analogous regimes, and season of most frequent occurrence of cases.

Regime	Regime Type	No Cases	Synoptic Forcing Strength	Composite PWAT (mm)	Other Regions	Season
R1-25th	cold front	4	strong	30-40	NE, SW	JJA
R1-75th	COL, orography	7	strong	35-45	NE, SE	JJA
R2-25th	mesohigh	4	weak	30-40	MDWST, NE	JJA
R3-25th	frontal	8	weak	30-40	MDWST, NGP	JJA
R3-75th	frontal	9	weak	35-45	MDWST	JJA
R4-25th	Synoptic, frontal	5	Strong, weak	25-40	NGP, MDWST, ROCK, SW	JJA
R2-75th	synoptic	11	strong	25-35	NGP	JJA
R5-25th	TC	7	strong	35-45	SE	SON
R4-75th	TC	8	strong	40-50	SE, SGP	JJA
R5-75th	TC	5	strong	40-50	SE	SON
R6-25th	Cold front	17	strong	20-40	SE, SGP	MAM
R7-25th	Warm front	8	strong	20-30	SE, SGP	MAM
R6-75th	synoptic	31	strong	25-40	SE, SGP, MDWST, NGP	MAM
R8-25th	ETC	9	strong	20-30	MDWST, NGP, ROCK, SGP	MAM
R9-25th	ETC	6	strong	15-20	NE	MAM
R10-25th	Cold front	8	strong	15-25	SE	SON
R7-75th	ETC	7	Strong	25-35	NE	MAM
R9-75th	ETC	14	strong	30-40	SE, SGP, MDWST, NE	DJF
R11-25th	Cold front	7	strong	10-20	ROCK, SW	DJF
R8-75th	AR, orography	6	Strong	20-25	PCST	DJF

Rockies. Among these, R6-75th and R6-25th stand out with the most frequent cases of occurrence (31 and 17, respectively), where the former includes a broad warm/moist sector with a widespread threat for excessive rainfall whereas the latter has moisture confined to the southern plains with limited areal extent for excessive rainfall risk. An AR regime in the winter was also identified, where events in the northwestern CONUS had good performance (R9-75th). One regime that was rarely identified was that of the monsoon in the western CONUS, as this period of study only included the summers of 2019 and 2020, both of which were two of the driest monsoon seasons on record for the southwestern CONUS

(NOAA 2020a; NOAA 2021). Thus, a period of study would need to include a more active monsoon season to determine if monsoon regimes would be among some of the best- or worst-performing forecasts.

Findings from this study can provide forecasters at WPC with additional assistance when using the CSU-MLP forecasts as a part of the forecasting process. In addition to a forecaster's own experience with excessive rainfall and flash floods among the different regimes identified in literature, days with regimes analogous with best- and worst-performing CSU-MLP forecasts can also help guide forecasters in determining how much weight the first-guess forecast should have in their ERO construction, especially given the tendency for CSU-MLP forecasts to have a higher frequency of categorical risk. For example, if a day is forecasted to resemble R6-25th, perhaps WPC would be more confident in lowering the initial categorical risk and/or coverage area. Correspondingly, if a day is forecasted to resemble R6-75th, perhaps forecasters would be more confident in issuing categorical risk and area more closely with the CSU-MLP first-guess. Given that best-performing CSU-MLP forecasts tend to score slightly higher than the ERO on average, the latter process poses the potential for increasing ERO skill. While this work specifically identifies synoptic regimes *given* best- and worst-performing CSU-MLP forecasts, it does not directly specify how CSU-MLP forecasts perform *given a synoptic regime*. For example, it could be that, given a regime such as R7-25th (8 cases), many CSU-MLP forecasts perform above the median daily BSS amongst CONUS or SGP forecasts, but none score above the 75th percentile. In this scenario, the average CSU-MLP BSS may be greater than the median but this regime in only identified for worst-performing forecasts. While beyond the scope of this study, the various synoptic regimes that occur among *all* forecasts of interest could be identified, and from these distributions of daily CSU-MLP BSS, investigate the performance among the regimes. However, the consistent pattern of regimes with strong forcing and lower moisture performing worse in comparison to those with slightly higher moisture, especially regimes with larger samples of cases such as R6-75th (31), R6-25th (17), or R9-75th (14), provides confidence that these regimes do have these characteristics for CSU-MLP to perform well/poorly.

This regime identification also benefits forecasters by providing a visualization of the general kinds of patterns that the RF model learns and the general variables it uses for individual forecast days. This would work to build more trust in machine learning, which people often perceive as "black-box" models (e.g. McGovern et al. 2019). While Herman and Schumacher (2018a) examine some of the spatial and temporal feature importance's of the CSU-MLP RF model, this work helps to elucidate the kinds of predictors that may have more influence for a given regime. For example, most identified worst-performing regimes have strong synoptic forcing but modest moisture, so it may be that predictors associated with dynamics (e.g MSLP, 10-m winds, vertical wind shear) are the more frequent variables used to traverse the various decision-trees in the RF model, with less emphasis on instability and moisture variables (e.g. CAPE, PWAT). Additional work would be needed to investigate this directly to examine, at a given location and forecast, which variables and thresholds are used to generate a prediction.

Challenges remain for future development and implementation of CSU-MLP "first-guess" forecasts. Beginning in September 2020, the GEFS underwent a major upgrade (NOAA 2020) that includes a new dynamical core, the Finite Volume Cubed-Sphere (FV3; Lin et al. 2017). Despite training on the older version, real-time forecasts have not shown substantial differences in forecast output. In fact, the last few weeks of this period of study included forecasts run on the updated GEFS. However, future development of CSU-MLP forecasts will require retraining using the FV3-GEFS. Additionally, WPC has recently redefined categorical probability thresholds (MRGL ≥ 5%; SLGT ≥ 15%; MDT ≥ 40%; HIGH ≥ 70%), so future CSU-MLP forecasts and evaluation would be needed using the new thresholds. CSU-MLP models have also been trained using forecast output from the NSSL-WRF (Hill and Schumacher 2021) and the HRRR (Trojiniak et al. 2021), both of which are convection-allowing models (CAMs). Although these models have not been recommended for use in operations, they do pose the potential to useful, particularly for warm-season regimes where excessive rainfall is highly dependent on convective evolution. Lastly, future work would include training and developing models for assistance at longer lead times, as WPC has begun issuing experimental day-4 and day-5 ERO forecasts.

REFERENCES

Adams, D. K., and A. C. Comrie, 1997: The North American Monsoon. *Bull. Amer. Meteor. Soc.*, **78**, 2197–2213, https://doi.org/10.1175/1520-0477(1997)078<2197:tnam>2.0.co;2.

Agel, L., M. Barlow, F. Colby, H. Binder, J. L. Catto, A. Hoell, and J. Cohen, 2019: Dynamical analysis of extreme precipitation in the US northeast based on large-scale meteorological patterns. *Climate Dynamics*, **52**, 1739–1760, https://doi.org/10.1007/s00382-018-4223-2.

Barlow, M., and Coauthors, 2019: North American extreme precipitation events and related large-scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics*, **53**, 6835–6875, https://doi.org/10.1007/s00382-019-04958-z.

Bradley, A. A., and J. A. Smith, 1994: The Hydrometeorological Environment of Extreme Rainstorms in the Southern Plains of the United States. *Journal of Applied Meteorology and Climatology*, **33**, 1418–1431, https://doi.org/10.1175/1520-0450(1994)033<1418:theoer>2.0.co;2.

Breiman, L., 2001: Random Forests. *Machine learning*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Clark, R. A., J. J. Gourley, Z. L. Flamig, Y. Hong, and E. Clark, 2014: CONUS-Wide Evaluation of National Weather Service Flash Flood Guidance Products. *Weather and Forecasting*, **29**, 377–392, https://doi.org/10.1175/waf-d-12-00124.1.

Doswell, C. A., H. E. Brooks, and R. A. Maddox, 1996: Flash Flood Forecasting: An Ingredients-Based Methodology. *Weather and Forecasting*, **11**, 560–581, https://doi.org/10.1175/1520-0434(1996)011<0560:fffaib>2.0.co;2.

Ely, L. L., Y. Enzel, and D. R. Cayan, 1994: Anomalous North Pacific atmospheric circulation and large winter floods in the southwestern United States. *Journal of Climate*, 977–987.

Erickson, M. J., J. S. Kastman, B. Albright, S. Perfater, J. A. Nelson, R. S. Schumacher, and G. R. Herman, 2019: Verification Results from the 2017 HMT–WPC Flash Flood and Intense Rainfall Experiment. *Journal of Applied Meteorology and Climatology*, **58**, 2591–2604, https://doi.org/10.1175/jamc-d-19-0097.1.

Erickson, M. J., B. Albright, and J. A. Nelson, 2021: Verifying and redefining the Weather Prediction Center's Excessive Rainfall Outlook forecast product. *Wea. Forecasting*, **36**, 325–340, https://doi.org/10.1175/waf-d-20-0020.1.

Fritsch, J. M., and R. E. Carbone, 2004: Improving Quantitative Precipitation Forecasts in the Warm Season: A USWRP Research and Development Strategy. *Bulletin of the American Meteorological Society*, **85**, 955–966, https://doi.org/10.1175/bams-85-7-955.

Hamill, T. M., and M. Scheuerer, 2018: Probabilistic precipitation forecast postprocessing using quantile mapping and rank-weighted best-member dressing. *Mon. Wea. Rev.*, **146**, 4079–4098, https://doi.org/10.1175/mwr-d-18-0147.1.

——, G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Y. Zhu, and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553–1565, https://doi.org/10.1175/bams-d-12-00014.1.

——, M. Scheuerer, and G. T. Bates, 2015: Analog probabilistic precipitation forecasts using GEFS reforecasts and climatology-calibrated precipitation analyses. *Mon. Wea. Rev.*, **143**, 3300–3309, https://doi.org/10.1175/mwr-d-15-0004.1.

Herman, G. R., and R. S. Schumacher, 2016: Extreme Precipitation in Models: An Evaluation. *Weather and Forecasting*, **31**, 1853–1879, https://doi.org/10.1175/waf-d-16-0093.1.

Herman, G. R., and R. S. Schumacher, 2018a: “Dendrology” in numerical weather prediction: What random forests and logistic regression tell us about forecasting extreme precipitation. *Mon. Wea. Rev.*, **146**, 1785–1812, https://doi.org/10.1175/mwr-d-17-0307.1.

Herman, G. R., and R. S. Schumacher, 2018b: Flash Flood Verification: Pondering Precipitation Proxies. *Journal of Hydrometeorology*, **19**, 1753–1776, https://doi.org/10.1175/jhm-d-18-0092.1.

——, and ——, 2018c: Money Doesn’t Grow on Trees, but Forecasts Do: Forecasting Extreme Precipitation with Random Forests. *Monthly Weather Review*, **146**, 1571–1600, https://doi.org/10.1175/mwr-d-17-0250.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hill, A. J., and R. S. Schumacher, 2021: Forecasting Excessive Rainfall with Random Forests and a Deterministic Convection-Allowing Model. *Weather and Forecasting*, **36**, 1693–1711, https://doi.org/10.1175/waf-d-21-0026.1.

Hou, D., 2014: Climatology-calibrated precipitation analysis at fine scales: Statistical adjustment of Stage IV toward CPC gauge-based analysis. *J. Hydrometeor.*, **15**, 2542–2557, https://doi.org/10.1175/jhm-d-11-0140.1.

Keene, K. M., and R. S. Schumacher, 2013: The Bow and Arrow Mesoscale Convective Structure. *Monthly Weather Review*, **141**, 1648–1672, https://doi.org/10.1175/mwr-d-12-00172.1.

Konrad, C. E., 1997: Synoptic-Scale Features Associated with Warm Season Heavy Rainfall over the Interior Southeastern United States. *Weather and Forecasting*, **12**, 557–571, https://doi.org/10.1175/1520-0434(1997)012<0557:ssfaww>2.0.co;2.

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**, 1131–1141, https://doi.org/10.1175/jhm-d-11-0108.1.

Lapenta, K. D., and Coauthors, 1995: The Challenge of Forecasting Heavy Rain and Flooding throughout the Eastern Region of the National Weather Service. Part I: Characteristics and Events. *Weather and Forecasting*, **10**, 78–90, https://doi.org/10.1175/1520-0434(1995)010<0078:tcofhr>2.0.co;2.

Lavers, D. A., and G. Villarini, 2013: Atmospheric Rivers and Flooding over the Central United States. *Journal of Climate*, **26**, 7829–7836, https://doi.org/10.1175/jcli-d-13-00212.1.

Lin, S.-J., W. Putman, L. Harris, and the G. F. Team, 2017: FV3: The GFDL finite-volume cubed-sphere dynamical core.

Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2, https://ams.confex.com/ams/Annual2005/techprogram/paper_83847.htm

Loken, E. D., A. J. Clark, A. McGovern, M. Flora, and K. Knopfmeier, 2019: Postprocessing next-day ensemble probabilistic precipitation forecasts using random forests. *Wea. Forecasting*, **34**, 2017–2044, https://doi.org/10.1175/waf-d-19-0109.1.

Maddox, R. A., C. F. Chappell, and L. R. Hoxit, 1979: Synoptic and Meso-α Scale Aspects of Flash Flood Events. *Bulletin of the American Meteorological Society*, **60**, 115–123, https://doi.org/10.1175/1520-0477-60.2.115.

Maddox, R. A., F. Canova, and L. R. Hoxit, 1980: Meteorological Characteristics of Flash Flood Events over the Western United States. *Monthly Weather Review*, **108**, 1866–1877.

Marquardt Collow, A. B., M. G. Bosilovich, and R. D. Koster, 2016: Large-Scale Influences on Summertime Extreme Precipitation in the Northeastern United States. *Journal of Hydrometeorology*, **17**, 3045–3061, https://doi.org/10.1175/jhm-d-16-0091.1.

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. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199, https://doi.org/10.1175/bams-d-18-0195.1.

Melhauser, C., and F. Zhang, 2012: Practical and Intrinsic Predictability of Severe and Convective Weather at the Mesoscales. *Journal of the Atmospheric Sciences*, **69**, 3350–3371, https://doi.org/10.1175/jas-d-11-0315.1.

Moore, B. J., K. M. Mahoney, E. M. Sukovich, R. Cifelli, and T. M. Hamill, 2015: Climatology and Environmental Characteristics of Extreme Precipitation Events in the Southeastern United States. *Monthly Weather Review*, **143**, 718–741, https://doi.org/10.1175/mwr-d-14-00065.1.

Mullens, E. D., 2021: Meteorological cause and characteristics of widespread heavy precipitation in the Texas Gulf watershed 2003–2018. *International Journal of Climatology*, **41**, 3743–3760, https://doi.org/10.1002/joc.7046.

Neiman, P. J., F. M. Ralph, B. J. Moore, M. Hughes, K. M. Mahoney, J. M. Cordeira, and M. D. Dettinger, 2013: The Landfall and Inland Penetration of a Flood-Producing Atmospheric River in Arizona. Part

I: Observed Synoptic-Scale, Orographic, and Hydrometeorological Characteristics. *Journal of Hydrometeorology*, **14**, 460–484, https://doi.org/10.1175/jhm-d-12-0101.1.

Nielsen, E. R., and R. S. Schumacher, 2016: Using Convection-Allowing Ensembles to Understand the Predictability of an Extreme Rainfall Event. *Monthly Weather Review*, **144**, 3651–3676, https://doi.org/10.1175/mwr-d-16-0083.1.

Nieto, R., M. Sprenger, H. Wernli, R. M. Trigo, and L. Gimeno, 2008: Identification and climatology of cut-off lows near the tropopause. *Ann N Y Acad Sci*, **1146**, 256–290, https://doi.org/10.1196/annals.1446.016.

NOAA, 2020a: Southwest Monsoon 2019 Review. Accessed 3 March 2022, https://www.weather.gov/psr/SouthwestMonsoon2019Review

NOAA, 2020b: NOAA upgrades Global Ensemble Forecast System. Accessed 3 March 2022, https://www.noaa.gov/media-release/noaa-upgrades-global-ensemble-forecast-system

NOAA, 2021: 2020 Monsoon Review. Accessed 3 March 2022, https://www.weather.gov/psr/2020MonsoonReview

NOAA, 2022a: NOAA National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters. Accessed 26 February 2022, https://www.ncei.noaa.gov/access/monitoring/billions/

——, 2022b: Weather Related Fatality and Injury Statistics. Accessed 26 February 2022, https://www.weather.gov/hazstat/

Novak, D. R., C. Bailey, K. F. Brill, P. Burke, W. A. Hogsett, R. Rausch, and M. Schichtel, 2014: Precipitation and Temperature Forecast Performance at the Weather Prediction Center. *Weather and Forecasting*, **29**, 489–504, https://doi.org/10.1175/waf-d-13-00066.1.

Parker, M. D., and R. H. Johnson, 2000: Organizational Modes of Midlatitude Mesoscale Convective Systems. *Monthly Weather Review*, **128**, 3413–3436, https://doi.org/10.1175/1520-0493(2001)129<3413:omommc>2.0.co;2.

Peters, J. M., and R. S. Schumacher, 2014: Objective Categorization of Heavy-Rain-Producing MCS Synoptic Types by Rotated Principal Component Analysis. *Monthly Weather Review*, **142**, 1716–1737, https://doi.org/10.1175/mwr-d-13-00295.1.

Rivera, E. R., F. Dominguez, and C. L. Castro, 2014: Atmospheric rivers and cool season extreme precipitation events in the Verde River basin of Arizona. *J. Hydrometeor.*, **15**, 813–829.

Schmidt, J. A., A. Anderson, and J. Paul, Spatially-variable, physically-derived flash flood guidance. *Preprints, 21st Conf. on Hydrology, San Antonio, TX, Amer. Meteor. Soc. B*, **6**.

Schumacher, R. S., 2017: Heavy Rainfall and Flash Flooding. https://doi.org/10.1093/acrefore/9780199389407.013.132.

——, and R. H. Johnson, 2005: Organization and Environmental Properties of Extreme-Rain-Producing Mesoscale Convective Systems. *Monthly Weather Review*, **133**, 961–976, https://doi.org/10.1175/mwr2899.1.

——, A. J. Hill, M. Klein, J. A. Nelson, M. J. Erickson, S. M. Trojniak, and G. R. Herman, 2021: From Random Forests to Flood Forecasts: A Research to Operations Success Story. *Bulletin of the American Meteorological Society*, **102**, E1742–E1755, https://doi.org/10.1175/bams-d-20-0186.1.

Shepherd, M., T. Mote, J. Dowd, M. Roden, P. Knox, S. C. McCutcheon, and S. E. Nelson, 2011: An Overview of Synoptic and Mesoscale Factors Contributing to the Disastrous Atlanta Flood of 2009. *Bulletin of the American Meteorological Society*, **92**, 861–870, https://doi.org/10.1175/2010BAMS3003.1.

Sukovich, E. M., F. M. Ralph, F. E. Barthold, D. W. Reynolds, and D. R. Novak, 2014: Extreme Quantitative Precipitation Forecast Performance at the Weather Prediction Center from 2001 to 2011. *Weather and Forecasting*, **29**, 894–911, https://doi.org/10.1175/waf-d-13-00061.1.

Trojniak, S., J. Correia Jr., and B. Albright, 2020: 2020 Flash Flood and Intense Rainfall Experiment: Findings and results. NOAA Rep., 99 pp., www.wpc.ncep.noaa.gov/hmt/Final_Report_2020_FFaIR_Experiment_Nov13.pdf

Viterbo, F., and Coauthors, 2020: A Multiscale, Hydrometeorological Forecast Evaluation of National Water Model Forecasts of the May 2018 Ellicott City, Maryland, Flood. *Journal of Hydrometeorology*, **21**, 475–499, https://doi.org/10.1175/jhm-d-19-0125.1.

Warner, M. D., C. F. Mass, and E. P. Salathé, 2012: Wintertime extreme precipitation events along the Pacific Northwest coast: Climatology and synoptic evolution. *Mon. Wea. Rev.*, **140**, 2021–2043.

Whan, K., and M. Schmeits, 2018: Comparing area probability forecasts of (extreme) local precipitation using parametric and machine learning statistical postprocessing methods. *Mon. Wea. Rev.*, **146**, 3651–3673, https://doi.org/10.1175/mwr-d-17-0290.1.

Xie, P., M. Chen, and W. Shi, CPC unified gauge-based analysis of global daily precipitation. *Preprints, 24th Conf. on Hydrology, Atlanta, GA, Amer. Meteor. Soc*, **2**.