

UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

APPLICATION OF MACHINE LEARNING FOR DRYLINE CLIMATOLOGIES
AND EXTENDED-RANGE SEVERE WEATHER FORECASTS

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

by MICHAEL J. HOSEK

Norman, Oklahoma

2025

APPLICATION OF MACHINE LEARNING FOR DRYLINE CLIMATOLOGIES
AND EXTENDED-RANGE SEVERE WEATHER FORECASTS

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Kimberly Hoogewind (Chair)

Dr. Aaron Hill

Dr. Kathy Pegion

Dr. Jason Furtado

Dr. Joseph Ripberger

©Copyright by MICHAEL J. HOSEK 2025

All Rights Reserved.

Acknowledgments

This dissertation is the culmination of effort by all of those around me who have supported me through each step of this journey. I want to first thank my advisors, Kim and Adam, for their guidance and encouragement from the early days of forming a project idea to preparing for what comes next after the PhD. I also want to thank my committee and all those in the NWC I learned from, both in and outside of classes.

I also want to thank the SoM staff and administrators who keep the School functioning and have gone out of their way to answer my many questions during my 6 years here as a student. Similarly, I want to thank the members of the Student Affairs Committee, both past and present, for all of the work they have done to foster a supportive environment and advance student interests. My three years serving on SAC helped me grow in many ways, and I'm proud to have contributed to its continuing legacy.

I would not be here at OU without the support of professors Cynthia Domack, Gordan Jones, and Brian Collett of Hamilton College, and WKTV Meteorologist Bill Kardas, all of whom went above and beyond supporting my meteorology goals and finding ways to help me begin research as an undergraduate.

The past six years have been challenging, even beyond the challenges of graduate school. I want to thank all of my SoM friends, who not only helped me in my learning, but who also become like family.

None of this would be possible without the love and support of my family. Even though they may be far away, their support is always a rock I could lean on. And I also want to thank my wonderful fiancé Rachael, who is an amazing friend, collaborator, and partner. And, last but not least, I want to thank my cat Pascal for his many excellent contributions to the methodology, analysis, writing, and overall time management.

This research was supported by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA21OAR4320204, U.S. Department of Commerce, the NOAA Research Global-Nest Initiative, and University of Oklahoma School of Meteorology Teaching Assistantships.

Table of Contents

Acknowledgments	iv
List of Tables	ix
List of Figures	x
Abstract	xvii
1 Introduction	1
2 A 16-year Climatology of WPC Analyzed Drylines and their Association with Severe Convection	5
2.1 Introduction	5
2.2 Data and Methods	10
2.2.1 WPC Surface Analysis Archive	10
2.2.2 Designating “Severe” and “Non-Severe” Drylines	11
2.2.3 Machine Learning Model and Training	13
2.3 Results	15
2.3.1 WPC Monthly and 3-Hourly Statistics	15
2.3.2 WPC Spatial Characteristics	17
2.3.3 WPC Annual Trend	25
2.4 Discussion	25
2.4.1 Annual Trend	25
2.4.2 Comparison to Machine Learning Model	29
2.5 Conclusion	36
3 Quantifying the Role of the Dryline in the Severe Thunderstorm Climatology of the Southern Great Plains	39
3.1 Introduction	39
3.2 Methods	41
3.2.1 Machine-Learning Dryline Model	41
3.2.2 SGP Severe Thunderstorm Climatology	42
3.2.3 Associating Severe Convection with Drylines	43
3.3 Results	45

3.3.1	Spatial Climatology	45
3.3.2	Report Count Climatology	50
3.4	Summary and Conclusion	51
4	Variability in the US Southern Great Plains Dryline Associated with the El Niño Southern Oscillation and Regional Drought	54
4.1	Introduction	54
4.1.1	Factors Influencing Dryline Frequency and Longitude	54
4.1.2	ENSO	56
4.1.3	Hypotheses and Outline	58
4.2	Methods	59
4.2.1	Dryline Identification	59
4.2.2	ENSO	61
4.2.3	Palmer Drought Severity Index	62
4.2.4	Environmental Composites	65
4.2.5	Statistical Testing	65
4.3	Results	68
4.3.1	ENSO	70
4.3.2	Drought	89
4.4	Discussion	115
4.5	Conclusions	120
5	Quantifying the Value of Extended Range Convection Allowing Global-Nested Ensembles for Severe Thunderstorm Forecasting	124
5.1	Motivation and Introduction	124
5.1.1	Motivation	124
5.1.2	Introduction	125
5.2	Data and Methods	127
5.2.1	Ensemble Forecasts	127
5.2.2	Random Forest Classifier	132
5.2.3	Explainable AI with TreeInterpreter package	134
5.2.4	Objective Verification Metrics	135
5.2.5	Forecast Dryline Identification	138
5.3	Results	140
5.3.1	Ensemble 500mb Skill	140
5.3.2	RF Forecast Statistics	141
5.3.3	Predictor Relative Contributions	145
5.3.4	Global-Nest Dryline Forecasts	148
5.4	Individual Forecast Case Studies	150
5.4.1	Day 3: 17 May 2017	151
5.4.2	Day 3: 24 May 2019	158
5.4.3	Day 8: 22 January 2017	163
5.5	Discussion	167

5.6 Conclusions	170
6 Summary and Conclusions	175
Reference List	178

List of Tables

1.1	List of all previous non-case study observed dryline studies with at least 3 years of data detailing the years, months, times, and region of study, as well as whether or not convection was included in any capacity. . . .	2
2.1	Same as Table 1.1, but the climatology developed in this Chapter is presented in the final row.	9
2.2	Training, validation, and testing configuration for each model in the cross-validation procedure	15
2.3	The number of individual non-severe WPC-analyzed drylines by region (row) and season (column). Total count for the entire season is shown in the top row. The relative frequency of individual drylines in each sub-region by meteorological season is in parenthetical (Fig. 2.4). Since drylines can be counted in multiple regions, the relative frequencies will not necessarily add up to 100%.	20
2.4	As in Table 2.3, but for severe WPC-analyzed drylines.	20
3.1	The percentage of all 2100–0300 UTC reports (middle column) and significant severe reports (right column) in the SGP associated with a dryline.	50
5.1	C-SHiELD ensemble configurations	129
5.2	List of all predictors, grouped in columns by category (top row). UH fields are only included for the Global-Nest RF. Predictors represent ensemble means at each grid point with the exception of UH NMEP and latitude/longitude.	133

List of Figures

2.1	The average number of “Non-Severe” (blue) and “Severe” (red) WPC-analyzed dryline days per month for the entire CONUS. The red-dashed line indicates the fraction of dryline days per month that are severe (right axis)	16
2.2	The number of individual “Non-Severe” (blue) and “Severe” (red) WPC-analyzed drylines by analysis hour for (a) meteorological spring, (b) summer, (c) autumn, and (d) winter for the entire CONUS. Red-dashed lines indicate the fraction of drylines that are severe for each analysis time (right axis).	17
2.3	Heatmap of the relative frequency of individual drylines per grid-box. All WPC-analyzed drylines are within the bounds of the figure. Note the colorscale is non-linear.	18
2.4	Delineation of the five sub-regions: Southern Great Plains, Northern Great Plains, Midwest, Southeast, and Southwest. Note: Portions of Mexico and Canada included in the polygons were not included in the analysis.	19
2.5	Median of average individual WPC-analyzed dryline longitude by month (color) and analysis hour (y-axis). Colors are grouped by meteorological season. A reference map is shown at top.	22
2.6	Median of average individual WPC-analyzed dryline latitude by month (color) and analysis hour (x-axis). Colors are grouped by meteorological season as in Fig. 2.5. A reference map is shown to the right.	24
2.7	Number of WPC-analyzed dryline days per year. The average number of dryline days for the period 2007–2015 (blue) and 2016–2022 (gold) are shown. The red-dashed line depicts the linear trend, and the equation for the trend is displayed in the upper left below the legend.	26

2.8	Number of (a) “Severe” and (b) “Non-Severe” WPC-analyzed dryline days per year. The average number of dryline days for the period 2007–2015 (purple) and 2016–2022 (green) are shown. The gray-dashed lines represent the linear trend, and the equation for the trend is displayed in the upper left below the legend.	27
2.9	WPC-analyzed Dryline counts by year for each 3-hour analysis time, from (a) 1200 UTC through (h) 0900 UTC. Blue dots represent the count for the given analysis time, while gray ‘x’ indicates the count from 2100 UTC, the reference time (panel d). Linear trends are shown for both the given and reference time in blue and gray dashed-lines, respectively.	28
2.10	Relative frequency heatmap of (a) ML-identified drylines and (b) the WPC-analyzed drylines. As in Fig. 2.3, the colorscale is nonlinear. The total number of drylines is in the upper left corner.	30
2.11	(a) A comparison between WPC-analyzed dryline days (gray) and ML-identified dryline days (purple) per year and (b) average days per month. The black-dashed line in (a) represents the linear trend in WPC-analyzed dryline day counts per year, while the purple-dash-dot line represents the linear trend in ML-identified dryline day counts per year. Equations for the linear trends are located below the legend in (a).	32
2.12	(a) Potential temperature (K; fill and light gray contours) and specific humidity gradient (green) contoured at 3, 6 g kg ⁻¹ per 100 km at 2100 UTC 5 May 2014. Black stippling indicates where the ML model identified a dryline. (b) The corresponding WPC surface analysis map at the corresponding analysis time.	34
3.1	(a) The average number of days in MAMJ with a 0000 UTC dryline search window, (b) the average number of days in MAMJ with at least one severe report between 2100 and 0300 UTC, and (c) the percentage of MAMJ severe days which are associated with a dryline. The value of each non-zero grid box is annotated, rounded to the nearest integer (note that grid boxes in (a) and (b) with more than 0 but less than 0.5 days per spring are annotated with “0”).	45
3.2	As in Fig. 3.1, but each row is an individual month. (a–c) March, (d–f) April, (g–i) May, and (j–l) June.	48

4.1	Mean-removed, linearly-detrended ENSO index versus SGP-average PDSI per month for 1940–2018 (color and marker type). Dashed lines represent the linear trend for the month corresponding to its color. The Kendall Tau correlation coefficient and p-value is reported for each month in the lower right. Solid vertical and horizontal lines denote the ENSO phase and PDSI phase thresholds used, respectively. The vertical red line represents the threshold for El Niño, and the vertical blue line represents the threshold for La Niña. The horizontal green line represents the threshold for wet, and the horizontal brown line represents the threshold for drought.	64
4.2	The number of years of (a) ENSO phases and (b) PDSI phases by month. The total number of drylines for (c) each ENSO phase and (d) each PDSI phase by month. Note that (a),(c) span 1940–2020 while (b),(d) span 1940–2018	69
4.3	(a), (c–f) The violin plots demonstrating the distribution of linearly-detrended mean-removed dryline frequency anomaly by month for each ENSO phase. The thick black line spans the 25th to 75th percentile of each distribution, with the white circle denoting the median and the red “x” representing the anomaly mean. The mean dryline frequency for each month is shown in upper right. The p-value of the Kruskal-Wallis test is shown in the upper left. (b) A table of p-values from the Bonferroni-corrected post-hoc Mann-Whitney U test. Each month is represented by a row, and each column represents the Mann-Whitney U test performed (El Niño–Neutral, La Niña–Neutral, and El Niño–La Niña, respectively). The table cells are shaded by statistical significance with green=significant, red=not significant, and yellow=significant only if no Bonferroni correction applied. The Mann-Whitney U tests are only performed for months where the Kruskal-Wallis test is significant. Months where no Mann-Whitney U tests are performed are shaded gray and denoted “N/A”.	71
4.4	As in Fig. 4.3, but for dryline median longitudes for each month. The number of drylines in each distribution is shown at the bottom of each.	73
4.5	February Dryline composite anomaly for (a) El Niño and (b) La Niña compared to (c) composite dryline day frequency during neutral ENSO, in terms of days per month. The Green box outlines the defined SGP region.	74
4.6	As in Fig. 4.5, but for March.	75

4.7	As in Fig. 4.5, but for April.	76
4.8	As in Fig. 4.5, but for May.	76
4.9	As in Fig. 4.5, but for June.	77
4.10	February (a) El Niño and (c) La Niña 500 mb height and wind composites. Solid black lines represent the composite height (m above mean sea level), and wind barbs denote the composite 500 mb wind vector in knots. Colored shading (scale beneath (e)) represents the 500 mb height anomaly compared to the 1940–2020 February climatology. The number of months in each composite is in the lower right. (e) The 500 mb composite height anomaly difference between (a) and (c). Stippling denotes statistical significance to the 95th percentile. Wind barbs represent the difference in 500 mb wind vector anomalies between (a) and (c), and are shown only where either the u or v component of the wind difference is statistically significant to the 95th percentile. The average number of dryline days per February for El Niño and La Niña is in the bottom right. (b), (d), and (f) as in (a), (c), and (e), but for Mean Sea Level Pressure (hPa) and 10-m wind.	80
4.11	As in Fig. 4.10, but for March.	82
4.12	As in Fig. 4.10, but composites consist only of dryline days for April and compare (a,b) La Niña to (c,d) neutral ENSO. The number of dryline days that go into each composite are in the lower right corner of (e) and (f).	84
4.13	As in Fig. 4.12, but for May. Composites are comparing (a,b) El Niño to (c,d) La Niña. Note that, due to the order of operations, for (e,f) lower 500 mb height or MSLP during La Niña is denoted by red shading, as opposed to blue shading in Fig. 4.12e,f.	86
4.14	As in Fig. 4.13, but for June.	88
4.15	As in Fig. 4.3, but for PDSI phase.	90
4.16	As in Fig. 4.4, but for PDSI phase.	92
4.17	As in Fig. 4.5, but for PDSI. (a) drought and (b) wet conditions dryline frequency anomaly composites compared to (c) the dryline frequency composite under normal conditions, in terms of days per month.	93

4.18	As in Fig. 4.17, but for March.	94
4.19	As in Fig. 4.17, but for April.	95
4.20	As in Fig. 4.17, but for May.	96
4.21	As in Fig. 4.17, but for June.	97
4.22	Same as Fig. 4.10, but for PDSI phases wet and drought.	99
4.23	Same as Fig. 4.22, but for March.	101
4.24	As in Fig. 4.23, but composites consist only of dryline days. The number of dryline days that go into each composite are in the lower right corner of (e) and (f).	103
4.25	Same as Fig. 4.22, but for April.	105
4.26	As in Fig. 4.24, but for April.	107
4.27	Same as Fig. 4.22, but for May.	109
4.28	Same as Fig. 4.24, but for May.	111
4.29	Same as Fig. 4.24, but for June.	114
5.1	The red grid denotes the subset of the NCEP 211 grid model data are interpolated to. The blue shading is the CONUS mask used in the random forest algorithm.	131
5.2	(a) Daily 500-mb height ensemble mean anomaly correlation coefficient and (b) average daily 500-mb height ensemble mean variance out to forecast day 15 for the Global (black) and Global-Nest (red) ensembles.	141

5.3	Attributes diagram for the (a) Global and (b) Global-Nest RF forecasts of severe convection. The number of forecasts of a given probability are shown in the inset figure in the upper left of each. Individual forecast days are colored from warm (day 1) to cool (day 15) colors, alternating between solid and dot-dashed lines by day. The black-dashed diagonal line represents perfect reliability, and the black-dot-dashed horizontal line represents the sample climatology probability of a severe report at all grid points. The solid black line is the line of no skill. Attributes are masked for any bin that does not contain at least 30 grid forecasts. . . .	142
5.4	(a) Brier Skill Score and (b) Area under the Receiver Operating Characteristic (ROC) curve by forecast lead day for the Global (black) and Global-Nest (red) RFs. Light-blue shading denotes lead days when the Global-Nest RF improvements in (a) BSS and (b) AUC is statistically significant to the 95th percentile.	143
5.5	Forecast (a) resolution and (b) reliability for the Global (black) and Global-Nest (red) RFs by forecast lead day. Light-blue shading denotes days where the difference between Global and Global-Nest RFs is statistically significant to the 95th percentile.	144
5.6	Mean relative contribution of each predictor group (a,b) and individual predictor (c,d) for the Global and Global-Nest RF, respectively. The solid lines in (a,b) represent the mean relative contribution for that predictor group, with the shading of the same color denoting the region between the individual predictor within that group with the highest relative contribution and lowest relative contribution at each lead day. Dashed lines denote the maximum and minimum relative contributions from an individual variable of that grouping. Since the predictors in (c,d) are not grouped, there is no max/min bounds. Note that lat/lon is the average of the latitude and longitude predictor, and is represented in (c,d) because the difference between the relative contributions from latitude and longitude predictors is minimal.	146
5.7	(a) POD and (b) FAR for a daily categorical forecast of a dryline for each Global-Nest ensemble member (thin colored lines) and the Global-Nest ensemble mean (thick black line) by forecast lead day.	148
5.8	As in Fig. 5.7, but for spatially-aggregated POD and FAR.	149

5.9	(a) the maximum day 1 categorical outlook issued by the SPC for 17 May 2017, and (b) scatterplot of all individual reports. The color and shape of the individual scatter points in (b) correspond to the type of severe report (bottom in-panel legend). Practically-Perfect Hindcast (PPH; Hitchens et al., 2013; Gensini et al., 2020) probabilities are shaded underneath and correspond to the legend below (b).	152
5.10	(a) Global and (f) Global-Nest RF day 3 severe convection forecasts for 17 May 2017. Shading denotes the forecasted severe probabilities (note that the scale max is at 0.5). The color-bar legend for total severe probability forecasts in (a) and (f) is to the right of (a). BSS and AUC for the Global (a) and Global-Nest (f) RF forecasts are in the bottom right corner. (b-e) Daily summed contributions to the Global RF forecast from individual groupings of variables: (b) thermodynamic, (c) kinematic, (d) significant tornado parameter, and (e) mean sea level pressure. (g-k) Daily summed contributions to the Global-Nest RF forecast from individual groupings of variables: (g) UH, (h) thermodynamic, (i) kinematic, (j) significant tornado parameter, and (k) mean sea level pressure. The color-bar legend for probability contribution is located at the bottom of the figure. Black squares denote each grid box with at least one severe report during this convective day.	154
5.11	The day 3 forecasted ensemble mean STP valid at 0000 UTC 18 May 2017 (convective day 17 May 2017) for (a) Global, (b) Global-Nest, (c) the difference between Global and Global-Nest. (d) The GEFS/R Zero-hour analysis STP for context. The color-bar legend for STP, both (a,b) forecasted and (d) analysis, is to the right of the figure. The color-bar legend for the difference in STP between the Global and Global-Nest is at the bottom of panel (c).	157
5.12	As in Fig. 5.9, but for 24 May 2019. Note there are 4 severe reports (2 hail, 2 wind) outside of this domain in northern California.	159
5.13	As in Fig. 5.10, but for the day 3 forecast for 24 May 2019	161
5.14	As in Fig. 5.9, but for 22 January 2017.	164
5.15	As in Fig. 5.10, but for the day 8 forecast for 22 January 2017	166

Abstract

The dryline is an important boundary for convection initiation in the central US. Although it has been studied for decades, there are no long-term climatological studies of both the dryline and its role in the severe thunderstorm climatology. In this dissertation, a machine learning (ML) model designed to identify fronts (FrontFinder AI) is applied to ERA5 reanalysis data to create an extended dryline climatology. The climatology for 16 years overlapping available NOAA Weather Prediction Center (WPC) surface analyses is compared for drylines across the entire contiguous US and for each month, expanding upon previous dryline climatologies which only considered a subset of months for specific regions. The ML dryline climatology corrects a non-meteorological change in annual dryline frequency observed in the WPC surface analyses. The ML model is then applied to ERA5 data back to 1940 to create a 26-year (1995–2020) climatology of dryline-associated severe thunderstorm reports in the Southern Great Plains, and to investigate dryline variability associated with drought and the El Niño–Southern Oscillation (ENSO) over 81 years (1940–2020). Drylines are associated with over 50% of springtime severe days from west Texas through southwest Oklahoma, and drylines are more frequent and further east during months of Southern Great Plains drought. The final chapter creates a random forest (RF) daily severe thunderstorm forecast model to evaluate whether a nested convection-allowing model (CAM) ensemble and its associated simulated storm attributes provide any additional skill over a convection-parameterized ensemble for extended-range forecasts (days 3–15). When both ensembles have daily synoptic-scale forecast skill (days 1–8), the CAM RF forecasts are significantly better. Preliminary results exploring forecasted dryline skill through day 15 are presented.

Chapter 1

Introduction

The dryline is an important mesoscale feature in the central US. Separating dry air from the high terrain of northern Mexico and the southwest US from the moisture-rich air over the Great Plains, drylines can serve as a focal point for the initiation of deep convection (e.g., Fujita, 1958; Rhea, 1966; Schaefer, 1986; Hobbs et al., 1996). Drylines are most common during spring—coincident with the seasonal peak in severe thunderstorms in the Great Plains—and are important factors in some significant severe thunderstorm outbreaks (e.g., McCarthy and Koch, 1982; Hoch and Markowski, 2005; Mitchell and Schultz, 2020; Chasteen and Koch, 2022).

Forecasting the dryline and dryline-initiated convection can be a challenge. The evolution of the dryline is modulated by both synoptic factors such as extratropical cyclones, as well as mesoscale factors such as soil moisture and the rate of daytime sensible heating (e.g., Hane, 2004; Grasso, 2000). Due to the potent atmospheric conditions often ahead of the dryline during spring in the central US (e.g., Gensini and Ashley, 2011), dryline-initiated convection can develop into supercells capable of very large hail and violent, long-track tornadoes. However, uncertainty in whether the dryline will initiate sustained deep convection can result in forecasts with a high ceiling (e.g., isolated supercells with large hail and long-track tornadoes), but a low floor (e.g., no storms occur). One example of such a forecast is 19 April 2023 in central Oklahoma: This event resulted in 18 tornadoes, including 2 EF-3, along with 65 severe hail reports in the NWS Norman County Warning Area (SPC, 2023). The 1200 UTC day 1 outlook

issued by the SPC had the region within a Slight Risk (level 2 of 5), and discussion highlighted the uncertainty of dryline convection initiation (“Storm development is uncertain along the dryline this afternoon/evening”) as well as the high ceiling of the event, if dryline-initiated storms developed (“If storms form, the environment favors a conditional threat of isolated very large hail... ..with some increase in the late evening tornado threat”; SPC, 2025).

Although the role—and uncertainty—of the dryline in the initiation of high-impact severe convection is well understood, few studies quantify this role beyond case studies or short-term climatologies (Table 1.1). The majority of past studies focus solely on drylines in the Southern Great Plains between April and June, and only one study (Hoch and Markowski, 2005) exceeds 20 years of data. Additionally, most studies do not include a climatological assessment of severe convection associated with the dryline.

Study	Years	Months	Times	Region	Convection?
Rhea (1966)	1959–1962	April–June	3-hourly	Great Plains	Yes
Schaefer (1974)	1966–1968	April–June	All	Great Plains	No
Hoch and Markowski (2005)	1973–2002	April–June	0000 UTC	Great Plains	No
Schultz et al. (2007)	2004–2005	April–June	0000 UTC	West Texas	No
Coffer et al. (2013)	2007–2011	April–June	0000 UTC	Southern High Plains	No
Bergmaier and Geerts (2008)	2010–2012	May–August	1500–0000 UTC, hourly	SE Wyoming	Yes
Duell and Van Den Broeke (2016)	1999–2013	All	All, 3-hourly	Mississippi River Valley	Yes
Johnson and Hitchens (2018)	2006–2015	April–June	0000 UTC	Central/Southern Great Plains	No
Mitchell and Schultz (2020)	2006–2015	April–June	1500 UTC–0300 UTC 3-hourly	Central/Southern Great Plains	Yes

Table 1.1: List of all previous non-case study observed dryline studies with at least 3 years of data detailing the years, months, times, and region of study, as well as whether or not convection was included in any capacity.

The motivation for this dissertation is to expand upon previous dryline climatolo-

gies to study variability and forecastability of drylines and dryline-initiated convection across all timescales from short-to-medium range forecasts, to subseasonal-to-seasonal (S2S), to future climate projections. While the scope of this goal is too large for a single dissertation, each Chapter represents a step towards this goal. Chapter 2 (Hosek et al., 2025) constructs a full calendar year, entire-contiguous US 16-year dryline climatology using NOAA Weather Prediction Center (WPC) surface analyses, and compares it to a climatology of drylines identified by FrontFinder AI, a machine learning (ML) front-detection algorithm applied to ERA5 (Justin et al., 2025). Chapter 3 (Hosek et al., 2026b) uses the ML-identified dryline climatology to determine the proportion of March–June severe storm days across the Southern Great Plains which are associated with drylines. Chapter 4, with the ML-identified dryline climatology extended back to 1940, evaluates monthly variability in dryline frequency and longitude associated with drought in the Southern Great Plains and the El Niño–Southern Oscillation. These three Chapters establish a detailed climatology of drylines and dryline-associated severe convection, and leverage the ML-identified drylines to provide a sample of years large enough to begin studying the relationship between drylines and known modes of climate variability.

An additional related goal of this dissertation is to investigate the viability of extended-range severe thunderstorm forecasts, including those not necessarily related to a dryline. Chapter 5 (Hosek et al., 2026a) compares severe thunderstorm forecasts generated with a Random Forecast Classifier from two nearly identical ensembles to evaluate whether convection-allowing models provide any additional skill compared to traditional convection-parameterized models out through forecast day 15. In addition to the ensemble comparison, FrontFinder AI is applied to the ensemble members to evaluate dryline forecast skill across lead times. This Chapter presents the preliminary stages of future work to evaluate potential forecast model biases impacting dryline

forecasts, the first step towards investigating how dryline forecast skill may affect convection initiation within the forecast ensembles.

Each Chapter is either published or currently in review, with the exception of Chapter 4. As such, each Chapter is relatively self-contained with its own introduction, methods, results, and conclusions sections. The only substantial changes from the published/submitted manuscripts is the addition of the application of FrontFinder AI to the forecast ensembles in Chapter 5. All other changes are minor to minimize repetition and enhance connections between Chapters.

Chapter 2

A 16-year Climatology of WPC Analyzed Drylines and their Association with Severe Convection

2.1 Introduction

The dryline is a mesoscale boundary that is common in the United States Great Plains where it separates dry air from the desert Southwest US and northern Mexico plateau from moist air from the Gulf of Mexico (e.g., Fujita, 1958; Rhea, 1966; Schaefer, 1986). The low-level convergence induced by the dryline's solenoidal circulation creates a region of lift along the boundary that has long been identified as a source of convection initiation (e.g., Fujita, 1958; Rhea, 1966; Ogura and Chen, 1977; Sun and Ogura, 1979; McCarthy and Koch, 1982; Ziegler and Hane, 1993; Ziegler et al., 1995; Grasso, 2000; Weiss and Bluestein, 2002; Hane, 2004). The dryline initiates discrete cellular convection more frequently than other common initiation mechanisms (e.g., cold fronts), which thus leads to an increased probability for dryline-initiated convection to develop into supercells given favorable environmental conditions (Bluestein and Parker, 1993; Dial et al., 2010). Therefore, many significant severe storm outbreaks, particularly events with significant supercellular tornadoes, have been at least in part attributed

to convection initiated along a dryline (e.g., McCarthy and Koch, 1982; Chasteen and Koch, 2022). In addition to severe convection, the dryline also is an important factor in the broader Great Plains warm-season hydroclimatology. The average longitude of the dryline influences the location of warm-season convective precipitation, which accounts for a significant portion of annual rainfall in the region, which then in turn influences the strength and location of the dryline (Fritsch et al., 1986; Ziegler et al., 1995; Shaw et al., 1997; Grasso, 2000; Weiss and Bluestein, 2002; Johnson and Hitchens, 2018).

The dryline is closely associated with the springtime severe weather “season” in the Great Plains (April—June). It is during this time of year and in this region that the dryline, and dryline-initiated convection, is found most frequently (Hoch and Markowski, 2005; Mitchell and Schultz, 2020). Drylines typically develop when synoptic-scale geostrophic flow bring together a dry and moist air mass through large-scale confluence, similar to the process for synoptic fronts (Hoskins and Bretherton, 1972; Schultz et al., 2007). The degree to which synoptic-scale processes dominate dryline evolution can vary case to case, as vertical mixing along a dryline can serve as a source of confluence on the same order of magnitude or exceeding synoptic confluence (i.e., synoptically quiescent drylines [Hane, 2004; Schultz et al., 2007]). Absent other forcing, from morning through late afternoon/evening a dryline will propagate eastward as a result of daytime vertical mixing processes (McCarthy and Koch, 1982; Sun and Wu, 1992; Ziegler and Hane, 1993; Parsons et al., 2000; Hane, 2004; Markowski and Richardson, 2010). The temperature and specific humidity gradient between the dry and moist side of the dryline will grow more distinct during this time, with the dry side heating up faster via large sensible heating (Sun and Ogura, 1979; Schaefer, 1986; Parsons et al., 2000; Markowski and Richardson, 2010). Deep boundary layer mixing on the west (dry) side of the dryline mixes westerly momentum aloft down throughout the boundary layer. Simultaneously, weak easterly winds on the east (moist) side of

the dryline are strengthened via the inland sea-breeze circulation, leading to increased convergence along the dryline which increases the strength of its moisture gradient (Sun and Ogura, 1979; Sun and Wu, 1992; Ziegler and Hane, 1993; Parsons et al., 2000; Weiss and Bluestein, 2002). As daytime mixing weakens, so does the downward mixing of westerly momentum, and the dryline’s eastward progression stalls and it eventually retreats back westward overnight due to the increased easterly winds to the east of the dryline (Schaefer, 1986; Sun and Wu, 1992; Parsons et al., 2000). This conceptual mesoscale evolution of the dryline may be augmented or interrupted by other synoptic forcings (e.g., a cold front overtaking the dryline [Schaefer, 1974]). From April to June, the dryline average longitude is found to retreat further west, hypothesized to be the result of the agricultural evapotranspiration from crop growth in the western Great Plains (Hoch and Markowski, 2005; Johnson and Hitchens, 2018).

Given its importance to regional severe convection and hydroclimatology, it is no surprise that many studies have sought to collect key dryline statistics to assess its characteristics and impact (Table 2.1). However, many of these studies are not 30-year climatologies. Only the Hoch and Markowski (2005) study is a 30-year dryline climatology, but their study, like many other non-case study analyses of observed drylines, only considers drylines in the central and southern Great Plains during April, May, and June (Rhea, 1966; Hoch and Markowski, 2005; Schultz et al., 2007; Coffey et al., 2013; Johnson and Hitchens, 2018; Mitchell and Schultz, 2020). While this spatial time frame and geographic region encompass the peak frequency for drylines and dryline-associated severe convection, the full annual cycle over the contiguous United States (CONUS) has not been explored. Only one study, Duell and Van Den Broeke (2016), considered a region outside of the Great Plains, but was restricted to the Mississippi River Valley. Expanding a climatological analysis to the full annual cycle without regional restriction may be useful for contextualizing research on long-term changes to

the dryline over time resulting from climate change, a topic of critical importance for the future hydroclimate of the central US and future severe convection climatology (e.g., Scaff et al., 2021). While our dryline climatology presented herein is limited to 16 years, it represents the first known dryline climatology covering the entire CONUS for the entire year, entire day (in 3-hourly increments), and includes a climatological assessment of dryline-associated severe convection (Table 2.1).

While the connection between drylines and convection initiation is well established, few non-case studies of drylines incorporate an accompanying analysis of dryline-associated convection (Table 2.1). Rhea (1966) initially identified the region within 200 nautical miles of the dryline as favorable for convection initiation, most typically in the late afternoon and evening. Duell and Van Den Broeke (2016) identified how frequently drylines in the Mississippi River Valley were associated with regional severe weather outbreaks, but their methodology did not specifically isolate convection associated with the dryline as opposed to other forcing mechanisms. Neither of these studies explicitly compared characteristics of drylines which served as an initiation mechanism to severe convection and those that did not. Bergmaier and Geerts (2015) described characteristics of drylines and their synoptic environments in relation to convection, including severe convection, but their study focused on May–August in southeastern Wyoming over 3 years. Mitchell and Schultz (2020) presented the most detailed analysis of dryline-initiated convection, presenting the frequency of dryline initiation and comparing composites of synoptic-scale fields for drylines which initiated convection and those that did not. However, their 10-year analysis of April–June and does not explicitly consider convection which produced severe weather hazards. Hoch and Markowski (2005), the only 30-year climatology of the dryline, does not include dryline-associated convection in their study.

This study seeks to fill these gaps in previous dryline studies by using an available

Study	Years	Months	Times	Region	Convection?
Rhea (1966)	1959–1962	April–June	All	Great Plains	Yes
Schaefer (1974)	1966–1968	April–June	All	Great Plains	No
Hoch and Markowski (2005)	1973–2002	April–June	0000 UTC	Great Plains	No
Schultz et al. (2007)	2004–2005	April–June	0000 UTC	West Texas	No
Coffer et al. (2013)	2007–2011	April–June	0000 UTC	Southern High Plains	No
Bergmaier and Geerts (2008)	2010–2012	May–August	1500–0000 UTC, hourly	SE Wyoming	Yes
Duell and Van Den Broeke (2016)	1999–2013	All	All, 3-hourly	Mississippi River Valley	Yes
Johnson and Hitchens (2018)	2006–2015	April–June	0000 UTC	Central/Southern Great Plains	No
Mitchell and Schultz (2020)	2006–2015	April–June	1500 UTC–0300 UTC 3-hourly	Central/Southern Great Plains	Yes
Hosek et al. (2025)	2007–2022	All	All, 3-hourly	CONUS	Yes

Table 2.1: Same as Table 1.1, but the climatology developed in this Chapter is presented in the final row.

16-year (2007–2022) archive of 3-hourly Unified Surface Analyses from the Weather Prediction Center (WPC) (NOAA, 2023). This dataset contains analyzed drylines for the entire CONUS and is created by expert forecasters from the WPC combining analyses from the Ocean Prediction Center, National Hurricane Center, and Honolulu Weather Forecast Office, as well as surface observations, satellite observations, and model analyses to draw surface mesoscale features (Berg et al., 2007; NOAA, 2013). Analyzed drylines from the WPC have been used in previous dryline climatology studies (Mitchell and Schultz, 2020), and the surface analyses more broadly have been used to train and test machine learning models for automated front and dryline detection (e.g., Justin et al., 2025).

WPC analyses are a subjective analysis of mesoscale and synoptic scale phenomena, and thus may be subject to human biases. Previous work has suggested that the dominant mode of drylines in the central United States has led to a systematic

misanalysis of drylines in the southeast US as cold fronts (Duell and Van Den Broeke, 2016). This motivates a need for automated methods to identify fronts and drylines, both for operational consistency and also for scientific work which seeks to understand the long-term climatological characteristics of these features. Considerable progress has been made in this area, particularly with machine learning techniques (e.g., Clark et al., 2015; Lagerquist et al., 2020; Justin et al., 2023, 2025). A significant component of developing these algorithms is having a sufficiently large dataset of “ground truth” data to train and test the model, and with the WPC surface analysis archive often being used in that capacity (e.g., Lagerquist et al., 2020; Justin et al., 2023, 2025). While any subjective biases may be a limitation of a model trained using the WPC archive, an understanding of these biases can be used to design methodologies to effectively mitigate their impacts.

This Chapter has two primary objectives:

1. Establish baseline climatology statistics for dryline frequency and position for drylines and those associated with severe convection, with a particular emphasis on non-spring and non-evening statistics which have not previously been reported.
2. Evaluate the WPC dryline archive for potential influence of non-meteorological factors/biases which may need to be accounted for when using this dataset as “ground truth” in model development or physical analysis.

2.2 Data and Methods

2.2.1 WPC Surface Analysis Archive

The modern WPC surface analysis archive for the CONUS spans from 2007–present. Analyses are produced every 3 hours (0000, 0300, ... , 2100 UTC). A 16-year period

(2007–2022) is used in this study (NOAA, 2023). Frontal data, including drylines, from these archived analyses are available in the form of a list of latitude and longitude coordinates. Line segments drawn from the archived dryline coordinates are used to “re-create” the drylines as a continuous line. The lines are then gridded to a $0.25^\circ \times 0.25^\circ$ grid, with any grid-box which the dryline intersects being considered a “dryline grid-box”. Gridding the dryline in this manner may obscure some details of small-scale features such as localized dryline bulges, which can be important to convection initiation (e.g., Hane, 2004). However, analysis of such small-scale dryline features is outside the scope of this study. A mask is applied over Canada, Mexico, and the ocean. In total, there are 6,997 individual drylines across 1,360 dryline days, defined as any daily period 1200–0900 UTC with at least one dryline present at one analysis time.

2.2.2 Designating “Severe” and “Non-Severe” Drylines

Every individual dryline is categorized as “Severe” if it is associated with severe convection and “Non-Severe” otherwise. “Non-Severe” drylines thus includes: 1) drylines which failed to initiate any convection at all; 2) drylines which initiated convection that remained sub-severe; or 3) drylines which initiated sub-severe convection which became severe outside of a predefined spatiotemporal window. A dryline is labeled “Severe” if a severe storm report—wind (damage or measured >50 kts), hail greater than 1 inch in diameter, or a tornado—from the NOAA Storm Prediction Center (SPC) Local Storm Reports Database (SPC, 2023) or a severe thunderstorm or tornado warning issued by a NWS forecast office (IEM, 2023) falls within a specified temporal and spatial proximity. For each individual grid-box a dryline intersects, a search window extends 1.5° (6 grid-boxes or ~ 150 km) east of the dryline grid-box, 0.5° (~ 50 km) west, and 0.25° (~ 25 km) north and south. The slight northward and southward extent of the search window reflects the slight northerly or southerly component to storm motion

common in the central US (e.g., VandenBerg et al., 2014). The ~ 150 km eastward extent of the search window is further east than past studies which have investigated dryline convection initiation. Mitchell and Schultz (2020) used 100 km east of the dryline as their eastward extent to identify initial occurrences of 40-dBZ radar echoes to define dryline-initiated convection, based on the work by Rhea (1966). This present study extends that threshold an additional ~ 50 km east to account for the time-lag between convection initiation and when a storm may produce severe hazards. With a typical storm motion on the order of tens of kilometers over tens of minutes (VandenBerg et al., 2014), this additional eastward extent of the search window provides storms initiated by the dryline an additional 30–60 minutes to intensify to severe criteria. A dryline is considered “Severe” if at least one report or warning—gridded to the same $0.25^\circ \times 0.25^\circ$ grid as drylines—is within any search window, and if the report or warning is issued within 1 hour prior to 2 hours after the dryline analysis time. This temporal window accounts for the 3 hours between WPC surface analysis maps while preventing reports or warnings from being double-counted to multiple drylines. The temporal window is off-centered with respect to the 3-hour gap between analyses in order to account for the potential temporal lag between the initiation of convection and its evolution into severe convection, in a similar manner to how the spatial search polygon is extended ~ 50 km further east than Mitchell and Schultz (2020). The ~ 50 km westward extent of the search window accounts for dryline-associated convection in the temporal window 1 hour prior to the analysis time, when the dryline and its associated convection may be further west. After categorizing every individual dryline as “Severe” or “Non-Severe”, the dataset is organized into dryline days (1200 UTC–0900 UTC, Chapter 2.2.1).

Although NWS-issued warnings are not immune to human biases and do not guarantee that a warned storm will physically produce a severe hazard (e.g., Barnes et al.,

2007; Barret, 2012; White and Stallins, 2017), they are included in the “Severe” criteria to supplement and help correct some of the well-known biases that affect the Local Storm Report database (e.g., Brooks et al., 2003; Ray et al., 2003; Doswell III et al., 2005; Paulikas, 2014; Allen and Tippett, 2015b). Of particular interest to this study was reducing errors of missing severe reports in regions of low population density (e.g., Brooks et al., 2003; Ray et al., 2003; Paulikas, 2014; Allen and Tippett, 2015b). There is some overlap of biases between warnings and reports: relationships between population density and warning issuance have been observed, although these relationships can vary by urban center and Weather Forecast Office (White and Stallins, 2017; Barret, 2012). While this combined dataset of warnings and reports is still not immune from human biases, using both in conjunction reduces the magnitude of these biases compared to using reports individually.

Reports are gridded to a single grid box. For tornado tracks, which may extend across multiple grid boxes for long-tracked tornadoes, the start location of the track is used. Warnings, which cover a spatial area and not just a single point, are polygons with multiple coordinates detailing their shape and extent. Every grid box which is intersected by a warning polygon is considered a “warned” grid-box (e.g., larger warning polygons led to more grid points flagged as warned).

2.2.3 Machine Learning Model and Training

Results from the WPC-analyzed drylines archive are compared to a machine learning (ML) model developed to identify fronts and drylines across the CONUS (Justin et al., 2025). The ML model is developed using the UNet 3+ architecture, a convolutional neural network for image segmentation (Huang et al., 2020), and predicts cold, warm, stationary, and occluded fronts in addition to drylines (Justin et al., 2025). Predictor variables include 10 native and derived quantities from ERA5 reanalysis (Hersbach

et al., 2018) on a $0.25^\circ \times 0.25^\circ$ grid at 5 vertical levels from 2-m AGL up to 850 hPa. Predictors are fed to the model at 3-hour increments synchronous with WPC surface analyses which are used as ground truth to train against (Justin et al., 2025). Using the predictor variables, the ML model predicts 6 probabilities at each 0.25° grid point:

$$P(\text{no front}) + P(\text{cold}) + P(\text{warm}) + P(\text{stationary}) + P(\text{occluded}) + P(\text{dryline}) = 100\%, \quad (2.1)$$

where P represents the probability of the given type of front, or no front. The output probabilities at each grid point are then calibrated to a 100 km neighborhood, representing the physical probability of each front type within 100 km of a given grid point (Justin et al., 2023).

For the purposes of this study, only dryline probabilities are considered. Any analysis time with at least one grid point of dryline probability exceeding 50% is deemed to have a dryline, and any convective day with at least one dryline identified is counted as a dryline day. Because the output probabilities are calibrated to a 100 km neighborhood, at any one grid point the sum of front type probabilities may no longer equal 100% as in Eq. (2.1). Analysis of grid points where both the calibrated dryline probability and probability of some other front type (or no front) exceed 50% is outside the scope of this study. The spatial coherence of probabilities exceeding 50%, and the possibility of an analysis time with multiple distinct drylines, is not considered as such occurrences were rare in WPC surface analyses.

The ML model is trained using WPC surface analyses 2007–2022, with an 8-fold cross-validation process. For each fold, 4 years are withheld from training, with 2 reserved for validation and 2 reserved for testing (Table 2.2). The 2 validation and testing years are intentionally non-consecutive in order to prevent model over- or under-

predicting drylines due to the hypothesized annual trend (Chapter 2.3.3).

Model Num.	Training Years	Validation Years	Testing Years
1	2007–2012, 2015–2020	2013, 2021	2014, 2022
2	2008–2013, 2016–2021	2014, 2022	2007, 2015
3	2009–2014, 2017–2022	2007, 2015	2008, 2016
4	2007, 2010–2015, 2018–2022	2008, 2016	2009, 2017
5	2007–2008, 2011–2016, 2019–2022	2009, 2017	2010, 2018
6	2007–2009, 2012–2017, 2020–2022	2010, 2018	2011, 2019
7	2007–2010, 2013–2018, 2021–2022	2011, 2019	2012, 2020
8	2007–2011, 2014–2019, 2022	2012, 2020	2013, 2021

Table 2.2: Training, validation, and testing configuration for each model in the cross-validation procedure

2.3 Results

2.3.1 WPC Monthly and 3-Hourly Statistics

The majority of WPC-analyzed dryline days are analyzed in meteorological spring (March–May) and June (Fig. 2.1). The fraction of dryline days which are “Severe” peaks close to 75% in May and July, with a steady increase from winter through spring and a steady decline from summer into winter. July and August have the fewest number of dryline days per month, yet these months also have among the highest fraction of dryline days which are associated with severe convection, consistent with the warm season peak in thunderstorm frequency (e.g., Changnon, 2001). There is a secondary peak in dryline day frequency in October, although its magnitude of ~ 5 days per month is only a fraction of the springtime peak of ~ 18 days per month in May. While April through June—the period when drylines are most commonly studied (Table 2.1)—have the most “Severe” dryline days per month, March has a similar number of dryline days

per month to June (Fig. 2.1).

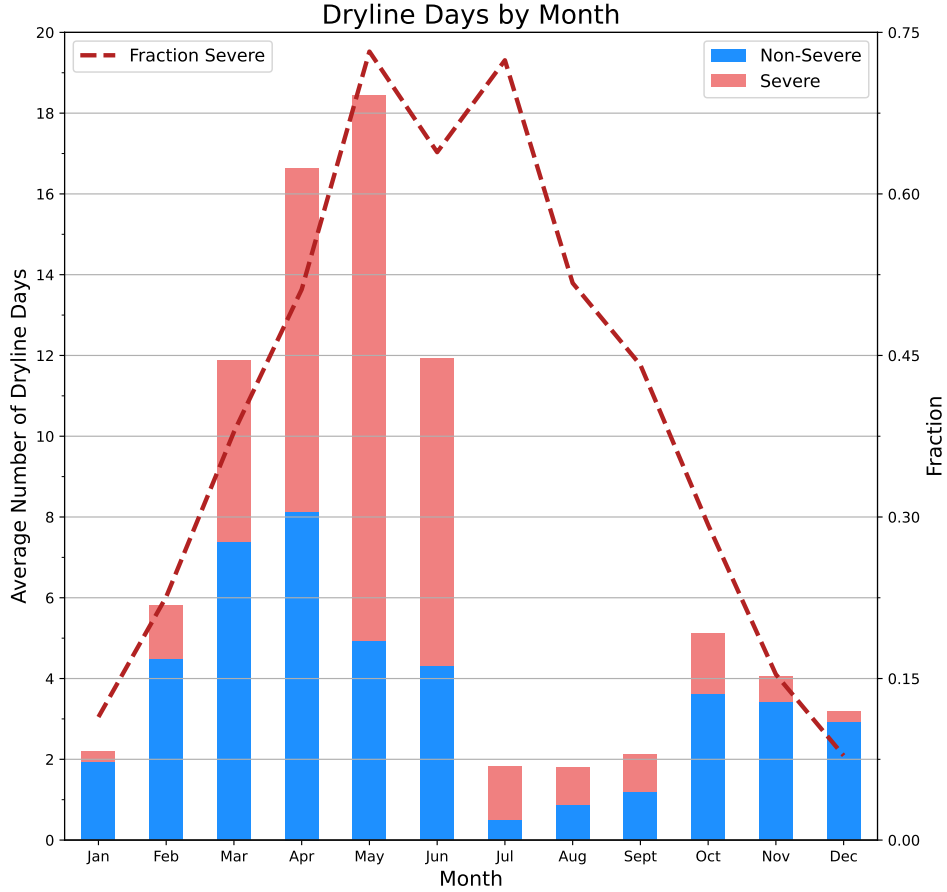


Figure 2.1: The average number of “Non-Severe” (blue) and “Severe” (red) WPC-analyzed dryline days per month for the entire CONUS. The red-dashed line indicates the fraction of dryline days per month that are severe (right axis)

The distribution of individual drylines analyzed per analysis time for each meteorological season is similar in both relative dryline frequency and the fraction of individual drylines which are “Severe” across all seasons (Fig. 2.2). The total number of drylines is marginally higher in the late-afternoon/early evening hours (2100–0300 UTC), consistent with the time when the dryline specific humidity and temperature gradients are strongest and most distinct from other boundary types (Chapter 2.1). Throughout each season, the highest frequency of “Severe” drylines is also in the late-afternoon/early evening, coincident with the diurnal severe convective maximum (e.g.,

Krocak and Brooks, 2020).

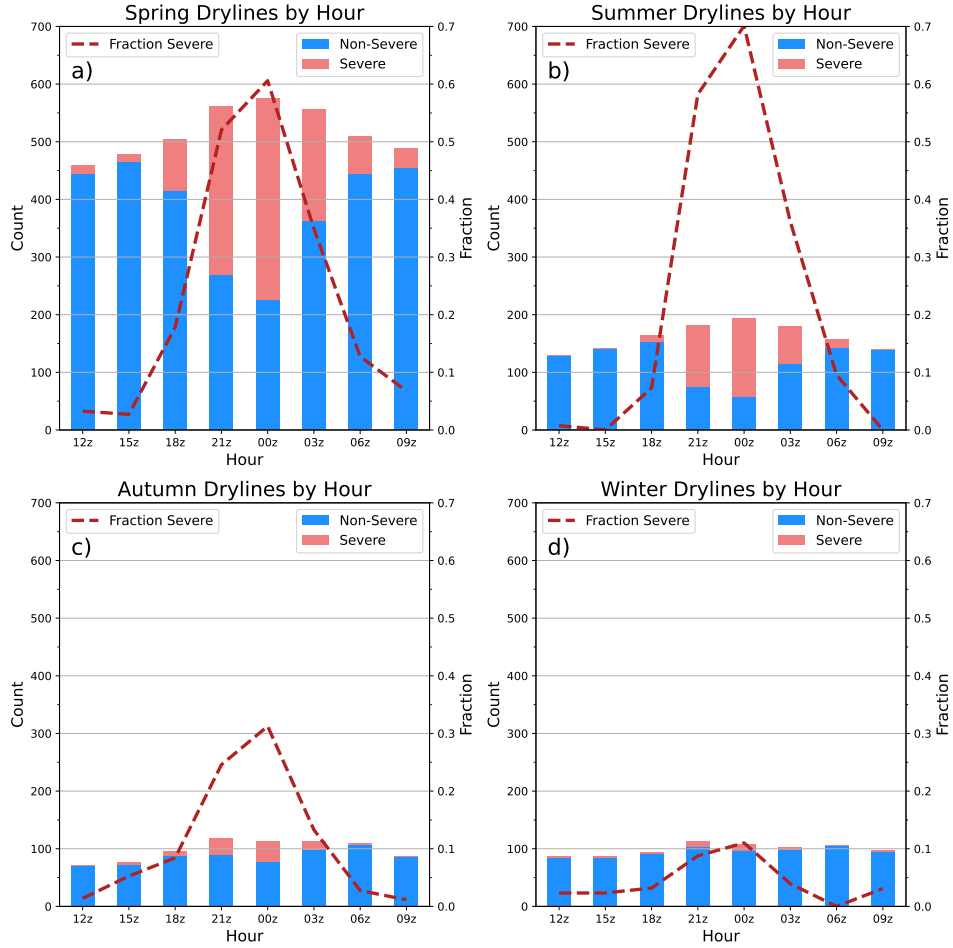


Figure 2.2: The number of individual “Non-Severe” (blue) and “Severe” (red) WPC-analyzed drylines by analysis hour for (a) meteorological spring, (b) summer, (c) autumn, and (d) winter for the entire CONUS. Red-dashed lines indicate the fraction of drylines that are severe for each analysis time (right axis).

2.3.2 WPC Spatial Characteristics

All WPC-analyzed drylines are located west of the Appalachian Mountains, with only a handful extending east of the Mississippi River. The majority of drylines are east of the Rocky Mountains, although there are several drylines analyzed in Arizona (Fig. 2.3). This large area is divided into five sub-regions based roughly on previous dryline

climatologies and regional climate/terrain, although our Southern Great Plains region extends further west than most previous climatologies (Fig. 2.4; Table 2.1). The extension of the Southern Great Plains region further west than some previous studies is intended to separate drylines identified in central/eastern New Mexico and Colorado along the eastern slopes of the Rocky Mountains from drylines identified further west in the Intermountain West and in Arizona (Fig. 2.3). All WPC-analyzed drylines in the CONUS are entirely within the defined sub-regions.

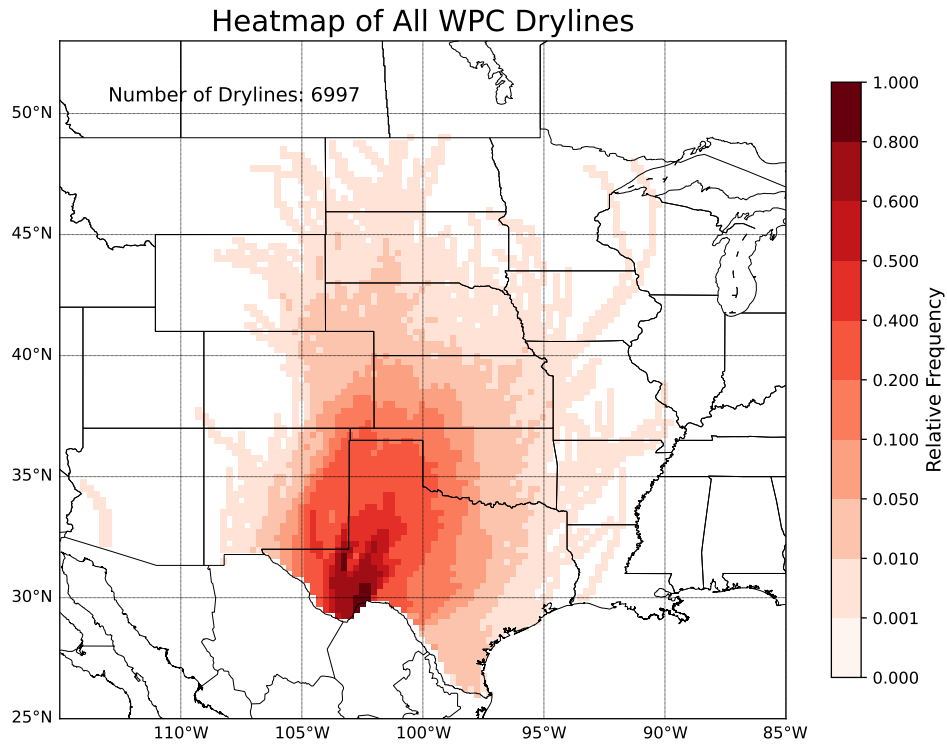


Figure 2.3: Heatmap of the relative frequency of individual drylines per grid-box. All WPC-analyzed drylines are within the bounds of the figure. Note the colorscale is non-linear.

An individual dryline is considered to be within a particular sub-region if *any single* dryline grid-box is within the sub-region, making it possible (and not infrequent) for drylines to be considered part of multiple sub-regions. “Severe” and “Non-Severe” individual dryline regional frequencies are calculated by sub-region for the entire year,



Figure 2.4: Delineation of the five sub-regions: Southern Great Plains, Northern Great Plains, Midwest, Southeast, and Southwest. Note: Portions of Mexico and Canada included in the polygons were not included in the analysis.

and also for each season (Tables 2.3, 2.4). The overwhelming majority of both “Severe” and “Non-Severe” drylines reside in the Southern Great Plains, both year-round and for each individual season. The Northern Great Plains is the second most frequent sub-region, with a peak in both “Severe” and “Non-Severe” frequency in summer. Few drylines are analyzed in the Midwest, Southeast, or Southwest. Drylines which progress east of the Great Plains typically do so after they are overtaken by and merge with an advancing cold front, and thus these boundaries are no longer drylines (e.g., Neiman and Wakimoto, 1999; Schultz, 2005). However, it has been hypothesized that the rate at which drylines are identified east of the Great Plains in WPC surface analyses is too low, potentially from a spatial bias which leads to these boundaries being misclassified

as cold fronts (Duell and Van Den Broeke, 2016).

Non-Severe	<i>All</i> (5479)	<i>Spring</i> (3083)	<i>Summer</i> (953)	<i>Autumn</i> (636)	<i>Winter</i> (757)
Southern Great Plains	5456 (99.6%)	3078 (99.8%)	939 (98.5%)	682 (99.4%)	757 (100%)
Northern Great Plains	430 (7.8%)	215 (7.0%)	131 (13.7%)	59 (8.6%)	25 (3.3%)
Midwest	35 (0.6%)	16 (0.5%)	0 (0.0%)	3 (0.4%)	16 (2.1%)
Southeast	12 (0.2%)	5 (0.2%)	0 (0.0%)	2 (0.3%)	5 (0.7%)
Southwest	51 (0.9%)	19 (0.6%)	20 (2.1%)	12 (1.7%)	0 (0.0%)

Table 2.3: The number of individual non-severe WPC-analyzed drylines by region (row) and season (column). Total count for the entire season is shown in the top row. The relative frequency of individual drylines in each sub-region by meteorological season is in parenthetical (Fig. 2.4). Since drylines can be counted in multiple regions, the relative frequencies will not necessarily add up to 100%.

Severe	<i>All</i> (1518)	<i>Spring</i> (1051)	<i>Summer</i> (335)	<i>Autumn</i> (96)	<i>Winter</i> (36)
Southern Great Plains	1507 (99.3%)	1050 (99.9%)	325 (97.0%)	96 (100%)	36 (100%)
Northern Great Plains	175 (11.5%)	89 (8.5%)	73 (21.8%)	13 (13.5%)	0 (0.0%)
Midwest	17 (1.1%)	15 (1.4%)	2 (0.6%)	0 (0.0%)	0 (0.0%)
Southeast	6 (0.4%)	6 (0.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Southwest	9 (0.6%)	8 (0.8%)	0 (0.0%)	1 (1.0%)	0 (0.0%)

Table 2.4: As in Table 2.3, but for severe WPC-analyzed drylines.

Spatially, there is no obvious difference between “Severe” and “Non-Severe” individual drylines (not shown). The highest frequency of drylines is in southwest Texas, with eastern New Mexico through western Oklahoma outlining a common area. Dryline frequency drops off rapidly north of Oklahoma and east of western Oklahoma/central Texas (Fig. 2.3).

The average furthest-east longitude for all drylines is consistent with the range of

longitudes found in previous studies (not shown) (Hoch and Markowski, 2005; Coffey et al., 2013; Johnson and Hitchens, 2018). Given the dominance of springtime Southern Great Plains drylines, it is unsurprising that the year-round full-CONUS average is similar to previous Great Plains climatologies.

The average dryline longitude by analysis hour distinctly shows the eastward progression of drylines during daytime mixing with an eastward progression of $1\text{--}1.5^\circ$ from 1200 to 2100/0000 UTC (Fig. 2.5). The reverse progression is observed during the overnight period from 0000 to 0900 UTC, consistent with characteristic dryline evolution (Chapter 2.1). This diurnal cycle is evident in every month of the year except for December, which has a more subtle diurnal cycle on the order of 0.25° . Superimposed on the diurnal cycle is an annual cycle where the dryline longitude changes by $3\text{--}4^\circ$ over the year (Fig. 2.5). Winter drylines are the furthest east, while summer drylines are the furthest west. This result is consistent with the observed westward shift in drylines observed between early and late spring (Hoch and Markowski, 2005; Johnson and Hitchens, 2018).

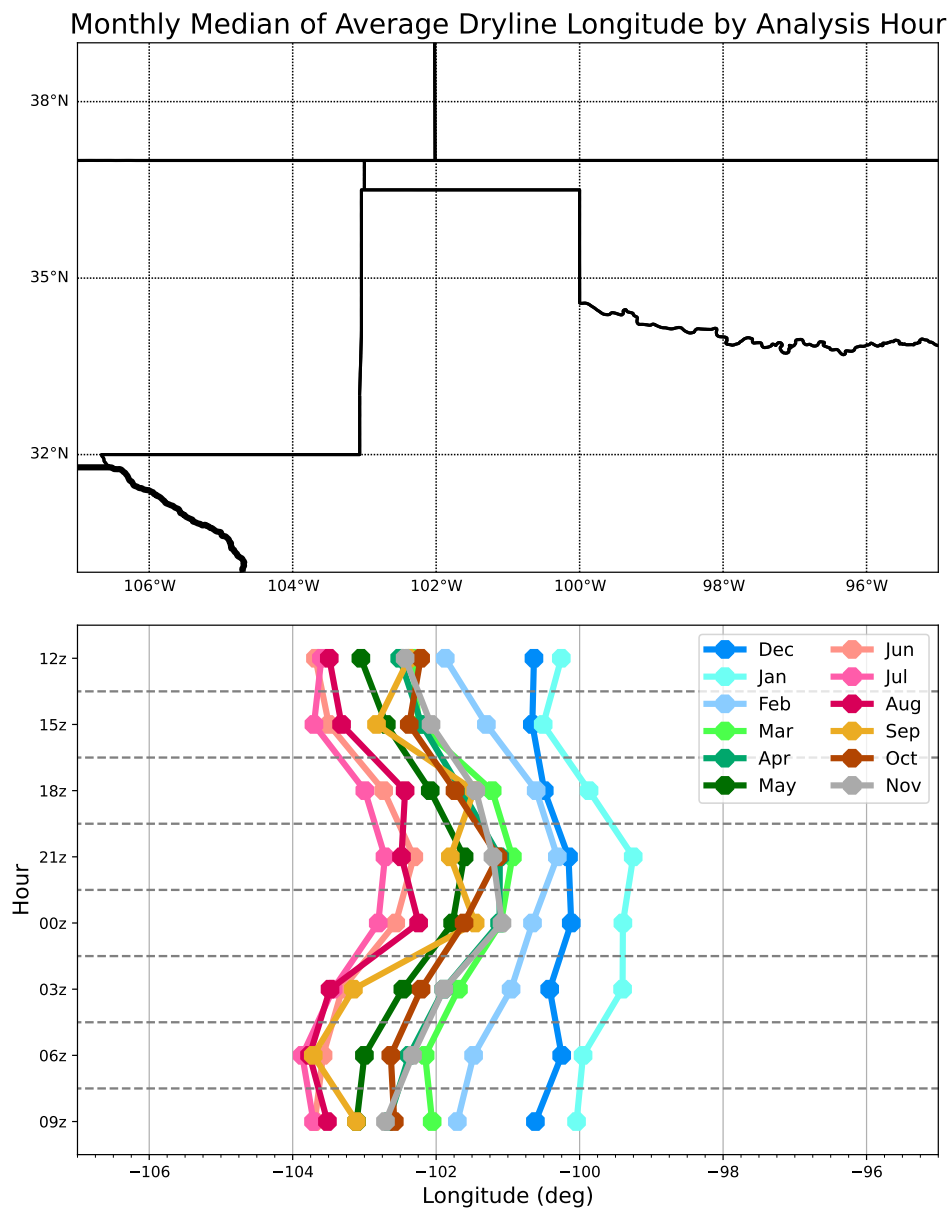


Figure 2.5: Median of average individual WPC-analyzed dryline longitude by month (color) and analysis hour (y-axis). Colors are grouped by meteorological season. A reference map is shown at top.

There is no clear diurnal cycle evident in dryline average latitude (Fig. 2.6). Across most months, dryline latitude decreases from 1800 or 2100 through 0600 UTC, with a decrease of just $0.25\text{--}0.5^\circ$ in spring and $0.75\text{--}1^\circ$ in November and winter. In spring months, northward extent is typically maximized at 1800 or 2100 UTC, while in summer it is maximized later at 0000 UTC, consistent with the longer days and later sunsets during summer. The large 2° northward shift from 1500 to 1800 UTC in August can be attributed to the influence of small sample sizes, with only 11 and 12 drylines analyzed by the WPC during those two analysis times, respectively. Of the 12 WPC-analyzed drylines at 1800 UTC in August, only 7 of them were preceded by a WPC-analyzed dryline at 1500 UTC (not shown). Similar small sample sizes effects are evident in the 2° southward shift from 0600 to 0900 UTC in August, and in the 1.5° northward shift from 1800 to 2100 UTC in July.

Annually, there is a subtle northward shift of approximately 1° from winter until June before a significant shift northward for July through September (Fig. 2.6). However, these three months have the fewest number of drylines and thus may be more influenced by individual outliers (Fig. 2.1).

2.3.3 WPC Annual Trend

There is an increasing trend in the number of WPC dryline days per year during the analysis time period (Fig. 2.7). While 2011 and 2015 stand out as high and low outliers, respectively, compared to their neighboring years, every year from 2016 onward has more dryline days than any non-2011 year from 2007–2015. A randomization test on the annual dryline day count with 10,000 iterations demonstrated the observed linear trend to be statistically higher than the randomly-shuffled trend distribution to the 99th percentile. This increasing annual trend is evident in both “Severe” and “Non-Severe” dryline days, both of which are also statistically significant using the same randomization test (Fig. 2.8). There is no significant trend in the ratio of “Severe” to “Non-Severe” dryline days. While 2015 has by far the fewest dryline days, this appears to primarily be the result of anomalously few “Non-Severe” dryline days, as 2007 and 2014 both have fewer “Severe” dryline days (Fig. 2.8).

2.4 Discussion

2.4.1 Annual Trend

The statistically significant increase in the number of WPC dryline days over this 16-year period is an unexpected result (Chapter 2.3.3). The increasing trend is evident in the total amount of individual drylines over this period as well as dryline days, with no change in the ratio of individual “Severe” to “Non-Severe” drylines (not shown). The increase is present in all seasons (not shown), although the relative rate of increase is highest in summer and winter due to these seasons having a lower total number of drylines compared to spring (Fig. 2.1).

The rate of increase in analyzed drylines is nearly equivalent across all analysis

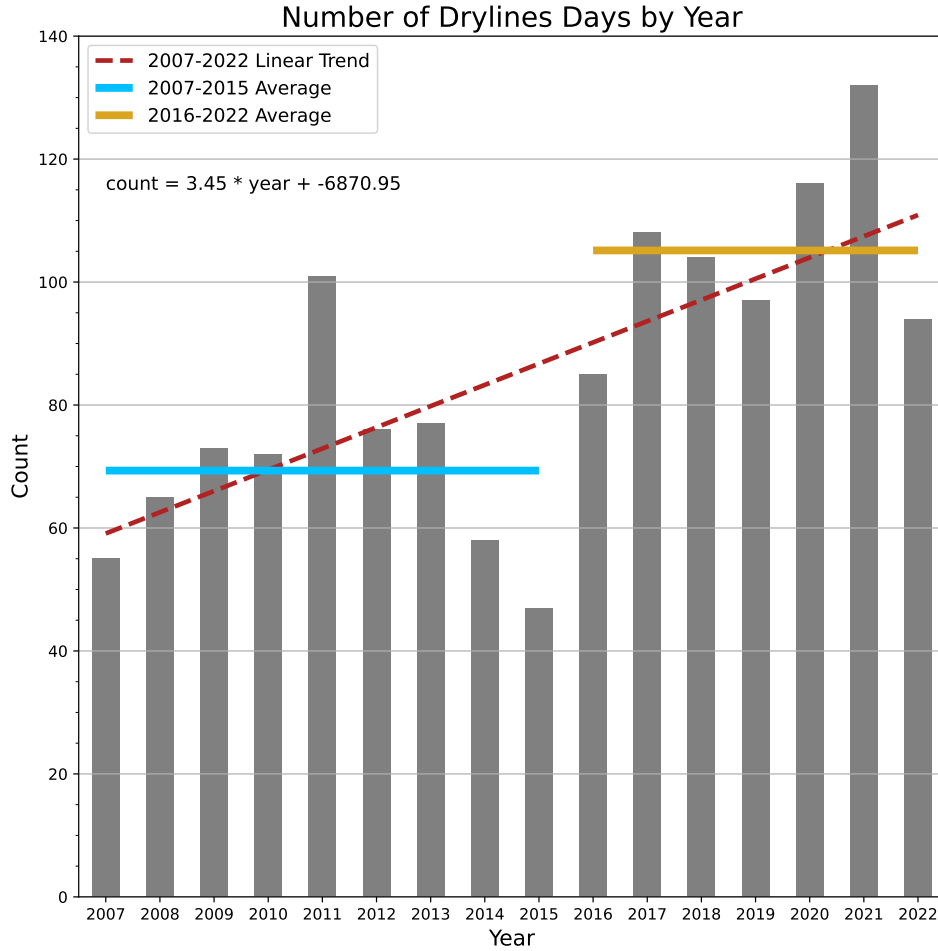


Figure 2.7: Number of WPC-analyzed dryline days per year. The average number of dryline days for the period 2007–2015 (blue) and 2016–2022 (gold) are shown. The red-dashed line depicts the linear trend, and the equation for the trend is displayed in the upper left below the legend.

times (Fig. 2.9), including 2100 UTC when it is thought that there would be the smallest deviation in drylines analyzed between different individual human forecasters due to the temperature and moisture gradients of the dryline being the strongest and most distinct from other boundaries at this time (Chapter 2.1).

Recent work by Andrews et al. (2024) found that the frequency of the Elevated Mixed Layer (EML) in the Great Plains is increasing on the order of 4–5 more EML days per decade in their 32-year climatology. Although the presence of an EML does

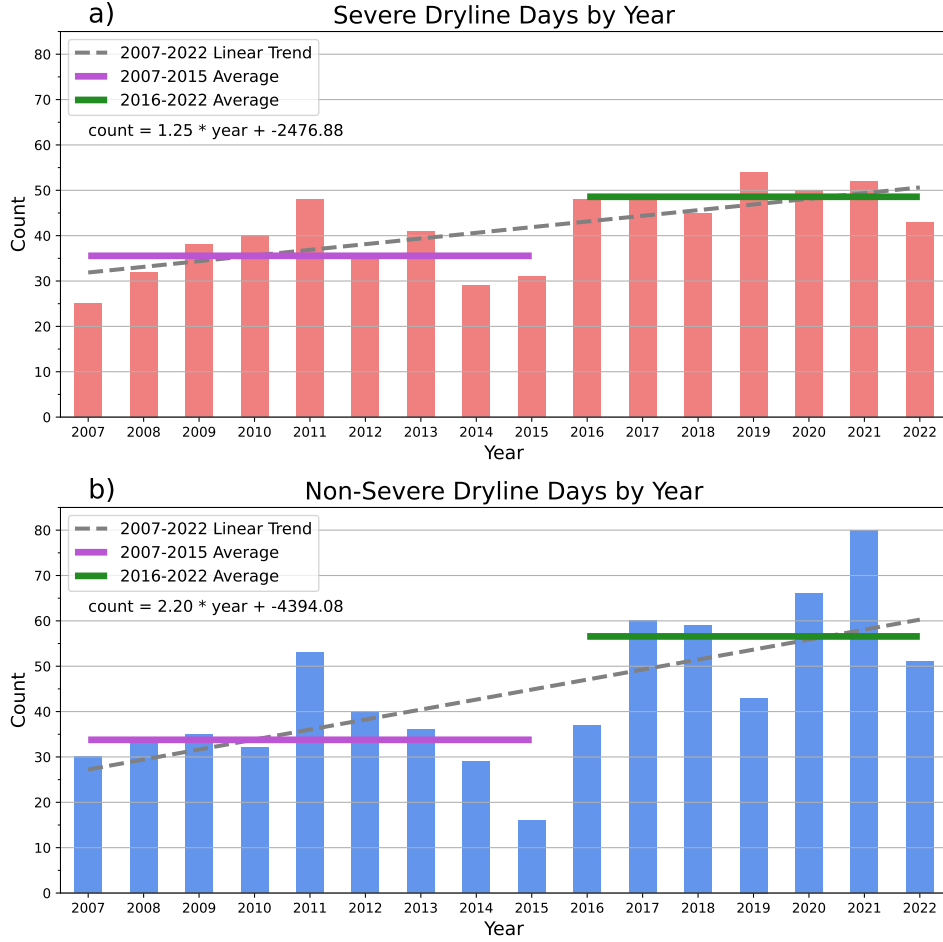


Figure 2.8: Number of (a) “Severe” and (b) “Non-Severe” WPC-analyzed dryline days per year. The average number of dryline days for the period 2007–2015 (purple) and 2016–2022 (green) are shown. The gray-dashed lines represent the linear trend, and the equation for the trend is displayed in the upper left below the legend.

not guarantee a dryline, drylines are typically associated with an EML (e.g., Parsons et al., 2000). Therefore, the increase in EML days could be evidence of a physical cause for the increase in dryline days evident in the WPC analyses.

However, the number of EML days east of 105°W —where we typically associate EMLs with drylines—does not show an increase in EML days over our 16-year period (not shown). Additionally, after performing a first-order differencing procedure to remove the trend, Kendall’s τ_B correlation was used to assess the monotonic relationship

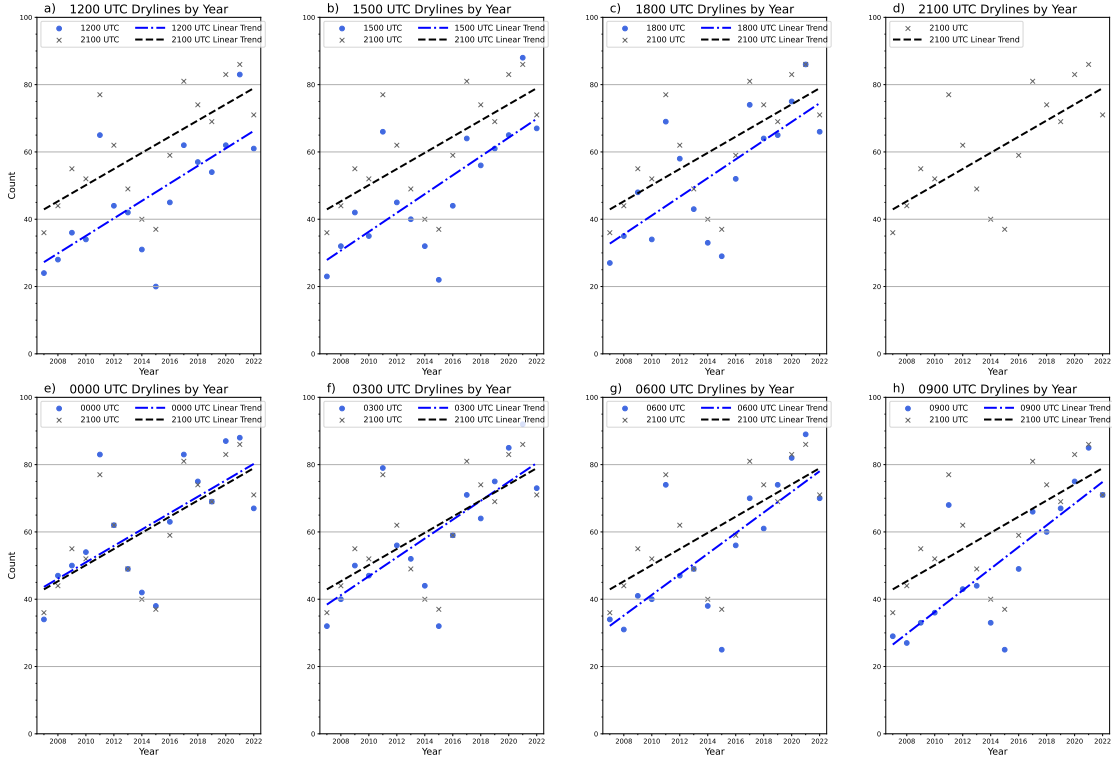


Figure 2.9: WPC-analyzed Dryline counts by year for each 3-hour analysis time, from (a) 1200 UTC through (h) 0900 UTC. Blue dots represent the count for the given analysis time, while gray ‘x’ indicates the count from 2100 UTC, the reference time (panel d). Linear trends are shown for both the given and reference time in blue and gray dashed-lines, respectively.

between the change in EML days and WPC dryline days per year and did not show a statistically significant relationship (not shown). It is likely our analysis does not show increasing EML day frequency due to our short analysis period and the high variability in year-to-year EML days (Andrews et al., 2024). Future work may explore the relationship between EML days and dryline days over a longer analysis period.

In discussions with several WPC forecasters, the forecasters unanimously hypothesized that the increasing trend in WPC analyzed drylines did not reflect a physical increase in dryline frequency (Greg Carbin, 2023, personal communication). There are many plausible non-meteorological factors that might explain the increase in analyzed drylines, including improvements to observing networks (satellites, mesonets/surface

stations, and real-time mesoanalyses), evolving criteria to define a dryline due to recent research, or even simply turnover in the forecasters responsible for producing these analyses and their potentially differing expert opinions on how they identify certain atmospheric features. The latter two were suggested by WPC forecasters, but we have no empirical evidence to support any of these hypotheses. An analysis of the variability of dryline identification frequency between different WPC forecasters is outside the scope of this paper.

2.4.2 Comparison to Machine Learning Model

2.4.2.1 Model Results

The spatial distribution of drylines identified by the ML model is similar to the WPC-analyzed distribution (Fig. 2.10). The ML model depicts the relative maximum in southwest Texas and the drop-off in frequency with northward and eastward extent. Few-to-no drylines are identified in the Midwest and Southeast. While consistent with the few drylines analyzed by the WPC in these regions (Fig. 2.10b), this may be a reflection of a spatial bias hypothesized by Duell and Van Den Broeke (2016) in WPC surface analyses propagating into the ML model. Further study is needed to properly assess how many boundaries in the Midwest and Southeast should be classified as drylines and how to design consistent definitions for such classification.

Our ML model, like any other automated or manually identification scheme, does not represent a perfect boundary identification scheme, because no such scheme can exist with human-defined categorical bins for the spectrum of atmospheric parameters we are seeking to define. Any identification methodology is subject to variability based on the explicit (i.e., researcher defined criteria) or implicit (i.e., matching the identification patterns of forecasters) decisions on how certain quantities should be

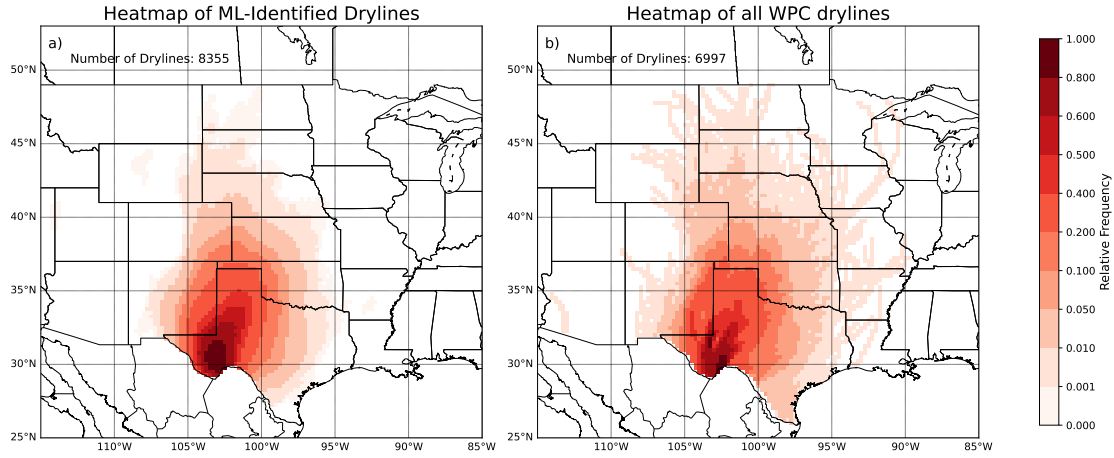


Figure 2.10: Relative frequency heatmap of (a) ML-identified drylines and (b) the WPC-analyzed drylines. As in Fig. 2.3, the colorscale is nonlinear. The total number of drylines is in the upper left corner.

considered (Thomas and Schultz, 2019a,b). What our ML model does do, however, is provide a *consistent* application of WPC expert forecaster criteria to identify drylines across the analysis period using a suite of 50 atmospheric variables (Justin et al., 2025). The marginal increasing trend in annual dryline days per year in the ML model is not statistically significant and cannot be determined to be independent from outliers (Fig. 2.11a), supporting the hypothesis that the increase observed in WPC record is the result of non-meteorological factors.

The ML model identifies significantly more dryline days compared to the WPC record from 2008–2015 (Fig. 2.11a). Despite this difference, the ML model appears to capture the interannual variability in the number of dryline days seen in the WPC record, with 2011 being a relative maximum for dryline day count and 2007 and 2015 being relative minima. Additionally, as we expect based on the physical relationship between the EML and the dryline, changes in EML days year-to-year are positively correlated with changes in ML dryline days year-to-year (i.e., if the number of EML days increase compared to the prior year, the number of ML dryline days increase as well). Evidence of this relationship suggests that the hypothesized non-meteorological

factors causing the significant increase in WPC analyzed drylines are masking the physical relationship between EML days and dryline days. Future work will be needed to explore the relationship between ML dryline days and EML days over a longer analysis period to determine if the increase in EML days observed by Andrews et al. (2024) is correlated with a long-term increase in the number of dryline days.

The number of ML-identified dryline days per year prior to 2015 is consistent with ML-identified and WPC-analyzed dryline days per year after 2015. While these results may seem to imply that the rate at which drylines are identified post-2015 is the “true” frequency of drylines, further study of the ML output beyond the scope of this paper should be undertaken to make this assessment. ML output probabilities using a 50 km neighborhood (not shown) depict a consistent 60–80 dryline days per year, more similar to the pre-2015 WPC annual dryline analysis rate.

Some difference in frequency aside, the ML model captures the same monthly cycle observed in the WPC record (Fig. 2.11b). The ML model identifies the secondary peak in dryline day frequency in October, similar to the WPC. However, the ML model identifies a higher frequency of dryline days in June than the WPC. WPC-analyzed dryline day frequency in June is on par with dryline day frequency in March, while the ML model depicts June to have a similar frequency of dryline days as April. In both the WPC record and the ML model, July is the month with the lowest dryline day frequency.

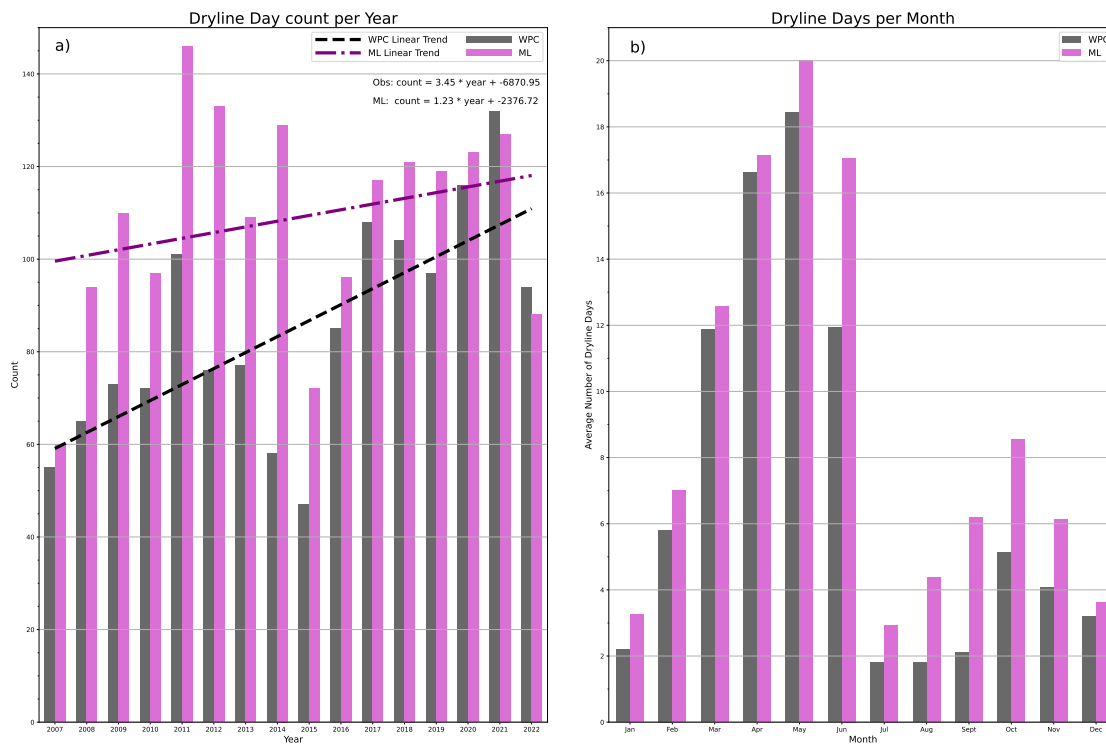


Figure 2.11: (a) A comparison between WPC-analyzed dryline days (gray) and ML-identified dryline days (purple) per year and (b) average days per month. The black-dashed line in (a) represents the linear trend in WPC-analyzed dryline day counts per year, while the purple-dash-dot line represents the linear trend in ML-identified dryline day counts per year. Equations for the linear trends are located below the legend in (a).

2.4.2.2 A Case Study: 2100 UTC 5 May 2014

The ML model identified a dryline at 1358 more surface analysis times than the WPC record throughout the analysis period. The year with the greatest difference between the number of dryline days analyzed by the WPC and the ML model is 2014, where the ML model identified 61 more days with drylines (Fig. 2.11a). To demonstrate the types of features the ML model is identifying as a dryline when the WPC does not analyze a dryline, surface potential temperature and specific humidity derived from ERA5 reanalysis are analyzed at 2100 UTC 5 May 2014 (Fig. 2.12). In previous studies, a 3 g kg^{-1} per 100 km specific humidity gradient co-located with an inverse potential temperature gradient (e.g., the drier side is warmer than the moist side) is the standard criteria used for dryline identification (Schaefer, 1986; Hoch and Markowski, 2005; Hoogewind, 2012; Coffer et al., 2013; Johnson and Hitchens, 2018).

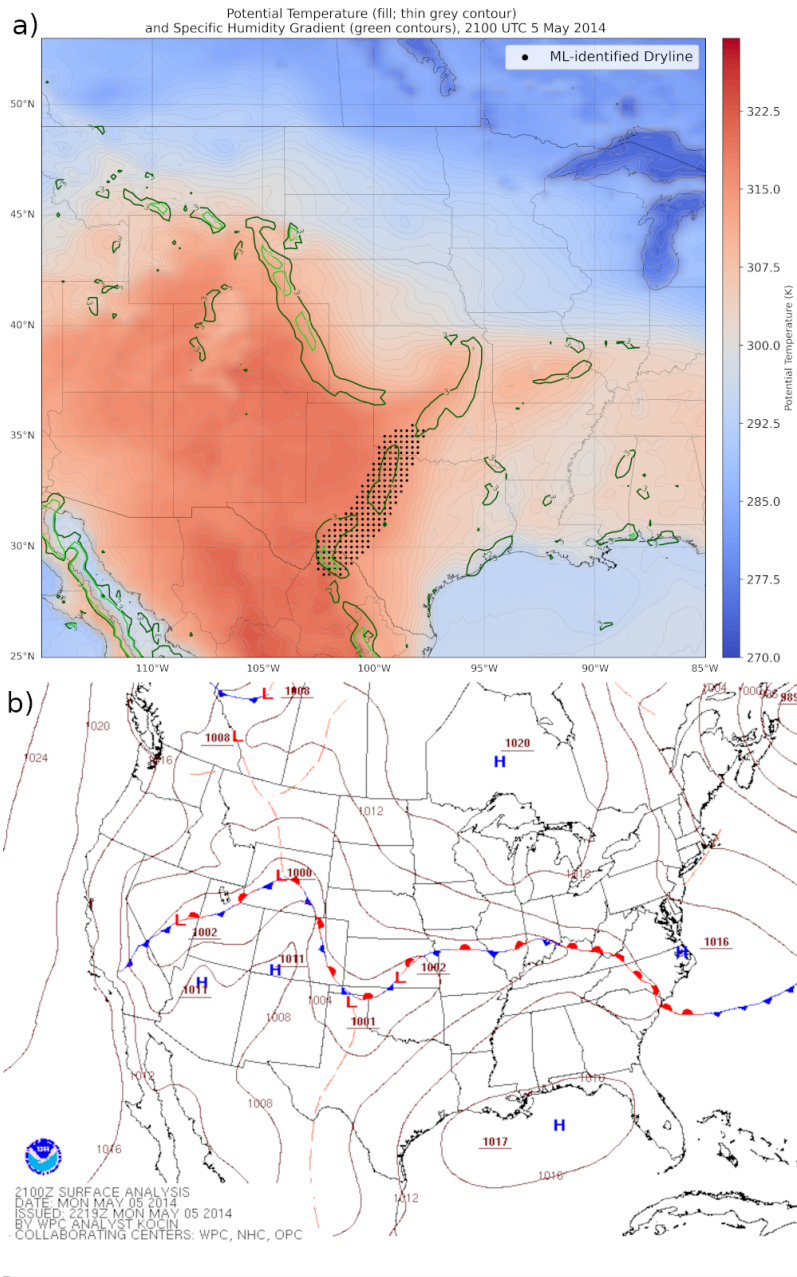


Figure 2.12: (a) Potential temperature (K; fill and light gray contours) and specific humidity gradient (green) contoured at 3, 6 g kg⁻¹ per 100 km at 2100 UTC 5 May 2014. Black stippling indicates where the ML model identified a dryline. (b) The corresponding WPC surface analysis map at the corresponding analysis time.

The ERA5 reanalysis shows a potential temperature boundary is evident across eastern Wyoming and Colorado, extending eastward through Indiana, consistent with the stationary boundary drawn in the WPC surface analysis (Fig. 2.12). Additionally, a meridionally-oriented potential temperature gradient is evident from eastern Kansas south–southwestward through central Texas (Fig. 2.12a). A 3 g kg^{-1} specific humidity gradient is evident along portions of that potential temperature gradient, with small gaps in southwest Oklahoma and central Texas where the specific humidity gradient is weaker. The stippling from the US–Mexico border through central Oklahoma depicts where the ML model identifies a dryline. Although the rigid 3 g kg^{-1} specific humidity gradient threshold is not met across the entire boundary, it appears as though there is enough of a moisture gradient evident to, along with the potential temperature gradient, reasonably classify this boundary as a dryline. In the WPC surface analysis, this boundary is identified as a mesoscale trough (brown-dashed lines) in the Texas panhandle through central/west Texas, slightly west of the moisture and potential temperature gradient in the ERA5 reanalysis (Fig. 2.12b).

This case study highlights the necessity of consistent methodologies to classify atmospheric features such as drylines. There may be reasonable and justifiable disagreements between expert forecasters regarding the classification of a boundary. Even if the authors themselves were to manually identify drylines throughout the entire period according to the established criteria—a dauntingly lengthy process for achieving a climatological dataset—the process would still be fraught with subjective decisions such as how to classify the feature of interest at 2100 UTC 5 May 2014 due to the fluid and non-discrete nature of the atmosphere. By using machine learning models such as Justin et al. (2025), we can apply a consistent criteria based on expert WPC forecasters using a wide variety of available atmospheric data. While the ML model may not initially correct all human biases, such as the spatial bias against analyzing

drylines in the Southeast or Midwest (Fig. 2.12), future study can assess the atmospheric structure of these identified boundaries and make the necessary adjustments to nudge the model accordingly towards an agreeable consensus solution.

2.5 Conclusion

A 16-year climatology of dryline frequency and location in the contiguous United States was created using the WPC surface analysis archive from 2007–2022. Unlike previous dryline studies, this climatology includes the entire year, entire day (in 3-hour increments), and entire CONUS. Severe storm reports and NWS-issued warnings were used to determine whether a dryline was associated with severe convection or not.

Unsurprisingly, drylines are most frequent in the Southern Great Plains from March–June, with the most drylines during May. July–September were the least active months for drylines, although the proportion of those drylines which were associated with severe convection was high, consistent with the annual peak in convection frequency during the warm-season. A small secondary peak in dryline frequency was identified in October.

Across nearly all months, the average dryline longitude shifts east during the day by 1–1.5° before retreating back west overnight, consistent with the conceptual model for daily dryline evolution. Superimposed on this daily evolution is a seasonal shift in dryline longitude of 3–4°, with the dryline located further east in winter and further west in summer.

An annual increasing trend in drylines analyzed by the WPC was identified across the 16-year analysis period. This increase was consistent at all analysis times, and occurred in equal proportion for both drylines associated with severe convection and those that were not. Although 16 years is too short to fully assess potential natural

decadal dryline variability, a machine learning model developed to identify fronts and drylines in ERA5 reanalyses (Justin et al., 2025) did not depict this annual increasing trend over the analysis period. The machine learning model represents a consistent application of dryline criteria learned from WPC analyses. Year-to-year changes in machine learning-identified drylines were also correlated with year-to-year changes in EML frequency, as would be expected given the physical relationship between drylines and the EML. This type of correlation was not evident between WPC analyzed drylines and the EML, further evidence that a different factor is masking any physical trends in dryline frequency over the analysis period. We therefore conclude that the annual increasing trend in drylines analyzed by the WPC does not reflect a physical increase in dryline frequency. We hypothesize that the trend could at least partially arise from an evolving definition of a dryline over time and the interpretations of different WPC forecasters. Additional work will be necessary to evaluate this hypothesis.

Since the machine learning model was trained using WPC surface analyses as “ground truth”, additional verification work will be necessary to determine if other potential human biases are present within the model’s identification of drylines. The spatial distribution of drylines identified by the machine learning model closely resembles the WPC record. Hypothesized spatial biases in the WPC record (e.g., Duell and Van Den Broeke, 2016) might influence the spectrum of atmospheric characteristics the machine learning model associates with drylines, reinforcing any spatial bias. A perfectly objective method of identifying drylines and other boundaries is an unobtainable goal due to the complexity of the atmosphere, our imperfect tools to measure and quantify it, and the narrow deterministic categorical bins we attempt to discretize its characteristics into. However, the automatic and consistent application of WPC forecaster front and dryline identification criteria by Justin et al. (2025) is unique. Whereas previous, non-machine learning automated or manual identification meth-

ods may consider only a limited number of atmospheric quantities and be sensitive to explicitly-defined categorization criteria selected by the researcher (Thomas and Schultz, 2019a,b), machine learning approaches like Justin et al. (2025) are able to remove an element of subjectivity by including a vast amount of information and letting the model architecture decide, based on the data itself, the classification criteria.

In Chapters 3 and 4, this ML model will be applied to the ERA5 reanalysis record back to 1940 in order to determine the proportion of severe thunderstorms in the Southern Great Plains attributable to drylines and study monthly dryline variability associated with the El Niño–Southern Oscillation and Southern Great Plains drought. Additional avenues for future work include expanding this type of climatology to all types of fronts, investigating the regional atmospheric characteristics for boundaries in the Midwest and Southeast to determine the extent of hypothesized misclassification of drylines as cold fronts, evaluating how a changing climate may affect the frequency and characteristics of drylines and fronts, expanding this type of climatological analysis beyond the CONUS, and evaluating feature-specific prediction in numerical weather prediction models, such as the preliminary work explored in portions of Chapter 5.

Chapter 3

Quantifying the Role of the Dryline in the Severe Thunderstorm Clima- tology of the Southern Great Plains

3.1 Introduction

A dryline is a mesoscale boundary separating dry and moist air, and typically develops from both synoptic and terrain influences (Fujita, 1958; Schaefer, 1986; Li et al., 2021). Drylines in the US are most common in the Southern Great Plains (SGP) during spring and early summer (Chapter 2). These months are also the most active for severe thunderstorms in the SGP, accounting for 44% of annual severe storm reports in the region from 1995–2020 (SPC, 2023). Across Kansas, Oklahoma, and Texas, March–June (MAMJ) severe thunderstorms account for on average \$3.1 billion USD worth of damage each year (NCEI, 2025), and tornadoes alone cause on average 9 fatalities and 131 injuries annually (SPC, 2023).

The dryline has long been recognized as a key mechanism for initiating convection in the SGP (e.g., Rhea, 1966; Mitchell and Schultz, 2020; Hosek et al., 2025). Dryline-initiated storms are often isolated cells (e.g., Bluestein and Parker, 1993; Ziegler and Rasmussen, 1998; Dial et al., 2010), which—given favorable environmental conditions

common in the SGP during spring (e.g., Gensini and Ashley, 2011)—can develop into supercells capable of very large hail and significant tornadoes (McCarthy and Koch, 1982; Smith et al., 2012).

Despite the well-recognized association between the dryline and SGP severe convection, few studies have quantified the relationship outside of case studies. Rhea (1966) evaluated the location of convection initiation relative to a dryline over 3 years for the months April–June. Mitchell and Schultz (2020) took a similar approach to 10 years (2006–2015) of April–June SGP drylines to determine the frequency of convective vs non-convective drylines and evaluate the synoptic-scale differences between them. However, neither Rhea (1966) nor Mitchell and Schultz (2020) explicitly considered severe convection. Non-case studies of drylines which do include an analysis of associated severe convection primarily focus on the frequency at which drylines serve as an initiation mechanism for severe events and characteristics of dryline-initiated severe events (Bergmaier and Geerts, 2015; Duell and Van Den Broeke, 2016; Hosek et al., 2025).

Dryline climatologies exceeding 20 years are rare, with only the Hoch and Markowski (2005) climatology of dryline frequency and longitude reaching 30 years (see Table 2.1). This is in part due to the time commitment of manual dryline identification (e.g., Hoch and Markowski, 2005; Coffey et al., 2013; Johnson and Hitchens, 2018) and also the limited archive of NOAA Weather Prediction Center (WPC) surface analysis maps (e.g., Mitchell and Schultz, 2020; Hosek et al., 2025). This limitation can be addressed via artificial intelligence/machine learning (AI/ML). The front-identification ML model developed by Justin et al. (2025) (FrontFinder AI) can be applied to reanalysis data to produce probabilistic identification of fronts and drylines. This application of AI allows for the development of longer climatologies of drylines than can be produced from observations.

In this Chapter, we quantify the fraction of SGP severe days in MAMJ that are associated with a dryline over a 26-year period (1995–2020). We evaluate the relative role of the dryline spatially across the SGP and how its role evolves through the spring and early summer.

3.2 Methods

3.2.1 Machine-Learning Dryline Model

Drylines are identified using a ML model developed to detect fronts and drylines across the contiguous United States (CONUS; Justin et al., 2025). The developed ML model uses the U-Net 3+ architecture, a convolutional neural network for image segmentation (Huang et al., 2020). The ML model uses 10 predictor variables from native and derived quantities from ERA5 reanalysis (Hersbach et al., 2018) at a horizontal grid spacing of 0.25° and five vertical levels from 2-m AGL to 850 hPa to predict drylines, cold, warm, stationary, and occluded fronts, and no-front (Justin et al., 2025). Front type probabilities are produced at each grid point and sum to 100%. These output probabilities are then calibrated to a 100-km neighborhood, and represent the probability of each front type within 100-km of a given point (Justin et al., 2025). The ML model is trained using NOAA Weather Prediction Center (WPC) 3-hourly surface analyses as ground-truth over the period 2007–2022, with an eightfold cross-validation process (Table 2.2). For 2007–2020, the predictions from each testing fold are used. The model is trained upon the entire 2007–2022 period and applied to all out-of-sample years prior to 2007. Dryline predictions were previously found to replicate the spatial and annual pattern of drylines from these expert-forecaster analyses (Chapter 2).

The ML model is applied to ERA5 reanalysis at 0000 UTC each day from 1940–

2020. Dryline probabilities are used to identify dryline objects, defined as any spatially continuous area of grid boxes with a dryline probability of at least 40% and an object-maximum probability of $\geq 60\%$. The maximum probability threshold eliminates features where the ML model is less confident to classify a dryline. The minimum probability threshold aids in maintaining spatial coherence of dryline objects. For example, using a 50% minimum threshold often led to multiple distinct dryline objects separated by a small number of grid points with probabilities slightly below the specified threshold. Dryline-objects which are at the 5th percentile or below in size (45 grid boxes) or axis ratio (the length of the major axis, in grid boxes, divided by the length of the minor axis; 1.6) are excluded. This procedure primarily culls out small dryline-objects along the southern border of Texas. No temporal trend exists in the amount of drylines eliminated by this method (not shown). Only a subset of this distribution of drylines (1995–2020; MAMJ) is used in this study.

3.2.2 SGP Severe Thunderstorm Climatology

The SGP region is defined similarly to Chapter 2 as 95–105°W and 30–40°N. The overwhelming majority of US drylines occur within this region (Chapter 2.3.2). Any dryline grid cell within this region is considered, even if a portion of the dryline extends beyond the region. Of the total sample of 3,172 days, drylines are present for 1,182 of them (37%), consistent with past climatological frequencies (Chapter 2; Hoch and Markowski, 2005; Mitchell and Schultz, 2020).

Severe storm reports (SPC, 2023) are used to identify severe thunderstorms. Reports are gridded to the NOAA National Centers for Environmental Protection (NCEP) 211 grid (NCEP, 2024) with ~ 81 -km grid spacing. Tornadoes are gridded as a single point object at the time and location of the start of the tornado track. Any grid cell with one or more gridded reports between 2100 and 0300 UTC is classified as a “severe

day”, while all others are “non-severe day”.

Storm reports are an imperfect dataset with well-documented non-meteorological influences, including a strong correlation between population density and reports and an increase in the number of reports over time due to increasing population, awareness, and accessibility of reporting tools (e.g., Anderson et al., 2007; Tuovinen et al., 2009; Paulikas, 2014; Allen and Tippett, 2015b). The binary classification of gridded severe days and relatively large 81-km grid were selected to mitigate some of these non-meteorological biases in the report database. Additionally, analysis is limited to 1995 onward. 1995 is the chosen cutoff because it represents the transition of operational radar networks to the WSR-88D, which substantially improved coverage and quality of radar data leading to increased efforts to verify severe thunderstorm and tornado warnings (e.g., Verbout et al., 2006; Allen and Tippett, 2015b). Annual trends in SGP dryline days, severe report days, and severe dryline days are not significant across the 26-year period (not shown).

The count of severe storm reports associated with drylines is presented in summary statistics (Chapter 3.3.2). Additionally, counts of significant severe reports—EF-2+ tornadoes, hail greater than 2 inches in diameter, or a measured wind gust greater than 65 kts—are also used to identify higher-end severe storms, which typically have the largest societal impacts. Trends in report counts are not assessed due to both the natural year-to-year variability in severe storm frequency as well as non-meteorological sources of variability (e.g., severe storm location relative to population centers).

3.2.3 Associating Severe Convection with Drylines

The procedure for identifying dryline-associated severe convection is adapted from Chapter 2.2.2. Drylines are gridded to the identical grid as reports, and from each dryline grid cell a “search window” extends 1 grid cell north, south, and west (~ 81 -

km) and 3 grid cells east (~ 240 -km). The north, south, and westward extension of the search window are included to account for both positional changes of the dryline over the 6-hour analysis period as well as slight northerly or southerly components to storm motion typical in the SGP (VandenBerg et al., 2014). The ~ 240 -km eastward extent of the search window is further east than past studies investigating dryline-initiated convection. Mitchell and Schultz (2020) defined dryline-initiated convection as an instance of 40-dBZ radar echo return within 100-km of a dryline, based on work by Rhea (1966). To account for the time lag between convection initiation and a storm producing severe hazards, and the typical storm motion on the order of tens of kilometers over tens of minutes (VandenBerg et al., 2014), an eastward extension of this 100-km window is necessary. The extension of the eastward search window to ~ 240 -km thus includes ~ 2 – 4 hours of storm motion from the 0000 UTC dryline location for a storm to intensify to severe criteria. The ability to refine the search window is limited by the relatively large grid spacing of ~ 81 km.

If a severe grid cell—a grid with at least one severe report between 2100 and 0300 UTC—is within the dryline search window, it (and all of its associated reports) is classified as dryline-associated severe. Since the search window is wide (~ 360 -km longitudinally across 6 hours), not all reports can be determined to be the direct result of dryline-initiated convection. Pre-existing outflow boundaries, trailing cold fronts which overtake the dryline, and warm fronts may contribute to convection initiation in the vicinity of a dryline. An analysis of satellite and radar objects, similar to Mitchell and Schultz (2020), would be necessary to classify the individual storms which produce the severe hazards as dryline-initiated. However, given the close spatiotemporal proximity to the dryline, it is likely that the dryline has at least influenced the mesoscale environment these storm reports occurred in.

3.3 Results

3.3.1 Spatial Climatology

The spatial climatology of dryline search window frequency (“dryline day”), severe day frequency, and the fraction of severe days associated with a dryline for MAMJ is shown in Fig. 3.1. Drylines are most frequent in southwest Texas (~ 42 out of 122 days per year), with >20 days per year across a region from southwest Texas through southwest Kansas, west/central Oklahoma, and central Texas (Fig. 3.1a), consistent with the relative maxima of dryline frequency identified in Chapter 2.3.2.

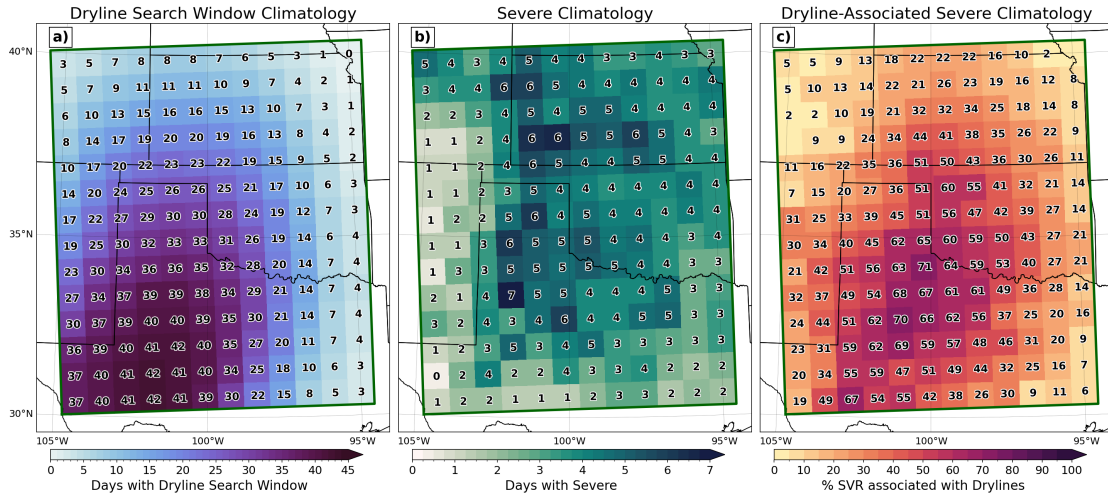


Figure 3.1: (a) The average number of days in MAMJ with a 0000 UTC dryline search window, (b) the average number of days in MAMJ with at least one severe report between 2100 and 0300 UTC, and (c) the percentage of MAMJ severe days which are associated with a dryline. The value of each non-zero grid box is annotated, rounded to the nearest integer (note that grid boxes in (a) and (b) with more than 0 but less than 0.5 days per spring are annotated with “0”).

Aside from portions of New Mexico, southern Colorado, and far southern Texas, the majority of the SGP has at least 3 severe days per spring, with a notable maximum of 5–6 days extending from eastern Colorado/western Kansas through the southern

Texas panhandle, with localized maxima in south-central Kansas and north Texas in the vicinity of major population centers (Wichita, KS, and Dallas-Fort Worth, TX, respectively; Fig. 3.1b). Drylines are associated with $>50\%$ of severe days across a wide region encompassing southwest Texas through western/central Oklahoma, with over 70% of the severe days in northwest Texas associated with a dryline (Fig. 3.1c). The largest proportion of the severe climatology associated with drylines is shifted northeast compared to the maximum in dryline day frequency (northwest Texas vs southwest Texas, respectively; Fig. 3.1a,c). Far southwest Texas and eastern New Mexico have a high frequency of dryline days (30–40 days per spring; Fig. 3.1a), yet a comparably low frequency of severe weather (<2 days; Fig. 3.1b) and low percentage of severe days associated with drylines (20–40%; Fig. 3.1c). By contrast, southwest Kansas through central Oklahoma has a lower dryline day frequency (20–30 days per spring; Fig. 3.1a), yet higher frequency of severe days (4–6; Fig. 3.1b) and higher proportion of dryline-associated severe days (40–60%; Fig. 3.1c).

Farther north, areas in western Kansas and northeastern Colorado have a relatively high frequency of MAMJ severe days (5–6; Fig 3.1b). However, the proportion of these severe days associated with drylines is relatively low ($<25\%$; Fig. 3.1c), attributable to the lower frequency of dryline days compared to the Texas panhandle and Oklahoma (Fig. 3.1a). In the far eastern SGP, dryline days are rare (<7 days per year; Fig. 3.1a). Yet aside from the northeast and southeast corners of the SGP domain, drylines are associated with 10–30% of eastern SGP severe days (Fig 3.1c).

On a monthly basis, March unsurprisingly has the fewest dryline days and severe days (Fig. 3.2a,b). However, over half of the severe days that do occur across much of west Texas, southeast Oklahoma, and the Oklahoma–Texas border are associated with drylines, including several grid cells with $>75\%$ of severe days associated with drylines (Fig. 3.2c). The highest frequency of severe days during March occurs in

eastern Oklahoma and east central Texas (>1 day per month), with between 20–50% of these days associated with drylines (Fig. 3.2b,c).

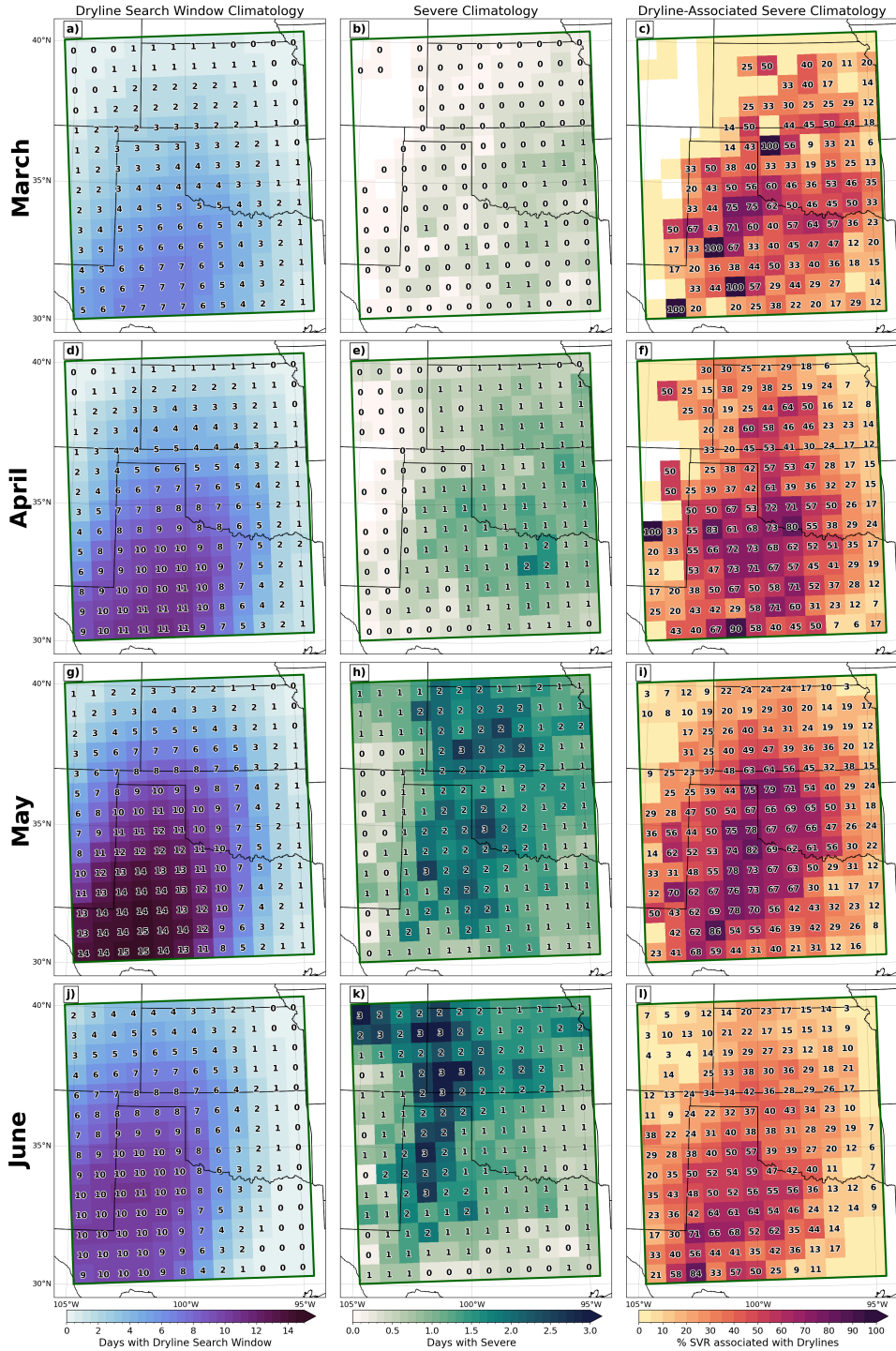


Figure 3.2: As in Fig. 3.1, but each row is an individual month. (a–c) March, (d–f) April, (g–i) May, and (j–l) June.

As spring progresses from March to June, dryline day frequency increases and shifts west and north (Fig. 3.2a,d,g,j). Dryline day frequency peaks in May in southwest Texas (Fig. 3.2g), and the seasonal progression is consistent with previous climatologies (Chapter 2; Hoch and Markowski, 2005). The number of severe days also increases each month, with a similar westward and northward shift as dryline days (Fig. 3.2b,e,h,k). Although there are more severe reports in May than June (19,723 vs 17,335, respectively), there is a widespread increase in the number of severe days during June, with 9 individual grid points with 2.5 days or more (Fig. 3.2k) compared to just 3 grid points averaging more than 2.5 days of severe weather during May (Fig. 3.2h). This can be attributed to May severe events producing more reports on a per-day basis compared to June.

The proportion of severe days associated with a dryline does not exhibit the same seasonal progression as dryline days and severe days (Fig. 3.2c,f,i,l). Each month, drylines are associated with the highest proportion of severe days in the Texas panhandle and western Oklahoma. In May, the month with the most severe reports and most dryline days, the dryline is associated with over 60% of severe days across an expansive area spanning southwest Texas through central/western Oklahoma, along with more than 75% in southwest Oklahoma and northwest Texas (Fig. 3.2i). The proportion of severe days associated with a dryline decreases across the SGP during June compared to May, yet drylines are still associated with more than 50% of severe days across most of the Texas panhandle (Fig. 3.2l). The increase in the number of severe days in Colorado and northwest Kansas during June is not associated with the increased northward frequency of drylines, with only 10–20% of severe days associated with drylines, similar to May (Fig. 3.2j–l).

3.3.2 Report Count Climatology

Despite the well-documented limitations of the severe report database, severe reports can offer insight to the overall impact of dryline-associated severe convection in the SGP. All gridded severe and significant severe report counts are summed across the entire SGP to mitigate significant spatial biases in reporting. Approximately 40% of all 2100–0300 UTC MAMJ severe reports from 1995–2025 are associated with a dryline, though this percentage varies on a monthly basis between roughly 30–50% (Table 3.1). The greatest proportion of all dryline-associated reports peak during May (48.0%), coincident with the seasonal cycle of severe weather in the region.

Month	% (number) of reports	% (number) of SigSevere reports
March	37.8% (1,293)	47.0% (174)
April	43.8% (4,339)	50.8% (637)
May	48.0% (9,461)	56.4% (1,667)
June	30.6% (5,302)	33.9% (778)
March–June	40.5% (20,395)	47.4% (3,256)

Table 3.1: The percentage of all 2100–0300 UTC reports (middle column) and significant severe reports (right column) in the SGP associated with a dryline.

In terms of significant severe weather, an even greater proportion can be attributed to the dryline, with just shy of 50% during the entire season (Table 3.1). The seasonal progression again peaks during May with over half (56.4%) of significant severe reports related to the dryline. Even in June, the month with the lowest proportion of dryline-associated significant severe reports, just over a third of all SGP significant severe reports (33.9%) are associated with drylines. Throughout the entire season, the dryline is a significant contributor to SGP severe weather, and even more so for high-impact severe weather.

3.4 Summary and Conclusion

Drylines at 0000 UTC from March–June are identified within the US Southern Great Plains using a ML model over a 26-year period (1995–2020). The characteristics of this ML-identified dryline climatology—a peak dryline frequency in May, with a gradual westward and northward shift during spring—is consistent with past dryline climatologies (Chapter 2; Hoch and Markowski, 2005). Gridded severe storm reports between 2100 UTC and 0300 UTC are used to evaluate the number of days per month with severe storms across the SGP. Severe days which occur in close proximity to a dryline are considered to be dryline-associated.

The dryline accounts for over 50% of severe weather days during MAMJ from southwest Texas through western/central Oklahoma, and is associated with at least 25% of severe days across the majority of the SGP. Despite the westward and northward shifts of both dryline days and severe days as the season progresses, the region from southwest Texas through western Oklahoma consistently has the highest fraction of severe days associated with drylines each month, with up to 82% of severe days in May associated with a dryline. Severe convection associated with the dryline is impactful, as determined using storm reports as a proxy for impacts. 40% of all MAMJ 2100–0300 UTC storm reports within the SGP are associated with a dryline, as well as 47% of significant severe reports. During May—the month with the most significant severe storm reports—56% of significant severe reports are associated with a dryline.

There are some limitations to the interpretation of the results presented. The climatology is limited temporally to 2100–0300 UTC, and while this time range corresponds with the majority of severe reports (Krocak and Brooks, 2020) and dryline-associated convection in the SGP (Rhea, 1966; Mitchell and Schultz, 2020; Hosek et al., 2025), it only represents approximately 63% of all MAMJ SGP severe reports. The

downstream and secondary effects of dryline-initiated convection (e.g., outflows from dryline-initiated storms serving as the initiation mechanism for additional convection) are an important component to the influence of the dryline on severe convection, but are outside the scope of this Chapter.

The non-meteorological biases of the severe storm report database and relative rareness of severe thunderstorms contribute to the occasional grid box with significantly higher (or lower) association with drylines compared to neighboring grid boxes. For example, it is not physically meaningful that one particular SW TX grid box has 100% of its March reports associated with drylines while its neighboring grid boxes universally have less than 60% of their reports associated with drylines (Fig. 3.2c). In addition to increasing the sample size, the lack of smoothness between grid boxes could be mitigated by also including National Weather Service-issued severe thunderstorm and tornado warnings to classify severe days (e.g., Hosek et al., 2025) or by using observed radar and satellite data to identify severe thunderstorms.

Overall, the dryline is shown to have a substantial role in the springtime severe weather climatology of the SGP. Therefore, variability in the dryline—its location and frequency—can potentially be linked to variability in severe thunderstorms across the SGP. Improving our understanding of dryline variability on subseasonal-to-seasonal (S2S) timescales may be a fruitful avenue towards improving longer-range S2S severe thunderstorm forecasting during MAMJ.

This work highlights the utility of FrontFinder ML front identification for climatological applications. Given the limited temporal availability of expert-identified drylines, the FrontFinder model effectively extended the period in which climatological analysis could be performed. Chapter 4 will take advantage of the FrontFinder AI model to evaluate variability in the dryline associated with SGP drought and the El Niño–Southern Oscillation from 1940–2020. Future work includes application of the

ML model to global climate models to assess future climate projections of the dryline.

Chapter 4

Variability in the US Southern Great Plains Dryline Associated with the El Niño Southern Oscillation and Regional Drought

4.1 Introduction

As established in Chapters 2 and 3, the dryline is associated with a significant portion of severe thunderstorms in the Southern Great Plains (SGP). Dryline variability on subseasonal-to-seasonal (S2S) timescales may therefore be a substantial contributor to SGP severe thunderstorm variability. In this Chapter, monthly dryline variability in frequency and longitude from February through June is examined with respect to the El Niño-Southern Oscillation (ENSO) and regional SGP drought.

4.1.1 Factors Influencing Dryline Frequency and Longitude

The dryline is modulated by both synoptic and mesoscale factors (Hane, 2004). At the synoptic scale, strong mid-troposphere westerly flow along with shortwave troughs crossing over the Rocky Mountains can develop drylines via confluence zones between

dry air from the west and moist air from the south/east. Air gradually descending from the high terrain of New Mexico and Colorado—an already dry airmass due to the desert terrain of the southwest US and northern Mexico— will warm and dry adiabatically. As a surface lee cyclone develops in eastern Colorado/northern New Mexico, moist air from the Gulf of Mexico/Gulf of America is advected northward. These dry and moist airmasses will converge along a confluence line extending south from the developing lee cyclone, leading to the development of a dryline (Hobbs et al., 1996).

The propagation, and thus longitude, of a dryline results from a combination of advection and mesoscale boundary layer (BL) processes. During daytime insolation, heating of the relatively shallow moist BL layer east of a dryline results in efficient vertical mixing. As the moist BL vertically mixes with the dry elevated mixed layer (EML) above it, the BL dries substantially. The moist BL is typically shallower to the west and deeper to the east due to the sloping terrain of the SGP, so the rapid drying and deep mixing of the BL will occur earlier in the west. As a result, the dryline will rapidly progress eastward as the shallow moist BL is mixed out ahead of it (Schaefer, 1974, 1986; Hane, 2004). The speed of dryline propagation is dependent on the depth and richness of BL moisture, the rate of BL sensible heating/mixing, and the amount of confluence and associated vertical lifting along the dryline (Sun and Wu, 1992; Hane, 2004). The rate of BL sensible heating can be influenced by mesoscale factors such as the fraction of cloud cover and amount of soil moisture (Ziegler et al., 1995; Grasso, 2000). A 10-year climatology of April–June dryline longitude and soil moisture found that 0000 UTC drylines were farther east when the soil moisture was lower (Johnson and Hitchens, 2018).

During strong lee cyclogenesis events associated with notable shortwaves and jet streaks, drylines will translate east-northeast with the lee cyclone, typically until the cold front associated with the cyclone catches up to and overtakes it (e.g., Hobbs et al.,

1996; Schultz, 2005). Drylines in these synoptically-driven environments may appear to lack the typical diurnal evolution where the dryline retreats westward overnight due to the strength of the east-northeast synoptic-scale advection (Hobbs et al., 1996; Hane, 2004). In exceptional cases, these synoptic conditions can result in drylines reaching as far east as the Mississippi River Valley prior to being overtaken by a cold front, where they are associated with significant regional outbreaks such as the 1974 Super Outbreak and the 26–27 April 2011 Outbreak (e.g., Locatelli et al., 2002; Chasteen and Koch, 2022; Duell and Van Den Broeke, 2016).

A combination of both synoptic pattern and SGP soil moisture influence the frequency of drylines. The presence of an east–west soil moisture gradient across the SGP has been identified as a prerequisite for dryline development (e.g., Sun and Wu, 1992; Ziegler et al., 1995; Grasso, 2000). The synoptic-scale pattern dictates the frequency of favorable atmospheric conditions for dryline development, such as the frequency and intensity of lee cyclogenesis events (Hobbs et al., 1996). Drought or pluvial events—themselves a result of a synoptic pattern—may affect the SGP soil moisture gradient, which in turn influences the frequency of drylines.

4.1.2 ENSO

ENSO describes patterns of area-averaged sea surface temperature (SST) anomalies in the equatorial eastern Pacific, which is correlated to the strength of subtropical upper tropospheric wind (Arkin, 1982). During warm SST anomalies (El Niño), the tropical atmosphere warms and tropical deep convection is enhanced. As a result, the Hadley cell strengthens and contracts, leading to an equatorward shift in the subtropical jet (Seager et al., 2003; Allen et al., 2015a; Cook et al., 2017; Schemm et al., 2018). The El Niño subtropical jet is also enhanced due to an increased temperature gradient between the tropics and midlatitudes, which induces a stronger westerly thermal wind response

(Seager et al., 2003). An opposite effect occurs during cool SST anomalies (La Niña): tropical deep convection is suppressed, and the Hadley cell is weakened and expands poleward. The jet, and cyclones associated with it, are shifted poleward (Seager et al., 2003; Allen et al., 2015a; Cook et al., 2017; Schemm et al., 2018). The influence of ENSO on CONUS weather is typically most prominent during winter and early spring, with minimal influence during summer (e.g., Lagerquist et al., 2020).

The placement and strength of the jet affects cyclone tracks (Eichler and Higgins, 2006; Schemm et al., 2018), front placement (Rudeva and Simmonds, 2015; Lagerquist et al., 2020), and severe storm frequency (e.g., Lee and Galway, 1956; Johns and Doswell III, 1992; Marzban and Schaefer, 2021; Cook and Schaefer, 2008; Allen et al., 2015a; Cook et al., 2017; Lepore et al., 2017; Molina et al., 2018; Moore, 2019; Tippet and Lepore, 2021; Nouri et al., 2021). During La Niña winters, enhanced zonal westerlies across the eastern Pacific and western US are favorable for lee cyclogenesis east of the Rocky Mountains, leading to stronger and more frequent lee cyclones across the central and northern US (Eichler and Higgins, 2006; Schemm et al., 2018). An increased frequency of warm fronts over the central US—associated with the increase in lee cyclones—represents the northward advection of warm, moist air from the Gulf of Mexico/Gulf of America into the southern and central Great Plains, a key ingredient in late winter/spring severe weather (Allen et al., 2015a; Cook et al., 2017; Lagerquist et al., 2020). During El Niño winters, the frequency of lee cyclogenesis in the central US decreases due to the enhanced subtropical jet being located farther south (Schemm et al., 2018). With cyclone tracks subsequently farther south, cold fronts push farther south through the Great Plains, scouring moisture out of the region and limiting potential for severe convection (Eichler and Higgins, 2006; Allen et al., 2015a; Cook et al., 2017; Schemm et al., 2018; Lagerquist et al., 2020).

ENSO has been linked to severe hail and tornado frequency during winter and

spring, with a broad increase in frequency in the SGP during La Niña (e.g., Cook and Schaefer, 2008; Allen et al., 2015a; Cook et al., 2017; Molina et al., 2018). Favorable jet dynamics and shear overlap with enhanced instability over the SGP, and the increase in lee cyclogenesis events also provides favorable low-level shear and initiation mechanisms (Hobbs et al., 1996; Allen et al., 2015a; Cook et al., 2017). In contrast, during El Niño, the more southerly track of cyclones results in a southward shift to severe storm frequency (Cook et al., 2017). Decreased instability and unfavorable jet dynamics result in fewer severe storm events during late winter and early spring, with events occurring farther south along the Gulf Coast and southeast US. By El Niño April, severe storm events do occur with more regularity in the SGP—likely due to the seasonal poleward progression of the jet stream—but the location of these events are shifted southward compared to La Niña April (Cook et al., 2017).

4.1.3 Hypotheses and Outline

This Chapter will address three hypotheses relating to dryline variability associated with ENSO and SGP drought:

1. Drylines will be more frequent and farther east during La Niña due to the increased frequency of lee cyclogenesis events and the relative decrease of scouring cold fronts in the SGP in late winter/early spring (Hobbs et al., 1996; Lagerquist et al., 2020).
2. Drylines will be more frequent and farther east during SGP drought conditions due to an increased soil moisture gradient and faster mesoscale dryline propagation.
3. The influence of ENSO on dryline frequency and longitude will be weakest in late spring due to the weakening teleconnection between Pacific SSTs and CONUS

synoptic-scale weather after late winter/early spring (e.g., Lagerquist et al., 2020). The influence of SGP drought on dryline frequency and longitude will be similar across all months.

Chapter 4.2 describes the identification of drylines, reiterating Chapters 2 and 3, as well as detailing ENSO and SGP drought definitions, the relationship between ENSO and SGP drought, analysis methods, and statistical testing methods. Chapter 4.3.1 presents the results of the ENSO analysis, while Chapter 4.3.2 presents the results of the SGP analysis. Chapters 4.4 and 4.5 synthesize the results and present conclusions.

4.2 Methods

4.2.1 Dryline Identification

Drylines are identified daily at 0000 UTC from 1940–2020 ERA5 reanalyses (Hersbach et al., 2018) using a machine learning (ML) model developed to identify fronts (Chapters 2–3; Justin et al., 2025; Hosek et al., 2025). The FrontFinder ML model is described in Chapters 2 and 3, and is restated here: The ML model is developed using the U-Net 3+ architecture, a convolutional neural network for image segmentation (Huang et al., 2020). It uses 10 predictor variables from native and derived quantities from ERA5 reanalysis at a horizontal grid spacing of 0.25 degrees and five vertical levels from 2 m AGL to 850 hPa to predict cold, warm, stationary, and occluded fronts in addition to drylines (Justin et al., 2025). Front type probabilities are produced at each 0.25 degree grid point and sum to 100%. These output probabilities are then calibrated to a 100-km neighborhood, and represent the probability of each front type within 100-km of a given point (Justin et al., 2025). The ML model is trained using NOAA Weather Prediction Center (WPC) 3-hourly surface analyses as ground-truth, and is

able to replicate the spatial and annual pattern of drylines from these expert-forecaster analyses (Chapter 2; Hosek et al., 2025).

As in Chapters 2 and 3, the ML model is trained using WPC surface analyses from 2007–2022 with an eightfold cross-validation process (Table 2.2). For 2007–2020, the predictions from each fold are used. Prior to 2007, the model is trained upon the entire 2007–2022 period and applied to all out-of-sample years.

Drylines are defined identically to Chapter 3, and the methodology is restated here: A dryline-object is defined as any spatially continuous area of grid cells with a dryline probability of at least 40% and an object-maximum probability of $\geq 60\%$. The maximum probability threshold eliminates features the ML model is less confident to classify as a dryline, and the minimum probability threshold aids in maintaining spatial coherence of dryline objects.

Dryline-objects which are below the 5th percentile in size (45 grid cells) or axis ratio (the length of the major axis, in grid cells, divided by the length of the minor axis; 1.6) are excluded. This procedure primarily culls out small dryline-objects along the southern border of Texas. There is no trend over time in the amount of drylines culled by this method each year (not shown). Of the 3,999 February–June dryline-objects identified by the ML model, 3,662 drylines are retained following the application of the coherence thresholds.

The Southern Great Plains region is defined as in Chapter 3—95–105°W and 30–40°N. For statistical testing and environmental field compositing, only the 3,444 drylines with a median latitude and longitude within the defined region are considered. The region threshold is not used for the dryline spatial composites (Chapters 4.3.1.2 and 4.3.2.2).

4.2.2 ENSO

ENSO phase is determined using 1940–2020 monthly ENSO 3.4 indices (Rayner et al., 2003). For each individual month, ENSO indices are mean-removed and linearly-detrended by year in order to remove any long-term trend. This linearly-detrended ENSO Index is used to categorize the ENSO phase for each individual month, with ± 0.5 as the threshold between phases.

Temporal sensitivity testing consisted of calculating statistics with lag-1 month ENSO (e.g., classify April using the March ENSO index) and using the seasonally-averaged index (March–April–May; MAM) to categorize an entire year (Feb–June). A lag-1 seasonally-averaged index (December–January–February; DJF) was also tested. Lagging of 1 month/season tests the response time of factors influencing the dryline—both atmospheric and land-surface—to tropical Pacific SSTs. Seasonally-averaging tests how the persistence of ENSO affects factors influencing the dryline, as well as implicitly testing against the strength of ENSO (e.g., a stronger ENSO index in one month is more likely to lead to a El Niño/La Niña for the season). Seasonal averaging decreases temporal resolution—it is possible for an ENSO neutral May to occur within an El Niño/La Niña MAM—but seasonal averaging is common for climatological variability studies involving ENSO (e.g., Allen et al., 2015a; Lagerquist et al., 2020). The lag-1 season test—using DJF average index to classify ENSO for February–June—combines the tests of both the lag-1 month and spring seasonal average. As with the individual month categorization, the lag-1 month and seasonally-averaged ENSO indices were mean-removed and linearly-detrended.

The results of the temporal sensitivity tests ranged from marginal (e.g., similar statistical significance compared to monthly ENSO) to detrimental (e.g., loss of some or all months of statistical significance). Across all tests, statistical significance relating to

dryline frequency remained similar to the individual month classification. In contrast, all tests showed substantially weaker significance between ENSO and dryline longitude compared to the individual month classification.

4.2.3 Palmer Drought Severity Index

The Palmer Drought Severity Index (PDSI; Palmer, 1965) is a metric used to measure meteorological drought. PDSI uses a basic water balance model considering sources influencing soil moisture loss (e.g., evapotranspiration), sources of soil moisture (e.g., precipitation), and the water holding capacity of the soil to determine the the water-deficit (or water-surplus) for a location. PDSI is calculated monthly, and is highly autocorrelated as the previous month’s PDSI is included in the calculation for the current month. While this formulation prevents PDSI from fluctuating drastically due to transient precipitation deficits/excesses, it limits the ability of PDSI to depict drought on shorter (<12 month) timescales (Dai and National Center for Atmospheric Research Staff (Eds.), 2023). The Self-Corrected PDSI (SC-PDSI, hereafter PDSI) developed by Wells et al. (2004) improved upon the Palmer (1965) formulation by generalizing the calculation to the local climatology and soil characteristics of any region, and is the version used in this Chapter and referred to as “PDSI”.

PDSI describes the severity of water deficit or surplus at a location. Ranging from -10 to +10, PDSI is calibrated such that an index of -1 to 1 is considered near normal, with ± 3 commonly used as the threshold for severe drought/wet (Palmer, 1965; Dai and National Center for Atmospheric Research Staff (Eds.), 2023). Monthly PDSI data on a 2.5° grid from 1940–2018 (Dai, 2017) is calculated from a combination of a variety of observations and reanalyses described in detail in Dai (2021). The average monthly PDSI across the SGP—defined as -105 to -95 W and 30 to 40 N (16 total grid points)—is used to determine whether a month is in “drought”, “normal”, or “wet” conditions.

As with ENSO described in Chapter 4.2.2, SGP-averaged PDSI is mean-removed and linearly-detrended by year.

A SGP-averaged PDSI threshold of -1 and 1 are used to define “drought” and “wet” months, respectively. Various thresholds between ± 0.5 and ± 4 were tested for sensitivity. The use of ± 0.5 as a threshold had only marginal effects. A threshold of ± 2 yielded weaker statistical results, and thresholds above ± 2 were not viable due to a lack of months with an SGP average PDSI exceeding ± 2 .

PDSI and ENSO index are well correlated for each month from Feb—June, with drought more likely during La Niña months and wet conditions more likely during El Niño months (Fig. 4.1). The linear trend in the indices is statistically significant for all months by the Kendall Tau correlation coefficient, a non-parametric rank correlation test (Kendall, 1938). The strength of the correlation is weak-to-moderate for all months, but the relationship is not symmetric between La Niña–drought and El Niño–wet. During La Niña, drought is slightly favored, but PDSI commonly ranges across normal and wet conditions as well. In contrast, during El Niño, the overwhelming majority of months are classified as wet (Fig. 4.1). Disentangling the relative influence of ENSO versus PDSI on dryline frequency and longitude would be insightful, but is outside the scope of this Chapter.

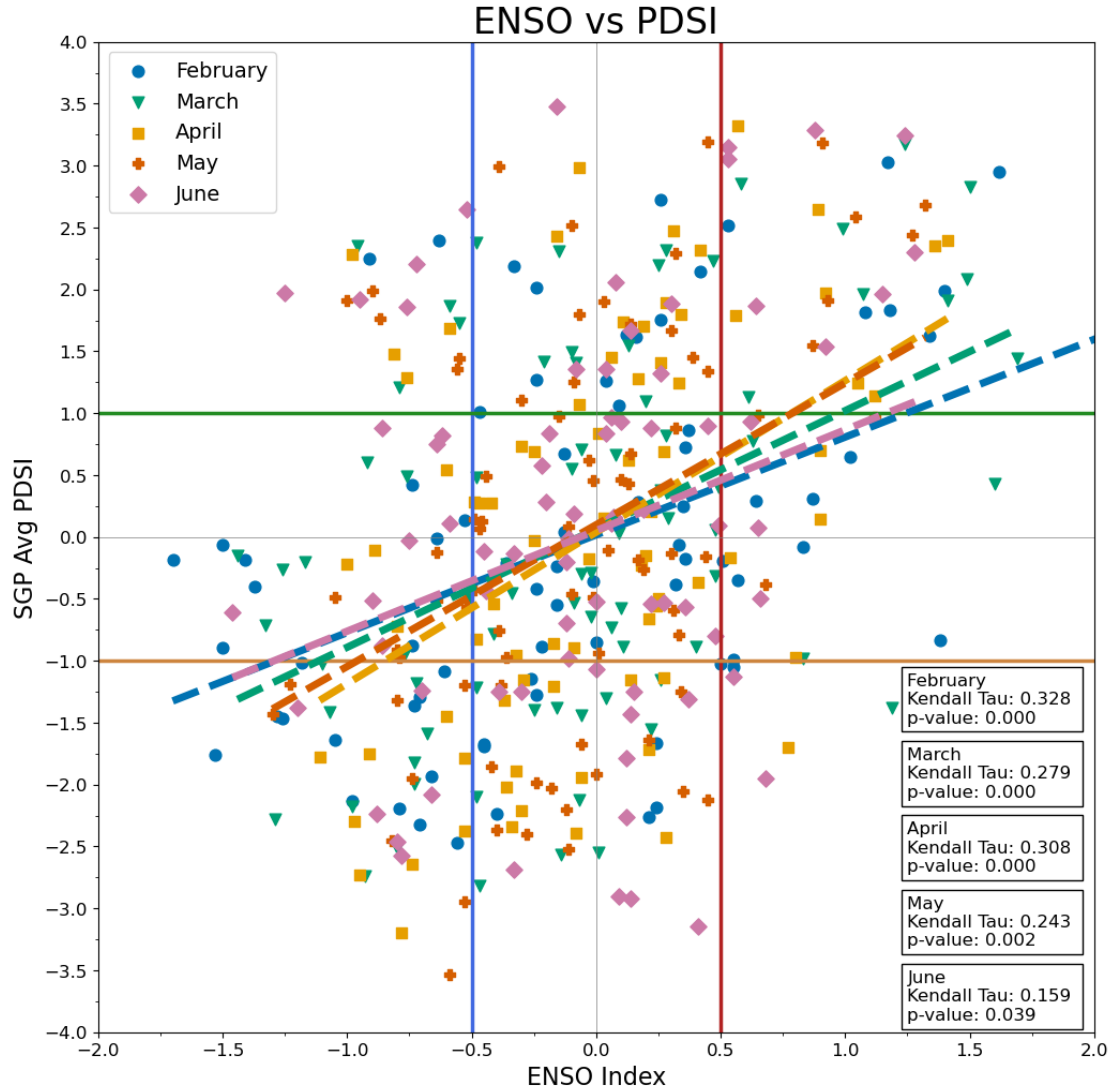


Figure 4.1: Mean-removed, linearly-detrended ENSO index versus SGP-average PDSI per month for 1940–2018 (color and marker type). Dashed lines represent the linear trend for the month corresponding to its color. The Kendall Tau correlation coefficient and p-value is reported for each month in the lower right. Solid vertical and horizontal lines denote the ENSO phase and PDSI phase thresholds used, respectively. The vertical red line represents the threshold for El Niño, and the vertical blue line represents the threshold for La Niña. The horizontal green line represents the threshold for wet, and the horizontal brown line represents the threshold for drought.

4.2.4 Environmental Composites

Composites of 500 mb height and wind, mean sea level pressure (MSLP), and 10-m wind at 0000 UTC are used to evaluate physical hypotheses for statistical relationships (see Chapter 4.2.5). For each individual month, environmental composites for each ENSO phase (La Niña, Neutral, and El Niño) and PDSI phase (drought, normal, and wet) are computed. Anomalies of these fields are relative to the 1940–2020 daily 0000 UTC climatology for the given month. An anomaly difference field between the two sets of composites is computed to compare environmental differences between phases which have statistically distinct dryline characteristics.

To explore differences in dryline frequency, environmental composites are generated across an entire month (e.g., all days of March in El Niño). These composites are assessed for differences which might explain why dryline frequency is higher (or lower) from a different phase. To explore differences in dryline longitude, environmental composites are generated only for days of a month with a dryline (e.g., all dryline days of March in El Niño).

4.2.5 Statistical Testing

4.2.5.1 Distributions

Monthly dryline frequency and median longitude distributions are statistically compared between La Niña–Neutral–El Niño and Drought–Normal–Wet. For each month, the variable (frequency or median longitude) is linearly-detrended and the mean is removed across the entire sample (1940–2020 for ENSO, 1940–2018 for PDSI).

The Kruskal-Wallis test is used to compare the distributions. Kruskal-Wallis is a non-parametric test of multiple distributions to evaluate whether any of the sample distributions originate from the same underlying population distribution (Kruskal and

Wallis, 1952). It is similar to the Mann-Whitney U test (Mann and Whitney, 1947), but it is capable of comparing more than 2 sample distributions. A limitation of the Kruskal-Wallis test is that it does not identify which sample distributions are significant different, and so post-hoc Mann-Whitney U tests are performed between each possible pairing of sample distributions (El Niño–Neutral, La Niña–Neutral, and El Niño–La Niña for ENSO; drought–normal, wet–normal, and drought–wet for PDSI) when the Kruskal-Wallis test identifies significance. The Mann-Whitney U test is a non-parametric test of whether the two distributions are from the same underlying population. Since the Mann-Whitney U tests are applied post-hoc, a Bonferroni correction is applied (e.g., Armstrong, 2014). Effectively, the returned p-value from each Mann-Whitney U test is multiplied by 3 (the number of hypotheses tested). While the Bonferroni correction is a crude technique which has been criticized for over-correcting—it increases the likelihood of type II error (false negative) at the expense of suppressing type I error (false positive)—its use in this Chapter provides additional confidence that the relationships with statistical significance are indeed significant (Perneger, 1998; Armstrong, 2014).

Statistical significance is evaluated as $p\text{-value} < 0.05$ for both sets of tests. Mann-Whitney U p-values above 0.05 but below 0.15 (e.g., p-values which would have been significant if not for the Bonferroni correction) are identified and communicated for the reader’s context given the limitations of the Bonferroni correction discussed above. However, these results are not considered statistically significant for the conclusions of this Chapter.

4.2.5.2 Spatial

Two sets of spatial fields are plotted and statistically tested: Dryline anomaly, and environmental composites.

Dryline anomaly plots are spatial composites of dryline frequency anomaly, in terms of number of days per month compared to neutral (ENSO) or normal (PDSI) dryline frequency. For ENSO, the anomalies span 1940–2020, while for PDSI, anomalies span 1940–2018. A randomization test with 10,000 iterations is used to determine significance at each grid point. For each iteration, a random subset of years (with replacement) are selected to create the spatial anomaly compared to neutral/normal. The number of years selected in each iteration is equal to the number of years of the phase which significance is tested for (e.g., there are 20 El Niño Februarys, so 20 random Februarys are selected for each iteration). At each grid cell, the distribution of 10,000 randomized anomalies are compared to the observed anomaly (e.g., February El Niño). The observed anomaly is determined to be statistically significant if it exceeds the 95th percentile of the distribution in either the positive or negative direction.

Environmental composite fields are tested using a similar methodology. The composite anomaly difference between two phases (e.g., El Niño minus La Niña 500 mb height anomaly composites) is tested via a randomization test of 10,000 iterations for each grid point where the sample years contributing to the two given phases are randomly re-assigned, with replacement. Similar to the dryline anomalies, statistical significance is determined as any grid point where the observed anomaly field exceeds the 95th percentile of the randomized distribution.

While there are statistically significant linear trends in environmental fields across 1940–2020 (not shown), the magnitude of these trends is small (e.g., 500 mb height increases on the order of 0.005 m) and thus physically meaningless for our composites and composite statistical tests.

4.3 Results

There is sufficient representation of the different ENSO and PDSI phases for each month (Fig. 4.2). Across all months included in this Chapter, neutral-phase ENSO and normal PDSI phases were the most common. Non-neutral ENSO phases were most frequent in February, with neutral ENSO dominating later in spring (Fig. 4.2a), consistent with expectations that ENSO weakens into the warm season. There is no substantial trend in PDSI phase by month (Fig. 4.2b), which is unsurprising since PDSI is highly autocorrelated (Chapter 4.2.3).

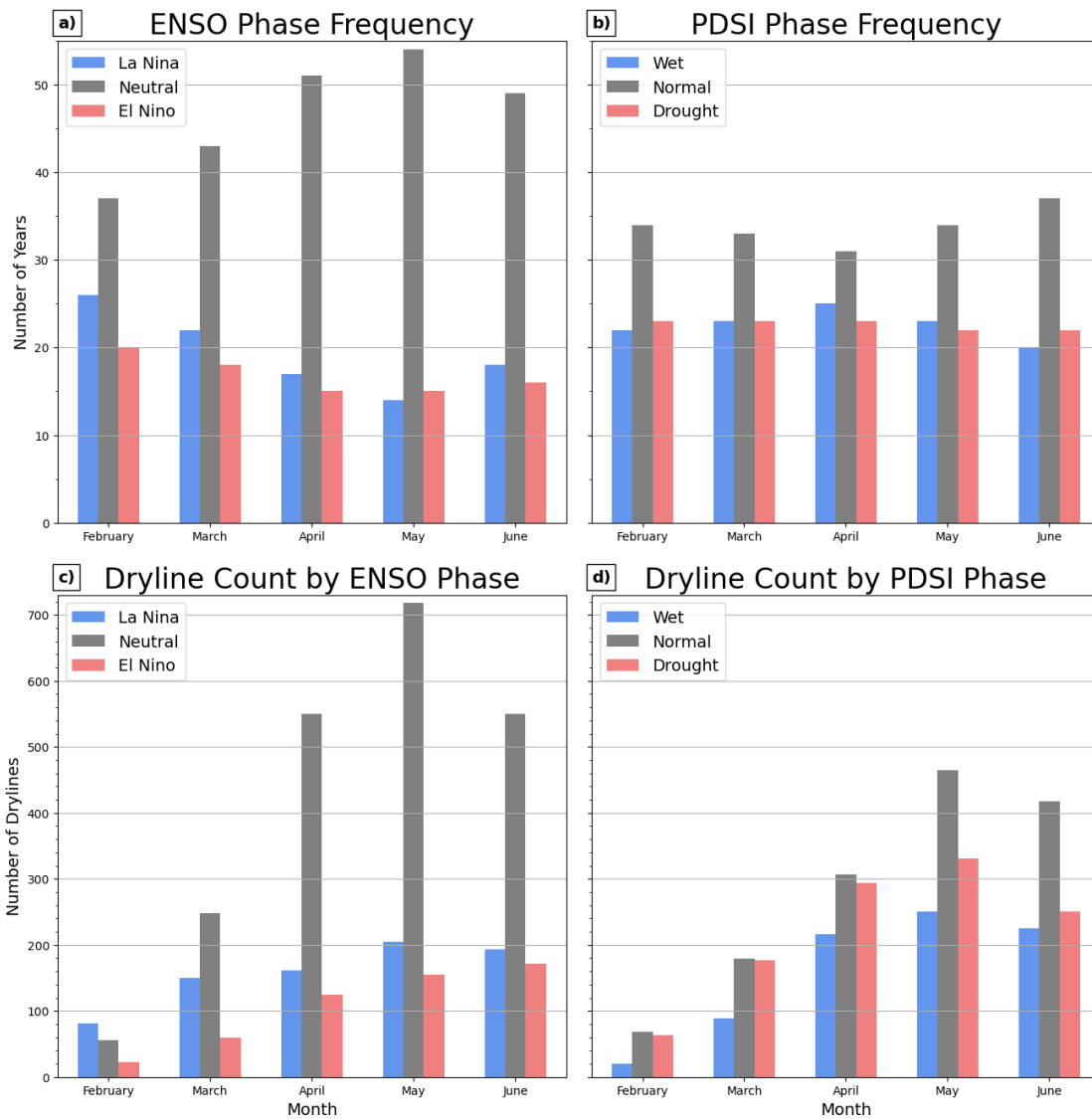


Figure 4.2: The number of years of (a) ENSO phases and (b) PDSI phases by month. The total number of drylines for (c) each ENSO phase and (d) each PDSI phase by month. Note that (a),(c) span 1940–2020 while (b),(d) span 1940–2018

With the exception of February, neutral ENSO drylines dominate the sample, with more neutral ENSO drylines every month than El Niño and La Niña combined (Fig. 4.2c). In February and March, there are substantially fewer drylines during El Niño compared to La Niña. Dryline counts across PDSI phases are more balanced, with only 100–200 more normal-phase drylines than either wet or drought phase drylines in late spring compared to neutral ENSO having 300+ more drylines than EL Niño or La Niña in May and June (Fig. 4.2c,d). While April–June is dominated by neutral ENSO, there are a sufficient number of years of El Niño/La Niña (10–20) with a sufficient number of drylines (100–200) to evaluate variability (Fig. 4.2a,c).

4.3.1 ENSO

4.3.1.1 Monthly Distributions

For each month, the distribution of dryline frequency anomaly for each ENSO phase is compared (Fig. 4.3). By the Kruskal-Wallis test, there is a significant difference in dryline frequency only in February and March. These two months have the lowest average dryline frequency (2.0 and 5.6 days per month, respectively, compared to >10 days per month for April–June). During February, drylines occur more frequency during La Niña than during neutral or El Niño conditions (Fig. 4.3a,b). The median dryline frequency anomaly for both February neutral and El Niño is negative, indicating that there are fewer than 2 dryline days per February during those phases. Comparatively, the majority of the February La Niña frequency anomaly distribution is above 0, indicating that most February La Niña events have more than 2 dryline days, although February La Niñas with less than 2 dryline days do occasionally occur (Fig. 4.3a).

Linearly Detrended Dryline Frequency verses ENSO Category, by Month

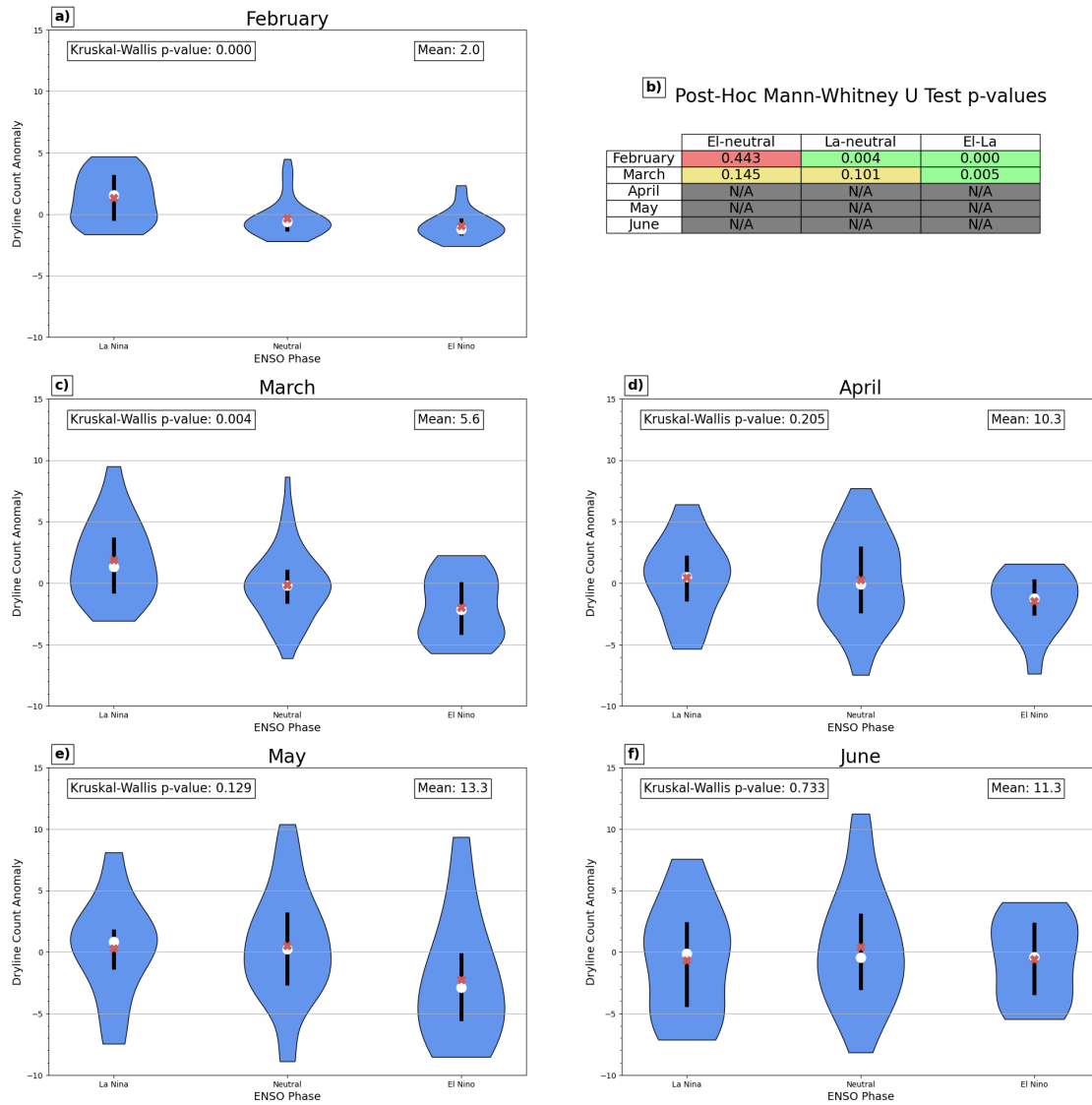


Figure 4.3: (a), (c-f) The violin plots demonstrating the distribution of linearly-detrended mean-removed dryline frequency anomaly by month for each ENSO phase. The thick black line spans the 25th to 75th percentile of each distribution, with the white circle denoting the median and the red “x” representing the anomaly mean. The mean dryline frequency for each month is shown in upper right. The p-value of the Kruskal-Wallis test is shown in the upper left. (b) A table of p-values from the Bonferroni-corrected post-hoc Mann-Whitney U test. Each month is represented by a row, and each column represents the Mann-Whitney U test performed (El Niño–Neutral, La Niña–Neutral, and El Niño–La Niña, respectively). The table cells are shaded by statistical significance with green=significant, red=not significant, and yellow=significant only if no Bonferroni correction applied. The Mann-Whitney U tests are only performed for months where the Kruskal-Wallis test is significant. Months where no Mann-Whitney U tests are performed are shaded gray and denoted “N/A”.

A similar trend to February is observed in March, with drylines appearing more frequently during March La Niña (Fig. 4.3c). During La Niña, there are on average ~ 2 more dryline days per month than usual. Dryline frequency during neutral ENSO is centered around 0, while on average ~ 2 fewer drylines occur during El Niño. After March, there is no statistically significant difference in dryline frequency by ENSO (Fig. 4.3d–f).

While ENSO phase is correlated with dryline frequency early in the analysis period (February–March), ENSO is correlated with median dryline longitude later in the season (April–June; Fig. 4.4). During May and June, La Niña drylines are on average $0.25\text{--}0.5^\circ$ farther east than El Niño drylines, although the 25th–75th percentile and tails of the distributions overlap substantially (Fig. 4.4e,f). However, during April, the median longitude distributions of La Niña and El Niño are similar. April neutral ENSO drylines are, on average, $\sim 0.25^\circ$ farther west than La Niña drylines, and this is the only month where La Niña longitude is significantly different from neutral ENSO longitude, but not significantly different from El Niño longitude (Fig. 4.4b,d). The uniqueness of this result may be an artifact of the larger sample of neutral ENSO drylines, which is $3\text{--}4\times$ the size of the La Niña and El Niño samples (Fig. 4.4d). However, a similar ratio of sample size is present for both May and June, and those distributions do not exhibit this same effect (Fig. 4.4e,f). Further work will be needed to evaluate whether sample size is impacting the statistics of April longitude, but for the purposes of this Chapter the result will be treated as physical.

Linear-Detrended Dryline Median Longitude versus ENSO Category, by Month

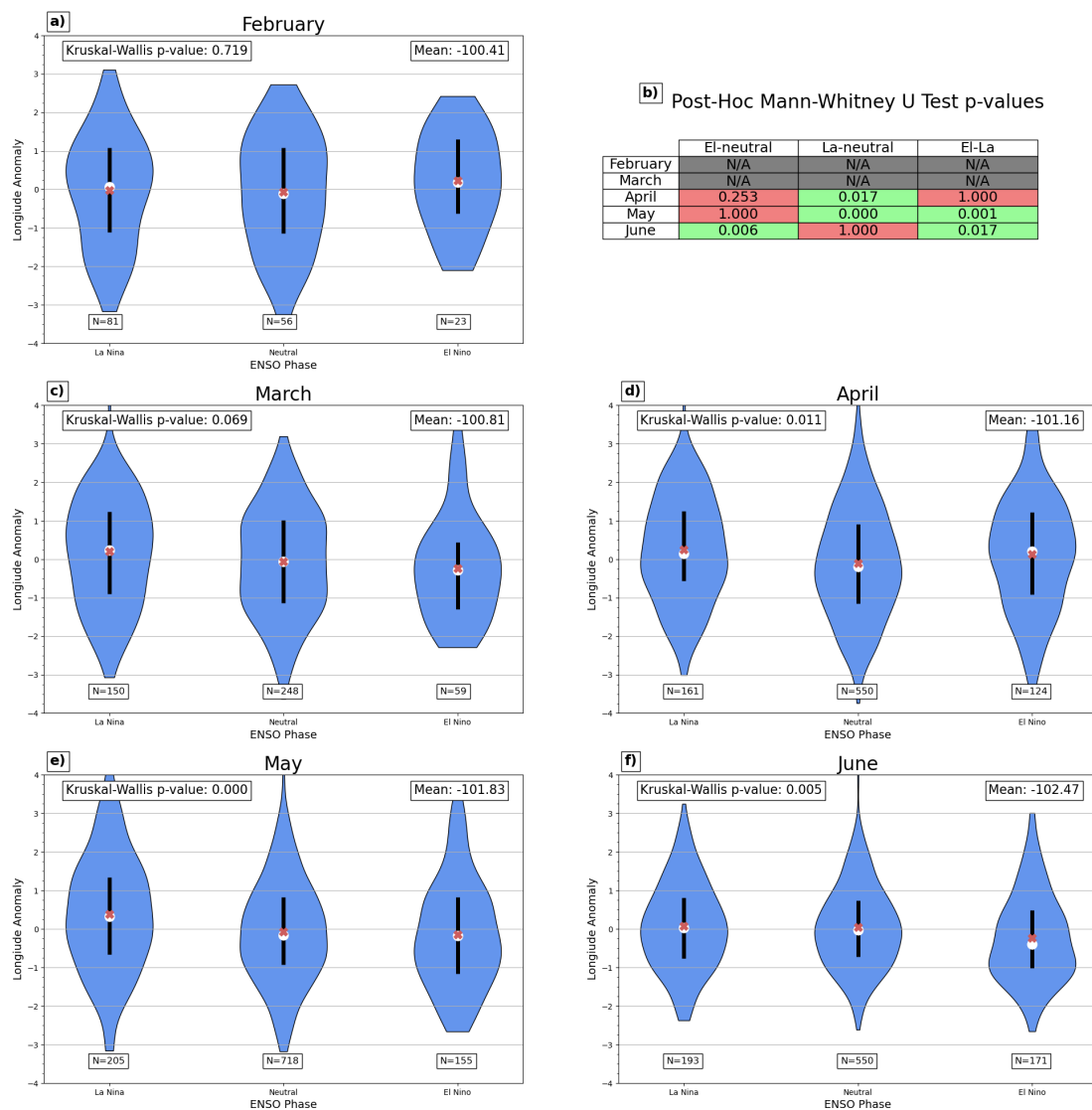


Figure 4.4: As in Fig. 4.3, but for dryline median longitudes for each month. The number of drylines in each distribution is shown at the bottom of each.

4.3.1.2 Dryline Composites

Composites of dryline frequency during El Niño and La Niña compared to neutral ENSO (see Chapter 4.2.5.2) support the statistical results found in Chapter 4.3.1.1. February dryline composite anomalies show a broad increase in dryline frequency across the SGP during La Niña compared to neutral ENSO, along with a broad decrease in dryline frequency during El Niño (Fig. 4.5). The strongest La Niña dryline frequency increases are in south-central Texas (Fig. 4.5b).

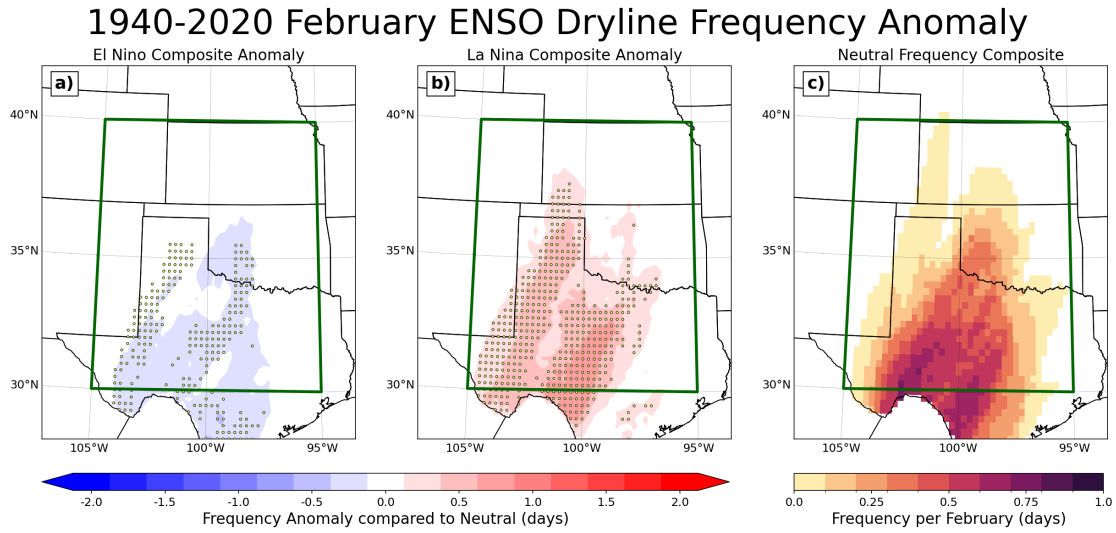


Figure 4.5: February Dryline composite anomaly for (a) El Niño and (b) La Niña compared to (c) composite dryline day frequency during neutral ENSO, in terms of days per month. The Green box outlines the defined SGP region.

March dryline frequency follows a similar pattern of broad increases during La Niña and decreases during El Niño (Fig. 4.6). The significant increases in frequency during La Niña—which correspond with significant decreases during El Niño—are across central Texas and Oklahoma. These significant anomalies are east of the peak dryline frequency during neutral ENSO in southwest Texas extending through the Texas panhandle and far northwestern Texas (Fig. 4.6c). The broad increase in dryline frequency during La Niña, and decrease in frequency during El Niño, for both February

and March is consistent with the statistical results from Chapter 4.3.1.1 (Fig. 4.3a–c).

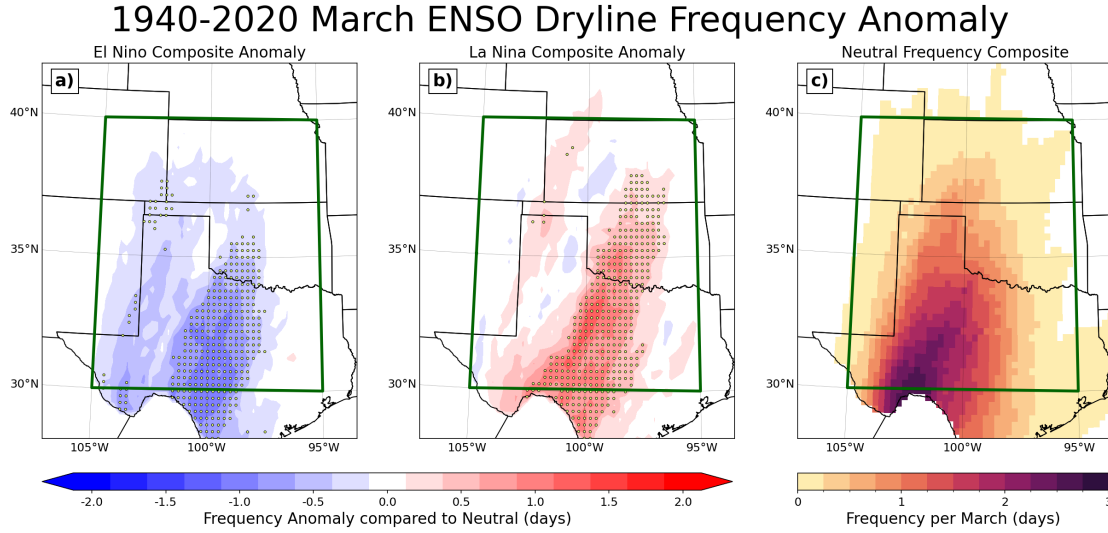


Figure 4.6: As in Fig. 4.5, but for March.

Unlike February and March, April ENSO phase is not significantly correlated with dryline frequency (Fig. 4.3d). April dryline frequency composite anomalies subsequently are not as broadly significant as February and March (Fig. 4.7). There are significant broadly negative anomalies during El Niño across west Texas through the Texas panhandle and western Kansas, but this same region is non-significantly negative during La Niña (Fig. 4.7a,b). La Niña frequency composite anomalies are longitudinally divided, with negative anomalies to the west and positive anomalies to the east (Fig. 4.7b), consistent with the significant eastward shift in La Niña longitude compared to neutral ENSO (Fig. 4.4b,d). However, aside from some isolated points in Kansas, Oklahoma, and Texas, the La Niña frequency anomalies are not spatially significant (Fig. 4.7b).

By May, however, the eastward shift in La Niña dryline longitude identified in Fig. 4.4e is amplified and statistically significant, with significance across both positive anomalies in central Texas and Oklahoma as well as negative anomalies in far southwest Texas and eastern New Mexico (Fig. 4.8b). The magnitude of the frequency anomalies

1940-2020 April ENSO Dryline Frequency Anomaly

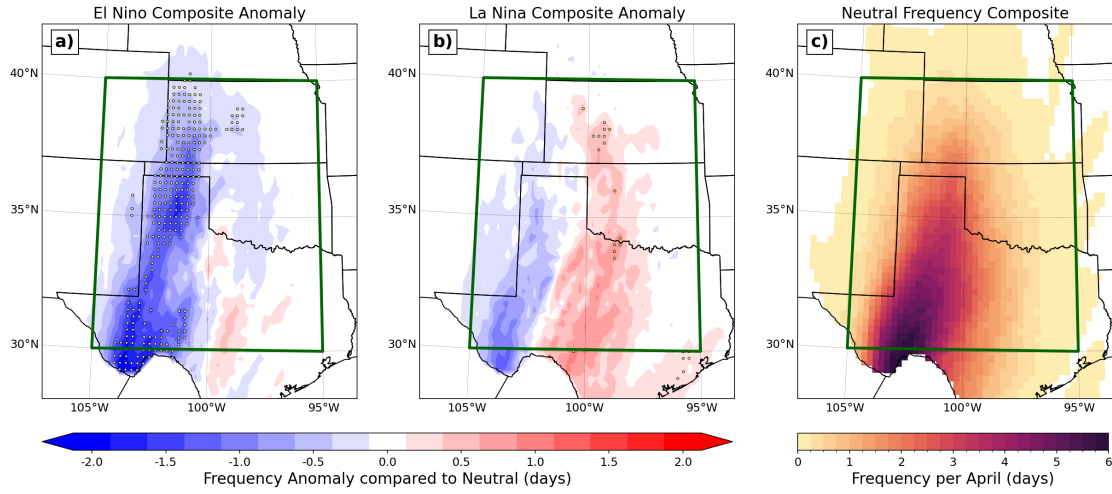


Figure 4.7: As in Fig. 4.5, but for April.

is higher than in April (Fig. 4.7b vs. Fig. 4.8b), but that may be related to the overall increase in dryline frequency in May versus April (Fig. 4.7c vs. Fig. 4.8c). May El Niño frequency anomalies are broadly negative and significant across most of the domain, although frequency was not found to be change significantly by ENSO phase (Fig. 4.3e).

1940-2020 May ENSO Dryline Frequency Anomaly

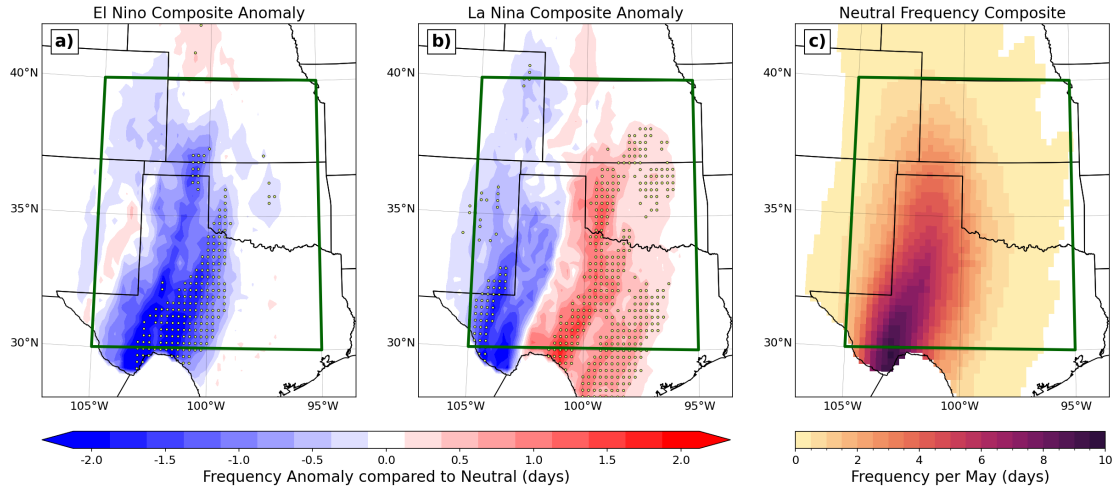


Figure 4.8: As in Fig. 4.5, but for May.

The pattern of June dryline frequency anomalies differs from April and May (Fig. 4.9). Both El Niño and La Niña composites show broadly negative frequency anomalies across the Texas panhandle and into western Kansas, with limited regions having significance (Fig. 4.9a,b). Significant positive anomalies north of the SGP during La Niña may be indicative of an earlier seasonal northward shift in dryline latitude (see Chapter 2.3.2), but further investigation into this feature is outside the scope of this Chapter. Although statistical significance is isolated, El Niño anomalies depict a dipole of decreased frequency in the Texas panhandle and increased frequency in eastern New Mexico, consistent with the significant westward shift in the longitude distribution compared to neutral ENSO and La Niña (Fig. 4.9a, Fig. 4.4b,f).

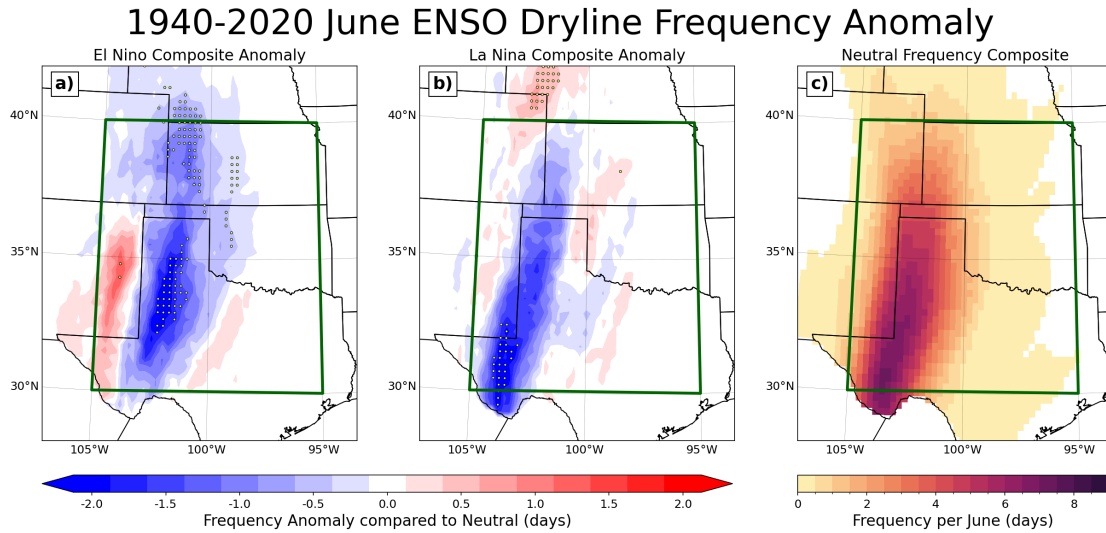


Figure 4.9: As in Fig. 4.5, but for June.

4.3.1.3 Environmental Composites

Environmental composite fields are compared between different ENSO phases to evaluate physical hypotheses for the statistical relationships found in Chapters 4.3.1.1 and 4.3.1.2 (see Chapter 4.2.4). Comparison of full month El Niño versus La Niña composites for February and March are used to evaluate synoptic changes that may explain the increase in dryline frequency during these months (Fig. 4.3a–c). Comparison of dryline-day only composites for El Niño versus La Niña for May and June—and comparison between La Niña and neutral ENSO for April, when the El Niño and La Niña distributions are not significantly different—are used to evaluate a physical explanation for the eastward shift in dryline longitude during La Niña compared to neutral ENSO and El Niño phases (Fig. 4.4b,d–f).

The February 500 mb height pattern features higher heights across the southern US during La Niña compared to El Niño (Fig. 4.10a,c,e). Additionally, 500 mb westerly winds are enhanced over much of the southern US during La Niña, with El Niño instead having enhanced westerly winds farther south over central/northern Mexico (Fig. 4.10e), consistent with the northward shift of the subtropical jet during La Niña (Chapter 4.1.2). There are minimal differences in MSLP or 10-m wind between El Niño and La Niña over the SGP (Fig. 4.10b,d,f). During La Niña, MSLP is higher over both the western and eastern US, and lower over southern central Canada and the northern Great Lakes (Fig. 4.10f). The lack of significant MSLP and 10-m wind differences in the SGP provides minimal insight for why drylines are more frequent during La Niña. It is possible that a synoptic signal is masked by the rarity of dryline days in February, with only 3.4 dryline days, on average, during La Niña. However, the higher MSLP across the far southern and southeast US along with higher 500 mb heights during La Niña may be indicative of fewer extratropical cyclones along the Gulf

Coast, consistent with the decrease in central/southern US cold front frequency during La Niña winters (Eichler and Higgins, 2006; Lagerquist et al., 2020). A decrease in Gulf Coast extratropical cyclone frequency, and associated decline in cold fronts in the SGP, may be more favorable for dryline development as relatively warm, moist air from the Gulf persists in the SGP. Further work would be needed to assess this hypothesis.

February

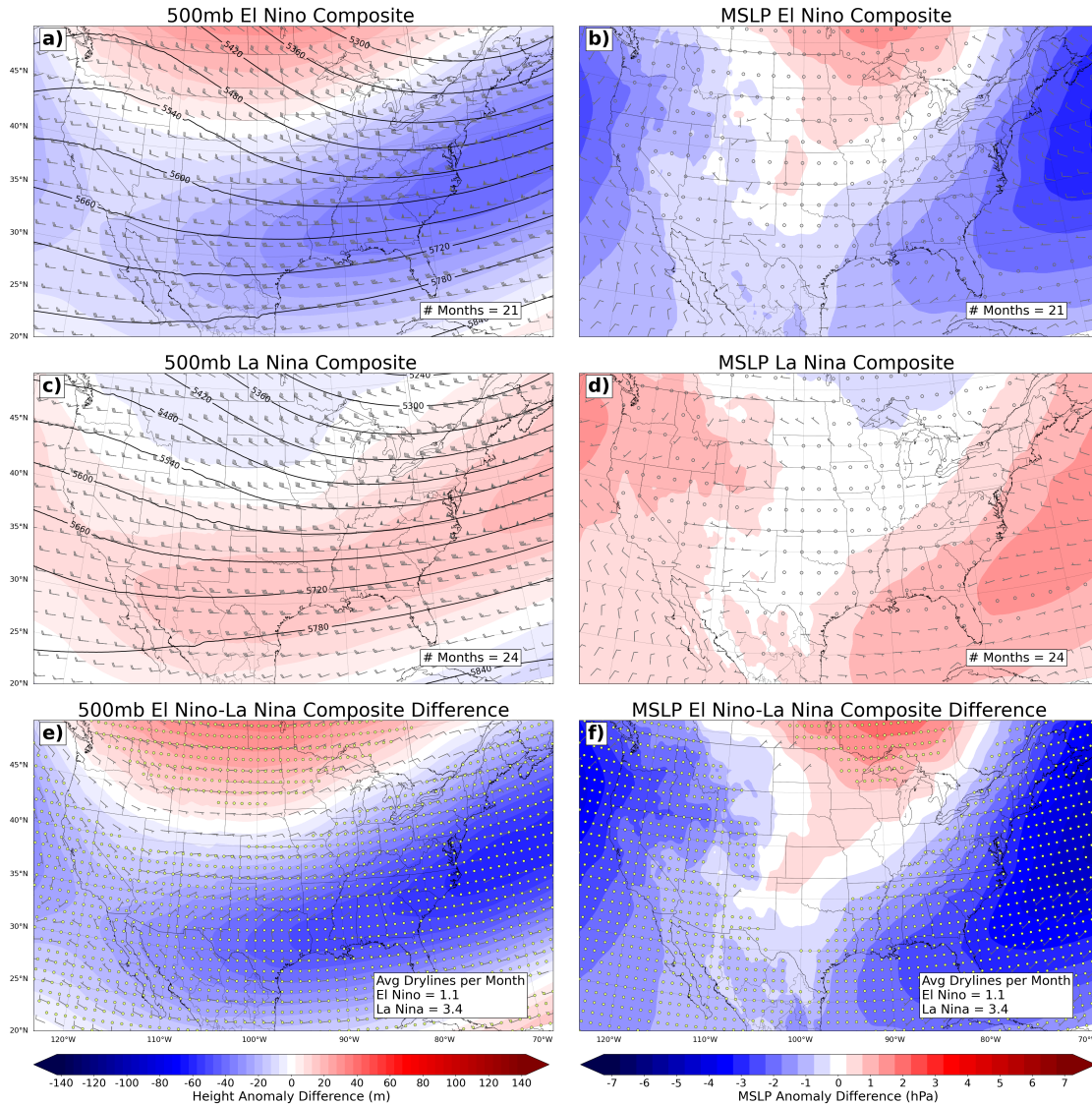


Figure 4.10: February (a) El Niño and (c) La Niña 500 mb height and wind composites. Solid black lines represent the composite height (m above mean sea level), and wind barbs denote the composite 500 mb wind vector in knots. Colored shading (scale beneath (e)) represents the 500 mb height anomaly compared to the 1940–2020 February climatology. The number of months in each composite is in the lower right. (e) The 500 mb composite height anomaly difference between (a) and (c). Stippling denotes statistical significance to the 95th percentile. Wind barbs represent the difference in 500 mb wind vector anomalies between (a) and (c), and are shown only where either the u or v component of the wind difference is statistically significant to the 95th percentile. The average number of dryline days per February for El Niño and La Niña is in the bottom right. (b), (d), and (f) as in (a), (c), and (e), but for Mean Sea Level Pressure (hPa) and 10-m wind.

The March environmental composites provide a clearer hypothesis for the increase in dryline frequency during La Niña (Fig. 4.11). Similar to February, La Niña 500 mb heights are higher across the southern US with stronger westerlies, particularly across southern Kansas and Colorado (Fig. 4.11e). At the surface during La Niña MSLP is lower across eastern Colorado, and the westerly component of 10-m winds across northern New Mexico is marginally increased (Fig. 4.11f). The concentrated region of lower pressure in eastern Colorado during La Niña is consistent with the location of lee cyclogenesis east of the Rocky Mountains (e.g., Hobbs et al., 1996), indicating either an increase in the frequency of lee cyclogenesis or an increased intensity (e.g., deeper low) of the lee cyclones. Past studies inform the hypothesis that this feature results from an increased frequency of lee cyclogenesis (Eichler and Higgins, 2006; Schemm et al., 2018), although future work will be needed to assess this. The hypothesized increased frequency of lee cyclogenesis events is attributable to the increased westerly 500 mb wind during La Niña over the Rocky Mountains of New Mexico and Colorado (on the order of 10 knots; Fig. 4.11e). As drylines are common with lee cyclones, an increase in lee cyclogenesis events would provide a clear hypothesis for the increase in dryline frequency during La Niña. It is not clear from the composites alone why February La Niña, which also featured increased westerlies across Colorado of a similar magnitude (Fig. 4.10e), did not similarly result in a significant feature suggesting an increase in lee cyclogenesis events. A smaller sample of February lee cyclogenesis events, potentially due in part to other factors influencing the frequency of lee cyclogenesis, may be obscuring the signal in the February composites.

March

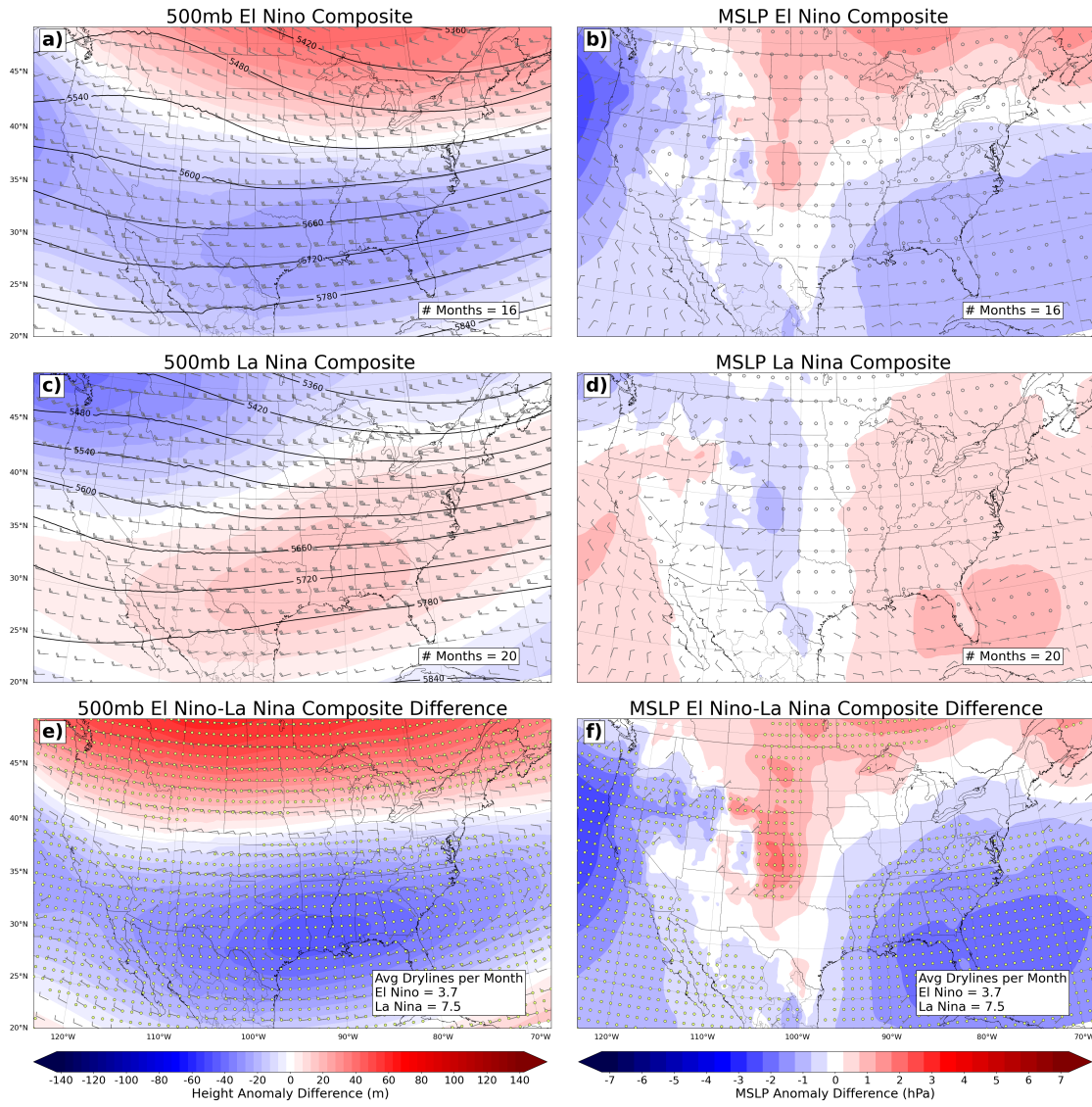


Figure 4.11: As in Fig. 4.10, but for March.

While ENSO does not have a statistical impact on the frequency of April–June drylines, it is associated with dryline median longitude (Figs. 4.3, 4.4). During April, La Niña drylines are on average $\sim 0.25^\circ$ farther east than drylines during neutral ENSO (Fig. 4.4d). Both La Niña and neutral ENSO April dryline days are defined by a 500 mb trough over the western US, a 500 mb ridge over the eastern US, and low MSLP over the central US (Fig. 4.12). La Niña dryline day 500 mb heights are lower across the majority of the CONUS, although the amplitude difference between the trough and ridge is similar for both La Niña and neutral ENSO dryline days (Fig. 4.12a,c,e). There are minimal differences in 500 mb flow. At the surface, MSLP is broadly lower across most of the CONUS during La Niña dryline days (Fig. 4.12b,d,f). Similar to 500 mb, there are no notable differences in 10-m flow. It is possible that the broadly lower 500 mb heights and lower MSLP during La Niña is a result of the much larger sample size of neutral ENSO dryline days (550) compared to La Niña dryline days (161) and the substantially more neutral ENSO Aprils (51) than La Niña Aprils (17) (Fig. 4.2a,c). In particular, it is notable that despite the broadly significant changes in 500 mb height and MSLP, there are minimal significant differences in wind vectors at both 500 mb and 10-m. If the composite difference is assumed to be a purely physical result, the deeper MSLP could be indicative of stronger extratropical cyclones during La Niña. Stronger extratropical cyclones could advect drylines farther east due to stronger synoptic-scale winds, explaining the eastward shift in La Niña dryline longitude. However, if this were the case, it is surprising that there is no reflection of stronger 10-m winds in the composite difference (Fig. 4.12f). Alternatively, mesoscale factors such as soil moisture and cloud cover, not captured in these composites, could be suitable hypotheses for the eastward shift in La Niña dryline longitude. Confidence in these physical hypotheses is low due to the possible non-physical influence of sample size on these composite comparisons.

April Dryline Days

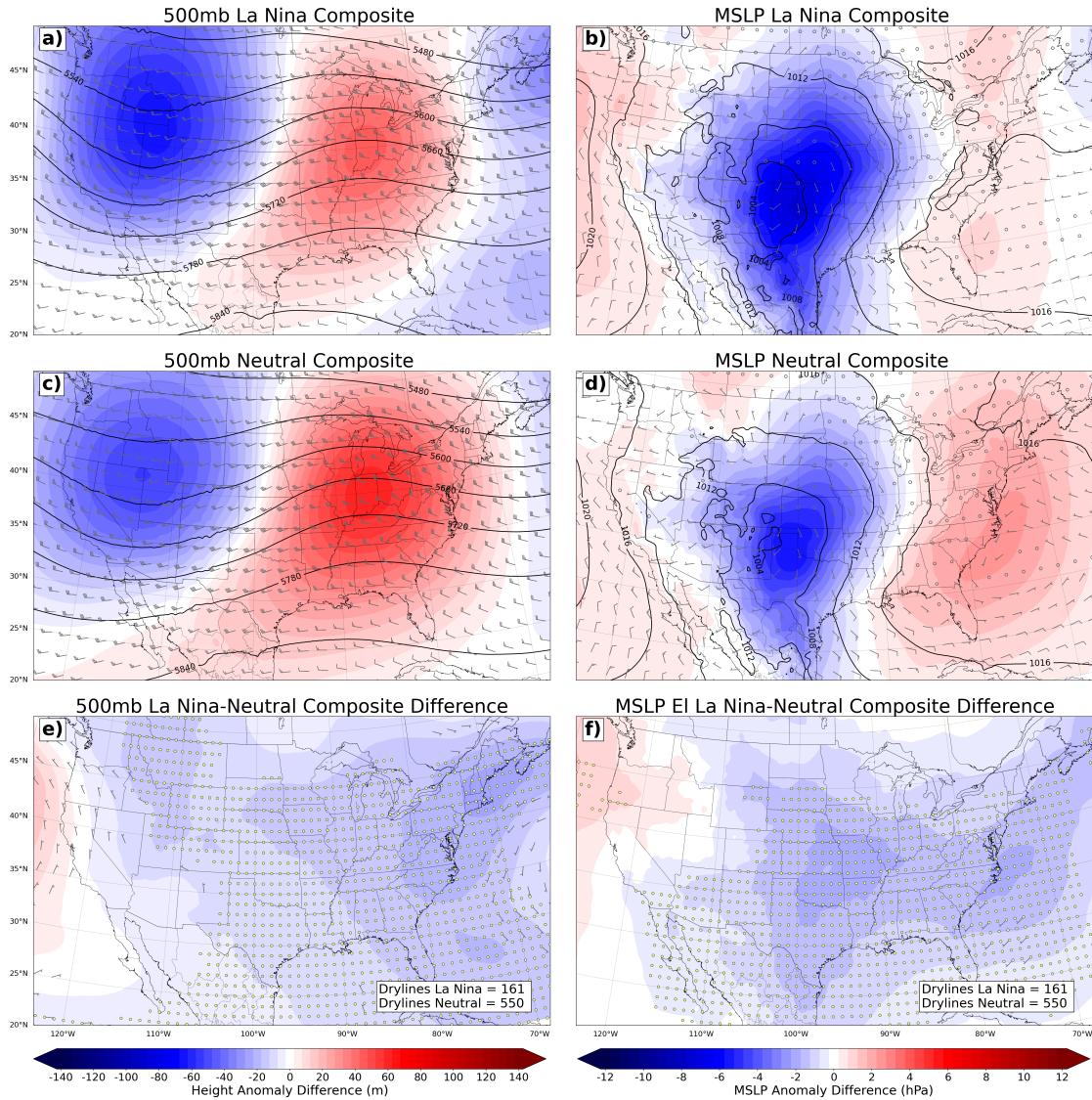


Figure 4.12: As in Fig. 4.10, but composites consist only of dryline days for April and compare (a,b) La Niña to (c,d) neutral ENSO. The number of dryline days that go into each composite are in the lower right corner of (e) and (f).

The comparison between composites of May El Niño dryline days and La Niña dryline days offers a more comparable sample size for both groups (155 dryline days and 205 dryline days, respectively; Fig. 4.2c). During La Niña dryline days, the base of the western US trough is shifted north compared to El Niño (Fig. 4.13a,c,e). The eastern US ridge is less prominent during La Niña dryline days, although the relative height anomaly between the lowest low of the trough and highest high of the ridge is equivalent between La Niña and El Niño dryline days (~ 100 m). While this is consistent with the northward shift of the jet stream expected during La Niña, La Niña dryline days also feature lower heights off the southeast US coast associated with a more amplified offshore trough. Although the magnitude of the 500 mb flow over the SGP is similar between El Niño and La Niña, the direction of the flow is more westerly during La Niña dryline days. While there are minimal differences in composite 10-m wind, MSLP for La Niña dryline days is broadly lower east of the Rocky Mountains (Fig. 4.13b,d,f). Lower MSLP over southeastern Colorado on the order of ~ 2 hPa may be an indication of stronger lee cyclones or an increased frequency of drylines associated with lee cyclones as opposed to synoptically-quiet drylines during La Niña. Stronger lee cyclones—potentially explained by an increased westerly component of 500 mb flow over the western SGP and Colorado (Fig. 4.13e)—could contribute to an eastward shift in dryline longitude due to a stronger advection component to dryline propagation, although, similar to the April composites (Fig. 4.12f), there is no reflection of stronger 10-m winds during La Niña (Fig. 4.13f). Additionally, the increased westerly component of flow aloft over the SGP—evident at both 500 mb and 700 mb (not shown)—may contribute to faster dryline propagation as increased westerly momentum is mixed down to the surface by deep vertical mixing on the dry side of the dryline (Chapter 4.1.1).

May Dryline Days

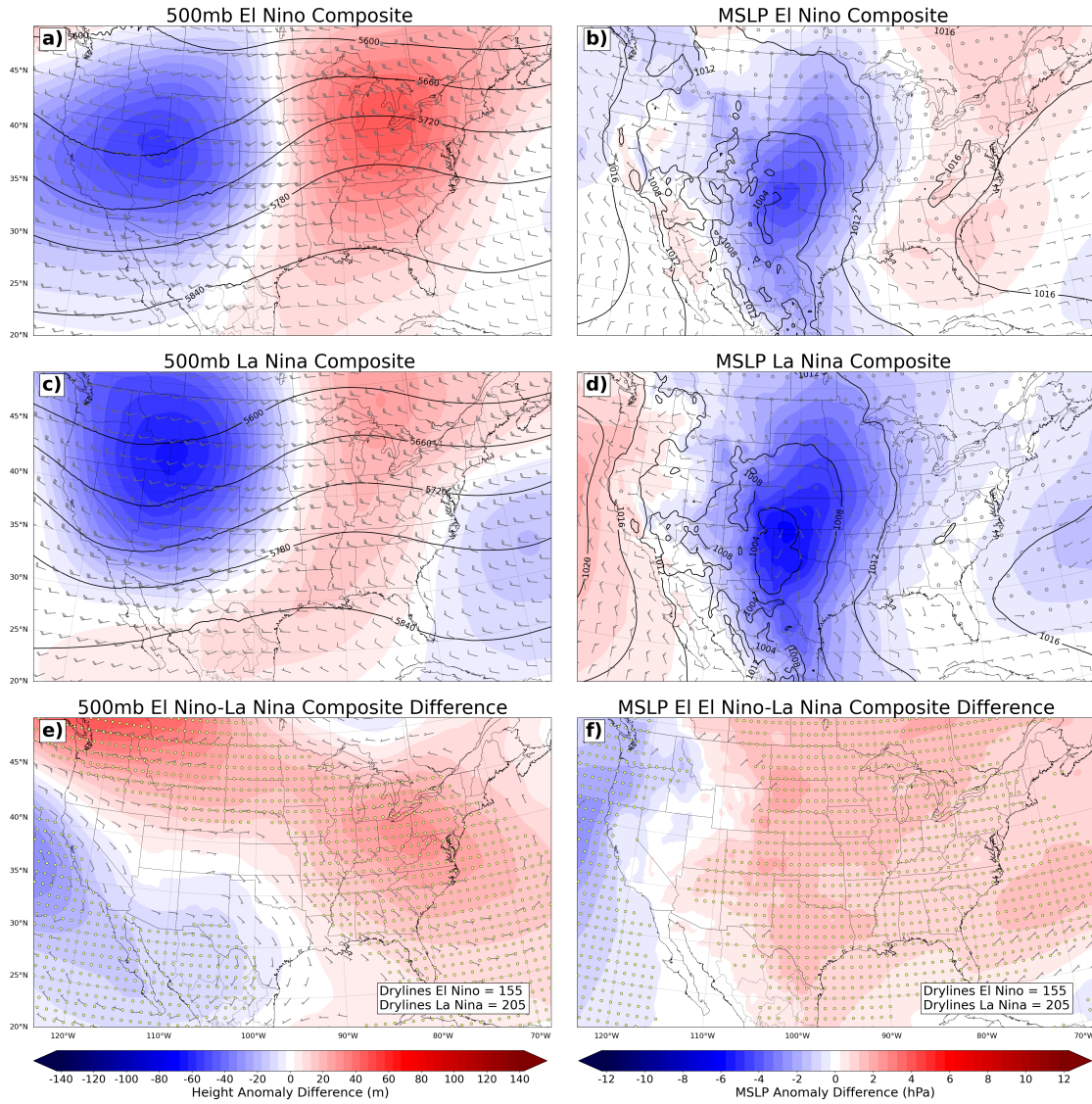


Figure 4.13: As in Fig. 4.12, but for May. Composites are comparing (a,b) El Niño to (c,d) La Niña. Note that, due to the order of operations, for (e,f) lower 500 mb height or MSLP during La Niña is denoted by red shading, as opposed to blue shading in Fig. 4.12e,f.

The composite differences for June dryline days are distinctly different compared to previous months (Fig. 4.14). The 500 mb trough–ridge pattern is significantly more amplified during La Niña compared to El Niño, with a relative height difference between trough and ridge of ~ 100 m during La Niña and ~ 60 m during El Niño (Fig. 4.14a,c,e). The amplified height pattern during La Niña is reflected in increased curvature of the 500 mb flow, with an increased southerly component to the flow over the SGP. The La Niña MSLP pattern reflects the 500 mb pattern, with lower MSLP in the central US from the northern Great Plains through southeast Colorado and the northern Texas panhandle, and higher MSLP over the eastern US (Fig. 4.14b,d,f). Along the Rocky Mountains from Montana through Colorado, westerly 10-m wind is stronger during La Niña. The MSLP and 10-m flow is consistent with an increased frequency/strength of extratropical cyclones over the northern Great Plains as well as in eastern Colorado. During June La Niña, drylines are more frequent north of the SGP, consistent with drylines attendant to extratropical cyclones in the northern Great Plains (Fig. 4.9b). The significant dryline longitude difference of $\sim 0.5^\circ$ is primarily the result of a westward shift in El Niño drylines (Fig. 4.4f), evident in increased dryline frequency in southwest Texas and eastern New Mexico compared to La Niña (Fig. 4.9a,b). The active synoptic pattern and extratropical cyclones over the northern Great Plains and eastern Colorado are favorable for drylines which are farther north as well as farther east via the advection component to dryline propagation. El Niño drylines, likely more frequently classifiable as synoptically-quiescent, are likely to be farther west, as eastward propagation is controlled primarily by mesoscale processes with minimal appreciable advection component. The near-normal frequency of drylines in southwest Texas and increased frequency in eastern New Mexico during El Niño (Fig. 4.9a) supports the hypothesis that June El Niño drylines are more often synoptically-quiescent.

June Dryline Days

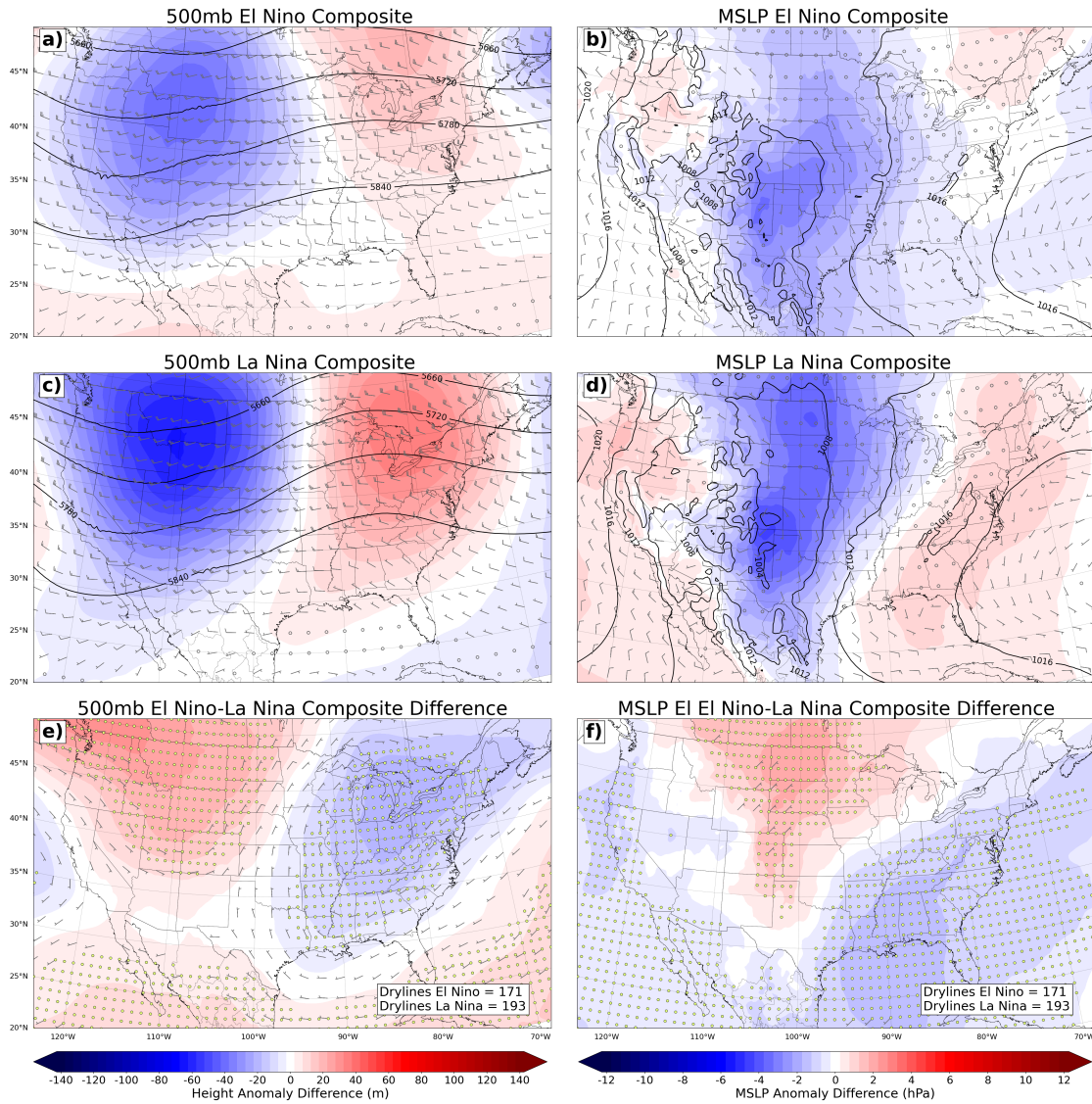


Figure 4.14: As in Fig. 4.13, but for June.

4.3.2 Drought

4.3.2.1 Monthly Distributions

PDSI is more strongly correlated with dryline frequency than ENSO. There is a statistically significant increase in dryline frequency during drought compared to wet for all months except for June (Fig. 4.15). Drought in particular drives these differences, as the distributions of dryline frequency for wet conditions and normal conditions are not statistically different (Fig. 4.15b). The difference in frequency distributions is substantial between drought and wet, with the 75th percentile of wet frequency anomaly falling below the 25th percentile of the drought frequency anomaly distribution for March and April (Fig. 4.15c,d).

Linearly Detrended Dryline Frequency verses PDSI Category, by Month

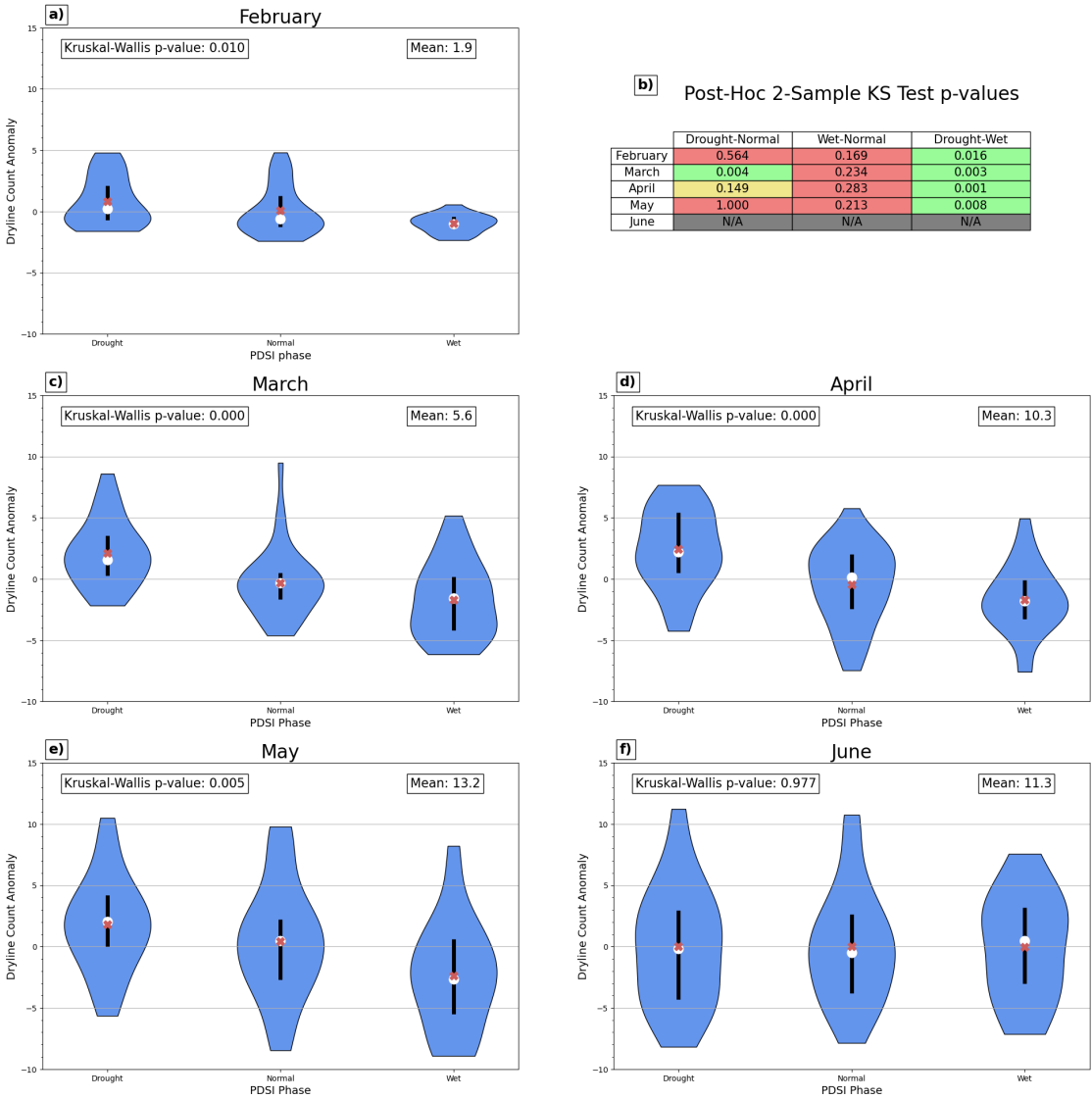


Figure 4.15: As in Fig. 4.3, but for PDSI phase.

For all months except February, drylines are farther east when conditions are drier (Fig. 4.16). Comparison of dryline longitude between drought and wet conditions are significant for March–June, with significance also between drought–normal and wet–normal for some of the months (Fig. 4.16b), indicative of a gradual eastward shift as conditions become drier. The difference in the median of dryline median longitude between drought and wet conditions is on the order of 0.5° across each individual month (Fig. 4.16c–f).

Linear-Detrended Dryline Median Longitude verses PDSI Category, by Month

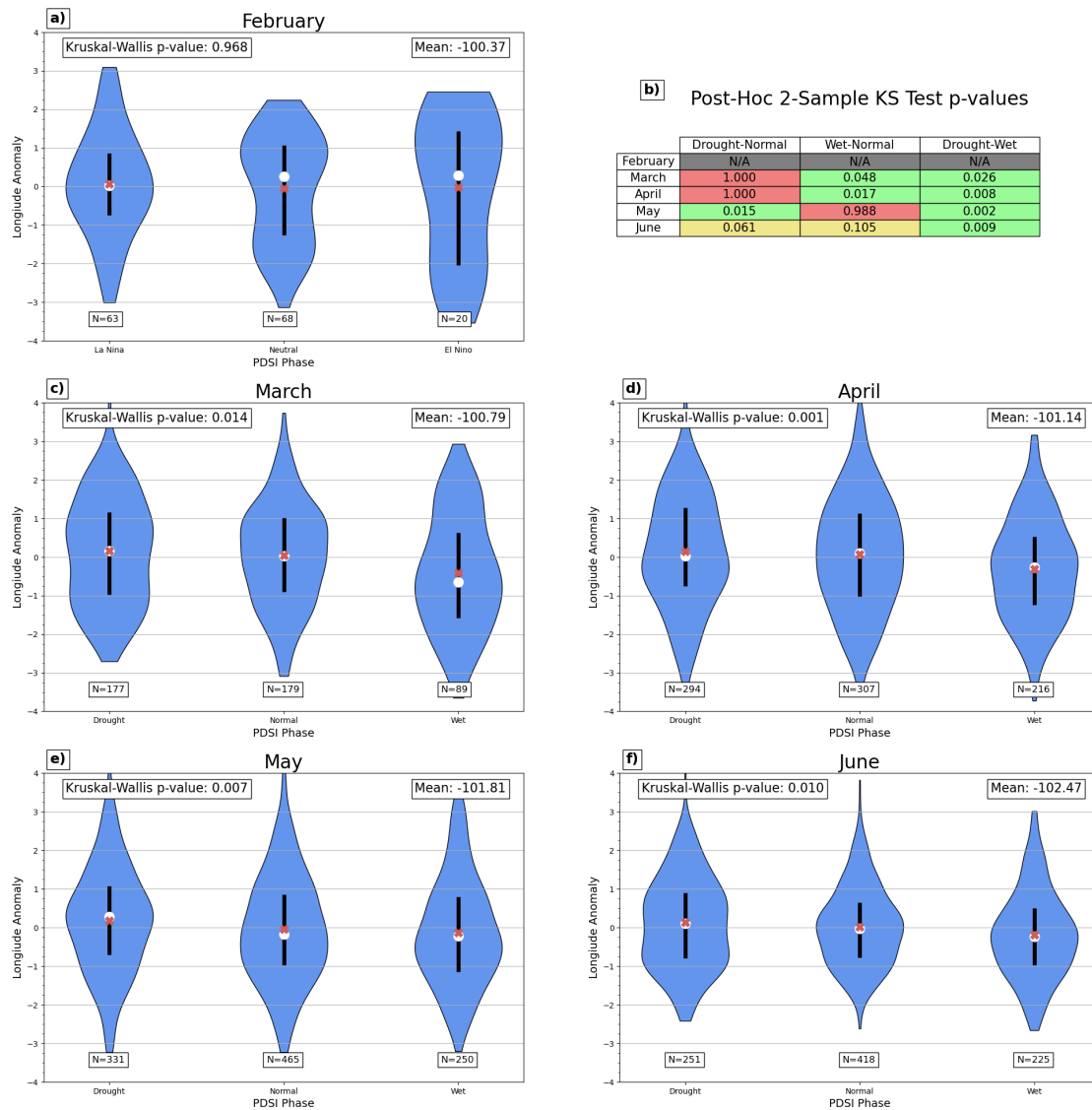


Figure 4.16: As in Fig. 4.4, but for PDSI phase.

4.3.2.2 Dryline Composites

Composites of drought and wet condition dryline frequency anomaly compared to normal (see Chapter 4.2.5.2) support the statistical significance found in Chapter 4.3.2.1. The February frequency anomaly composites depict a broad increase in dryline frequency during drought along with a broad decrease in frequency during wet conditions (Fig. 4.17). Across the portions of central Texas with the highest dryline frequency during normal conditions, there is minimal change in dryline frequency during drought (Fig. 4.17a,c). Increases in dryline frequency during drought span both far east/south-east SGP as well as western Texas and western Oklahoma (Fig. 4.17a).

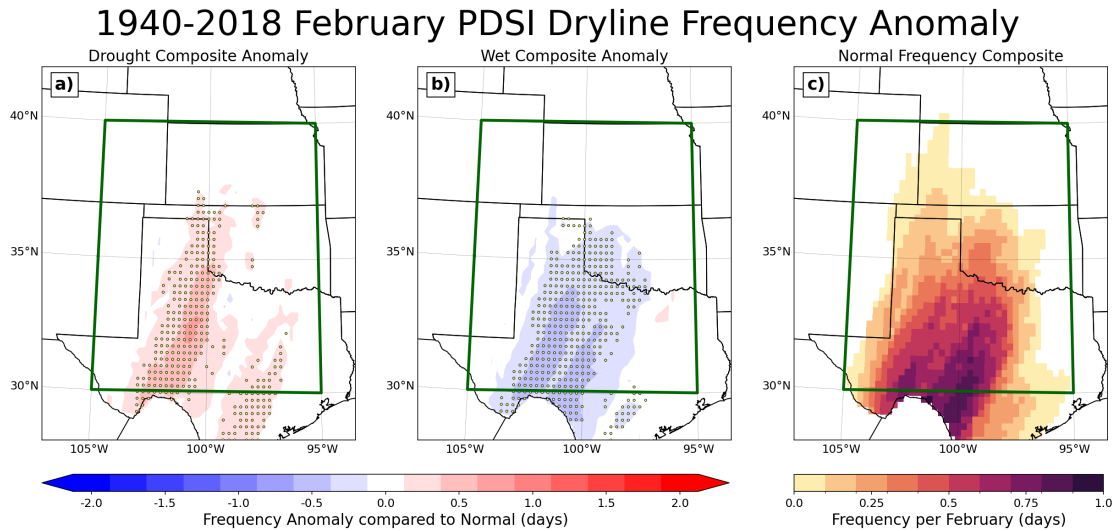


Figure 4.17: As in Fig. 4.5, but for PDSI. (a) drought and (b) wet conditions dryline frequency anomaly composites compared to (c) the dryline frequency composite under normal conditions, in terms of days per month.

The March frequency anomaly composites show a broad increase in dryline frequency during drought compared to normal conditions, and a broad decrease in dryline frequency during wet conditions (Fig. 4.18). Dryline frequency in the western Texas panhandle—on the order of 1 day per month—is not significantly changed between PDSI phases (Fig. 4.18). During normal conditions, drylines are most frequent

across west/central Texas (1.5–2 days per month; Fig. 4.18c). During drought, dryline frequency increases across this same region (Fig. 4.18a), consistent with the lack of difference between the distribution of dryline longitude between normal conditions and drought (Fig. 4.16b,c). Drylines are significantly farther west during wet conditions, compared to both normal and drought, due to the significant decrease in drylines across west/central Texas (Fig. 4.18b). This trend may be the result of feedbacks from increased dryline convection initiation during wet conditions. When drylines initiate convection, the mesoscale mixing processes responsible for eastward propagation are slowed due to cloud cover and precipitation east of the dryline, thus drylines may not propagate as far east. Investigating the role of dryline convection initiation compared to synoptic or antecedent soil conditions is outside the scope of this Chapter.

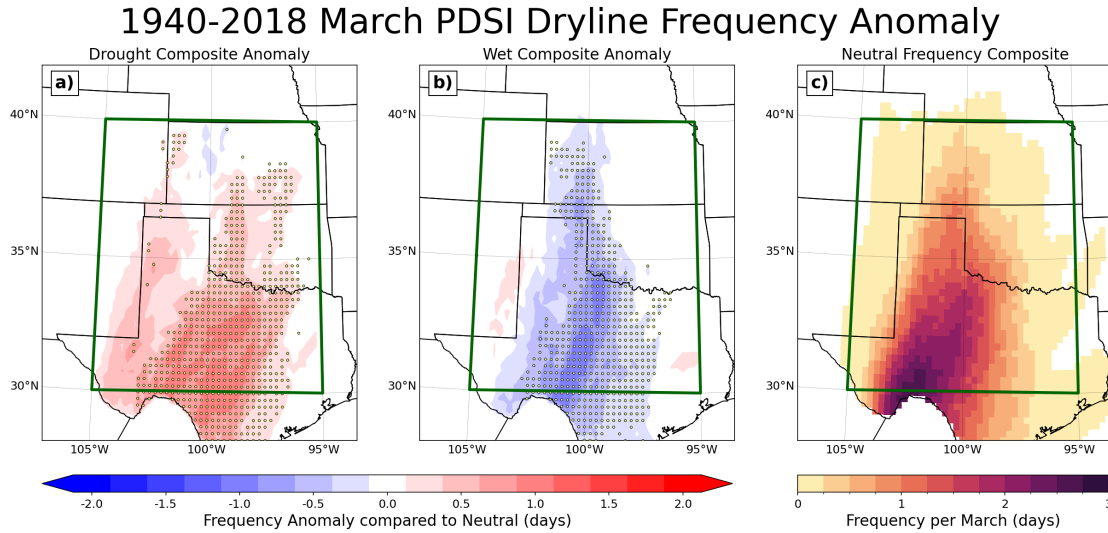


Figure 4.18: As in Fig. 4.17, but for March.

The spatial dryline frequency anomalies in April are similar to the pattern in March (Fig. 4.19). During drought, there is a widespread increase in dryline frequency across much of the SGP, coupled with a widespread decrease during wet conditions (Fig. 4.19a,b). In southwest Texas—where drylines are most frequency during normal conditions—there is no significant change in dryline frequency during wet conditions

(Fig. 4.19b,c), consistent with the lack of significant difference in dryline frequency distributions between normal and wet conditions (Fig. 4.15b,d). The decrease in drylines across central Texas and central/western Oklahoma contributes to the westward shift in dryline longitude during wet conditions compared to both normal and drought conditions (Fig. 4.16b,d), as well as the decrease in frequency during wet conditions versus drought (Fig. 4.15b,d).

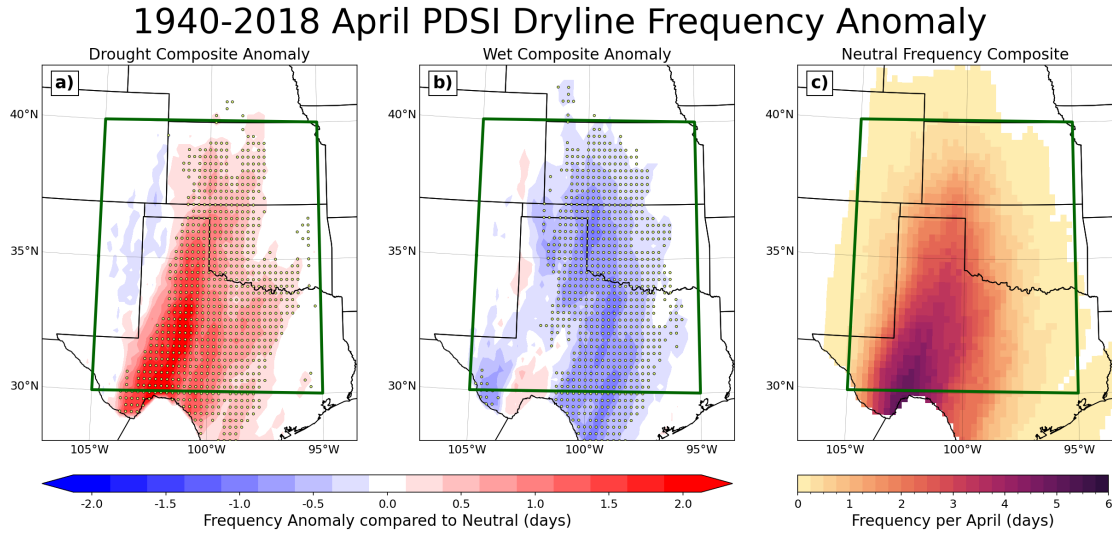


Figure 4.19: As in Fig. 4.17, but for April.

The May pattern of dryline frequency anomalies is similar to both April and March (Fig. 4.20). The magnitude of negative frequency anomalies during wet conditions is higher than in April (Figs. 4.19b vs. 4.20b), although this may be attributed to the higher number of dryline days overall in May versus April (~ 10 days vs. ~ 6 days on average, respectively; Figs. 4.19c, 4.20c). Significant decreases in frequency in wet conditions overlap with increases in frequency during drought across the Texas panhandle, western Oklahoma, and western Kansas (Fig. 4.20a,b). During drought, a marginally significant dipole of frequency is evident in the NW SGP, with positive frequency anomalies across the Texas and Oklahoma panhandles and western Kansas adjacent to significant negative frequency anomalies in southeast Colorado and north-

east New Mexico (Fig. 4.20a). These dryline frequency anomalies are consistent with the increase in dryline frequency and eastward shift in dryline longitude during drought compared to wet conditions (Figs. 4.15b,e, 4.16b,e).

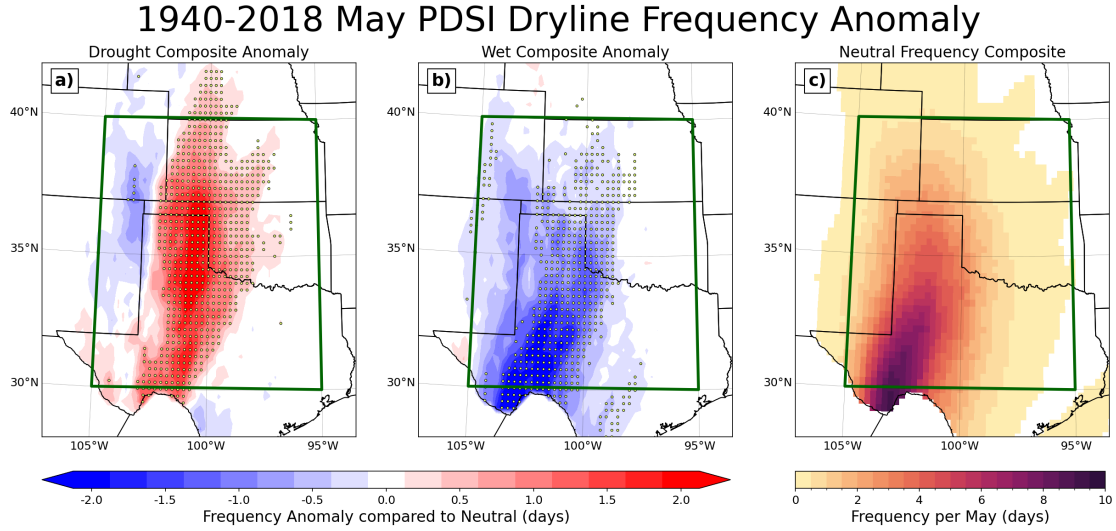


Figure 4.20: As in Fig. 4.17, but for May.

The pattern of dryline frequency anomalies during June is different from previous months (Fig. 4.21). During drought, marginally significant negative dryline frequency anomalies are present across the Texas panhandle and southwest Texas, where overall dryline frequency during normal conditions is maximized (Fig. 4.21a,c). Positive dryline frequency anomalies extend across western Kansas and north of the SGP, with isolated significance in west/central Oklahoma, eastern Colorado, and New Mexico (Fig. 4.21a). This pattern reflects a northward shift in dryline frequency during drought, with the marginal positive significance across Oklahoma and New Mexico likely the result of a low underlying dryline frequency during normal conditions (Fig. 4.21a,c). During wet conditions, dryline frequency anomalies are not significant across most of the SGP with the exception of negative frequency anomalies in western Kansas. Although these wet negative anomalies overlap the positive frequency anomalies during drought, the magnitude of the anomalies are relatively weak, reflected by the lack of

significant change in dryline frequency by PDSI category (Fig. 4.15e). The region of positive dryline frequency anomalies during drought extends slightly east of longitude of maximum dryline frequency during normal conditions (Fig. 4.21a,c). Combined with marginally negative frequency anomalies in the Texas panhandle (Fig. 4.21a), the June drought composite anomaly is consistent with the slight but significant eastward shift in dryline longitude compared to wet conditions (Fig. 4.16b,e).

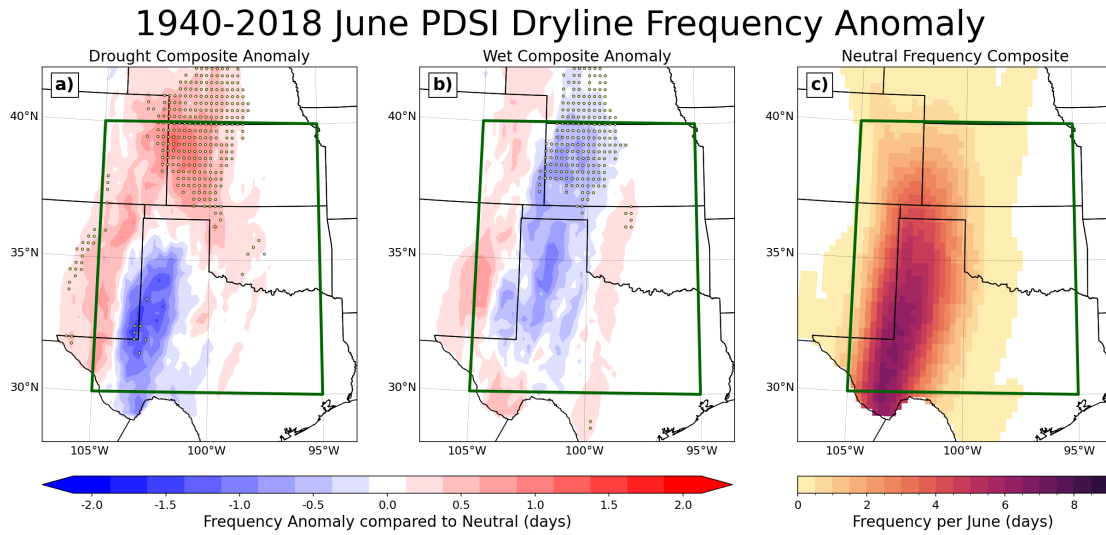


Figure 4.21: As in Fig. 4.17, but for June.

4.3.2.3 Environmental Composites

Environmental composite fields are compared between different PDSI phases to evaluate physical hypotheses for the statistical relationships found in Chapters 4.3.2.1 and 4.3.2.2 (see Chapter 4.2.4). Comparison of full month drought versus wet composites for February–May are used to evaluate synoptic changes that may explain the increase in dryline frequency during these months (Fig. 4.15a–e). Comparison of dryline-day only composites for drought versus wet for March–June are used to physically evaluate the significant eastward shift in dryline longitude during drought compared to wet (Fig. 4.16b–f).

The February 500 mb height pattern during wet conditions shows lower-than-climatological heights across most of the western and southern US, along with higher-than-climatological heights over southern Canada and the Great Lakes region (Fig. 4.22a). This latitudinal dipole is reversed during drought, with higher heights across the southern US and lower heights over the Great Lakes (Fig. 4.22c). The tighter 500 mb height gradient during drought results in faster westerlies stretching from the Pacific Northwest through the eastern US (Fig. 4.22e). Composites of MSLP highlight an inverse pattern between drought and wet conditions, with higher MSLP over the western US and Gulf Coast/Atlantic along with lower MSLP across southern Canada through the northern SGP during drought (Fig. 4.22b,d,e). The MSLP wet–drought composite difference is similar to the February El Niño–La Niña MSLP composite difference (Figs. 4.22e, 4.10e). Increased westerly flow aloft and lower MSLP over the northern SGP during drought could be indicative of favorable conditions for drylines attendant to extratropical cyclones over the northern SGP or farther south. Similar to La Niña, the drought composites are indicative of fewer extratropical cyclones along the Gulf Coast (Eichler and Higgins, 2006).

February

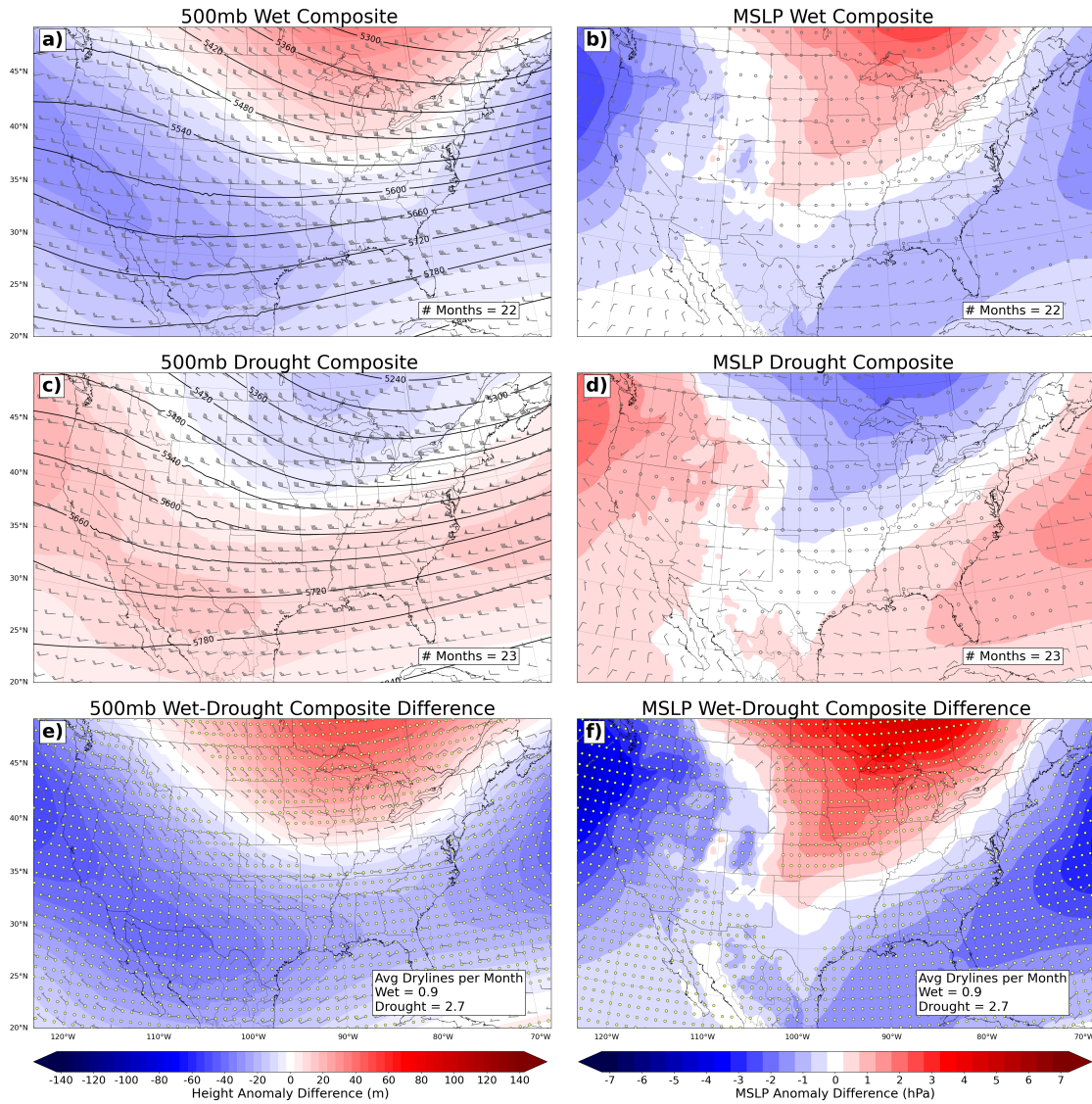


Figure 4.22: Same as Fig. 4.10, but for PDSI phases wet and drought.

The environmental composite comparison between March wet and drought (Fig. 4.23) is similar to the environmental composite comparison between March El Niño and La Niña (Fig. 4.11). During drought, 500 mb height contours are more tightly packed over the US, as evidenced by both increased heights over the southern US and northern Mexico paired with lower-than-climatology heights over the northern US as well as the increased westerly flow over the US during drought (Fig. 4.23a,c,e). Over the SGP, westerly 500 mb flow is 5–10 kts faster during drought compared to wet conditions, particularly over the northern SGP (Fig. 4.23e). During drought, MSLP is lower over eastern Colorado (Fig. 4.24f), indicating either an increase in the frequency of lee cyclogenesis events or stronger lee cyclogenesis events. As discussed in Chapters 4.1.1 and 4.3.1.3, drylines are a component of the conceptual model of lee cyclones east of the Rocky Mountains (Hobbs et al., 1996). Therefore, the increase in drylines during SGP drought in March (Fig. 4.15b,c) suggest that this MSLP pattern is the result of an increase in lee cyclogenesis events. Increased westerly flow across the Rocky Mountains of northern New Mexico and Colorado at both 500 mb and 10-m is favorable for lee cyclone development (Fig. 4.23e,f). The seeming contradiction between SGP drought and an increase in frequency and/or strength of lee cyclones, which would suggest more synoptic forcing for SGP precipitation events, is discussed in detail in Chapter 4.4.

March

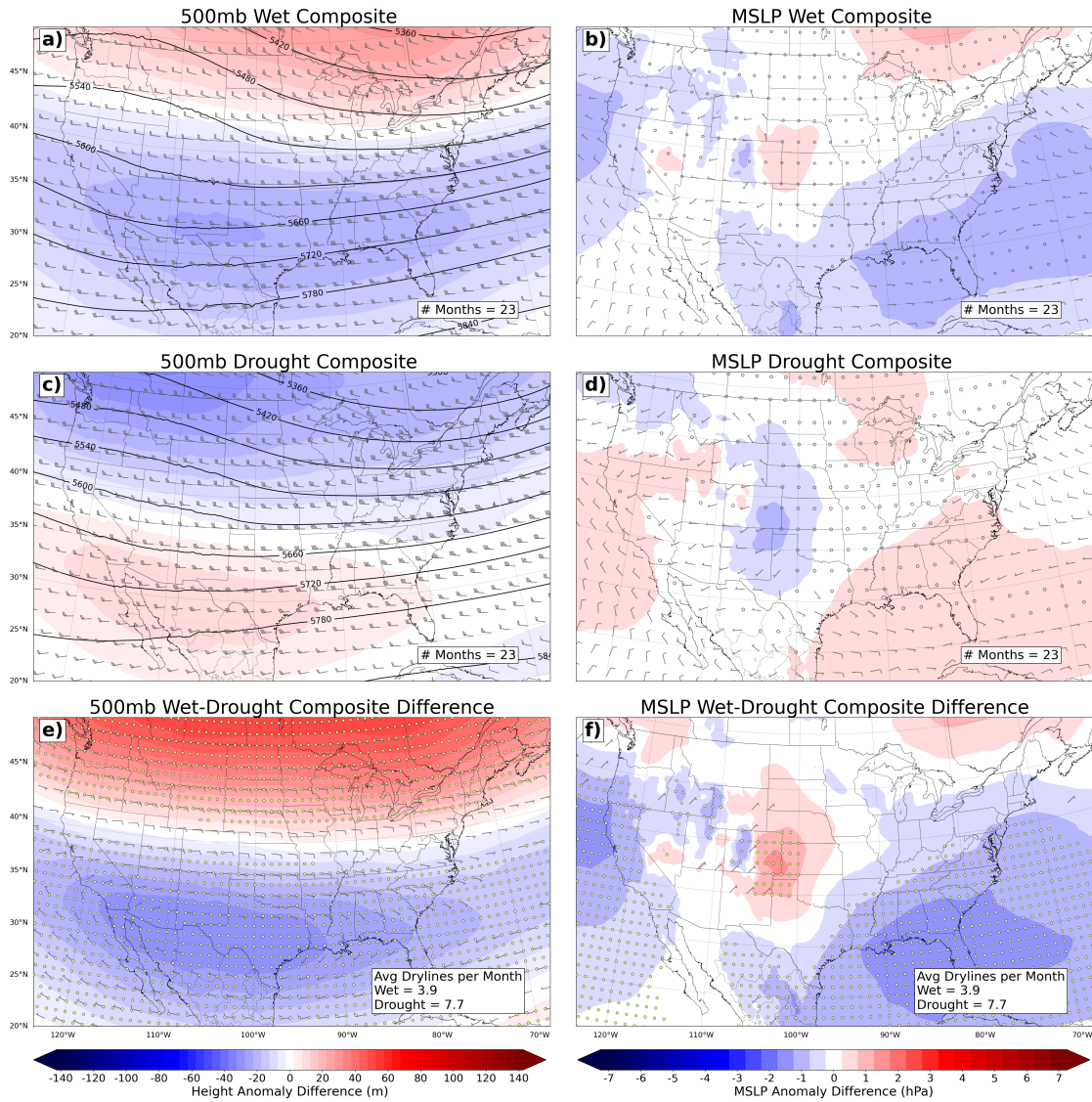


Figure 4.23: Same as Fig. 4.22, but for March.

March drylines during drought are, on average, $\sim 0.75^\circ$ farther east than March drylines during wet conditions (Fig. 4.16c). At 500 mb, both drought and wet condition March dryline days are defined by a trough over the western US and a ridge over the eastern US (Fig. 4.24a,c). The latitude of the trough axis during drought is farther north than during wet conditions, resulting in higher 500 mb heights over the southwestern US, including the SGP, during drought dryline days (Fig. 4.24e). Flow anomalies at 500 mb during drought are anticyclonic around the higher height centered in the southwest US, resulting in increased westerly flow over Colorado and an increased westerly component to the flow over SGP during drought. There are minimal MSLP differences between drought and wet dryline days in the SGP (Fig. 4.24f). The increased westerly component of flow aloft during drought may contribute to the eastward shift in dryline longitude. Deep vertical mixing on the dry side of the dryline mixes momentum down to the surface, and increased westerly flow increasing the westerly surface flow would increase confluence along the dryline, enhancing vertical mixing and thus increasing the speed of eastward propagation (see Chapter 4.1.1). Some marginal increases in the westerly component of 10-m wind during drought are evident in the Texas panhandle and northern New Mexico (Fig. 4.24b,d,f), which could be a representation of the downward mixing of the increased westerly flow.

March Dryline Days

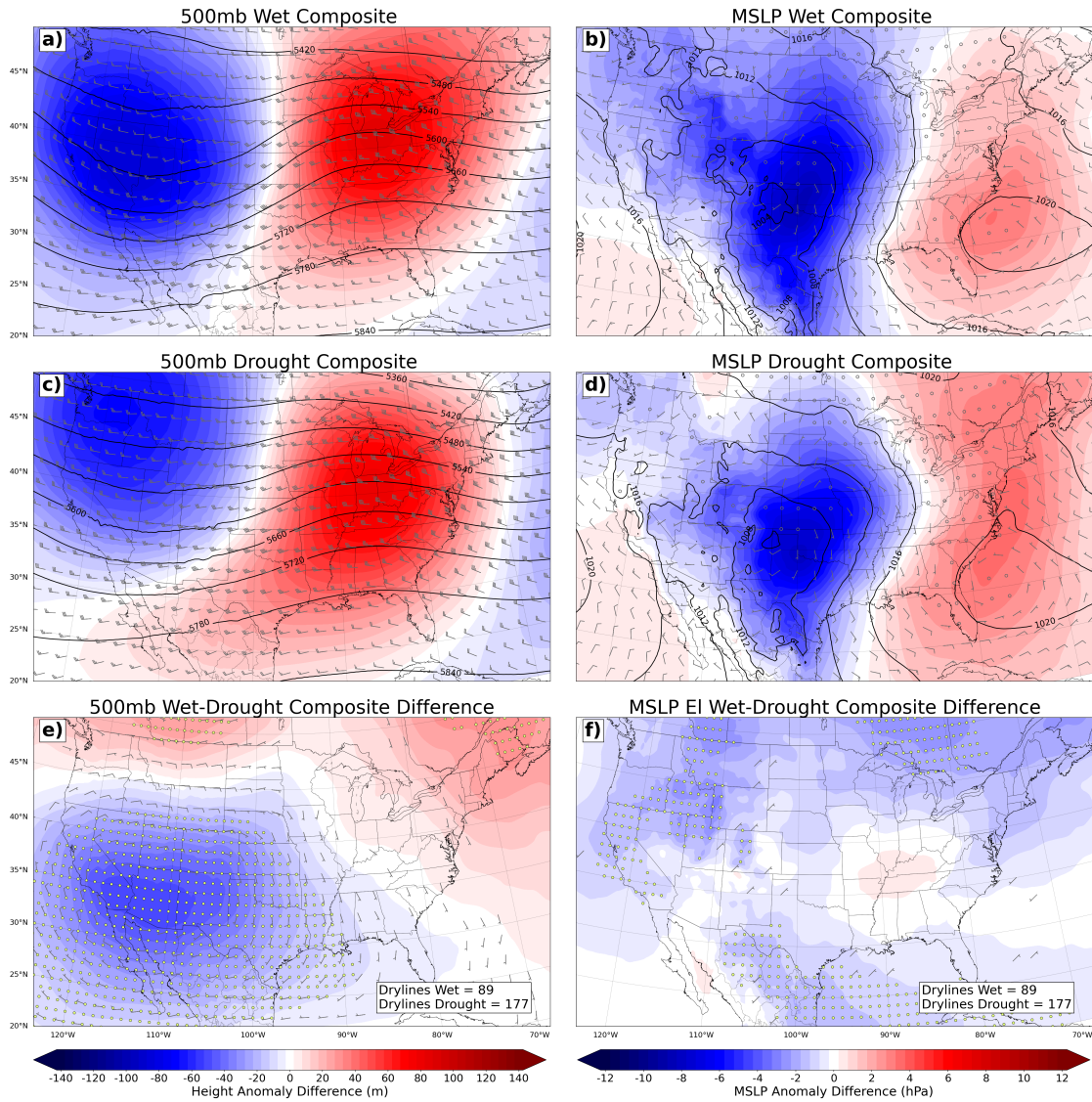


Figure 4.24: As in Fig. 4.23, but composites consist only of dryline days. The number of dryline days that go into each composite are in the lower right corner of (e) and (f).

The environmental composites of April drought and wet condition days (Fig. 4.25) are similar to the March composites (Fig. 4.23). During drought, 500 mb heights are higher over the southwestern US and northern Mexico (Fig. 4.25a,c,e). Unlike March, these higher heights are concentrated in the southern US and do not extend zonally across the entire domain. 500 mb flow in drought has the same pattern as March—during drought, westerly 500 mb flow is faster over the northern SGP, including over the New Mexico and Colorado Rocky Mountains, with an increased westerly component of the flow over much of the SGP (Fig. 4.25e). The response at the surface is more prominent during April than during March. During April drought, MSLP is lower east of the Rockies as well as across most of the central US (Fig. 4.25b,d,f). Westerly 10-m winds over the New Mexico and Colorado Rocky Mountains are faster during drought (Fig. 4.25f). The increased westerly flow both aloft and near the surface is consistent with the conditions favorable for lee cyclogenesis, which would explain the lower MSLP in eastern Colorado and the SGP, and thus explain the increase in dryline frequency during April drought (Chapter 3.1.1).

April

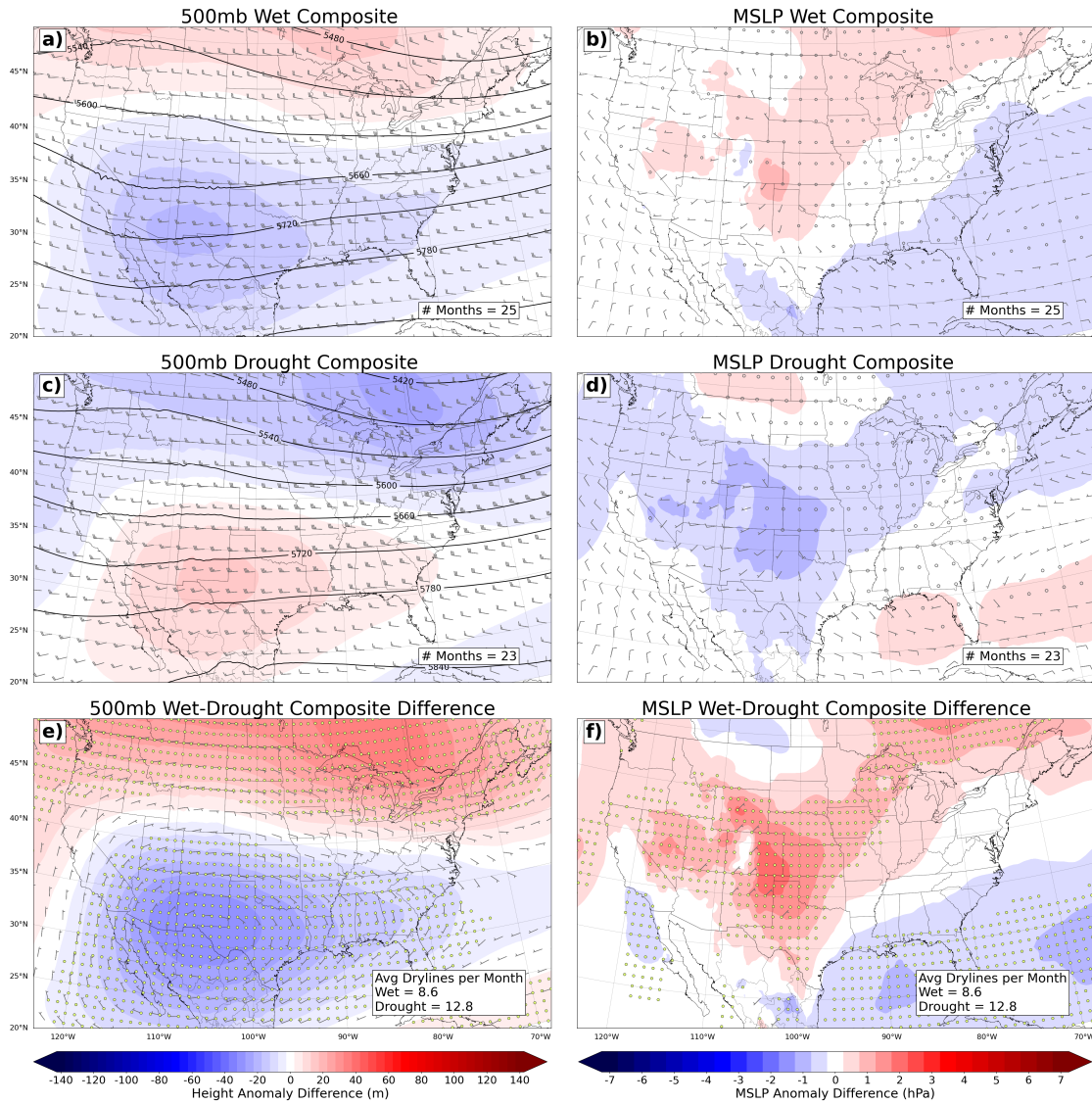


Figure 4.25: Same as Fig. 4.22, but for April.

Environmental composites of April dryline days (Fig. 4.26) are similar to composites of March dryline days (Fig. 4.24). The magnitude of the 500 mb height anomaly of the western US trough is not as large during April (Fig. 4.24a,c vs. (Fig. 4.26a,c). Similar to March, during drought the negative height anomalies associated with the western trough are shifted north, resulting in higher heights over the southwestern US and SGP during drought dryline days compared to wet dryline days and an increased westerly component to 500 mb flow over the SGP during drought (Fig. 4.26e). However, the downstream 500 mb synoptic pattern during April is distinctly different. During March dryline days, there were minimal differences in 500 mb height or flow over the eastern US (Fig. 4.24e). However, during April, drought dryline days feature significantly lower 500 mb height over the northeast US, the result of increased troughing off the east coast (Fig. 4.26c,f). The trough-ridge-trough pattern during drought dryline days could be indicative of a synoptic pattern defined by shorter-wavelength progressive waves. The MSLP pattern marginally supports a hypothesis that SGP extratropical cyclones associated with drylines propagate eastward quicker during drought. During drought, MSLP is marginally lower in the northeast SGP and eastward into the Ohio River Valley (Fig. 4.26b,d,f). Drylines which are attendant to an extratropical cyclone will be advected eastward along with the cyclone until a cold front overtakes the dryline or other processes diffuse the boundary (Chapter 4.1.1). Thus, extratropical cyclones propagating farther east during drought dryline days would explain the $\sim 0.5^\circ$ average eastward shift in dryline longitude during drought compared to wet conditions (Fig. 4.16d). Similar to March, April drought dryline day 10-m winds along the western SGP have a marginally increased westerly component, which could be related to increased westerly flow aloft and indicate a faster propagation speed of the dryline (Fig. 4.26e,f).

April Dryline Days

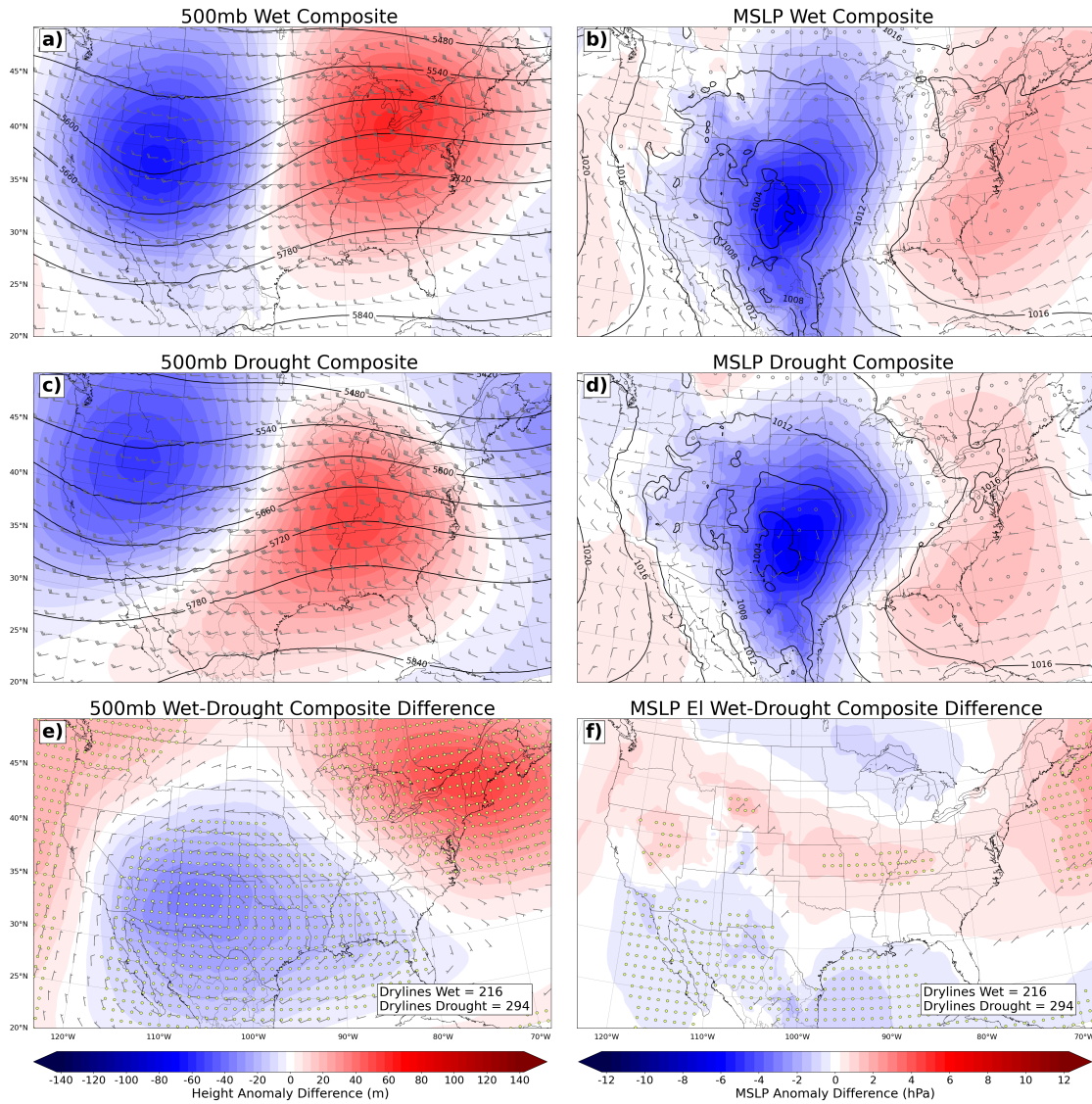


Figure 4.26: As in Fig. 4.24, but for April.

The May environmental composite comparison between drought and wet conditions is similar to the patterns in March and April (Fig. 4.27). For both drought and wet composites, the spacing of 500 mb height contours is wider and westerly flow across the entire domain is weaker compared to previous months, consistent with the seasonal progression of the upper-level jet stream from late winter to May (Fig. 4.27a,c). However, the difference between the wet and drought 500 mb composites is similar to past months, with higher heights over the southern US and lower heights over the northern US contributing to faster westerly flow across the SGP during drought (Fig. 4.27e). The westerly component of 500 mb flow over the New Mexico and Colorado Rocky Mountains is only marginally faster during drought, but the significantly lower MSLP over eastern Colorado indicative of increased and/or stronger lee cyclogenesis events is still evident (Fig. 4.27e,f). Similar to past months, the increase in dryline frequency during May drought can be attributed to the increase in lee cyclogenesis events due to an increase in westerly flow both aloft and at the surface over the Rocky Mountains (Fig. 4.27e,f).

May

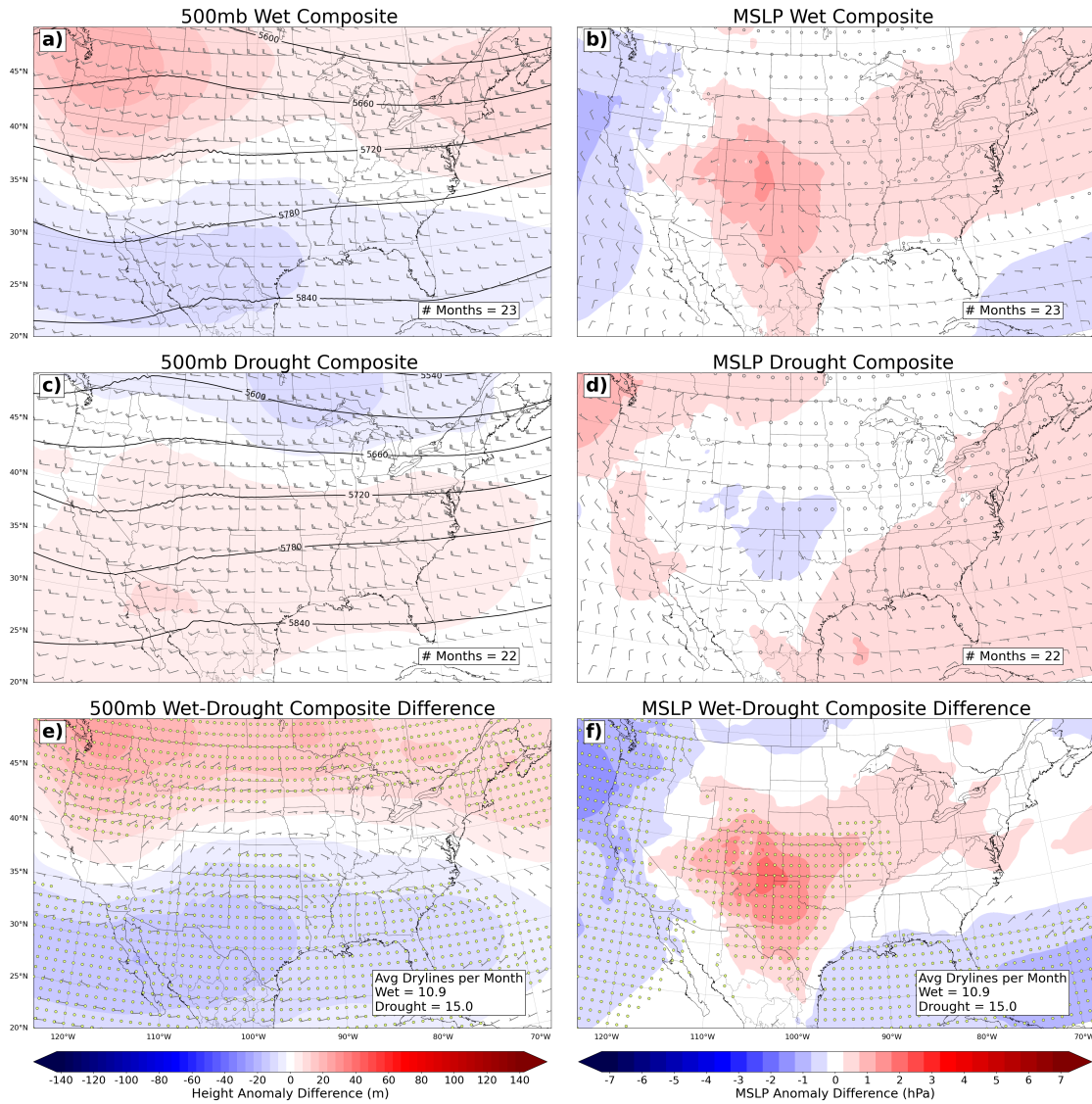


Figure 4.27: Same as Fig. 4.22, but for May.

The 500 mb pattern for May dryline days is similar to April (Fig. 4.28a,c,e). The amplitude of height anomalies is diminished compared to April, but May dryline days during drought are characterized by a western US trough base which is farther north, higher heights over the southwest and SGP, and lower heights off the northeast US. Compared to April, the maximum height increases during drought dryline days compared to wet dryline days is shifted east. It is centered over southwest Texas and eastern New Mexico, compared to central/western New Mexico in April (Fig. 4.28e vs. Fig. 4.26e). Increases in the westward component of westerly flow at 500 mb during drought are over Kansas and the far eastern SGP, displaced from the Texas panhandle and western Oklahoma where drylines are most frequent during May (Fig. 4.20). While there is no difference in 10-m wind between drought and wet dryline days, MSLP across eastern New Mexico/Colorado and the Oklahoma and Texas panhandles is lower during drought (Fig. 4.28b,d,f). There are multiple hypotheses which could explain this: First, lower MSLP during drought indicates that extratropical cyclones in the SGP with attendant drylines are stronger than during wet conditions. This does not directly explain why May drylines are $\sim 0.5^\circ$ farther east on average during drought compared to wet conditions (Fig. 4.16e), but a stronger extratropical cyclone would induce stronger westerly near-surface winds to its southwest, which would assist in the farther eastward propagation of a dryline. However, there is no significant difference in 10-m wind between drought and wet dryline days in the SGP (Fig. 4.28f). A second possible explanation could be that, during wet conditions, there is an increased frequency of synoptically-quiescent drylines. With fewer drylines associated with an extratropical cyclone, the MSLP composite for dryline days would be weaker, as is seen. Mesoscale dryline processes dominate the evolution of synoptically-quiescent drylines (Hane, 2004), which, compared to an extratropical cyclone where advection combines with mesoscale processes to propagate a dryline (Chapter 4.1.1), could result

in synoptically-quiescent drylines not propagating as far east. Further investigation into drought and wet condition dryline days would be necessary to determine the relative rate of synoptically-quiescent versus synoptically-active drylines.

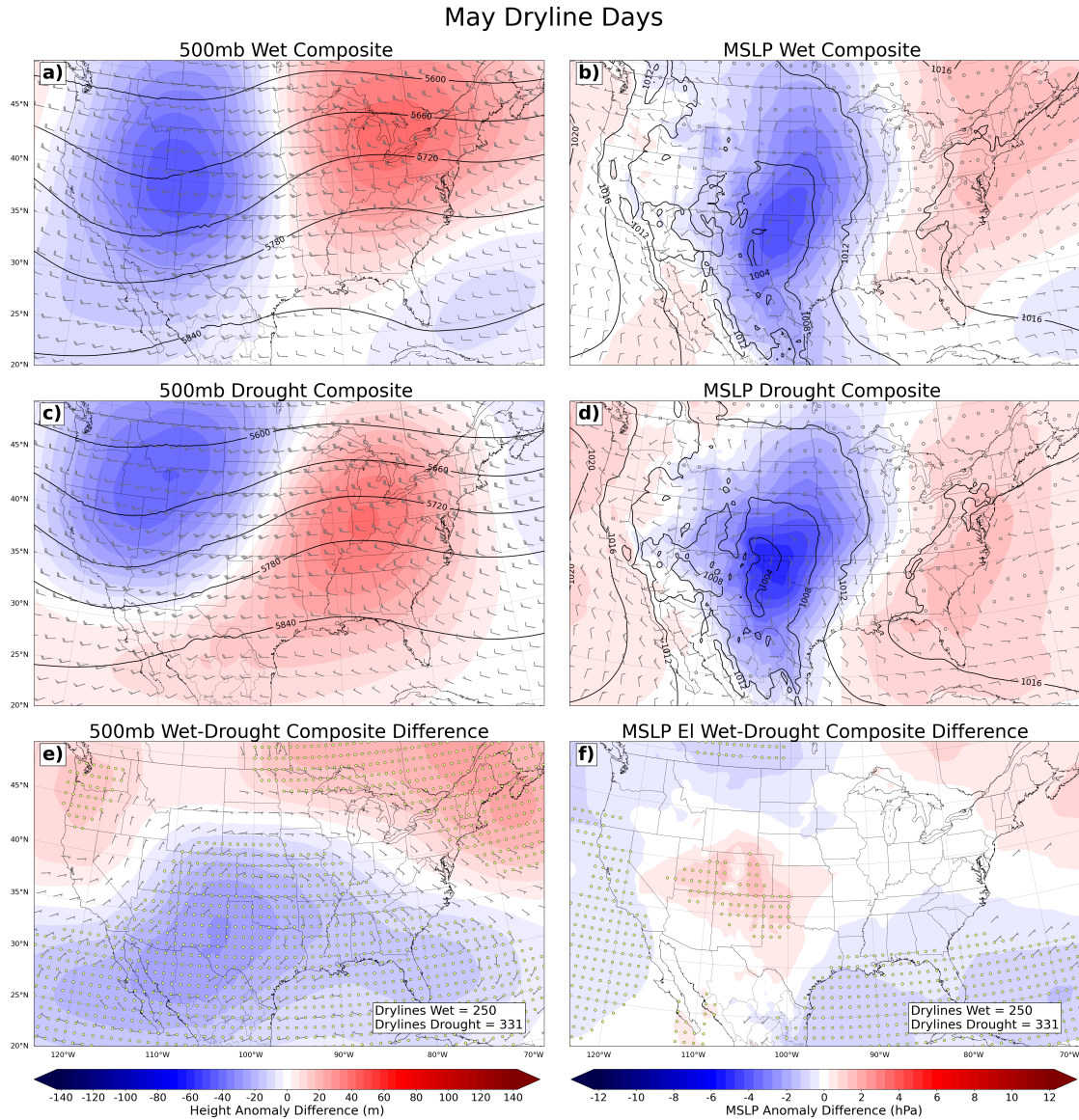


Figure 4.28: Same as Fig. 4.24, but for May.

During drought in June, drylines are on average $\sim 0.5^\circ$ farther east than during wet conditions (Fig. 4.16f), with the greatest difference in dryline distribution being a significant increase in drylines in Kansas during drought compared to wet conditions (Fig. 4.21). While the general synoptic pattern for June dryline days is similar to previous months—a trough over the western US, ridge over the eastern US, and, during drought, a trough off the east coast—the pattern from the difference in wet and drought 500 mb composite is substantially different (Fig. 4.29a,c,e). Rather than a shift in latitude of the western trough that was evident in past months, the difference in 500 mb pattern is longitudinal between June drought and wet conditions (Fig. 4.29e). The eastern US ridge during wet conditions is weaker and shifted to the northeast US, compared to a stronger ridge centered over the Great Lakes region during drought. Negative 500 mb height anomalies associated with the western trough are centered over Wyoming during wet conditions, while the stronger trough during drought has its negative 500 mb height anomalies centered over Idaho. The longitudinal shift in the 500 mb height pattern results in a composite difference that highlights higher heights over the entire central US during drought, with lower heights over the Pacific northwest and off the northeast coast. The westerly flow at 500 mb is slower over the far southern SGP and more southwesterly across central/northern Colorado, but is otherwise unchanged compared to wet conditions over the SGP (Fig. 4.29e).

Differences in June MSLP drought and wet dryline day composites are also unique compared to previous months. During drought dryline days, there is a clear negative MSLP anomaly centered over southeastern Colorado consistent with lee cyclones, similar to previous months (Fig. 4.29d). However, during wet dryline days, negative MSLP anomalies are less concentrated (Fig. 4.29b). The lowest anomalies are in eastern central New Mexico, extending into the Texas panhandle, substantially farther south than the drought dryline day composite. As a result, during drought dryline

days, there is significantly lower MSLP across Colorado, Kansas, Nebraska, Wyoming, South Dakota, and into southeastern Montana (Fig. 4.29f). While an extratropical cyclone center over southeastern Colorado would not be conducive to dryline development over western Kansas—where dryline frequency during drought is increased (Fig. 4.21a)—the northward extension of lower MSLP over the northern Great Plains may be indicative of increased extratropical cyclone frequency across South Dakota and Nebraska, which would be favorable for dryline development in Kansas. Another substantial difference comparing June dryline days to previous months is the widespread higher MSLP over much of the eastern US during drought dryline days (Fig. 4.29f). It is unclear if this directly influences June dryline longitude, but this is consistent with the increased amplitude of the 500 mb trough-ridge pattern during drought compared to wet conditions (Fig. 4.29a,c).

June Dryline Days

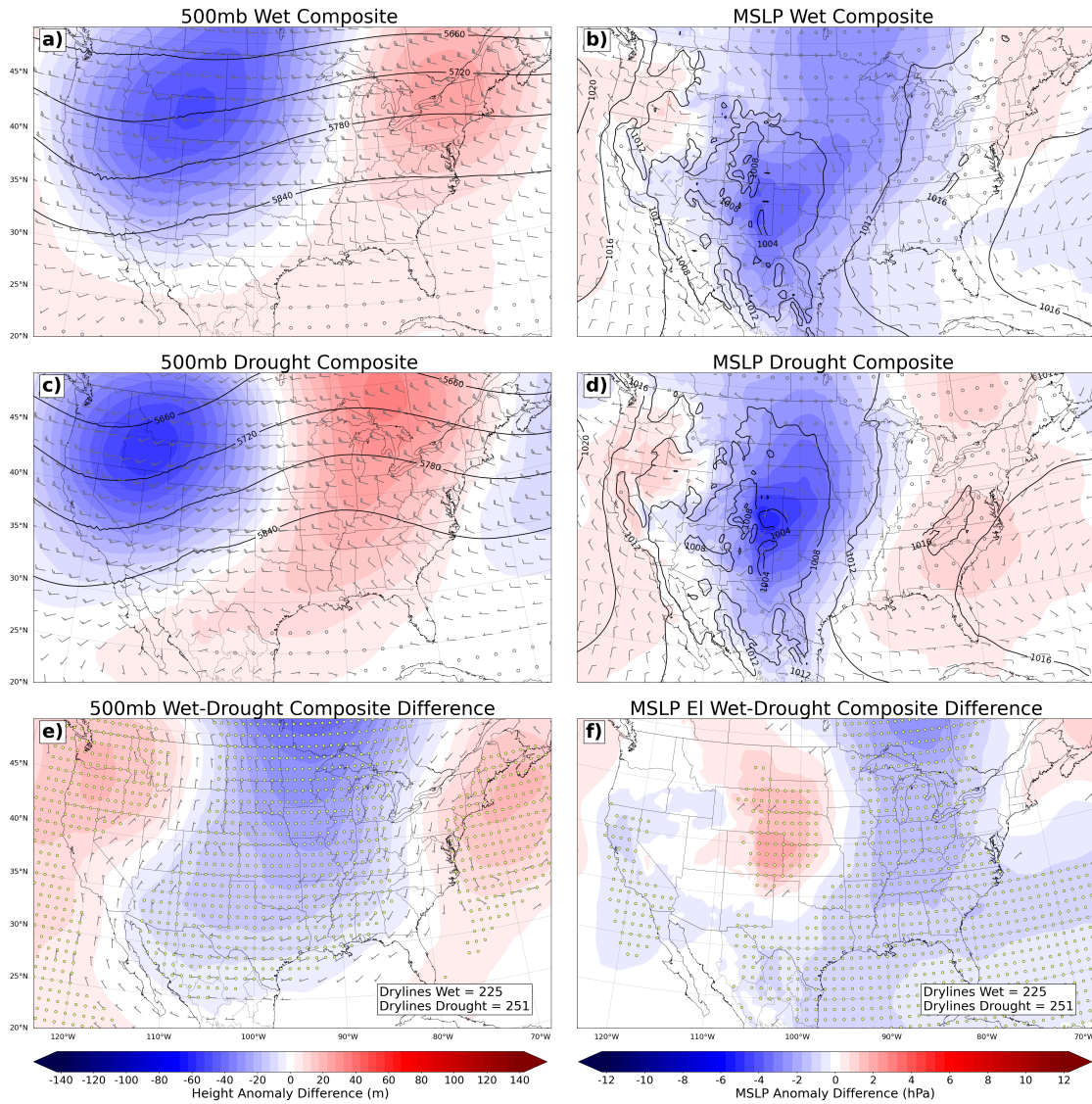


Figure 4.29: Same as Fig. 4.24, but for June.

4.4 Discussion

SGP drought, as defined by averaged PDSI, is correlated with both an increase in dryline frequency (February–May; Fig. 4.15) and an eastward shift in longitude (March–June; Fig. 4.16). Compared to wet conditions, SGP drought months are defined by increased westerly flow aloft over the Rocky Mountains and SGP, lower MSLP in the central US, particularly eastern Colorado, and marginally increased westerly 10-m winds in Colorado and New Mexico (Chapter 4.3.2.3). Lower composite MSLP in eastern Colorado is consistent with an increase in frequency and/or intensity of lee cyclones, which develop in this region downwind of the Rocky Mountains in New Mexico and Colorado (e.g., Smith, 1986; Hobbs et al., 1996). The increased westerly flow—perpendicular to the Rocky Mountains—both aloft and at the surface are factors which favor lee cyclogenesis (Clark, 1990; Karyampudi et al., 1995). The faster westerly winds in the SGP drought composites may also be an indication of an increased frequency of jet streaks, which provide the necessary dynamical trigger for lee cyclogenesis events (Achter and Horn, 1986; Mattocks and Bleck, 1986; Bell and Bosart, 1994; Karyampudi et al., 1995). A subjective analysis of daily 500 mb height and MSLP for 5 years of drought and wet Aprils identified a substantial increase in lee cyclone frequency during drought (not shown). An increase in lee cyclogenesis events during SGP drought would be consistent with both the increase in dryline frequency and eastward shift in longitude seen in Chapters 4.3.2.1 and 4.3.2.2 (Marshment and Horn, 1986; Hobbs et al., 1996; Schultz et al., 2007).

Any causality between SGP drought and the frequency of lee cyclogenesis events is unclear. Although substantial differences in soil moisture can directly influence low-to-mid tropospheric processes like dryline evolution and convection initiation (e.g., Grasso, 2000; henderson et al., 2022), it is unlikely that SGP drought influenced the

location, frequency, and strength of upper air jet streaks over the Rocky Mountains. Rather, the inverse is the more likely hypothesis: drought conditions are present in the SGP due to a synoptic upper-level pattern that both reduces precipitation and also is favorable for lee cyclogenesis events. However, with an increased dryline frequency during SGP drought, and drylines closely associated with severe convection (and thus precipitation) in the SGP (Chapter 3), this hypothesis seems incongruous. To this end, two additional hypotheses emerge: A physical hypothesis could be that while dryline-initiated convection provides significant localized precipitation, the isolated nature of this convection results in an overall soil moisture deficit over the entire SGP region. Additionally, the region west of these more-frequent drylines would face a water budget deficit due to drier near-surface air, less precipitation, and increased solar radiation. Although only a portion of the SGP would experience these drier conditions, the lower PDSI values west of the dryline may be sufficient to drop the SGP-averaged PDSI below -1, given normal or slightly-dry PDSI over the rest of the SGP. An alternative, non-physical hypothesis relates to the nature of PDSI used to define drought: Due to the high autocorrelation of PDSI month-to-month (Chapter 4.2.3), the SGP is classified as “drought” largely due to the influence of factors many months prior to spring, and the increased frequency of drylines (and, possibly, dryline-initiated convection) is not sufficient to remove the SGP region from drought despite a monthly surplus of precipitation. Further work to explore these hypotheses would include compositing monthly precipitation, compositing precipitation across different common synoptic patterns, assessing the amount of convective precipitation between dryline days in drought and wet conditions, identifying large-scale factors which influence the frequency of synoptic factors favorable for lee cyclogenesis events, and investigating how the spatial distribution of PDSI correlates with monthly precipitation. Preliminary composites of 2-m specific humidity indicate drier near-surface conditions across the majority of the SGP

during drought, particularly west of central Oklahoma (not shown). Drier atmospheric conditions during drought could impact the amount of precipitation from a lee cyclone.

ENSO, and its well-documented seasonal influence on the jet stream across the CONUS (e.g., Seager et al., 2003; Allen et al., 2015a; Cook et al., 2017; Schemm et al., 2018), would seem to be a prime candidate to potentially explain the synoptic pattern favorable for lee cyclogenesis during SGP drought. ENSO is correlated to SGP drought both in past literature (e.g., Schubert et al., 2004) as well as with PDSI directly (Fig. 4.1). However, the strength of this correlation is limited. While the majority of wet SGP months occur during El Niño, there is a greater distribution of SGP PDSI values during La Niña (Fig. 4.1). Much of the literature exploring central US droughts have similarly found limitations using solely ENSO, finding greater explained variance in multidecadal SST patterns across the Atlantic and entire Pacific (e.g., Schubert et al., 2004; McCabe et al., 2004).

The limited correlation between ENSO and SGP drought lead to a different relationship between ENSO and drylines. Generally, the correlation between ENSO and drylines is weaker (Figs. 4.3, 4.4). ENSO is only correlated with dryline frequency for February and March, when drylines are more frequent during La Niña. The environmental composites for La Niña are similar for those of SGP drought, highlighting faster westerly wind over the SGP and Rocky Mountains aloft, and lower MSLP in eastern Colorado consistent with an increase in lee cyclogenesis events (Chapter 4.3.1.3).

For April–June, La Niña is correlated with an eastward shift in dryline longitude. This shift is most prominent during May, where La Niña dryline frequency features a dipole of higher frequency east of the central Texas panhandle and lower frequency west (Fig. 4.8b). This frequency dipole may not have been as visible in drought frequency composites due to the underlying increase in dryline frequency (Chapter 4.3.2.2). Environmental composites comparing El Niño and La Niña dryline days with

drought and wet dryline days are distinctly different (Chapter 4.3.1.3 vs. 4.3.2.3). For example, during SGP drought May dryline day 500 mb height is higher across the entire southern US compared to the wet condition composite (Fig. 4.28e). However, May La Niña dryline day 500 mb height is only higher over northern Mexico, and is broadly lower across the east and southeast US associated with a relatively deeper offshore trough (Fig. 4.13e). These differences suggest that ENSO influences drylines in a manner that is unique from the way SGP drought (and factors directly related to it) correlate with drylines.

The increase in February and March dryline frequency during La Niña is consistent with an increase in tornado counts in the SGP compared to El Niño (Cook et al., 2017). Given the substantial role of the dryline in the March SGP climatology (Chapter 3), it is reasonable to hypothesize that dryline-initiated convection is at least in part responsible for the increase in tornado reports in the SGP during La Niña. During MAM La Niña, hail reports increase significantly in north Texas and Oklahoma (Allen et al., 2015a), consistent with the eastward shift of dryline longitude leading to increased dryline frequency in north Texas and Oklahoma (e.g., fig. 4.8). The decrease in El Niño dryline frequency in April and May is not significant, so the decrease in El Niño MAM hail reports may be related to synoptic environment changes less conducive to severe storms or convection initiation, or a relative decrease in the frequency of dryline-initiated convection when the dryline is farther west.

Literature on the relationship between SGP drought—when drylines are more frequent and farther east—and severe weather is scarce compared to the work that has been done understanding the connection of ENSO to severe storms. Galway (1979) found a weak positive correlation between seasonal precipitation and tornado frequency in Oklahoma and Kansas. However, they also found that out of three years in their sample considered to be extremely dry, two of them featured above-normal tornado

activity. Their methodology of defining drought via seasonal precipitation anomaly is not analogous to the long response time of PDSI to seasonal precipitation surplus/deficits. Shepherd et al. (2009) used a similar precipitation anomaly methodology to define drought, but correlated fall–winter drought to spring tornadoes, which may be more comparable to the spring PDSI used in this Chapter. While they found a robust decrease in spring tornado frequency when the previous fall–winter is characterized by drought, their study focused on the southeast US, which may not be analogous to the SGP. Drylines have been attributed to significant southeast US severe outbreaks (e.g., April 26–27 2011; Chasteen and Koch, 2022), but the overall frequency of drylines in the southeast is substantially lower than the SGP (Chapter 2), and the dryline likely does not play as significant of a role in the annual severe climatology of the southeast. Additionally, both Galway (1979) and Shepherd et al. (2009) focus solely on tornadoes, which may further limit comparisons between their findings and inferences from Chapters 3 and 4, where all severe convective hazards are considered.

Given the substantial role of the dryline in the SGP annual severe climatology (Chapter 3), it is reasonable to hypothesize that the increase in dryline frequency during SGP drought could result in an increase in SGP severe storms. However, this hypothesis might be sensitive to the definition of drought employed (PDSI vs seasonal/monthly precipitation anomaly). Future work to test this hypothesis would include quantifying the role of the dryline in the severe storm climatology during drought years specifically, finding the difference in dryline-initiation frequency between drought and non-drought years, and testing the sensitivity of these results, along with the results from this Chapter, to varying methodologies of defining drought.

4.5 Conclusions

Drylines are identified daily at 0000 UTC from 1940–2020 using ERA5 reanalysis and a machine-learning model. Dryline characteristics—their frequency and median longitude—are compared across different ENSO phases and drought classification for the SGP monthly for February–June. The spatial distribution of dryline frequency anomaly for each month is used to evaluate the statistical relationships identified in the distributions. Monthly environmental composites of 500 mb height and wind, mean sea level pressure, and 10-m wind are used to identify physical synoptic-scale factors which may explain the statistical and spatial variations in dryline characteristics between ENSO phases and drought classifications. The conclusions to the three hypotheses stated in Chapter 4.1.3 can be summarized with the following:

1. During La Niña, drylines are more frequent in February and March, and are shifted farther east during April–June. Lower MSLP in eastern Colorado supports the hypothesis that this difference is the result of increased lee cyclone frequency, along with an increase in extratropical cyclones in the northern Great Plains during June.
2. Drylines are more frequent and farther east during SGP drought. However, synoptic factors favoring an increased frequency in lee cyclogenesis contribute at least partially to this result. The effect of soil moisture on dryline propagation, not explicitly tested in this Chapter, may still have a role, but its relative influence compared to synoptic factors is unknown.
3. For ENSO, while dryline frequency is only impacted early in the season (February and March), dryline longitude is significantly affected through June. SGP drought influences both dryline frequency and longitude throughout the entire

season, with the exception of longitude in February and frequency in June. While the eastward shift in dryline longitude is similar across months, during June both La Niña and SGP drought most prominently result in a northward shift in dryline frequency.

These results can be combined with the quantification of the dryline’s role in the SGP severe storm climatology (Chapter 3) to create hypotheses about how ENSO and drought influence severe storm frequency in the SGP. During La Niña, increases in dryline frequency early in the season (February and March) along with an eastward shift in dryline frequency in April and May are consistent with increases in hail and tornado frequency in the SGP (Allen et al., 2015a; Cook et al., 2017). The effect of drought on the SGP severe storm climatology is less clear, but the increase in dryline frequency and eastward shift should influence at least the location of severe storms within the SGP. Further work will be needed to parse the relationship between drought and dryline-initiated severe storms.

There are several limitations to the methodologies employed in this Chapter. The identification of drylines uses a ML model trained on WPC surface analysis maps from 2007–2020. While it is demonstrated in Chapter 2 that this ML model resolves a non-physical increase in WPC-identified dryline frequency across that period, other non-physical attributes arising from the inherently subjective nature of boundary identification may still be present within the model, such as a bias against dryline identification in the Mississippi Valley (Duell and Van Den Broeke, 2016). The drylines identified by the ML model represent a consistent application of criteria from a variety of WPC forecasters.

The reanalysis data this ML is applied to also has its own limitations. While ERA5 currently represents state-of-the-art reanalysis extending back to 1940, its biases may affect both the identification of drylines and interpretation of environmental compos-

ites. For example, in a comparison of ERA5 with observations from 2005–2019, Chen et al. (2024) found that while 10-m wind is represented well, ERA5 tends to overestimate low wind speeds and underestimate high winds speeds. These errors were higher over the complex terrain of the Rocky Mountains and during the warm season, when prevalent small-scale convective processes may not be represented well. Without a truth to compare to, it is unknown whether conclusions drawn from the 10-m wind composites in Chapters 4.3.1.3 and 4.3.2.3 are affected by these biases. Sensitivity tests using other reanalysis models (which may have similar biases of their own) and comparisons to observations (which have their own spatial and temporal limitations, in addition to instrument errors) may be useful in evaluating the sensitivity of these conclusions to reanalysis biases.

There are also limitations in the interpretation of environmental composites. As they represent the mean state, differences such as the decreased MSLP in eastern Colorado during drought could be the result of deeper lee cyclones or a better collocation of lee cyclones during drought, rather than an increase in lee cyclones. While a manual analysis of 5 years of drought and wet April’s showed an increase in subjectively-identified lee cyclones during drought (not shown), applying a manual approach to the entire dataset is prohibitive. An analysis method such as Self-Organizing Maps (SOMs; Kohonen, 2001) would be useful in identifying the frequency of days classified into nodes with a synoptic pattern consistent of a lee cyclone. SOMs could also be useful in classifying drylines as synoptically-quiescent or synoptic-active, as defined by Hane (2004).

The strengths and limitations of PDSI have been discussed at length in Chapters 4.2.3 and 4.4. One further limitation of the drought classification used in this Chapter not yet discussed is the use of SGP-average PDSI. PDSI can vary considerably across the 16 grid cells spanning the SGP. For example, the SGP average PDSI for April 1946

is -0.43, or “normal”. However, 8 of the grid cells have a PDSI below -1 (“drought”), including 7 of the 8 grid cells west of 100°W (not shown). The average PDSI is raised by three grid cells in the far eastern SGP with PDSI exceeding 2. Since the longitudinal gradient of soil moisture is a significant factor in dryline development (e.g., Grasso, 2000), the dryline characteristics of April 1946 may be significantly different from April 1976, another “normal” month (average PDSI of 0.6; not shown), but where 3 of the 8 grid cells west of 100°W have a PDSI near 2, and only 1 of the 8 grid cells has a PDSI below -1. SOMs of the PDSI maps could be used to group years with more physical consistency.

Despite the above limitations, this Chapter has demonstrated relationships between dryline frequency and longitude with both ENSO and SGP drought. Future work could create distinct climatologies for synoptically-active and synoptically-quiescent drylines, explore how ENSO and drought relationships influence the frequency and location of severe storms, and explore the utility of S2S dryline frequency/longitude predictions as a method of S2S severe storm forecasting in the SGP.

Chapter 5

Quantifying the Value of Extended Range Convection Allowing Global-Nested Ensembles for Severe Thunderstorm Forecasting

5.1 Motivation and Introduction

5.1.1 Motivation

The previous chapters of this dissertation focus on developing the climatology of dryline-associated severe convection and the subseasonal-to-seasonal relationship between large-scale factors and dryline characteristics. However, those chapters only indirectly address the underlying motivation of this dissertation to explore avenues for improvement in severe convection forecasts. To that end, this chapter presents the early phase of work to explore whether experimental Convection-allowing model (CAM) ensembles improve severe convection forecasts at extended ranges compared to traditional convection-parameterized ensembles. Due to time constraints, this chapter primarily addresses general severe convection forecasting skill. It will be left to

future work to explore the difference in predictability of drylines—and how dryline predictability translates to severe convection predictability—between the two experimental extended-range ensembles compared in this Chapter.

5.1.2 Introduction

CAMs are high-resolution forecast models that are able to explicitly depict individual convective storms and organized storm systems such as Mesoscale Convective Systems (MCSs). Although CAMs do not explicitly simulate the occurrence of severe convective hazards—tornadoes, hail greater than 1 inch in diameter, or wind gusts greater than 50 knots and/or thunderstorm-produced damage—CAMs are able to improve operational forecasting of these hazards through their more detailed depiction of storm timing, mode, intensity, and evolution (e.g., Done et al., 2004; Kain et al., 2006; Gallus et al., 2008; Duda and Gallus, 2010; Coniglio et al., 2010; Smith et al., 2012; Thompson et al., 2012) as well as storm attributes (e.g., updraft helicity; Kain et al., 2008, 2010; Sobash et al., 2011, 2016a,b). CAMs supply additional information to the forecast process that convection-parameterized models, which do not explicitly resolve convection, cannot provide. Using only convection-parameterized models, severe thunderstorm forecasting is largely based on ingredients-based methodologies incorporating environmental variables and synoptic-scale pattern recognition and climatology (Johns and Doswell III, 1992; Doswell III et al., 1993). The chaotic nature of convection and the two-way interactions between convective storms and their environments are not explicitly depicted in convection-parameterized models, placing additional limits on their ability to predict the details of deep convection beyond their parameterization scheme (e.g., Lilly, 1990; Baldwin et al., 2002; Wilson and Roberts, 2006; Weisman et al., 2008).

Despite the improvements that CAMs can offer severe thunderstorm forecasting, these models are currently limited operationally to the short-term (<60 hours) primar-

ily due to high computational cost. The computational cost of medium- (days 4-8) and long-range (days 8+) forecasts scales exponentially due to both forecast length and the necessity of employing an ensemble. Deterministic CAM runs have inherent limited predictability at increasing lead time associated with the growth of small-scale error propagation over time (Zhang et al., 2003, 2007). A CAM ensemble is more useful to a severe weather forecast because it provides a probabilistic framework to quantitatively assess uncertainty (Gneiting and Raftery, 2005; Xue et al., 2007). Operational model ensembles of the contiguous United States (CONUS) at the medium- and long-range are currently limited to convection-parameterized models. But with continued improvements in computing, a long-range ensemble of CAMs is becoming more feasible (Harris et al., 2020). In this Chapter, we seek to quantify how much, if at all, a CAM ensemble in the medium- and long-range can improve severe thunderstorm forecasts generated from machine learning compared to a typical convection-parameterized ensemble system. Our comparison uses two ensembles from the NOAA Geophysical Fluid Dynamics Laboratory (GFDL)—a global convection-parameterized ensemble (Global), and an identical global ensemble with a 2-way nested CAM over the CONUS (Global-Nest; Harris et al., 2020). Both ensembles are detailed further in Chapter 5.2.1.

Since neither CAMs nor convection-parameterized models explicitly forecast severe convective hazards, we evaluate the performance of a random forest (RF; Breiman, 2001) designed to produce daily probabilistic forecasts of severe convection using ensemble output. Machine learning and artificial intelligence (ML/AI) techniques such as RFs can synthesize large amounts of atmospheric information and identify complex nonlinear relationships between numerous variables and occurrences of severe convective hazards (e.g., Lagerquist et al., 2017; Sobash et al., 2020; Burke et al., 2020; Flora et al., 2021; McGovern et al., 2023). ML/AI techniques have been developed to forecast severe convective hazards for both short lead times using storm attribute information

from CAMs (e.g., Loken et al., 2020, 2022), as well as extended-range forecasts using solely environmental information (e.g., Hill et al., 2023; Clark et al., 2025). When properly developed, ML/AI algorithms can both objectively and subjectively perform competitively with human-generated severe convection forecasts (e.g., Loken et al., 2020; Hill et al., 2023; Clark et al., 2023a).

The use of RFs to create severe convection forecasts enables an objective comparison of performance between the Global and Global-Nest ensembles. Additionally, the contributions from individual variables to each RF are compared using explainable-AI techniques to identify which variables are the most important to the RF forecast (e.g., Loken et al., 2022). The difference in relative importance of predictors between the Global RF and Global-Nest RF reflects the impact of including storm attributes, as well as any indirect effects of environmental modification from simulated convection.

This Chapter also provides an opportunity to explore ensemble skill forecasting drylines by applying the FrontFinder AI (Justin et al., 2025) to ensemble members. Given the significant role of the dryline in springtime severe convection in the Southern Great Plains (Chapter 3), the ability of a forecast model to accurately depict the frequency and location of the dryline should correlate with its RF severe thunderstorm forecasting skill. Preliminary work evaluating dryline forecasts is presented in Chapter 5.3.4.

5.2 Data and Methods

5.2.1 Ensemble Forecasts

Dynamical forecast model data are from the GFDL System for High-resolution prediction on Earth-to-Local Domains for the contiguous United States (C-SHiELD; Harris

et al., 2019, 2020), which is based on the Finite-Volume Cubed-Sphere Dynamical Core (FV3; Putman and Lin, 2007). Details of the ensemble configuration can be found in Table 5.1. Two separate 11-member ensembles with this same core are compared: Global and Global-Nest. Both ensembles have a horizontal grid spacing of 16-km spanning the entire globe. The Global-Nest additionally includes a 2-way nested CAM over the CONUS. The nested CAM has 5-km grid spacing, and the deep convection parameterization from the Arakawa-Schubert scheme is turned off (the shallow convection parameterization remains active). In addition to influencing environmental fields through its 2-way nest, the Global-Nest also outputs hourly-maximum storm attribute fields of 2–5 km updraft helicity (UH) and maximum composite reflectivity, as well as 10-m wind speed (Kain et al., 2010). UH is the only storm attribute used in this study, as the inclusion of reflectivity and 10-m wind speed do not meaningfully change RF performance over the full sample of cases (not shown).

The Global ensemble produces output at 6-hour increments (1200, 1800, 0000, 0600 UTC). The Global forecasts serve as lateral boundary conditions for the 5-km nest, which produces output hourly. For consistency, Global-Nest forecast output is only considered at the same 6-hour increments as the Global. Within each individual ensemble member, environmental fields are representative at the respective forecast time. The hourly-maximum UH field from the nest is post-processed to be a 6-hourly maximum. The UH fields thus represent the maximum value over the previous 6 hours. For example, in the forecast valid at 1800 UTC, the UH of a given grid cell is the hourly-maximum UH within that cell between 1200 and 1759 UTC.

Forecasts are run out for 720 hours and are initialized at 0000 UTC on the 1st, 8th, 15th, and 22nd of each month from October–May, as well as June 1st, for 2016–2020, for a total of 153 forecast initializations. Since forecasts are initialized once per week, the sample of events for each individual lead day is different (e.g., the day 3 sample

Model Parameter	Configuration
Initial Conditions	Global Ensemble Forecast System (GEFS) v11 (Zhou et al., 2017)
Horizontal Grid spacing	16 km (5 km two-way CONUS nest) (Harris and Lin, 2013)
Vertical Levels	63—identical to Global Forecast System (GFS) v15
Microphysics	Inline GFDL (Harris et al., 2020)
PBL Scheme	Yonsei University PBL scheme (Hong, 2010)
Radiation	Rapid Radiative Transfer Model for GCMs (Clough et al., 2005)
Land surface	Noah land surface model (Ek et al., 2003)
Ocean	Mixed-layer ocean model (Pollard et al., 1973)
Convection parameterization	Scale-aware simplified Arakawa-Schubert scheme (Han et al., 2017)
Dynamical Core	Finite-Volume Cubed-Sphere (FV3; Putman and Lin, 2007)

Table 5.1: C-SHiELD ensemble configurations

includes the 3rd, 10th, 17th, and 24th of each month while the day 4 sample includes the 4th, 11th, 18th, and 25th of each month). Forecasts are run retrospectively using the first 11 ensemble members of the Global Ensemble Forecast System (GEFS) as initial conditions. Despite having forecast output through day 30, our analysis only extends out through day 15, as daily RF forecast skill and ensemble forecast skill by day 15 is minimal both in past studies (Clark et al., 2025) and in our verification (Chapters 5.3.1, 5.3.2).

Computational limitations were important considerations in the design of this C-SHiELD configuration and the choices for forecast initialization. The nested CAM has coarser horizontal grid spacing than typical operational CAMs (5-km vs 3-km or less)

in an effort to reduce computational costs while retaining enough horizontal resolution to adequately depict the bulk properties of convection. Forecasts are only run for the cool season because ML forecasts generated from dynamical forecast ensembles have demonstrated low skill in severe forecasting at extended lead times during the summer (e.g., Clark et al., 2025), when severe events are less likely to be associated with strong synoptic-scale systems compared to the cool season (e.g., Brooks, 2009). Initializing forecasts 7 days apart 4 times a month saves considerable computational power compared to initializing forecasts daily. While this precludes an analysis of the evolution of the forecast as lead time decreases ($d\text{prog}/dt$: Hamill, 2003), daily initializations do not functionally increase the sample size of severe events.

Ensemble output is post-processed similar to Clark et al. (2025). Commonly-used fields and indices such as bulk shear and the significant tornado parameter (STP; Thompson et al., 2003) are computed using native model fields and pressure level data. Predictors are interpolated to a subset of the NOAA National Centers for Environmental Prediction (NCEP) 211 grid (NCEP, 2024) with approximately 81 km grid spacing encompassing the CONUS (Fig. 5.1). All predictors are bilinearly interpolated with the exception of precipitation, which is conservatively interpolated in order to ensure the total value remains the same as it was prior to regridding (Jones, 1999), and UH, where the maximum UH value within the ~ 81 km grid is used. Our domain is consistent with the Storm Prediction Center (SPC) definition of severe weather in their forecasts (i.e., within a 40 km radius of any point) and the smaller number of grid points reduces computational costs (Clark et al., 2025). Grid cells outside the CONUS are masked.

UH is a measure of rotation in the simulated storm updraft (e.g., the mesocyclone), and has been demonstrated to be a useful proxy for the severe potential of a convective storm (Kain et al., 2008; Sobash et al., 2011). Unlike environmental variables like

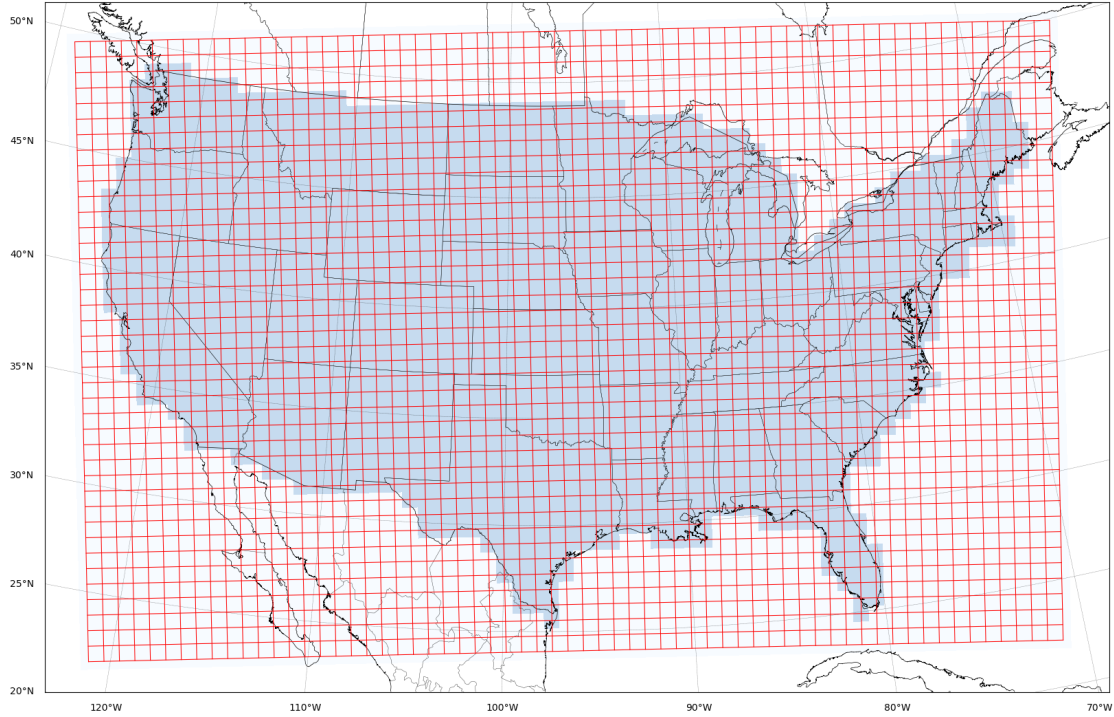


Figure 5.1: The red grid denotes the subset of the NCEP 211 grid model data are interpolated to. The blue shading is the CONUS mask used in the random forest algorithm.

temperature, an ensemble mean of 6-hourly maximum UH without additional context does a poor job conveying the spread of individual ensemble members. For example, a low ensemble mean UH at a particular grid point can occur when either 1) only a few members depict a high-UH storm at that location, 2) many members depict a low-UH storm at that location, or 3) many members depict high-UH storms in the general region, but due to ensemble spread only a few depict the storm at that exact given grid point. The interpretation of CAM ensemble output varies considerably depending on which of these three scenarios led to the output ensemble mean UH. Therefore, to better leverage the information provided by the UH ensemble, six smoothed Neighborhood Maximum Ensemble Probability (NMEP; Schwartz and Sobash, 2017) fields are calculated from the 6-hourly maximum UH. The RFs trained with NMEP UH outperformed RFs trained with ensemble mean UH or ensemble max UH over the entire

sample.

The NMEP field represents the fraction of ensemble members with UH exceeding a given threshold (e.g., $85 \text{ m}^2\text{s}^{-2}$) at each individual grid cell. These fields are then smoothed using a Gaussian smoother with $\sigma=120$ km. The 6 UH thresholds chosen for the NMEP fields are the 99.5th, 99.7th, 99.8th, 99.9th, 99.95th, and 99.99th percentiles of all daily maximum forecasted UH on the 5-km grid for each day of all 153 forecast initializations. The use of UH percentiles rather than fixed values is motivated by substantial variations in UH climatologies between different models (Sobash et al., 2016b; Potvin et al., 2019; Gallo et al., 2020). These high thresholds represent when a simulated storm may be most likely to be severe and cover a wide range of simulated UH values from $20 \text{ m}^2\text{s}^{-2}$ to $150 \text{ m}^2\text{s}^{-2}$.

5.2.2 Random Forest Classifier

Severe convection probabilities are calculated using RF classifiers from the Python module Scikit-Learn (Pedregosa et al., 2011). RF hyperparameters are set identical to Clark et al. (2025): 200 trees, maximum tree depth of 15 levels, minimum of 20 samples per leaf node, minimization of entropy for splits, and the square root of the total number of predictors is evaluated at each node (Loken et al., 2019, 2020; Clark and Loken, 2022). Past studies with similar RFs, albeit only forecasting day 2 severe weather, have found minimal sensitivity to hyperparameter tuning (Loken et al., 2019), so no additional tuning is performed here (Clark and Loken, 2022; Clark et al., 2025). Severe storm reports are used to classify each domain grid cell as severe/non-severe based on whether one or more severe reports (tornado, hail, or wind) are recorded at any time within that grid cell during the SPC convective day (1200–1159 UTC).

A combination of native and post-processed fields are used as predictor variables (Table 5.2). These predictors are computed on both the Global and Global-Nest. With

the exception of the UH NMEP fields and latitude/longitude, the ensemble mean is used for all predictors. Environmental predictors are the instantaneous values valid at the forecast output time with the exception of precipitation amount, which is the total accumulated precipitation over the previous 6 hours.

Kinematic	Thermodynamic	Other	Updraft Helicity
0–3 km storm-relative helicity (m^2s^{-2})	surface-based CAPE (Jkg^{-1})	precipitation amount (kgm^{-2})	UH 150 m^2s^{-2} NMEP
0–1 km bulk shear (ms^{-1})	surface-based CIN (Jkg^{-1})	mean sea-level pressure (Pa)	UH 85 m^2s^{-2} NMEP
0–3 km bulk shear (ms^{-1})	surface-based LCL (m)	Fixed-layer significant tornado parameter (unitless)	UH 60 m^2s^{-2} NMEP
0–6 km bulk shear (ms^{-1})	most-unstable CAPE (Jkg^{-1})	latitude ($^{\circ}$)	UH 40 m^2s^{-2} NMEP
10 m u-wind (ms^{-1})	2 m temperature (K)	longitude ($^{\circ}$)	UH 30 m^2s^{-2} NMEP
10 m v-wind (ms^{-1})	2 m water vapor mixing ratio (gkg^{-1})		UH 20 m^2s^{-2} NMEP
10 m wind (ms^{-1})	700 mb–500 mb lapse rate (Ckm^{-1})		

Table 5.2: List of all predictors, grouped in columns by category (top row). UH fields are only included for the Global-Nest RF. Predictors represent ensemble means at each grid point with the exception of UH NMEP and latitude/longitude.

The RFs are trained to produce daily severe convection forecasts out to day 15 using ensemble data from each convective day. In the Global RF, there are 76 total predictors (19 environmental variables $\times$ 4 output times per day), and in the Global-Nest RF there are 100 total predictors (19 environmental variables plus 6 UH NMEP fields $\times$ 4 output times per day). Predictions are spatially independent with no knowledge of the predictors (or final prediction) at adjacent grid points. A RF is trained for each lead day, for a total of 30 independent RFs—two (Global and Global-Nest) for each lead day 1–15. 9-fold cross-validation with randomized training and testing sets by forecast initialization date is used (e.g., Platt, 1999). In each fold, 136 forecasts (188,360 individual grid points) are used for training while the remaining 17 (23,545 individual grid points) are used for testing. Due to the limited number of forecast initializations,

there are not separate validation and testing datasets. The same set of training and testing forecast initializations are used for both the Global and Global-Nest ensembles.

5.2.3 Explainable AI with TreeInterpreter package

A benefit of RFs is that the contribution of individual variables to the output can be easily determined in a human-understandable way. For all generated forecasts, the Python package TreeInterpreter is used to decompose each forecast into the contributions from each individual variable (Saabas, 2015). For each forecast at each grid point, the contributions from each variable are returned in the form

$$forecast = bias + var_1 + var_2 + \dots + var_n \quad (5.1)$$

where $var_1, \dots, var_n$ represent the individual contributions from all n predictor variables and $bias$ represents the mean of the training set. Individual variable contributions can be positive or negative. For each testing forecast, TreeInterpreter sums each predictor’s contributions across all nodes. For further discussion of RF interpretability with TreeInterpreter, see Loken et al. (2022).

For each predictor, the mean contribution to severe probabilities is found by summing the absolute value of contributions from each forecast grid point across each forecast for a given lead day (Loken et al., 2022; Mazurek et al., 2025). The contributions from each predictor are then summed together to create a dimensionless quantity which is used to determine what fraction of the overall contribution is from each individual variable. Predictors are categorized as in Table 5.2, with “kinematic”, “thermodynamic”, and “UH” contributions represented as both an average of the relative contribution from individual variables (Chapter 5.3.3) and as the sum of relative contributions from each individual variable in that category for individual case studies

(Chapter 5.4). Predictors categorized as “other” are represented as individual variables, aside from latitude and longitude which are combined akin to the groupings above.

TreeInterpreter is one of several explainable-AI techniques used to interpret machine learning models (Flora et al., 2024). TreeInterpreter shares similarities with a more generalizable technique known as Shapely additive explanations (SHAP; Shapley, 1953; Molnar, 2022). Both techniques enable a local explainability of how much individual predictors are contributing to an output probability, but unlike TreeInterpreter, SHAP is generalizable to ML methods beyond RFs (Flora et al., 2024). However, TreeInterpreter is much less computationally expensive to run, and so it is the technique of choice in this study. For further discussion of explainable-AI techniques see Flora et al. (2024).

5.2.4 Objective Verification Metrics

Similar to the evaluation in Clark et al. (2025), the RF severe convection forecasts are objectively evaluated using attributes diagrams (Hsu and Murphy, 1986), Brier Skill Score (BSS; e.g., Wilks, 2011), and the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC; Mason, 1982). Attribute diagrams compare binned forecast probabilities to the observed frequency of storm reports within a given bin. Our attributes diagrams are constructed with probability bins of [0%–5%), [5%–10%) ... and [95%–100%]. Any bin with less than 20 forecasts is excluded to reduce excess noisiness associated with small sample sizes. A 1991–2020 30-year 3-day spatially-varying rolling climatology of severe storm reports is used to determine the threshold for forecast skill.

BSS is determined by comparing the Brier Score (BS) of a forecast to the BS of a reference forecast (BS_{ref}) in the form

$$BSS = 1 - \frac{BS}{BS_{ref}} \quad (5.2)$$

(Wilks, 2011). BSS has a range of $-\infty$ to 1, with 1 being a perfect forecast and 0 being no skill relative to the reference forecast—the 30-year climatology of severe storm reports. BS is commonly decomposed into three components: reliability (BS_{rel}), resolution (BS_{res}), and uncertainty (UNC), which can be written as

$$BS = BS_{rel} - BS_{res} + UNC \quad (5.3)$$

(Murphy, 1973). Reliability is a measure of how well the forecasted probability of severe convection aligns with the observed frequency. For a forecast with perfect reliability, across all instances of forecasts of 30% probability, severe convection would be observed 30% of the time. Resolution is a measure of how different the observed frequency of severe convection for a given forecast is compared to the underlying climatology (Murphy, 1993). For a forecast with perfect resolution, the observed frequency of severe convection for a given forecast would be 100%—as far away from the overall climatology of severe convection ($\sim 0.5\%$) as possible. When the reference forecast is climatology, BS_{ref} simplifies because both resolution and reliability terms are 0 for forecasts of climatology. Thus, eqn. (5.2) can be written simply in the terms of the BS decomposition from eqn. (5.3) as

$$BSS = \frac{BS_{rel} - BS_{res}}{UNC} \quad (5.4)$$

(Mason, 2004; Wilks, 2011). In addition to evaluating the BSS of the Global and Global-Nest RFs, we also directly compare the resolution and reliability of both sets of RF forecasts.

AUC measures the forecasts' ability to discriminate between events and non-events.

It is computed by calculating the probability of detection (POD) versus the probability of false detection (POFD) for given forecast probability thresholds. AUC is the area under the curve connecting the POD-POFD curve following the trapezoidal approximation described in Wandishin et al. (2001). AUC ranges from 0 (worst) to 1 (best), with 0.5 representing the line of no skill, and 0.7 considered the lower limit for a useful forecast (e.g., Buizza et al., 1999). The same probability thresholds as the attributes diagrams are used in the calculation of AUC.

For both BSS and AUC, forecast day averages are computed by aggregation (e.g., summing the elements that make up each measure for all forecasts) rather than averaging the individual score for each of the 153 forecasts. Statistical significance of the difference between the Global and Global-Nest RF BSS and AUC is assessed via a resampling test with 10,000 iterations. Forecast resampling is performed following the recommendations of Hamill (1999). After randomizing the 153 forecast initializations, BSS and AUC are re-calculated by aggregation, and the difference in BSS and AUC between the Global-Nest and Global RFs is calculated. This step is repeated 10,000 times for each lead day to create a randomized distribution of BSS and AUC differences. The BSS and AUC differences from our non-randomized sample are then compared to the randomized distribution, and if our sample BSS or AUC difference exceeds the 95th percentile of the distribution (or is less than the 5th percentile) it is considered statistically significant. The same resampling process is repeated to test for statistical significance in the difference between Global and Global-Nest RF BS_{res} and BS_{rel} .

In addition to evaluating the severe convection forecasts, daily ensemble mean 500-mb height forecasts are used to evaluate the synoptic-scale pattern skill of both ensembles over the CONUS. Average daily 500-mb height ensemble variance is calculated to assess the rate of error growth (e.g., Eckel and Mass, 2005). Anomaly correlation

coefficient (ACC; e.g., Wilks, 2011) of daily average 500-mb height anomalies is calculated to assess forecast skill. ACC ranges from 1 (forecast matches observation) to -1 (forecast is perfect inverse of observation), with 0 representing a forecast with no value and 0.6 considered the threshold for the forecast being valuable for synoptic-scale features compared to climatology (Hollingsworth et al., 1979; Murphy and Epstein, 1989). A centered 31-day rolling climatology of daily-averaged 1200–1200 UTC (1200, 1800, 0000, 0600 UTC) 500-mb heights from the Global Ensemble Forecast System (GEFS) 00-hour analyses over 20 years (2000–2019) serves as the climatology from which anomalies are computed (Hamill et al., 2022).

5.2.5 Forecast Dryline Identification

Drylines are identified in the Global-Nest forecast ensemble members using the FrontFinder AI model (see Chapters 2–4; Justin et al., 2025). This ML model has been discussed in previous Chapters, and will be briefly reiterated here: It uses a convolutional neural network for image segmentation (Huang et al., 2020) to create predictions on front type, including drylines, from 10 predictor variables. The ML model is trained using ERA5 reanalyses (Hersbach et al., 2018) with a horizontal grid spacing of 0.25° and five vertical levels from 2-m AGL to 850 hPa with NOAA Weather Prediction Center (WPC) 3-hourly surface analyses from 2007–2022 as ground-truth.

The ML dryline identification model is applied to all forecasts of each member of the Global-Nest ensemble. The necessary inputs to the ML model are bilinearly interpolated to 0.25° horizontal spacing, and the max-min normalization parameters are adjusted based on values in the Global-Nest forecasts. Output dryline probabilities are considered for every March–May forecast initialization out to forecast day 15.

Dryline probability of detection (POD) and false alarm rate (FAR) are assessed both spatially and daily categorically at 0000 UTC each lead day. Spatial POD and FAR are

determined by comparing the 0.25° output of each ensemble member to the 0.25° output from ERA5 (Chapter 2). Daily categories compare the maximum forecast dryline probability anywhere in CONUS—upscaled to the ~ 81 -km grid—to the maximum probability from the 0.25° ERA5 predictions. For both evaluations, 50% is used as the threshold to denote a grid point or day as having a dryline.

In addition to assessing the 11 ensemble members individually, POD and FAR is also calculated for an ensemble mean forecast using two techniques. For spatial POD/FAR, an ensemble mean of the 11 ensemble members is found at each grid point, similar to how the ensemble mean of other atmospheric variables such as temperature are found. Hits, misses, and false alarms are counted on a gridpoint-by-gridpoint basis for all forecasts of the given lead day. For the daily categorical POD/FAR, the ensemble mean forecast is a binary yes/no dryline forecast based on how many ensemble members identified a dryline for that day. If 6 or more ensemble members (out of 11 total) had a maximum dryline probability above 50%, the categorical ensemble mean dryline forecast is 1, else it is 0.

The two assessments of POD and FAR test different skill from the Global-Nest ensemble. The spatial calculation tests the skill of ensemble members forecasting the dryline location. Forecasts with a dryline in an incorrect location are double-penalized—once for missing the dryline at a given location, and again for forecasting a dryline at an incorrect location. The ensemble mean evaluates the spread of the individual ensemble members. The daily categorization does not punish spatial errors in dryline identification, and comparison with the spatial POD/FAR can provide inferences to the types of errors in both the individual forecast members as well as the ensemble as a whole.

At this time, assessment of Global-Nest dryline forecasts are not compared to Global-Nest RF severe thunderstorm forecast performance. Future work will evalu-

ate how dryline forecast skill—both the presence of a dryline as well as its location—correlate with severe thunderstorm forecast skill. Future work will also expand forecasted dryline analysis to include drylines from the Global ensemble forecasts.

5.3 Results

5.3.1 Ensemble 500mb Skill

Both the Global and Global-Nest ensembles have near-identical skill in synoptic-scale forecasting through day 15 as shown through their daily ensemble mean 500-mb height ACC (Fig. 5.2a). ACC for both ensembles falls below 0.6—the typical threshold used for when a synoptic forecast is marginally useful relative to climatology (e.g., Hollingsworth et al., 1980)—after day 9. Both ensembles have a high ACC above 0.95 through day 4 before a steady and mostly linear decline in ACC by forecast lead day, corresponding to the slow rate of forecast spread growth through day 4 (Fig. 5.2b). The forecast spread in both ensembles grows at a near-identical rate throughout the 15-day period. While the ACC and ensemble variance are nearly identical between both ensembles, the exact errors in both ensemble forecasts which result in these ACC values may be different.

Regardless of the accuracy of small-scale physical processes in the nested CAM, forecast skill of the Global-Nest will be bounded by the synoptic-scale skill of the larger convection-parameterized model which serves as the lateral boundary conditions to initialize it. While the upscaling process may smooth out small-scale errors in 500-mb height associated with storm-scale processes in the CAM, the similarities in ACC and forecast spread between the Global and Global-Nest indicate that storm-scale errors do not rapidly grow upscale to synoptic-scale pattern errors which negatively

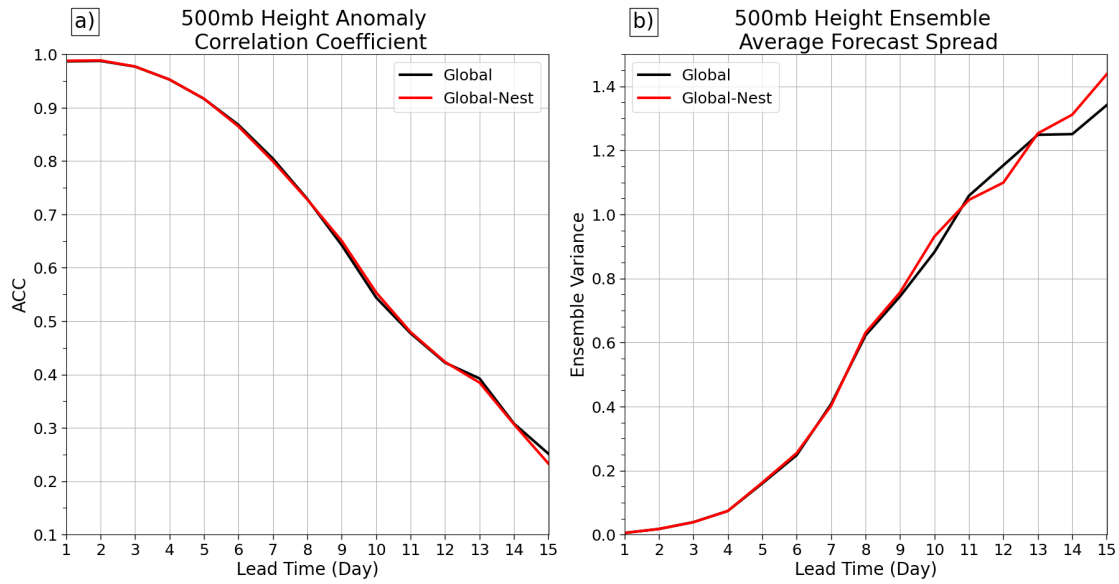


Figure 5.2: (a) Daily 500-mb height ensemble mean anomaly correlation coefficient and (b) average daily 500-mb height ensemble mean variance out to forecast day 15 for the Global (black) and Global-Nest (red) ensembles.

affect Global-Nest synoptic-scale predictability compared to the Global through day 15.

5.3.2 RF Forecast Statistics

Attributes diagrams of the severe probability forecasts by both the Global and Global-Nest RFs show similar forecast reliability and skill (Fig. 5.3). In both the Global and Global-Nest RFs, severe forecast probabilities rarely exceed 0.5, with most day 8+ forecasts rarely having probabilities above 0.3. The Global-Nest RF has slightly more frequent occurrences of higher-probability forecasts compared to the Global RF, particularly for days 1–5 (Fig. 5.3a,b, inset). Forecast days 1–5 have a slight under-forecasting bias across all forecast probability bins for both RF models. At some longer lead times (e.g., day 6), the Global-Nest RF forecasts have improved reliability at higher-probability forecasts compared to the Global RF (Fig. 5.3). In both sets of RFs, forecasts tend to become unskillful at the higher-end of their forecasted probabilities

beyond forecast day 5.

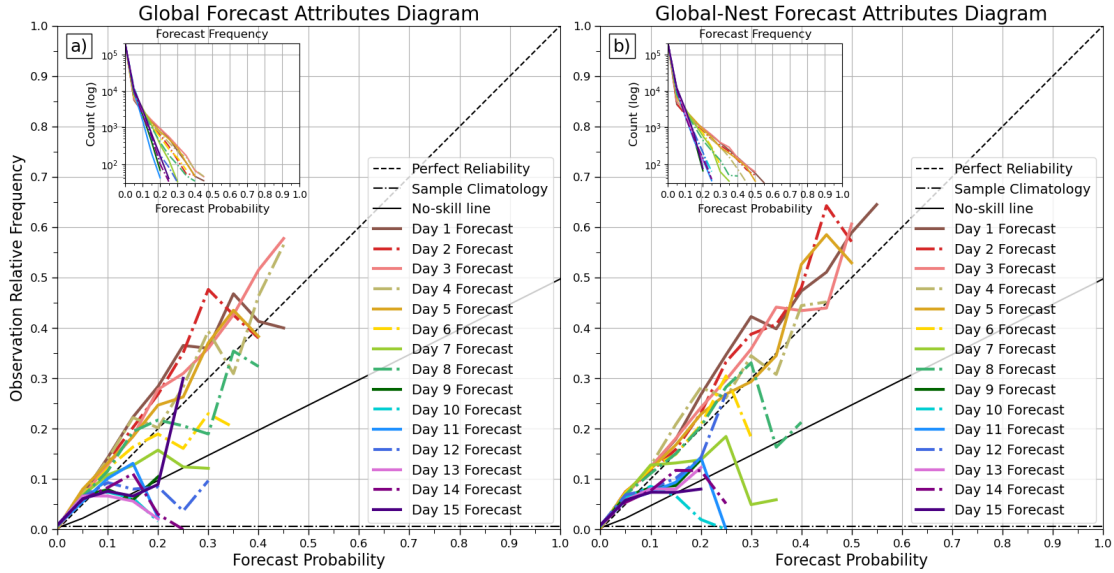


Figure 5.3: Attributes diagram for the (a) Global and (b) Global-Nest RF forecasts of severe convection. The number of forecasts of a given probability are shown in the inset figure in the upper left of each. Individual forecast days are colored from warm (day 1) to cool (day 15) colors, alternating between solid and dot-dashed lines by day. The black-dashed diagonal line represents perfect reliability, and the black-dot-dashed horizontal line represents the sample climatology probability of a severe report at all grid points. The solid black line is the line of no skill. Attributes are masked for any bin that does not contain at least 30 grid forecasts.

BSS and AUC (Fig. 5.4) are evaluated as a function of forecast lead day. In both metrics, forecast skill is the highest on day 1 and generally decreases through day 9, after which skill does not change substantially. The BSS of ~ 0.03 after day 9 indicates that those long-range forecasts are only about as good as climatology. In both BSS and AUC, the Global-Nest RF exhibits statistically significant improvement over the Global RF for most lead times through day 7. After day 7, the objective performance of the Global-Nest and Global RFs is similar. For BSS, the Global-Nest RF is significantly higher for days 1–6, except for day 4 (Fig. 5.4a). For AUC, the Global-Nest RF is significantly better for days 2, 4, 6, and 7 (Fig. 5.4b). Some forecast skill being evident through day 7 in the Global-Nest RF is consistent with subjective

analyses of FV3-based CAM ensembles (Clark et al., 2023b).

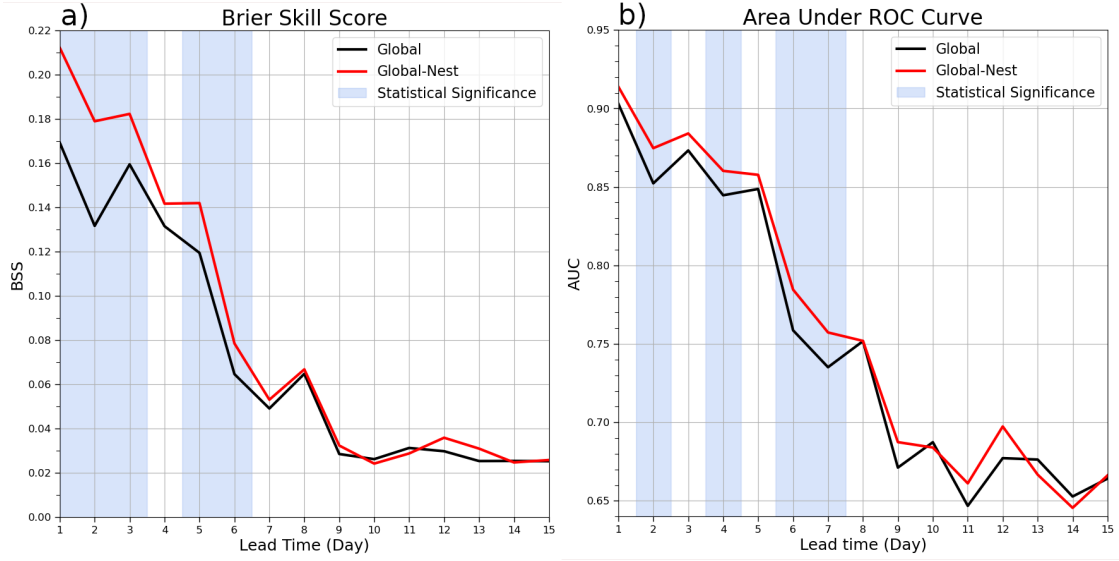


Figure 5.4: (a) Brier Skill Score and (b) Area under the Receiver Operating Characteristic (ROC) curve by forecast lead day for the Global (black) and Global-Nest (red) RFs. Light-blue shading denotes lead days when the Global-Nest RF improvements in (a) BSS and (b) AUC is statistically significant to the 95th percentile.

The overall decrease in forecast skill through day 9 is largely attributable to a decrease in the BS_{res} (Fig. 5.5a). BS_{res} in both RFs follows the same pattern as BSS and AUC: generally falling through day 9, and then flattening out afterwards. The Global-Nest RF BS_{res} is slightly better than the Global RF BS_{res} through day 7, with the improvement being statistically significant for days 1, 2, 6, and 7. BS_{rel} is similar for both RFs. The Global RF has statistically better reliability for day 6, and the Global-Nest RF is statistically better for day 13, but it is unclear if this represents a meaningful improvement (Fig. 5.5b). The overall improvement of both Global and Global-Nest RF BS_{rel} as lead time increases—a smaller reliability is associated with a smaller difference between forecast probabilities and the frequency of occurrence (Chapter 5.2.4)—can be attributed to the decrease in the number of forecast bins populated as lead time increases (Fig. 5.3), rather than a practical improvement in

BS_{rel} .

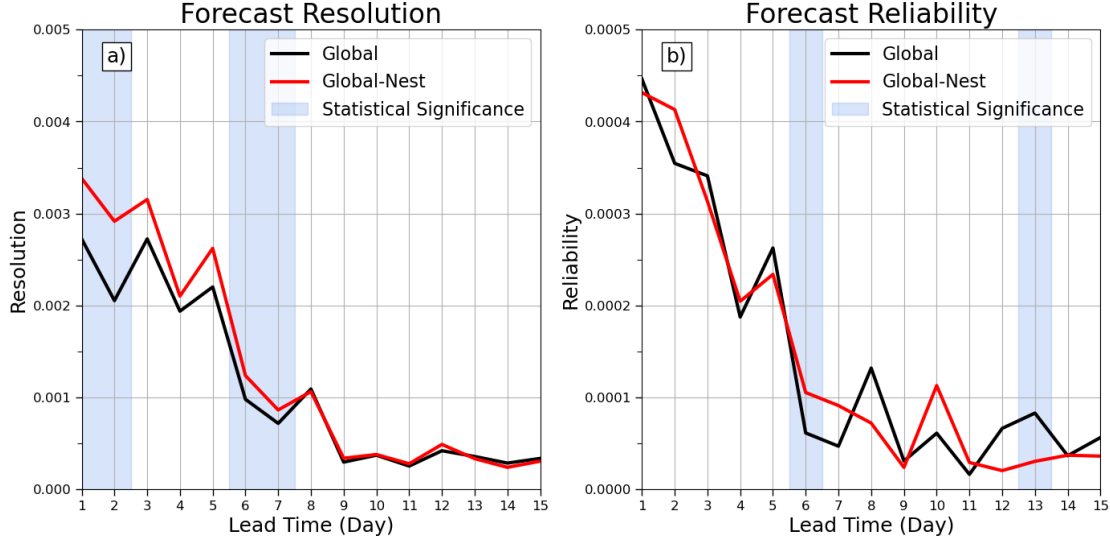


Figure 5.5: Forecast (a) resolution and (b) reliability for the Global (black) and Global-Nest (red) RFs by forecast lead day. Light-blue shading denotes days where the difference between Global and Global-Nest RFs is statistically significant to the 95th percentile.

The general decline in RF forecast skill through day 9 mirrors the decrease in 500-mb ACC in both the Global and Global-Nest (Figs. 5.2a, 5.4). Day 8 is the furthest lead time where daily 500-mb ACC exceeds 0.6 (Fig. 5.2), which is the furthest lead time where Global and Global-Nest BSS exceeds ~ 0.03 , the baseline BSS value for each day beyond day 8 (Fig. 5.4a). The consistency between these results suggests that the RFs ability to predict severe storms is directly associated with the ensemble's ability to correctly forecast the large-scale synoptic pattern. The lack of skill added to the Global-Nest RF compared to the Global RF beyond day 8 may therefore be a reflection that the underlying ensemble forecast has no synoptic-scale skill, not that a CAM couldn't provide additional value at these extended lead times.

Statistical significance for individual lead days for BSS and AUC (Fig. 5.4) and resolution and reliability (Fig. 5.5) should be judged with caution due to the day-to-day jaggedness of the curves, which is likely a consequence of our small sample size

and forecast initialization strategy. Since forecasts are only initialized 4 times a month, not every severe event is represented at each lead time. Severe events that the RFs perform better (or worse) with may inflate (or depress) forecast skill for a given lead day relative to the RF skill of neighboring lead times, as is the case with both RF day 3 forecasts verifying better than their respective day 2 RF forecasts (Fig. 5.4). A larger sample size or daily forecast initializations would likely produce smoother curves (e.g., Clark et al., 2025).

5.3.3 Predictor Relative Contributions

The relative probability contribution of each category of predictors to forecasts of severe convection is assessed for forecast lead days 1–15 according to Chapter 5.2.3, and normalized by the number of predictors within each group (Fig. 5.6). For the Global RF, kinematic variables are the most important, providing on average 6.5% of the total probability per predictor (across 7 individual predictors) through day 4 (Fig. 5.6a). Thermodynamic predictors provide on average 5% of the total probability per predictor (across 7 individual predictors), approximately the same as STP and higher than the other individual predictors (Fig. 5.6a,c). Across all lead times, the least-contributing predictor of both the thermodynamic and kinematic grouping contribute approximately the same amount. However, the maximum-contributing predictor of the kinematic grouping contributes over 12% of the relative contributions through day 4, before dropping to $\sim 10\%$ by day 7 (Fig. 5.6a). The maximum-contributing predictor of the thermodynamic grouping comparatively contributes under 8% of the total contribution.

Relative Fraction of Mean Contribution by Variable Type

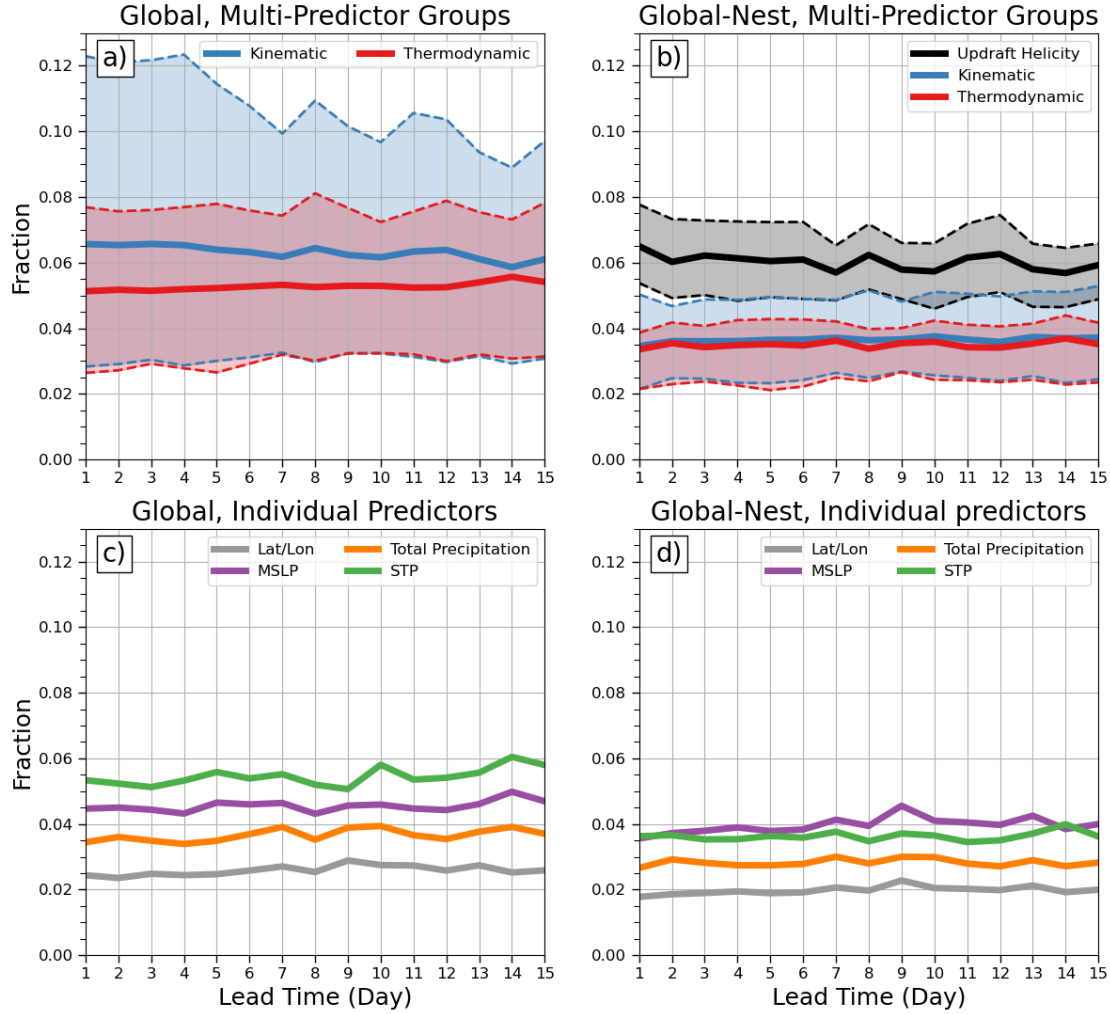


Figure 5.6: Mean relative contribution of each predictor group (a,b) and individual predictor (c,d) for the Global and Global-Nest RF, respectively. The solid lines in (a,b) represent the mean relative contribution for that predictor group, with the shading of the same color denoting the region between the individual predictor within that group with the highest relative contribution and lowest relative contribution at each lead day. Dashed lines denote the maximum and minimum relative contributions from an individual variable of that grouping. Since the predictors in (c,d) are not grouped, there is no max/min bounds. Note that lat/lon is the average of the latitude and longitude predictor, and is represented in (c,d) because the difference between the relative contributions from latitude and longitude predictors is minimal.

As forecast lead time increases to day 15, the average relative contribution from the kinematic grouping decreases slightly, but it remains the primary contributor to severe probabilities (Fig. 5.6a,c). The average relative contribution from thermodynamic predictors increases slightly, but remains below the kinematic average. The contribution from individual predictors increases slightly with increasing lead time, with STP contributing the most among the individual predictors and contributing approximately as much as the average of the thermodynamic grouping (Fig. 5.6a,c).

For the Global-Nest RF, the UH grouping is the dominant contributor at all forecast lead times (Fig. 5.6b). The average of all 6 UH contributions is higher than the maximum relative contribution from any of the other individual predictors, and the UH predictor which individually contributes the least still contributes more than any other individual predictor at all lead times except for days 10 and 13–15 (Fig. 5.6b,d). Similar to the Global RF, the maximum-contributing individual kinematic predictor contributes more than the maximum-contributing individual thermodynamic predictor, but the overall average contribution from each kinematic and thermodynamic predictor is approximately the same across all lead times (Fig. 5.6b). Of the individual variables, MSLP and STP contribute a similar amount, unlike the Global RF where STP contributed $\sim 0.5\%$ more than MSLP at all lead times (Fig. 5.6c,d). Both MSLP and STP contribute approximately as much as the average from the kinematic and thermodynamic grouping.

Since both ensembles displayed remarkably similar synoptic-scale forecast skill (Fig. 5.2), the improvement of the Global-Nest RF severe convection forecast skill (Fig. 5.4) can be attributed to contributions from UH. Even as RF forecast performance decreased with increasing lead time (Fig. 5.4), the relative contribution to forecasted severe probabilities do not change substantially (Fig. 5.6).

5.3.4 Global-Nest Dryline Forecasts

Daily categorical short-term forecasts of March–May drylines in the Global-Nest have a low FAR, but they also have a low POD of less than 0.6 (Fig. 5.7). POD is higher from days 3–8 than it is for days 1–2, before generally declining in all members. FAR generally increases with increasing lead day. The jagged day-to-day variability in POD/FAR is attributable to the varying sample size for each day. The number of drylines for each lead day varies between 20 and 55 (not shown).

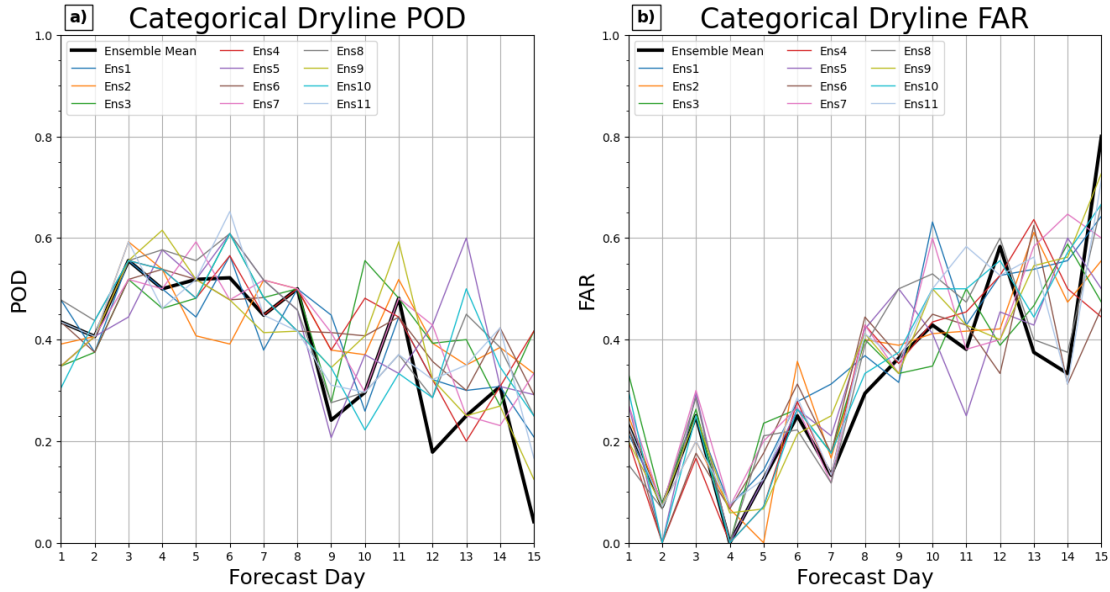


Figure 5.7: (a) POD and (b) FAR for a daily categorical forecast of a dryline for each Global-Nest ensemble member (thin colored lines) and the Global-Nest ensemble mean (thick black line) by forecast lead day.

The spatial aggregate POD is lower and FAR is higher than the categorical dryline identification (Fig. 5.8). For all ensemble members, POD peaks at day 2 below 0.3. FAR is generally high, even at short lead times, and increases to above 0.8 for most ensemble members beyond day 9. The ensemble mean FAR and POD are at the lower end of the envelope of individual ensemble members, and both drop to 0.0 beyond day 9. This can be attributed to the growing dispersion in dryline probabilities among the

ensemble members with increasing lead time. As forecast spread of dryline probabilities increase, the ensemble mean dryline probability at any given grid point decreases. Beyond day 9, the dryline probability dispersion across all ensemble members is so great that there are no longer any grid points in the ensemble mean with dryline probabilities above 50%. Therefore, with no drylines forecasted in the ensemble mean, POD and FAR are both 0.0.

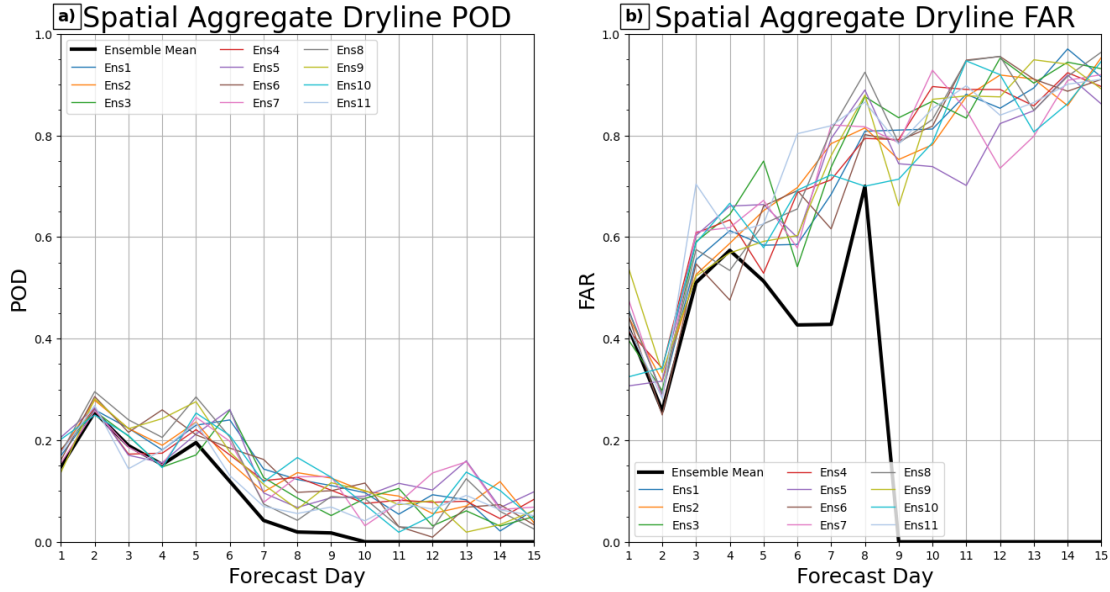


Figure 5.8: As in Fig. 5.7, but for spatially-aggregated POD and FAR.

The lower POD and higher FAR in the spatially-aggregated dryline probabilities compared to the daily non-spatial categorization highlights the spatial errors in Global-Nest dryline forecasts. For a binary yes/no forecast of a dryline present anywhere in the CONUS, the Global-Nest ensemble has mediocre performance through day 8, with a POD between 0.4 and 0.6 and a FAR generally below 0.3 (Fig. 5.7). Forecasting daily drylines beyond day 8 is prohibitive due to declining POD and increasing FAR, consistent with the decline in synoptical-scale skill in 500-mb height forecasts (Chapter 5.3.1). The decrease in POD, and increase in FAR, in the spatially-aggregated verification is the result of dryline probabilities being misplaced spatially compared to the

observed dryline probabilities.

Spatial errors in forecasted dryline placement could be the result of systematic errors in the forecast system. Errors representing physical processes relevant for dryline propagation, such as soil moisture, the mixing of the boundary layer, and wind speed and direction near the surface could result in a systematic faster (or slower) propagation speed of the dryline during the diurnal cycle. Such errors have been documented previously in other dynamical forecast model systems (e.g., a warm and dry bias in the HRRR 2-m temperature; Min et al., 2021). Further work would be necessary to determine if a systematic pattern exists in the spatial errors in dryline probabilities.

Further work can additionally explore the relationship between forecasted dryline errors and the Global-Nest RF severe thunderstorm forecasts. Since the dryline is associated with a significant portion of springtime severe convection in the US Southern Great Plains (Chapter 3), it is hypothesized that spatial errors in dryline placement may correlate with spatial errors in severe thunderstorm forecasts. Further work can additionally evaluate dryline forecasts within the Global ensemble. One potential explanation for an increase in Global-Nest RF skill may be improved forecasts of drylines. However, it is unclear at this time how RF skill correlates with dryline forecast skill and how Global and Global-Nest RF skill differs during spring dryline days.

5.4 Individual Forecast Case Studies

All Global RF and Global-Nest RF forecasts for lead days 3–9 were subjectively assessed. Consistent with the objective verification statistics (Chapter 5.3.2), forecast skill decreased with increasing lead time, with very little skill evident beyond day 9. Differences between the Global RF and Global-Nest RF forecasts are most stark at shorter lead times (day 3), with most, but not all, day 3 Global-Nest RF forecasts

being subjectively superior to Global RF forecasts.

By day 8, differences between the Global and Global-Nest RF are more marginal. The number of false alarms (forecasts with moderately high severe probabilities, yet no severe reports) and misses (forecasts with low severe probability, yet widespread severe reports) increase dramatically compared to day 3 (not shown). However, even at this lead time, valuable forecasts can still be identified.

Three representative individual forecasts from the Global and Global-Nest RF are subjectively compared in this section—The day 3 forecast for 17 May 2017, the day 3 forecast for 24 May 2019, and the day 8 forecast for 22 January 2017. These forecasts are highlighted because they reflect forecast performance characteristics common among all day 3 and day 8 forecasts, respectively. Forecast probabilities, and the contributions to those probabilities from predictors determined via TreeInterpreter, are summed across the convective day (1200 UTC, 1800 UTC, 0000 UTC, and 0600 UTC) to present a daily probability and contribution.

5.4.1 Day 3: 17 May 2017

The convective day of 17 May 2017 is associated with 372 severe storm reports (SPC, 2023), primarily across Iowa, Illinois, and Wisconsin, from discrete supercells as well as storm clusters and linear segments (Fig. 5.9b). The majority (299) of reports are associated with severe convective winds, but there are also 15 reported tornadoes with the strongest being rated an EF-2 on the Enhanced Fujita scale (Fujita, 1971; WSEC, 2006; Edwards et al., 2013, 2021) and 58 severe hail reports. The maximum categorical day 1 outlook issued by the SPC is an Enhanced Risk across much of Iowa and extending eastward into southwest Wisconsin and northwest Illinois (Fig. 5.9a).

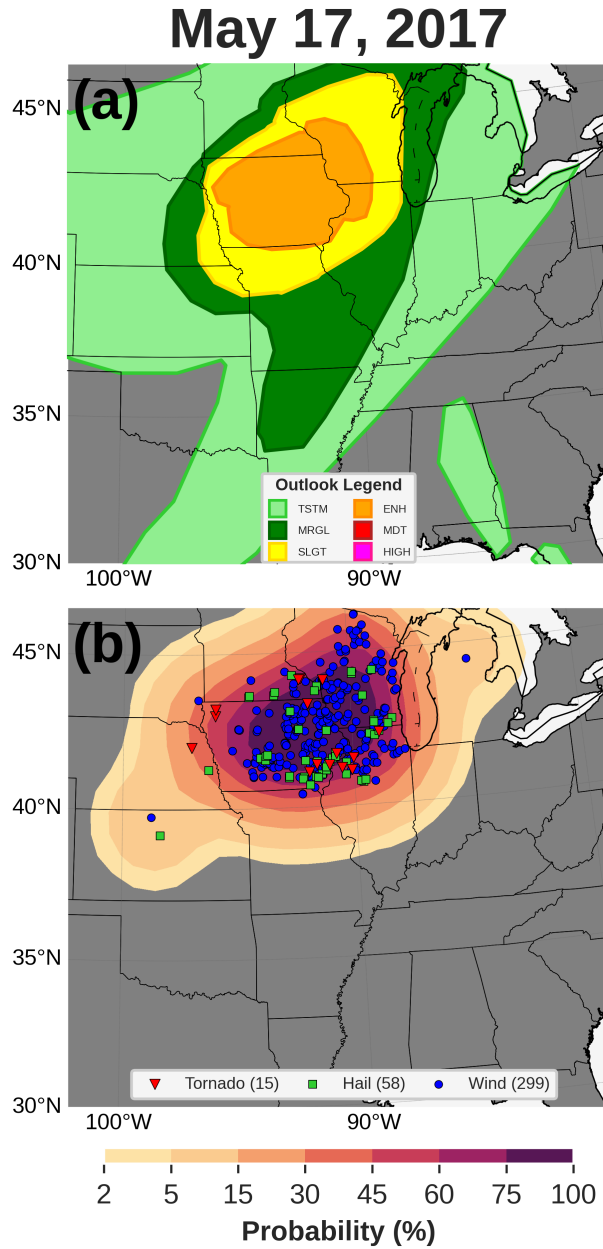


Figure 5.9: (a) the maximum day 1 categorical outlook issued by the SPC for 17 May 2017, and (b) scatterplot of all individual reports. The color and shape of the individual scatter points in (b) correspond to the type of severe report (bottom in-panel legend). Practically-Perfect Hindcast (PPH; Hitchens et al., 2013; Gensini et al., 2020) probabilities are shaded underneath and correspond to the legend below (b).

The Global RF forecast has elevated severe probabilities across much of the Midwest and Mississippi River region (Fig. 5.10a). While these higher probabilities do include most of the region of the Midwest where severe reports occurred, much of the area forecasted with higher probabilities do not have any severe reports, including a large swath from east Texas to western Ohio. The Global-Nest RF forecast has a concentrated region of the Midwest where it has higher severe probabilities, and this region is coincident with the majority of locations with severe reports (Fig. 5.10f). While the Global-Nest RF forecast did have severe probabilities extending from east Texas to western Ohio like the Global RF forecast, probabilities for the Global-Nest RF forecast are lower (25% in the Global RF vs 10% in the Global-Nest RF). In addition to having a smaller region of higher severe probabilities better co-located with severe reports, the Global-Nest RF forecast also has a higher maximum forecasted severe probability. Across all day 3 forecasts, spatial improvements in the Global-Nest RF compared to the Global RF like those highlighted in this case are common. These spatial improvements, combined with higher forecast probabilities, are reflected in the higher BSS and AUC for the Global-Nest is this case (Fig. 5.10a,f), as well as aggregated across all day 3 forecasts (Fig. 5.4).

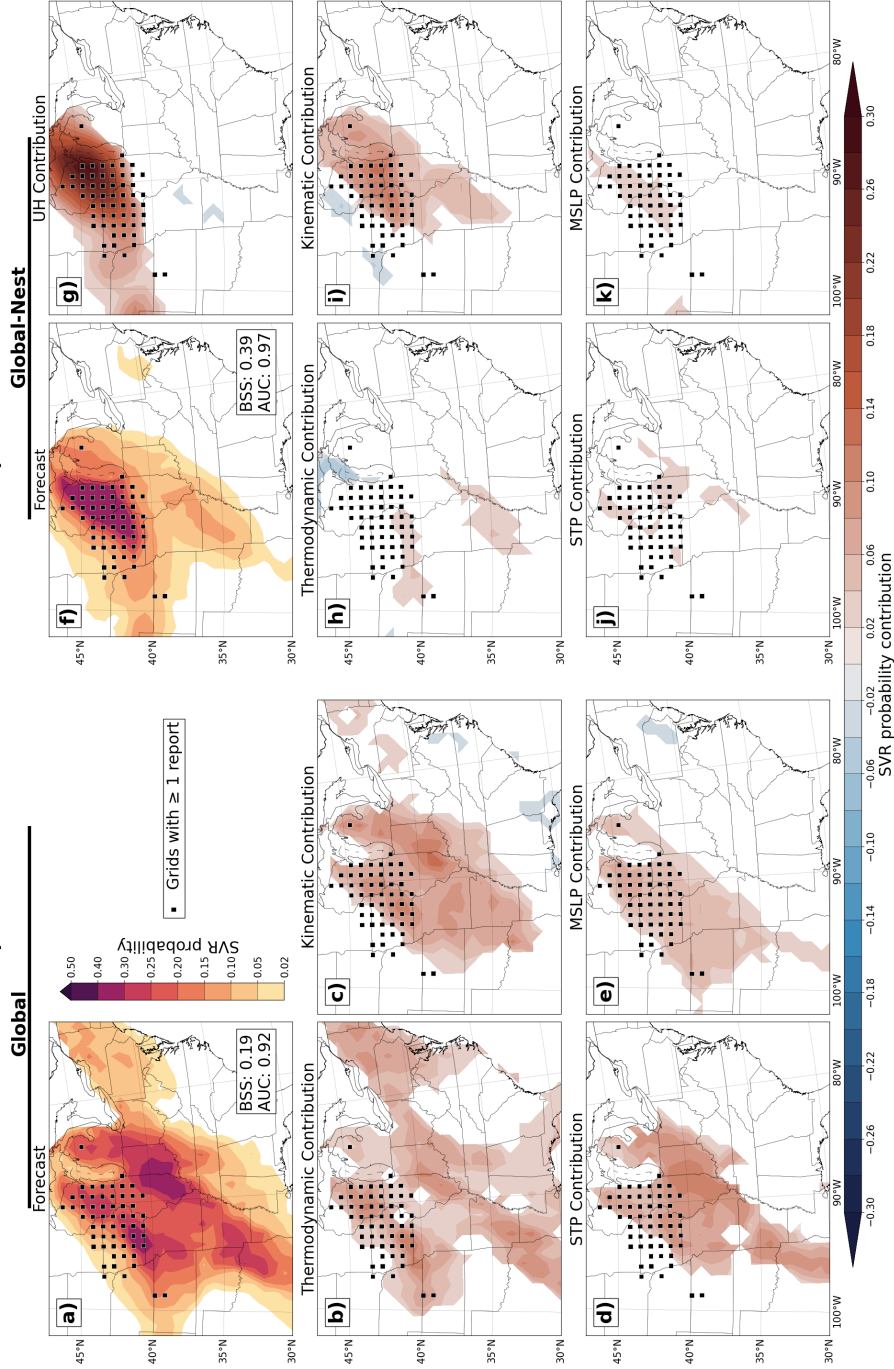


Figure 5.10: (a) Global and (f) Global-Nest RF day 3 severe convection forecasts for 17 May 2017. Shading denotes the forecasted severe probabilities (note that the scale max is at 0.5). The color-bar legend for total severe probability forecasts in (a) and (f) is to the right of (a). BSS and AUC for the Global (a) and Global-Nest (f) RF forecasts are in the bottom right corner. (b-e) Daily summed contributions to the Global RF forecast from individual groupings of variables: (b) thermodynamic, (c) kinematic, (d) significant tornado parameter, and (e) mean sea level pressure. (g-k) Daily summed contributions to the Global-Nest RF forecast from individual groupings of variables: (g) thermodynamic, (h) kinematic, (i) significant tornado parameter, and (j) mean sea level pressure. The color-bar legend for probability contribution is located at the bottom of the figure. Black squares denote each grid box with at least one severe report during this convective day.

The spatial improvement of the Global-Nest RF forecast is attributable to the substantial role of UH as a contributor. Strongly positive UH contributions across Wisconsin into Iowa and northern Illinois dominate neutral to weakly positive contributions from thermodynamic variables, STP, and MSLP (Fig. 5.10h,j,k). In contrast, the Global RF forecast features widespread positive contributions across much of the central and eastern CONUS from each category of environmental predictors, including portions of the south-central and northeast US where no reports occurred (Fig. 5.10b-e). The contribution from STP, a multi-variable parameter incorporating thermodynamic and kinematic variables passed to the RF as a single predictor, to the Global RF forecast is on the same order of magnitude as the collective summed contributions from the seven predictors in the kinematic and thermodynamic groupings.

Precipitation amount, which in convection-parameterized models can be used as a rough proxy to identify regions where convection may occur, provides minimal contributions in either the Global or Global-Nest RF forecasts (not shown). The lack of importance of 6-hour accumulated precipitation amount, particularly for the Global RF, is surprising. Some possible explanations for this are that the Global RF has sufficient information from the other predictors, rendering precipitation amount redundant, or that the RF devalues precipitation amount due to the abundance of precipitation that is not associated with severe convective hazards thus leading to a low correlation between precipitation and severe reports. The lack of a notable negative contribution to severe probabilities where there is no forecasted precipitation—what should be a very clear signal that there would be no possibility for severe reports—may be due to errors in modeled precipitation leading to instances in training where severe reports occur at locations with no forecasted precipitation. Further investigation into the cause of this unexpected result is outside the scope of this study. Latitude and longitude also did not contribute substantially to either forecast (not shown).

The large differences in contributions from STP—minimal contributions in the Global-Nest RF and widespread positive contributions in the Global RF (Fig. 5.10d,j)—might imply that the STP forecasts within these two ensembles differ substantially. However, a comparison of ensemble mean STP at 0000 UTC 18 May 2017 shows similarity between the two forecasts (Fig. 5.11). Both ensemble means have a corridor of enhanced STP from east Texas through the Midwest. Zero-hour analysis from the GEFSv12 Reforecast Dataset (GEFS/R; Hamill et al., 2022; Zhou et al., 2022) shows a corridor of enhanced STP in a similar location as forecasted, although the magnitude of STP in the analysis is higher than forecasted in either ensemble (Fig. 5.11d). The relatively low STP over Iowa and Wisconsin is likely due to the forecasted severe weather having already occurred in this area by 0000 UTC. While the maximum STP is lower and slightly further east in the Global-Nest compared to the Global, both forecast ensembles have very similar forecasted STP (Fig. 5.11c). The GEFS/R STP analysis shows elevated values of STP across both regions where the Global forecasted STP is higher than the Global-Nest and vice versa (Fig. 5.11c,d). The similarity in forecasted STP between the Global and Global-Nest at 0000 UTC 18 May 2017, as well as the other three output times of this convective day (not shown), indicate that the difference in STP contribution between the Global-Nest RF and Global RF is due to a decreased importance of STP as a predictor in the Global-Nest RF.

Day 3 Forecast STP Comparison for:
0000 UTC 18 May 2017

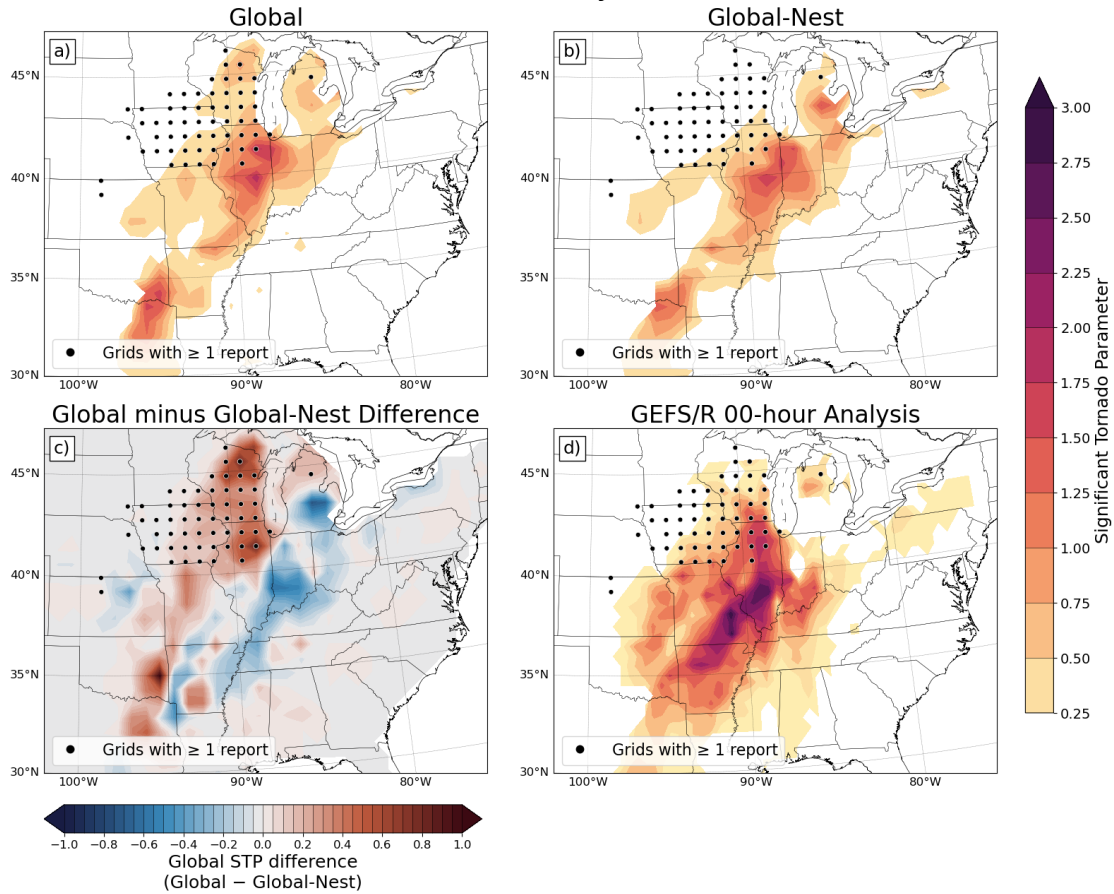


Figure 5.11: The day 3 forecasted ensemble mean STP valid at 0000 UTC 18 May 2017 (convective day 17 May 2017) for (a) Global, (b) Global-Nest, (c) the difference between Global and Global-Nest. (d) The GEFS/R Zero-hour analysis STP for context. The color-bar legend for STP, both (a,b) forecasted and (d) analysis, is to the right of the figure. The color-bar legend for the difference in STP between the Global and Global-Nest is at the bottom of panel (c).

5.4.2 Day 3: 24 May 2019

While the majority of Global-Nest RF day 3 forecasts are subjectively better than the corresponding Global RF forecast in ways shown and discussed in the first case, this is not true for all day 3 forecasts, including the 24 May 2019 event. This convective day is associated with 151 severe storm reports (SPC, 2023) primarily from southwest Texas through Illinois and southern Wisconsin, as well as a cluster of reports in eastern North Dakota and northwest Minnesota (Fig. 5.12b). The majority (87) of these reports are associated with severe convective winds, but there are also 47 hail reports and 17 EF-0/EF-1 tornadoes. The main corridor of severe convection from southwest Texas through Illinois included both discrete supercells as well as clusters/linear segments. The maximum day 1 categorical outlook issued by the SPC outlined this corridor with a Slight Risk, with localized areas in Texas, Oklahoma, and Illinois/Iowa/Missouri of Enhanced Risk (Fig. 5.12a).

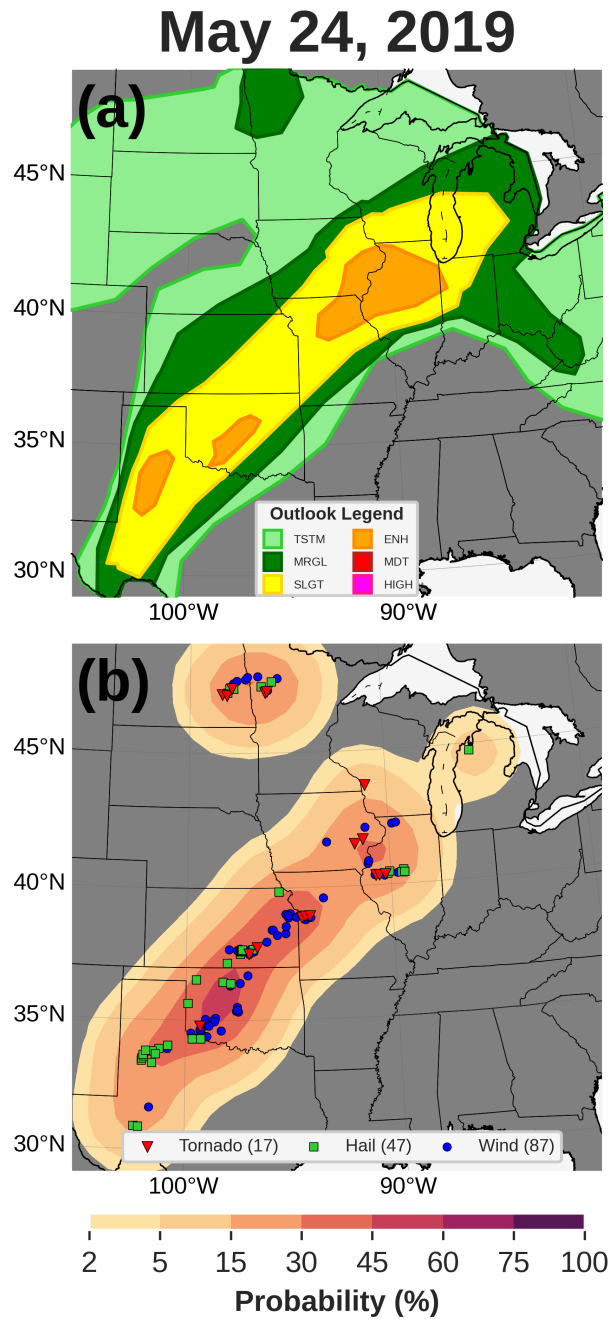


Figure 5.12: As in Fig. 5.9, but for 24 May 2019. Note there are 4 severe reports (2 hail, 2 wind) outside of this domain in northern California.

The day 3 forecast for 24 May 2019 is an example where the Global RF forecast is superior, both objectively—the Global RF BSS is 0.42 and AUC is 0.87 compared to the Global-Nest RF BSS of 0.38 and AUC of 0.70—and subjectively (Fig. 5.13a,f). Both forecasts have a similar maximum probability and both forecasts highlight the corridor from southwest Texas through Illinois. However, the Global RF forecast has superior spatial coverage of higher ($>15\%$) probabilities across the entire corridor, while the Global-Nest RF forecast has a notable minima ($<5\%$) across Oklahoma, Kansas, and Missouri where severe convection did occur (Fig. 5.13a,f). Neither the Global nor Global-Nest RF forecasts highlight severe convection potential as far north as the cluster of severe reports in eastern North Dakota and northwest Minnesota.

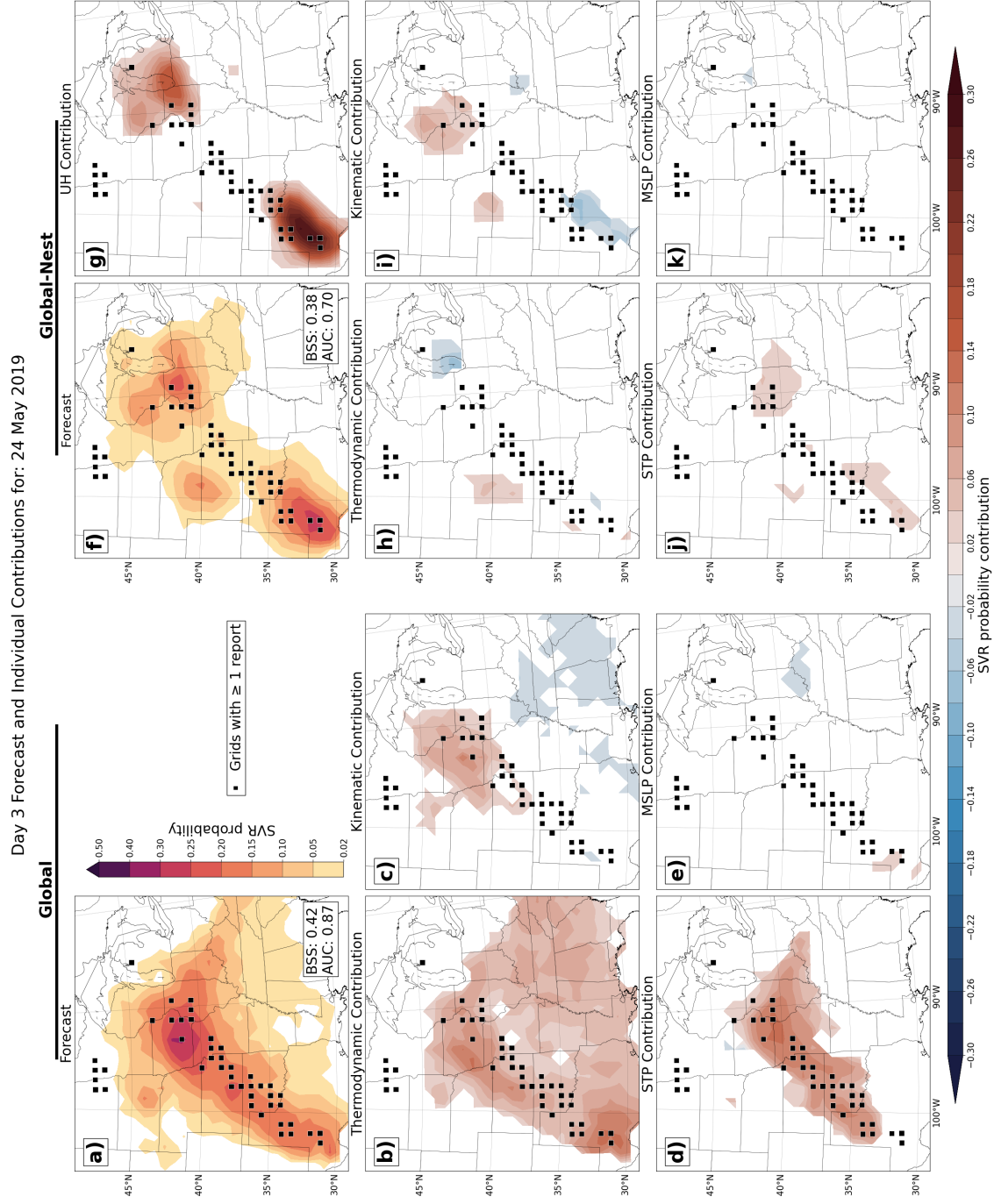


Figure 5.13: As in Fig. 5.10, but for the day 3 forecast for 24 May 2019

As with the day 3 forecast for 17 May 2017, UH is the primary contributor to severe probabilities in the Global-Nest RF, though its spatial coverage does not span the full area where severe weather occurred (Fig. 5.13g-k). UH provide no contribution to severe probabilities across Oklahoma, Kansas, and Missouri, and have a distinct maximum in west Texas and weaker maximum in the Midwest (Fig. 5.13g). While there are some weak and isolated positive contributions in the Global-Nest RF from STP, kinematic, and thermodynamic variables in Oklahoma and Kansas, the very minimal contributions result in low severe probabilities along this corridor where severe convection did occur (Fig. 5.13h-j). For the majority of day 3 forecasts, the UH contributions correspond well with regions where severe convection occurs. However, this case demonstrates a forecast where the UH predictors, and thus the resulting Global-Nest RF forecast, do not represent the occurrence of severe storms very well.

The Global RF day 3 forecast for 24 May 2019 features strong contributions from STP, kinematic, and thermodynamic variables (Fig. 5.13b-d). Contributions from STP outline the entire region from west Texas through Illinois where severe weather occurred, which is where the forecast from the Global-Nest RF is generally lacking higher probability (Fig. 5.13d). Contributions of similar magnitude from thermodynamic variables occur across wide swaths of the eastern CONUS (Fig. 5.13b). Contributions from kinematic variables are weaker and centered over Iowa, however the kinematic variables also provide negative contributions across much of the southeast to offset to some extent the widespread positive contributions from thermodynamic variables in this region (Fig. 5.13c). This balance of widespread positive contributions from some environmental variables and negative contributions from others results in an overall Global RF forecast with higher severe probabilities across the entire corridor from southwest Texas through Illinois, where severe convection occurred.

5.4.3 Day 8: 22 January 2017

By forecast day 8, statistical differences in objective performance between the Global and Global-Nest RF severe convection forecasts are marginal (Figs. 5.4 and 5.5). Subjectively, the number of false alarms (over-forecasts with moderately high severe probabilities, yet few if any severe reports) and failures-to-detect (under-forecasts with low severe probability, yet widespread severe reports) increase dramatically compared to day 3 (not shown). Failures-to-detect in particular become increasingly common, consistent with the decreasing frequency of higher magnitude probability forecasts at longer lead times (Fig. 5.3). However, there are still occasional day 8 forecasts which could be considered to provide useful forecast guidance, such as 22 January 2017. This high-impact convective day is associated with 158 severe storm reports (SPC, 2023) across Alabama, Georgia, and Florida, with 15 hail reports and 19 tornado reports with the strongest being two EF-3 tornadoes in Georgia (Fig. 5.14b). The severe weather resulted in 6 fatalities and 90 injuries, mainly across southern Georgia (SPC, 2023). The maximum day 1 categorical outlook issued by the SPC is a High Risk for northern Florida and southern Georgia (Fig. 5.14a).

January 22, 2017

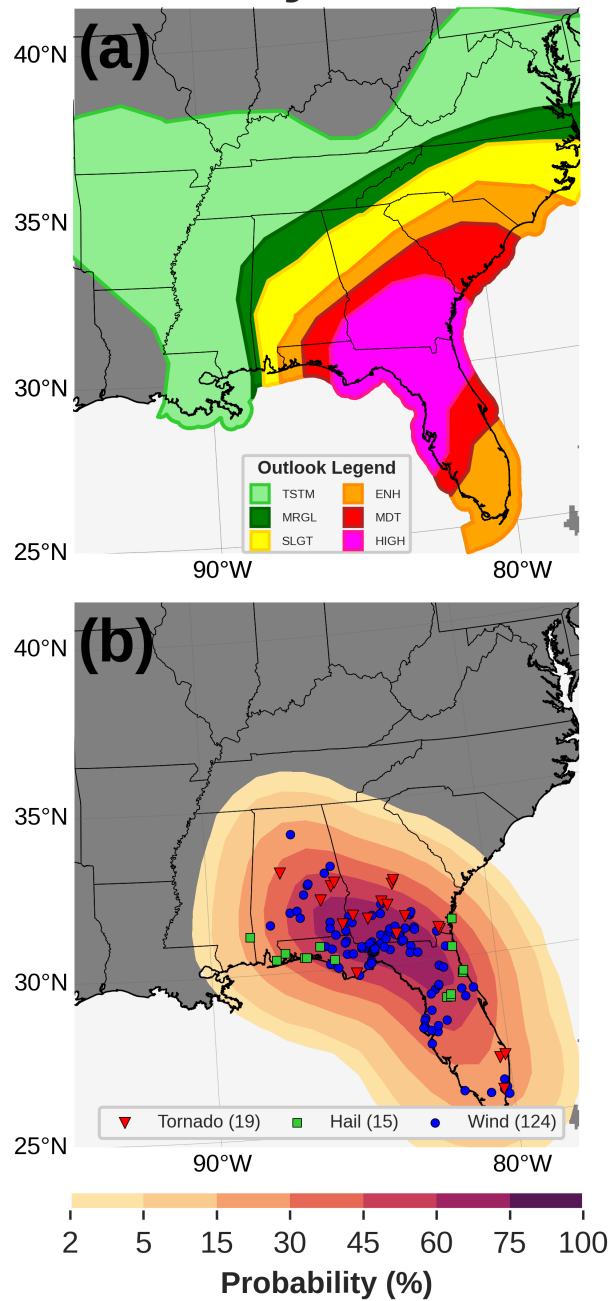


Figure 5.14: As in Fig. 5.9, but for 22 January 2017.

Both Global and Global-Nest RF forecasts have relatively high severe probabilities for a day 8 forecast across the southeast (Fig. 5.15a,f), with the Global-Nest RF maximum of 30% being high for any lead time (Fig. 5.3). While both forecasts would provide useful guidance, especially at this lead time, the Global-Nest RF is objectively and subjectively superior with its higher maximum forecast probabilities and the placement of its maximum across southern Georgia and northern Florida, where the most significant severe storms occurred. The Global RF forecast has both a lower maximum probability, and these probabilities are displaced farther west over primarily Alabama.

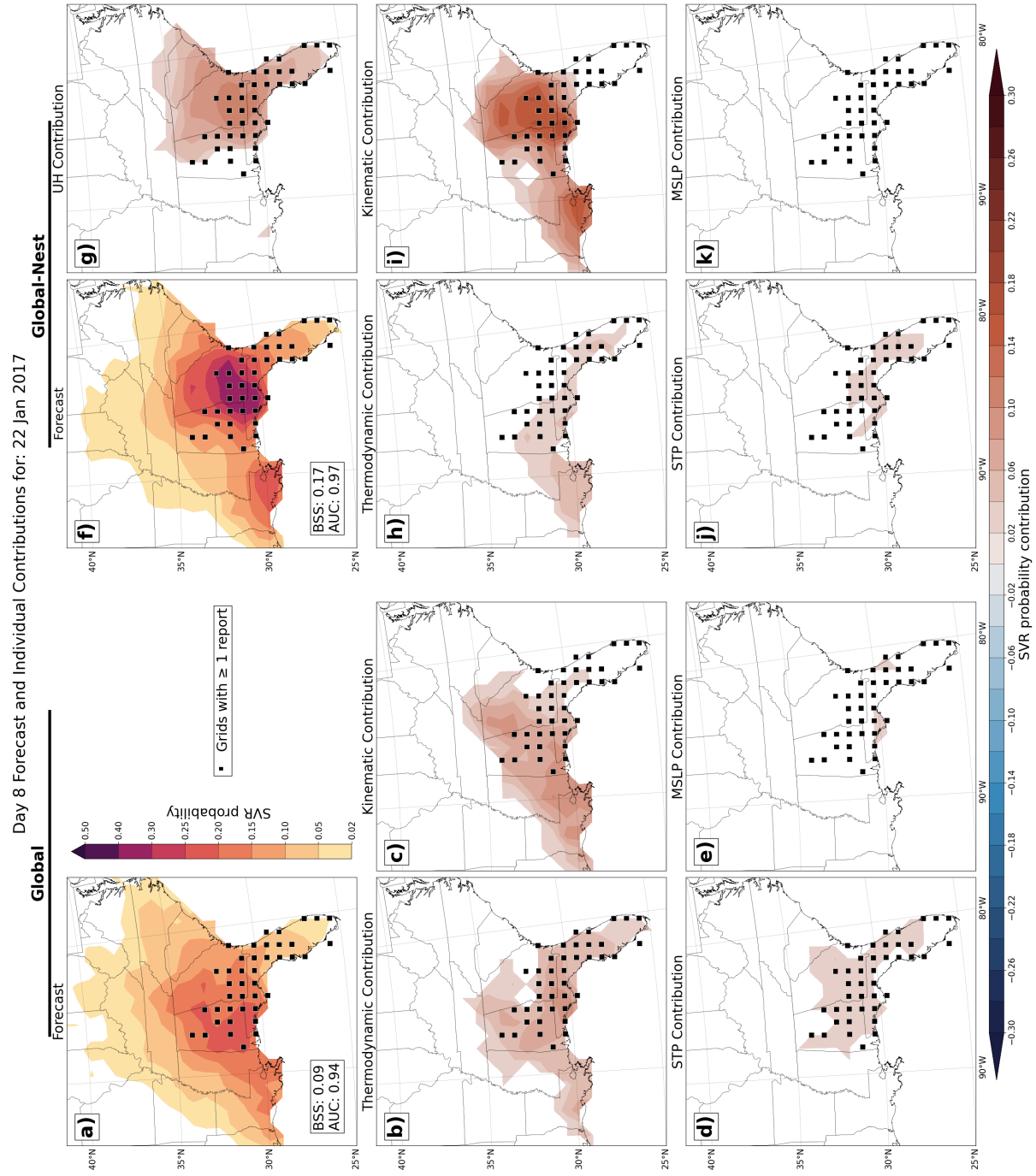


Figure 5.15: As in Fig. 5.10, but for the day 8 forecast for 22 January 2017

Unlike the day 3 forecasts highlighted in Chapter 5.4.1 and 5.4.2, UH is not the primary variable group contributing to the improvement of the Global-Nest forecast. While UH does contribute positive probability, it is the kinematic variables in the Global-Nest which provide the majority of severe probabilities (Fig. 5.15g,i). The other variable groupings contribute minimally to the Global-Nest RF forecast (Fig. 5.15h,j,k).

In the Global RF forecast, thermodynamic and kinematic variables contribute a relatively equal amount to the forecast and are the primary contributors (Fig. 5.15b,c). Unlike the Global RF day 3 forecasts previously highlighted, there are lesser contributions from STP to this Global RF forecast (Fig. 5.15d).

Although day 8 forecasts of this higher quality are rare in this dataset, this case may represent the occasional “forecast of opportunity”—a forecast with greater predictability than typical at extended lead times due to teleconnections and large-scale factors (e.g., Gensini et al., 2019; Mariotti et al., 2020). The higher probabilities and improved spatial component of the Global-Nest RF highlights the improvement to this forecast not just from the UH predictors, but also from how other predictors can contribute in the Global-Nest RF. Further investigation into how well either RF captures forecasts of opportunity at day 8 or beyond are outside the scope of this study.

5.5 Discussion

Across all lead times of the Global RF, but particularly for lead days 1–7, the relative contribution from kinematic predictors is higher than that from thermodynamic predictors (Fig. 5.6a). This result contrasts with Mazurek et al. (2025), who found that for the day 2 severe thunderstorm forecasts of the Colorado State University Machine Learning Probabilities (CSU-MLP) thermodynamic predictors contributed just

as much as kinematic predictors during the cool season. The CSU-MLP is trained using GEFS/R over approximately 9 years, with predictor inputs every 3 hours and separate RFs for 3 defined CONUS regions. The limitation of our RFs to the cool and spring seasons, when environments typically have higher shear, lower CAPE, and severe events are more likely to be forced by strong synoptic systems compared to summer (e.g., Johns and Doswell III, 1992; Doswell III et al., 1993; Brooks, 2009), may explain why our RFs more heavily weigh kinematic predictors compared to other RFs trained over the full year.

In the CSU-MLP, accumulated precipitation is a meaningful predictor (Mazurek et al., 2025), while it contributed less than the average kinematic or thermodynamic predictor for our RFs (Fig. 5.6). One plausible explanation for this drastic difference in predictor importance may be that the GEFS/R has improved forecast skill of precipitation compared to C-SHiELD. Additionally, while our predictor list was similar to the CSU-MLP, it was not identical. Some predictors may have lesser importance because they are redundant—their relationship with severe convection is better captured with a different predictor—or outsized importance because they partially represent a relationship not captured by any other predictor. Direct comparison of predictor importance between RFs with different training sets, predictors, and underlying forecast data should be performed with caution.

It is difficult to disentangle importance to RF probabilities (or lack thereof) deriving from the forecast skill of a given variable against its intrinsic ability to discriminate severe environments. While RFs, and ML techniques generally, can efficiently identify complicated non-linear relationships between multiple predictors, if our predictors are not perfectly forecasted the relationships the RF learns may not necessarily be physical. Our RFs—a proof-of-concept trained on forecast output—are therefore ill-suited to compare to studies investigating the skillfulness of different variables in predicting

severe hazards using observed atmospheric variables (e.g., Coffey et al., 2019; Coniglio and Parker, 2020) or (re)analysis model output (e.g., Kuchera and Parker, 2006; Nixon et al., 2023).

Despite the limitations in comparing our RFs to other similar forecast systems or physical relationships between atmospheric variables and severe storms, our RFs do enable a direct comparison of the impact of including a nested CAM in a global ensemble. The inclusion of UH in the Global-Nest RF resulted in UH predictors dominating the probability contributions (Fig. 5.6b). Most other predictors contributed a similar amount (3-4% per predictor), including the kinematic and thermodynamic groups, while each UH predictor contributed over 5% with the exception of a couple of lead days after day 10. Even as overall RF forecast skill decreased with increasing lead time (Fig. 5.4), UH remained the dominant predictor. When UH was removed as a predictor grouping in the Global-Nest RF, its severe weather forecast skill was nearly identical to that of the Global RF (not shown).

From our subjective analysis of individual forecasts (Chapter 5.4), the primary improvements in the Global-Nest RF forecasts came from improved spatial component of the forecasts, including both better co-location of maximum severe probabilities with reports and the limiting of widespread probabilities beyond where reports occurred, as well as a higher maximum probability. The inclusion of an extended-range CAM and associated UH output may therefore be able to help forecaster's narrow down where severe convection is likely to occur in extended-range forecasts and improve forecaster confidence. Even at day 8, when our RF forecasts collectively had minimal skill, we are still able to identify subtle improvements in individual RF forecasts, indicating that CAMs may be able to assist severe convection forecasting at day 8 and beyond.

Our objective evaluation of RF forecast skill is limited to a spatially-strict daily probability forecast. While this provides an easily defined and understandable met-

ric, this approach may have limited operational applicability for extended-range severe forecasts. At longer lead times, due to the expected overall degradation of forecast skill, forecasters and end-users may find more value in forecasts with less-strict spatiotemporal bounds (e.g., is there going to be severe weather across the southeast US on days 13–15?). Our daily forecasts provide minimal insight on whether a nested CAM could provide additional value to extended-range forecasts across different spatiotemporal scales. Consultation with forecasters, social/behavioral scientists, and end-users is a necessary component in the development of operational medium- to long-range severe weather forecast guidance (e.g., White et al., 2017), and is outside the scope of this study.

5.6 Conclusions

Daily forecasts of the probability of severe convection out to day 15 are created using two random forests (RFs) of predictor variables: one from a convection-parameterized Global ensemble, and the other from a Global-Nest ensemble which features a CAM configuration nested over the contiguous United States. Both ensembles come from the GFDL C-SHiELD, with the only difference between these two ensembles being the nested CAM. This comparison enables a direct quantification of the potential improvement to extended-range severe convection forecasts. The daily RF forecasts are compared both objectively and subjectively over the full 15-day forecast period.

From a subjective analysis, the Global-Nest RF forecasts are often superior to the Global RF forecasts through day 4 due to their improved spatial placement of severe probabilities relative to observed severe reports as well as their higher forecast probability magnitudes. Much of this improvement in skill is derived from the NMEP UH fields, which are the dominant contributor and, more often than not, are better

co-located with the observed severe reports than any other environmental predictor. As lead time increases to day 8, the differences between the Global and Global-Nest RF forecasts lessen and forecasts are far less skillful, beyond a few exceptional cases. UH continues to be an important predictor, but it no longer dominates the contributions to nearly every forecast like it did in shorter-range forecasts. Beyond day 8, differences between the Global and Global-Nest RF forecasts are negligible and most forecasts do not appear to have much subjective value.

Objective measures of severe weather forecast skill—Brier Skill Score (BSS) and Area Under the receiver operating characteristic Curve (AUC)—collaborate the subjective analysis. These objective metrics show a statistically-significant improvement in the Global-Nest RF forecasts compared to the Global RF forecasts are most lead times through day 7. Beyond day 8, neither the Global RF nor Global-Nest RF forecasts have much skill or are significantly different from each other, consistent with the low skill of the ensemble 500 mb pattern forecasts. The lack of improvement in the Global-Nest RF compared to the Global RF beyond day 7 is likely a reflection of neither of the underlying ensembles having sufficient synoptic-scale skill to provide useful forecasts beyond that lead time, rather than a practical limit on the range at which a CAM ensemble adds value.

TreeInterpreter is used to decompose the RF forecasts into contributions from individual predictors. Across all forecast lead days, UH has the greatest contribution to the Global-Nest RF severe convection probabilities. The relative contribution of UH to Global-Nest RF probabilities does not change substantially even as severe weather forecast skill declines with increasing lead time. Kinematic predictors are generally more important than thermodynamic predictors, particularly in the Global RF. The increased importance of kinematic predictors may result from our training only during the cool season, when kinematic variables tend to be stronger and severe events

are more closely associated with strong synoptic-scale systems (e.g., Brooks, 2009), or it may also be a representation of improved forecast skill of these predictors in the C-SHiELD ensemble compared to thermodynamic predictors.

Although our methodology follows from previous literature on machine learning forecasts of extended-range severe convection (e.g., Hill et al., 2023; Clark et al., 2025), it may be possible to improve upon the severe convection forecast model used in this study, either through use of different predictor variables, using a different machine learning approach entirely, or comparing ensemble systems with different configurations and/or CAM dynamical cores. The results presented here thus do not define the ceiling of how much severe convection forecasts can be improved at medium- to long-range lead times from using ensemble CAM output instead of output from a convection-parameterized ensemble.

There are limitations to this study that provide context to our results and indicate avenues for additional research. The comparison presented here is limited to forecasts initialized once per week from October to early June with a sample size of only 153 unique forecast initializations over 4 years. The small sample size—a consequence of computational limitations—leads to some unintuitive results in the objective verification, such as improved RF forecast skill at day 3 compared to day 2. We hypothesize that over a larger sample the objective verification would smooth out and skill would decrease with each increasing lead day. Because of the jumps in day-to-day verification metrics, statistical significance for any individual lead day should be viewed cautiously. However, given the preponderance of lead days where the Global-Nest RF is better than the Global RF between days 1 and 7, we are confident that the Global-Nest RF is an improvement over the Global RF through day 7. An analysis of whether this improvement quantified here justifies the additional expenses of running an extended-range CAM ensemble operationally is outside the scope of this study.

Another limitation to our study is our binary classification of each grid point on a given convective day using the presence of at least one severe storm report. While the spatial coherence of severe grid points can indicate a widespread severe event, the RFs otherwise have no information as to the scope of the severe convection—a grid box with a single severe weather report is treated equally to a grid box with multiple, and perhaps significant, severe reports. Although not feasible with our current sample size, future work applying this methodology to compare RFs trained to specifically forecast the probability of severe storm outbreaks (e.g., Doswell III et al., 2006) in the medium- to long-range would be insightful.

There are additional avenues for future work that can further explore the question of whether a CAM ensemble can provide value to extended-range severe convection forecasts. Some examples include performing verification of model output of relevant environmental fields (e.g., MUCAPE) against reanalysis fields, identifying potential synoptic-scale factors that may result in improved (or degraded) extended-range forecasts (e.g., Brannan et al., 2022), and comparing the position and strength of convection-initiating boundaries (e.g., drylines) and how errors in a forecasted boundary influence downstream predictability.

Using FrontFinder AI to identify fronts, the Global-Nest ensemble shows marginal skill forecasting the occurrence of drylines through day 8 from March–May. However, the ensemble members have a low POD and high FAR, even at short lead times, when spatially verifying forecasted dryline probabilities compared to probabilities from applying the ML dryline model to ERA5. Further work is needed to determine if the spatial errors are the result of a systematic bias in the forecast system, including an assessment of the displacement between Global-Nest and ERA5 dryline-objects and expanding the evaluation to 1200, 1800, and 0600 UTC to evaluate the diurnal cycle of forecasted dryline-objects. Additionally, applying the ML dryline identification model

to the Global ensemble and comparing RF severe thunderstorm forecast skill between the Global and Global-Nest can provide insight to the effect boundary placement on both short and extended-range severe thunderstorm forecasts.

Chapter 6

Summary and Conclusions

Each Chapter of this dissertation represents a step towards the overarching motivation of improving dryline-initiated severe convection forecasts across short-to-medium range forecasts, S2S forecasts, and future climate projections. Chapter 2 built a 16-year observed dryline climatology across the entire year for the entire contiguous United States. In addition to serving as a baseline for dryline studies in the US outside of the springtime Southern Great Plains (SGP), Chapter 2 also established the utility of FrontFinder AI for dryline climatological work. Chapter 3 builds off of this climatology by quantifying the relationship of dryline-associated convection with the springtime peak in severe storms in the SGP. With over 40% of all evening SGP severe reports associated with a dryline, and over 60% of all severe days across west/northwest Texas and southwest Oklahoma associated with a dryline, it may be feasible to produce long-range and S2S severe storm forecasts using dryline variability factors.

Chapter 4 explores dryline variability from 1940–2020 related to SGP drought and the El Niño–Southern Oscillation (ENSO). Drylines are both more frequent and further east for most spring months with SGP drought. While the relationship with ENSO is weaker, early season La Niña is associated with increased dryline frequency. Future work will be needed to determine if the early season increases in tornadoes, and particularly hail, in the SGP during La Niña (Allen et al., 2015a; Cook et al., 2017) is the result of dryline-initiated convection. Chapter 4 only represents a small portion of the work possible with the ML-identified dryline climatology. Future work can include

assessment of the dryline in relation to other forms of seasonal, annual, and decadal climate variability, assessing how factors such as SGP drought affect dryline diurnal evolution by including additional daily analysis times, and performing similar studies for different regions of the US (e.g., northern Great Plains; Mississippi River Valley) and for different seasons.

Chapter 5 addresses the motivation with a different approach from the previous Chapters. In Chapter 5, daily Random Forest severe thunderstorm forecasts from two experimental forecast ensembles—the Global and Global-Nest, from GFDL’s C-SHiELD—are compared out through forecast day 15. The Global-Nest RF, leveraging forecasted storm attribute output from its nested Convection-Allowing Model (CAM) over the CONUS, has improved forecasts compared to the Global RF through day 8. These results highlight the utility of CAMs for extended-range severe thunderstorm forecasts. The next step in this work is to examine the skill of these ensembles in forecasting drylines, and how dryline forecast skill correlates with RF severe thunderstorm forecast skill. Preliminary work applying the FrontFinder AI model to forecast data showed that the Global-Nest ensemble members were not very skillful in correctly predicting the location of the dryline at 0000 UTC, even for day 1 forecasts. Even when removing the spatial component of verification, the ensemble members only correctly forecasted about 50% of dryline days at short lead times.

Additional work to continue the goals of this dissertation would be to evaluate dryline forecast skill and potential systematic biases in current operational CAMs and convection-parameterized models, construct and test a S2S severe thunderstorm forecast system incorporating aspects of dryline variability, and apply the FrontFinder AI model to global climate models to compare future projections against the observed dryline climatology. The entire year, full-CONUS climatology built in Chapter 2 can be used to compare to future dryline projections in other studies such as Scaff et al.

(2021). Given the significant role of the dryline in the current severe thunderstorm climatology of the SGP, exploring how dryline frequency and longitude may change in future climate scenarios may be an insightful avenue to explore potential changes in the future severe storm climatology. Some studies have projected the central US to become more arid throughout the 21st century (e.g., Seager et al., 2018), which, based on the results from Chapter 4, could lead to drylines being more frequent and propagating further east. This trend could influence the seasonal frequency and location of dryline-initiated severe thunderstorms, and is a ripe avenue for future work.

While this dissertation is focused primarily on the relationship between drylines and severe storms, the dryline is also important for the broader hydroclimate of the central US. Additional applications that can be explored in future work include seasonal and future climate projections of the water budget, energy demand, and effects on agriculture.

Reference List

- Achter, T. H., and L. H. Horn, 1986: Spring season Colorado cyclones. part I: Use of composites to relate upper and lower tropospheric wind fields. *J. Appl. Meteor. Climatol.*, **25**, 732–743, [https://doi.org/10.1175/1520-0450\(1986\)025<0732:SSCCPI>2.0.CO;2](https://doi.org/10.1175/1520-0450(1986)025<0732:SSCCPI>2.0.CO;2).
- Allen, J., M. Tippett, and A. Sobel, 2015a: Influence of the El Niño/Southern Oscillation on tornado and hail frequency in the United States. *Nat. Geosci.*, **8**, 278–283, <https://doi.org/10.1038/ngeo2385>.
- Allen, J. T., and M. K. Tippett, 2015b: The characteristics of United States hail reports: 1955–2014. *Electronic J. Severe Storms Meteor.*, **10**, 1–31, <https://doi.org/10.55599/ejssm.v10i3.60>.
- Anderson, C. J., C. K. Wickle, Q. Zhou, and J. A. Royle, 2007: Population influences on tornado reports in the United States. *Wea. Forecasting*, <https://doi.org/10.1175/WAF997.1>.
- Andrews, M. S., V. A. Gensini, A. M. Haberlie, W. S. Ashley, A. C. Michaelis, and M. Taszarek, 2024: Climatology of the elevated mixed layer over the contiguous United States and northern Mexico using ERA5: 1979–2021. *J. Climate*, **37**, 1833–1851, <https://doi.org/10.1175/JCLI-D-23-0517.1>.
- Arkin, P. A., 1982: The relationship between interannual variability in the 200 mb tropical wind field and the Southern Oscillation. *Mon. Wea. Rev.*, **110**, 1393–1404, [https://doi.org/10.1175/1520-0493\(1982\)110<1393:TRBIVI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<1393:TRBIVI>2.0.CO;2).
- Armstrong, R. A., 2014: When to use the Bonferroni correction. *Ophthalmic Physiol Opt*, **34**, 502–508, <https://doi.org/10.1111/opo.12131>.
- Baldwin, M. E., J. S. Kain, and M. P. Kay, 2002: Properties of the convection scheme in NCEP’s Eta Model that affect forecast sounding interpretation. *Wea. Forecasting*, **17**, 1063–1079, [https://doi.org/10.1175/1520-0434\(2002\)017<1063:POTCSI>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<1063:POTCSI>2.0.CO;2).
- Barnes, L. R., E. C. Gruntfest, M. H. Hayden, D. M. Schultz, and C. Benight, 2007: False alarms and close calls: A conceptual model of warning accuracy. *Wea. Forecasting*, **22**, 1140–1147, <https://doi.org/10.1175/WAF1031.1>.
- Barret, K. M., 2012: The spatial distribution of contiguous united states thunderstorm

- related short-fuse severe weather warnings. Ph.D. thesis, Texas State university, 248 pp., [Available online at <https://digital.library.txstate.edu/handle/10877/4293>].
- Bell, G. D., and L. F. Bosart, 1994: Midtropospheric closed cyclone formation over the southwestern United States, the eastern United States, and the Alps. *Mon. Wea. Rev.*, **122**, 791–813, [https://doi.org/10.1175/1520-0493\(1994\)122<0791:MCCFOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<0791:MCCFOT>2.0.CO;2).
- Berg, R., J. Clark, D. Roth, and T. Birchard, 2007: The national weather service unified surface analysis. *22nd Conf. on Weather and Analysis Forecasting/18th Conf. on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 6A.4, https://ams.confex.com/ams/22WAF18NWP/techprogram/paper_124199.htm.
- Bergmaier, P. T., and B. Geerts, 2015: Characteristics and synoptic environment of drylines occurring over the higher terrain of southeastern Wyoming. *Wea. Forecasting*, **30**, 1733–1748, <https://doi.org/10.1175/WAF-D-15-0061.1>.
- Bluestein, H. B., and S. S. Parker, 1993: Modes of isolated, severe convective storm formation along the dryline. *Mon. Wea. Rev.*, **121**, 1354–1372, [https://doi.org/10.1175/1520-0493\(1993\)121<1354:MOISCS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1993)121<1354:MOISCS>2.0.CO;2).
- Brannan, A. L., S. Gertonson, I. L. Jirak, B. Robers, A. D. Lyons, and J. L. Guyer, 2022: Predictability of large-scale patterns that lead to tornado outbreaks at lead times out to 14 days. *30th Conf. Severe Local Storms*, Santa Fe, NM, Amer. Meteor. Soc., 9.2.
- Breiman, L., 2001: Random forests. *Machine Learning*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Brooks, H. E., 2009: Proximity soundings for severe convection for Europe and the United States from reanalysis data. *J. Atmos. Res.*, **93**, 546–553, <https://doi.org/10.1016/j.atmosres.2008.10.005>.
- Brooks, H. E., C. A. Doswell III, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640, [https://doi.org/10.1175/1520-0434\(2003\)018<0626:CEOLDT>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0626:CEOLDT>2.0.CO;2).
- Buizza, R., A. Hollingsworth, F. Lalaurette, and A. Ghelli, 1999: Probabilistic predictions of precipitation using the ECMWF ensemble prediction system. *Wea. Forecasting*, **14**, 168–189, [https://doi.org/10.1175/1520-0434\(1999\)014<0168:PPOPOT>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0168:PPOPOT>2.0.CO;2).
- Burke, A., N. Snook, S. Gagne, D. J. II McCorkle, and A. McGovern, 2020: Calibration

- of machine learning-based probabilistic hail predictions for operational forecasting. *Wea. Forecasting*, **35**, 149–168, <https://doi.org/10.1175/WAF-D-19-0105.1>.
- Changnon, S. A., 2001: Thunderstorm rainfall in the conterminous United States. *Bull. Amer. Meteor. Soc.*, **82**, 1925–1940, [https://doi.org/10.1175/1520-0477\(2001\)082<1925:TRITCU>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<1925:TRITCU>2.3.CO;2).
- Chasteen, M. B., and S. E. Koch, 2022: Multiscale aspects of the 22–27 April 2011 tornado outbreak. part II: Environmental modifications and upscale feedbacks arising from latent processes. *Mon. Wea. Rev.*, **150**, 337–368, <https://doi.org/10.1175/MWR-D-21-0014.1>.
- Chen, T.-C., F. Collet, and A. Di Luca, 2024: Evaluation of ERA5 precipitation and 10-m wind speed associated with extratropical cyclones using station data over North America. *International Journal of Climatology*, **44**, 729–747, <https://doi.org/10.1002/joc.8339>.
- Clark, A. J., and E. D. Loken, 2022: Machine learning-derived severe weather probabilities from a Warn-on-Forecast system. *Wea. Forecasting*, **37**, 1721–1740, <https://doi.org/10.1175/WAF-D-22-0056.1>.
- Clark, A. J., A. MacKenzie, A. McGovern, V. Lakshmanan, and R. A. Brown, 2015: An automated, multiparameter dryline identification algorithm. *Wea. Forecasting*, **30**, 1781–1794, <https://doi.org/10.1175/WAF-D-15-0070.1>.
- Clark, A. J., and Coauthors, 2023a: The third real-time, virtual spring forecasting experiment to advance severe weather prediction capabilities. *Bull. Amer. Meteor.*, **104**, E456–E458, <https://doi.org/10.1175/BAMS-D-22-0213.1>.
- Clark, A. J., and Coauthors, 2023b: The first hybrid NOAA hazardous weather testbed spring forecasting experiment for advancing severe weather prediction. *Bull. Amer. Meteor.*, **104**, E2305–E2307, <https://doi.org/10.1175/BAMS-D-23-0275.1>.
- Clark, J., A. K. A. Hoogewind, A. J. Hill, E. D. Loken, and M. J. Hosek, 2025: Extended range machine-learning severe weather guidance based on the operational GEFS. *Wea. Forecasting*, **40**, 1325–1343, <https://doi.org/10.1175/WAF-D-24-0116.1>.
- Clark, J. H. E., 1990: An observational and theoretical study of Colorado lee cyclogenesis. *J. Atmos. Sci.*, **47**, 1541–1561, [https://doi.org/10.1175/1520-0469\(1990\)047<1541:AOATSO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1990)047<1541:AOATSO>2.0.CO;2).
- Clough, S. A., M. W. Shephard, E. J. Mlawer, J. S. Delamere, M. J. Iacono, K. Cady-

- Pereira, S. Boukabara, and P. D. Brown, 2005: Atmospheric radiative transfer modeling: A summary of the AER codes. *J. Quant. Spectrosc. Radiat. Transf.*, **91**, 233–244, <https://doi.org/10.1016/j.jqsrt.2004.05.058>.
- Coffer, B. E., L. C. Maudlin, P. G. Veals, and A. J. Clark, 2013: Dryline position errors in experimental convection-allowing NSSL-WRF model forecasts and the operational NAM. *Wea. Forecasting*, **28**, 746–761, <https://doi.org/10.1175/WAF-D-12-00092.1>, URL <https://journals.ametsoc.org/doi/10.1175/WAF-D-12-00092.1>.
- Coffer, B. E., M. D. Parker, R. L. Thompson, B. T. Smith, and R. E. Jewell, 2019: Using near-ground storm relative helicity in supercell tornado forecasting. *Wea. Forecasting*, **34**, 1417–1435, <https://doi.org/10.1175/WAF-D-19-0115.1>.
- Coniglio, M. C., K. L. Elmore, J. S. Kain, S. J. Weiss, M. Xue, and M. L. Weisman, 2010: Evaluation of WRF model output for severe weather forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment. *Wea. Forecasting*, **25**, 408–427, <https://doi.org/10.1175/2009WAF2222258.1>.
- Coniglio, M. C., and M. D. Parker, 2020: Insights into supercells and their environments from three decades of targeted radiosonde observations. *Mon. Wea. Rev.*, **148**, 4893–4915, <https://doi.org/10.1175/MWR-D-20-0105.1>.
- Cook, A. R., L. M. Leslie, D. B. Parsons, and J. T. Schaefer, 2017: The impact of El Niño–Southern Oscillation (ENSO) on Winter and early Spring U.S. tornado outbreaks. *J. Appl. Meteor. Climatol.*, **56**, 2455–2478, <https://doi.org/10.1175/JAMC-D-16-0249.1>.
- Cook, A. R., and J. T. Schaefer, 2008: The relation of El Niño–Southern Oscillation (ENSO) to Winter tornado outbreaks. *Mon. Wea. Rev.*, **136**, 3121–3137, <https://doi.org/10.1175/2007MWR2171.1>.
- Dai, A., 2017: Dai global Palmer Drought Severity Index (PDSI). Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, <https://doi.org/10.5065/D6QF8R93>.
- Dai, A., 2021: Hydroclimatic trends during 1950–2018 over global land. *Clim. Dyn.*, **56**, 4027–4049, <https://doi.org/10.1007/s00382-021-05684-1>.
- Dai, A., and National Center for Atmospheric Research Staff (Eds.), 2023: The Climate Data Guide: Palmer Drought Severity Index (PDSI). Accessed: 2025-05-01, <https://climatedataguide.ucar.edu/climate-data/palmer-drought-severity-index-pdsi>.
- Dial, G. L., J. P. Racy, and R. L. Thompson, 2010: Short-term convective mode evo-

- lution along synoptic boundaries. *Wea. Forecasting*, **25**, 1430–1446, <https://doi.org/10.1175/2010WAF2222315.1>.
- Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmosph. Sci. Lett.*, **5**, 110–117, <https://doi.org/10.1002/asl.72>.
- Doswell III, C. A., H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Wea. Forecasting*, **20**, 577–595, <https://doi.org/10.1175/WAF866.1>.
- Doswell III, C. A., R. Edwards, R. L. Thompson, J. A. Hart, and K. C. Crosbie, 2006: A simple and flexible method for ranking severe weather events. *Wea. Forecasting*, **21**, 939–951, <https://doi.org/10.1175/WAF959.1>.
- Doswell III, C. A., S. J. Weiss, and R. H. Johns, 1993: *In The tornado: Its structure, dynamics, prediction, and hazards*, Monograph Series, Vol. 79. Amer. Geophys. Union, <https://doi.org/10.1029/GM079p0557>.
- Duda, J. D., and W. A. Gallus, 2010: Spring and summer midwestern severe weather reports in supercells compared to other morphologies. *Wea. Forecasting*, **25**, 190–206, <https://doi.org/10.1175/2009WAF2222338.1>.
- Duell, R. S., and M. S. Van Den Broeke, 2016: Climatology, synoptic conditions, and misanalyses of Mississippi River Valley drylines. *Mon. Wea. Rev.*, **144**, 927–943, <https://doi.org/10.1175/MWR-D-15-0108.1>.
- Eckel, F. A., and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350, <https://doi.org/10.1175/WAF843.1>.
- Edwards, R., H. E. Brooks, and H. Cohn, 2021: Changes in tornado climatology accompanying the Enhanced Fujita Scale. *J. Appl. Meteor. Climatol.*, **60**, 1465–1482, <https://doi.org/10.1175/JAMC-D-21-0058.1>.
- Edwards, R., J. G. LaDue, J. T. Ferree, K. Scharfenberg, C. Maier, and W. L. Coulbourne, 2013: Tornado intensity estimation: Past, present, and future. *Bull. Amer. Meteor. Soc.*, **94**, 641–653, <https://doi.org/10.1175/BAMS-D-11-00006.1>.
- Eichler, T., and W. Higgins, 2006: Climatology and ENSO-related variability of North American extratropical cyclone activity. *J. Climate*, **19**, 2076–2093, <https://doi.org/10.1175/JCLI3725.1>.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and

- J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, <https://doi.org/10.1029/2002JD003296>.
- Flora, M. L., C. K. Potvin, A. McGovern, and S. Handler, 2024: A machine learning explainability tutorial for atmospheric sciences. *Artif. Intell. Earth Syst.*, **3**, <https://doi.org/10.1175/AIES-D-23-0018.1>.
- Flora, M. L., C. K. Potvin, P. S. Skinner, S. Handler, and A. McGovern, 2021: Using machine learning to generate storm-scale probabilistic guidance of severe weather hazards in the Warn-on-Forecast system. *Mon. Wea. Rev.*, **149**, 1535–1557, <https://doi.org/10.1175/MWR-D-20-0194.1>.
- Fritsch, J. M., R. J. Kane, and C. R. Chelius, 1986: The contribution of mesoscale convective weather systems to the warm-season precipitation in the United States. *J. Climate Appl. Meteor.*, **25**, 1333–1345, [https://doi.org/10.1175/1520-0450\(1986\)025<1333:TCOMCW>2.0.CO;2](https://doi.org/10.1175/1520-0450(1986)025<1333:TCOMCW>2.0.CO;2).
- Fujita, T., 1958: Structure and movement of a dry front. *Bull. Amer. Meteor. Soc.*, **39**, 574–582, <https://doi.org/10.1175/1520-0477-39.11.574>.
- Fujita, T. T., 1971: Proposed characterization of tornadoes and hurricanes by area and intensity. Tech. rep., University of Chicago SMRP Research Paper 91, 42 pp.
- Gallo, B. T., and Coauthors, 2020: Exploring convection-allowing model evaluation strategies for severe local storms using the Finite-Volume Cubed-Sphere (FV3) model core. *Wea. Forecasting*, **36**, 3–19, <https://doi.org/10.1175/WAF-D-20-0090.1>.
- Gallus, W. A., N. A. Snook, and E. V. Johnson, 2008: Spring and summer severe weather reports over the midwest as a function of convective mode: A preliminary study. *Wea. Forecasting*, **23**, 101–113, <https://doi.org/10.1175/2007WAF2006120.1>, URL <https://journals.ametsoc.org/doi/10.1175/2007WAF2006120.1>.
- Galway, J. G., 1979: Relationship between precipitation and tornado activity. *Water Resour. Res.*, **15**, 961–964, <https://doi.org/10.1029/WR015i004p00961>.
- Gensini, V. A., and W. S. Ashley, 2011: Climatology of potentially severe convective environments from North American regional reanalysis. *Electronic J. Severe Storms Meteor.*, **6**, 1–40, <https://doi.org/10.55599/ejssm.v6i8.35>.
- Gensini, V. A., D. Gold, J. T. Allen, and B. S. Barrett, 2019: Extended U.S. tornado outbreak during late May 2019: A forecast of opportunity. *Geophysical Research Letters*, **46**, 10 150–10 158, <https://doi.org/10.1029/2019GL084470>.

- Gensini, V. A., A. M. Haberlie, and P. T. Marsh, 2020: Practically perfect hindcasts of severe convective storms. *Bull. Amer. Meteor. Soc.*, **101**, E1259–E1278, <https://doi.org/10.1175/BAMS-D-19-0321.1>.
- Gneiting, T., and A. E. Raftery, 2005: Weather forecasting with ensemble methods. *Science*, **310**, 248–249, <https://doi.org/10.1126/science.1115255>, URL <https://www.science.org/doi/10.1126/science.1115255>.
- Grasso, L. D., 2000: A numerical simulation of dryline sensitivity to soil moisture. *Mon. Wea. Rev.*, **128**, 2816–2834, [https://doi.org/10.1175/1520-0493\(2000\)128<2816:ANSODS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2816:ANSODS>2.0.CO;2).
- Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, [https://doi.org/10.1175/1520-0434\(1999\)014<0155:HTFENP>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2).
- Hamill, T. M., 2003: Evaluating forecasts’ rules of thumb: A study of $d(\text{prog})/dt$. *Wea. Forecasting*, **18**, 933–937, [https://doi.org/10.1175/1520-0434\(2003\)018<0933:EFROTA>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0933:EFROTA>2.0.CO;2).
- Hamill, T. M., and Coauthors, 2022: The reanalysis for the Global Ensemble Forecast System, Version 12. *Mon. Wea. Rev.*, **150**, 59–79, <https://doi.org/10.1175/MWR-D-21-0023.1>.
- Han, J., W. Wang, Y. C. Kwon, S.-Y. Hong, V. Tallapragada, and F. Yang, 2017: Updates in the NCEP GFS cumulus convection schemes with scale and aerosol awareness. *Wea. Forecasting*, **32**, 2005–2017, <https://doi.org/10.1175/WAF-D-17-0046>.
- Hane, C. E., 2004: Quiescent and synoptically-active drylines: A comparison based upon case studies. *Meteorology and Atmospheric Physics*, **86**, 195–211, <https://doi.org/10.1007/s00703-003-0026-y>.
- Harris, L., and Coauthors, 2020: GFDL SHIELD: A unified system for weather-to-seasonal prediction. *Journal of Advances in Modeling Earth Systems*, **12**, <https://doi.org/10.1029/2020MS002223>.
- Harris, L. M., and S.-J. Lin, 2013: A two-way nested global-regional dynamical core on the cubed-sphere grid. *Mon. Wea. Rev.*, **141**, 283–306, <https://doi.org/10.1175/MWR-D-11-00201.1>.
- Harris, L. M., S. L. Rees, M. Morin, L. Zhou, and W. F. Stern, 2019: Explicit prediction of continental convection in a skillful variable-resolution global model. *Journal of Advances in Modeling Earth Systems*, 1847–1869, <https://doi.org/>

10.1029/2018MS001542.

- henderson, D. S., J. A. Otkin, and J. R. Mecikalski, 2022: Examining the role of the land surface on convection using high-resolution model forecasts over the southeastern United States. *Journal of Geophysical Research: Atmospheres*, **127**, e2022JD036563, <https://doi.org/10.1029/2022JD036563>.
- Hersbach, H., and Coauthors, 2018: ERA5 hourly data on pressure levels from 1940 to present. Copernicus Climate Change Service climate data Store, accessed 15 March 2021, <https://doi.org/10.24381/cds.bd0915c6>.
- Hill, A. J., R. S. Schumacher, and I. L. Jirak, 2023: A new paradigm for medium-range severe weather forecasts: Probabilistic random forest-based predictions. *Wea. Forecasting*, **38**, 251–272, <https://doi.org/10.1175/WAF-D-22-0143.1>.
- Hitchens, N. M., H. E. Brooks, and M. P. Kay, 2013: Objective limits on forecasting skill of rare events. *Wea. Forecasting*, **28**, 525–534, <https://doi.org/10.1175/WAF-D-12-00113.1>.
- Hobbs, P. V., J. D. Locatelli, and J. E. Martin, 1996: A new conceptual model for cyclones generated in the lee of the Rocky Mountains. *Bull. Amer. Meteor. Soc.*, **77**, 1169–1178, [https://doi.org/10.1175/1520-0477\(1996\)077<1169:ANCMFC>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<1169:ANCMFC>2.0.CO;2).
- Hoch, J., and P. Markowski, 2005: A climatology of springtime dryline position in the U.S. Great Plains Region. *J. Climate*, **18**, 2132–2137, <https://doi.org/10.1175/JCLI3392.1>.
- Hollingsworth, A., K. Arpe, M. Tiedtke, M. Capaldo, and H. Savijarvi, 1980: The performance of a medium-range forecast model in winter-impact of physical parameterizations. *Mon. Wea. Rev.*, **108**, 1736–1773, [https://doi.org/10.1175/1520-0493\(1980\)108<1736:TPOAMR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1736:TPOAMR>2.0.CO;2).
- Hollingsworth, A., K. Arpe, M. Tiedtke, M. Capaldo, H. Savijarvi, O. Akesson, and J. A. Woods, 1979: Comparison of medium range forecasts made with two parameterizations. Tech. rep. 13, ECMWF.
- Hong, S.-Y., 2010: A new stable boundary-layer mixing scheme and its impact on the simulated East Asian summer monsoon. *Quart. J. Royal Meteorol. Soc.*, **136**, 1481–1496, <https://doi.org/10.1002/qj.665>.
- Hoogewind, K. A., 2012: A proposed method for objectively identifying and characterizing frontal zones. M.S. thesis, Purdue University, 267 pp., [Available online at <https://docs.lib.purdue.edu/dissertations/AAI1530135>].

- Hosek, M. J., K. A. Hoogewind, A. J. Clark, L. Harris, K.-Y. Cheng, and L. Zhou, 2026a: Quantifying the value of convection-allowing global-nested ensembles for extended-range severe weather forecasting. *Wea. Forecasting.*, In Press.
- Hosek, M. J., K. A. Hoogewind, A. J. Clark, and A. D. Justin, 2026b: Application of machine learning to quantify the role of the dryline in the severe thunderstorm climatology of the US Southern Great Plains. *Wea. Forecasting*, Submitted Oct. 1 2025.
- Hosek, M. J., K. A. Hoogewind, A. J. Clark, A. D. Justin, and J. T. Allen, 2025: A 16-year climatology of WPC-analyzed drylines and their association with severe convection. *J. Appl. Meteor. Climatol.*, **64**, 383–399, <https://doi.org/10.1175/JAMC-D-24-0124.1>.
- Hoskins, B. J., and F. P. Bretherton, 1972: Atmospheric frontogenesis models: Mathematical formulation and solution. *J. Atmos. Sci.*, **29**, 11–37, [https://doi.org/10.1175/1520-0469\(1972\)029<0011:AFMMFA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029<0011:AFMMFA>2.0.CO;2).
- Hsu, W.-R., and A. H. Murphy, 1986: The attributes diagram: A geometrical framework for assessing the quality of probability forecasts. *Int. J. Forecasting*, **2**, 285–293.
- Huang, H., and Coauthors, 2020: UNet 3+: A full-scale connected unet for medical image segmentation. *arXiv*, <https://doi.org/10.48550/arxiv.2004.08790>.
- IEM, 2023: Archived NWS watch/warnings. Iowa Environmental Mesonet, accessed 15 May 2023, <https://mesonet.agron.iastate.edu/request/gis/watchwarn.phtml>.
- Johns, R. H., and C. A. Doswell III, 1992: Severe local storms forecasting. *Wea. Forecasting*, **7**, 588–612, [https://doi.org/10.1175/1520-0434\(1992\)007<0588:SLSF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1992)007<0588:SLSF>2.0.CO;2).
- Johnson, Z. F., and N. M. Hitchens, 2018: Effects of soil moisture on the longitudinal dryline position in the Southern Great Plains. *J. Hydrometeor.*, **19**, 273–287, <https://doi.org/10.1175/JHM-D-17-0091.1>.
- Jones, P. W., 1999: First- and second-order conservative remapping schemes for grids in spherical coordinates. *Mon. Wea. Rev.*, **127**, 2204–2210, [https://doi.org/10.1175/1520-0493\(1999\)127<2204:FASOCR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2).
- Justin, A. D., A. McGovern, and J. T. Allen, 2025: FrontFinder AI: Efficient identification of frontal boundaries over the continental United States and NOAA’s Unified Surface Analysis domain using the UNET3+ model architecture. *Artif. Intell. Earth Sys.*, **4**, e240 043, <https://doi.org/10.1175/AIES-D-24-0043.1>.

- Justin, A. D., C. Willingham, A. McGovern, and J. T. Allen, 2023: Toward operational real-time identification of frontal boundaries using machine learning. *Artif. Intell. Earth Sys.*, **2**, e220052, <https://doi.org/10.1175/AIES-D-22-0052.1>.
- Kain, J. S., S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high-resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536–1542, <https://doi.org/10.1175/2010WAF2222430.1>.
- Kain, J. S., S. J. Weiss, J. J. Levit, M. E. Baldwin, and D. R. Bright, 2006: Examination of convection-allowing configuration of the WRF model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Wea. Forecasting*, **21**, 167–181, <https://doi.org/10.1175/WAF906.1>.
- Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, <https://doi.org/10.1175/WAF2007106.1>.
- Karyampudi, V. M., M. L. Kaplan, S. E. Koch, and R. J. Zamora, 1995: The influence of the Rocky Mountain on the 13–14 April 1986 severe weather outbreak. part I: Mesoscale lee cyclogenesis and its relationship to severe weather and dust storms. *Mon. Wea. Rev.*, **123**, 1394–1422, [https://doi.org/10.1175/1520-0493\(1995\)123<1394:TIOTRM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<1394:TIOTRM>2.0.CO;2).
- Kendall, M. G., 1938: A new measure of rank correlation. *Biometrika*, **30**, 81–93, <https://doi.org/10.1093/biomet/30.1-2.81>.
- Kohonen, T., 2001: The basic SOM. *Self-Organizing Maps*, Springer Series in Information Sciences, vol 30. Springer, Berlin, Heidelberg., 105–176, https://doi.org/10.1007/978-3-642-56927-2_3.
- Krocak, M. J., and H. E. Brooks, 2020: An analysis of subdaily severe thunderstorm probabilities for the United States. *Wea. Forecasting*, **35**, 107–112, <https://doi.org/10.1175/WAF-D-19-0145.1>.
- Kruskal, W. H., and W. A. Wallis, 1952: Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, **47**, 583–621, <https://doi.org/10.1080/01621459.1952.10483441>.
- Kuchera, E. L., and M. D. Parker, 2006: Severe convective wind environments. *Wea. Forecasting*, **21**, 595–612, <https://doi.org/10.1175/WAF931.1>.
- Lagerquist, R., J. T. Allen, and A. McGovern, 2020: Climatology and variability of

- warm and cold fronts over North America from 1979 to 2018. *J. Climate*, **33**, 6531–6554, <https://doi.org/10.1175/JCLI-D-19-0680.1>.
- Lagerquist, R., A. McGovern, and T. Smith, 2017: Machine learning for real-time prediction of damaging straight-line convective wind. *Wea. Forecasting*, **32**, 2175–2193, <https://doi.org/10.1175/WAF-D-17-0038.1>.
- Lee, J. T., and J. G. Galway, 1956: Preliminary report on the relationship between the jet at the 200-mb level and tornado occurrence. *Bull. Amer. Meteor. Soc.*, **37**, 327–332, <https://doi.org/10.1175/1520-0477-37.7.327>.
- Lepore, C., M. K. Tippett, and J. T. Allen, 2017: ENSO-based probabilistic forecasts of March–May U.S. tornado and hail activity. *Geophys. Res. Lett.*, **44**, 9039–9101, <https://doi.org/10.1002/2017GL074781>.
- Li, F., D. R. Chavas, K. A. Reed, N. Rosenbloom, and D. T. Dawson II, 2021: The role of elevated terrain and the Gulf of Mexico in the production of severe local storm environments over North America. *J. climate*, **34**, 7799–7819, <https://doi.org/10.1175/JCLI-D-20-0607.1>.
- Lilly, D. K., 1990: Numerical prediction of thunderstorms—has its time come? *Quart. J. Royal Meteor. Soc.*, **116**, 779–798, <https://doi.org/10.1002/qj.49711649402>.
- Locatelli, J. D., M. T. Stoeling, and P. V. Hobbs, 2002: A new look at the Super Outbreak of tornadoes on 3–4 April 1974. *Mon. Wea. Rev.*, **130**, 1633–1651, [https://doi.org/10.1175/1520-0493\(2002\)130<1633:ANLATS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1633:ANLATS>2.0.CO;2).
- Loken, E. D., A. J. Clark, and C. D. Karstens, 2020: Generating probabilistic next-day severe weather forecasts from convection-allowing ensembles using random forests. *Wea. Forecasting*, **35**, 1605–1631, <https://doi.org/10.1175/WAF-D-19-0258.1>.
- Loken, E. D., A. J. Clark, and A. McGovern, 2022: Comparing and interpreting differently designed random forests for next-day severe weather hazard prediction. *Wea. Forecasting*, **37**, 871–899, <https://doi.org/10.1175/WAF-D-21-0138.1>.
- 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>.
- Mann, H. B., and D. R. Whitney, 1947: On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.*, **18**, 50–60, <https://doi.org/10.1214/aoms/1177730491>.

- Mariotti, A., and Coauthors, 2020: Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull. Amer. Meteor. Soc.*, **101**, E608–E625, <https://doi.org/10.1175/BAMS-D-18-0326.1>.
- Markowski, P. M., and Y. P. Richardson, 2010: *Mesoscale Meteorology in Midlatitudes*, Wiley-Blackwell, 407 pp.
- Marshment, R. A., and L. H. Horn, 1986: Spring season Colorado cyclones. part II. composites of atmospheric moisture and moist static stability. *J. Appl. Meteor. Climatol.*, **25**, 744–752, [https://doi.org/10.1175/1520-0450\(1986\)025<0744:SSCCPI>2.0.CO;2](https://doi.org/10.1175/1520-0450(1986)025<0744:SSCCPI>2.0.CO;2).
- Marzban, C., and J. T. Schaefer, 2021: The correlation between U.S. tornadoes and Pacific sea surface temperatures. *Mon. Wea. Rev.*, **129**, 884–895, [https://doi.org/10.1175/1520-0493\(2001\)129<0884:TCBUST>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0884:TCBUST>2.0.CO;2).
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Mason, S. J., 2004: On using “climatology” as a reference strategy in the Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **132**, 1891–1895, [https://doi.org/10.1175/1520-0493\(2004\)132<1891:OUCAAR>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1891:OUCAAR>2.0.CO;2).
- Mattocks, C., and R. Bleck, 1986: Jet streak dynamics and geostrophic adjustment processes during the initial states of lee cyclogenesis. *Mon. Wea. Rev.*, **114**, 2033–2056, [https://doi.org/10.1175/1520-0493\(1986\)114<2033:JSDAGA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<2033:JSDAGA>2.0.CO;2).
- Mazurek, A. C., A. J. Hill, R. S. Schumacher, and H. J. McDaniel, 2025: Can ingredients-based forecasting be learned? disentangling a random forest’s severe weather predictions. *Wea. Forecasting*, **40**, 237–258, <https://doi.org/10.1175/WAF-D-23-0193.1>.
- McCabe, G. J., M. A. Palecki, and J. L. Betancourt, 2004: Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *Proc. Natl. Acad. Sci.*, **101**, 4136–4141, <https://doi.org/10.1073/pnas.0306738101>.
- McCarthy, J., and S. E. Koch, 1982: The evolution of an Oklahoma dryline. Part I: A meso- and subsynoptic-scale analysis. *J. Atmos. Sci.*, **39**, 225–236, [https://doi.org/10.1175/1520-0469\(1982\)039<0225:TEOAOD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1982)039<0225:TEOAOD>2.0.CO;2).
- McGovern, A., R. J. Chase, M. Flora, D. J. Gagne II, R. Lagerquist, C. K. Potvin, N. Snook, and E. Loken, 2023: A review of machine learning for convective weather. *Artif. Intell. Earth Syst.*, **2**, <https://doi.org/10.1175/AIES-D-22-0077.1>.

- Min, L., D. R. Fitzjarrald, Y. Du, B. E. J. Rose, J. Hong, and Q. Min, 2021: Exploring sources of surface bias in HRRR using New York State Mesonet. *Journal of Geophysical Research: Atmospheres*, **126**, e2021JD034989, <https://doi.org/10.1029/2021JD034989>.
- Mitchell, T., and D. M. Schultz, 2020: A synoptic climatology of spring dryline convection in the Southern Great Plains. *Wea. Forecasting*, **35**, 1561–1582, <https://doi.org/10.1175/WAF-D-19-0160.1>.
- Molina, M. J., J. T. Allen, and V. A. Gensini, 2018: The Gulf of Mexico and ENSO influence on subseasonal and seasonal CONUS winter tornado variability. *J. Appl. Meteor. Climatol.*, **57**, 2439–2463, <https://doi.org/10.1175/JAMC-D-18-0046.1>.
- Molnar, C., 2022: *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed. Molnar, 318 pp., <https://christophm.github.io/interpretable-ml-book/>.
- Moore, T. W., 2019: Seasonal frequency and spatial distribution of tornadoes in the United States and their relationship to the El Niño/Southern Oscillation. *Annals of the American Association of Geographers*, **109**, 1033–1051, <https://doi.org/10.1080/24694452.2018.1511412>.
- Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600, [https://doi.org/10.1175/1520-0450\(1973\)012<0595:ANVPOT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2).
- Murphy, A. H., 1993: What is a good forecast? an essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, [https://doi.org/10.1175/1520-0434\(1993\)008<0281:WIAGFA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2).
- Murphy, A. H., and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.*, **117**, 572–582, [https://doi.org/10.1175/1520-0493\(1989\)117<0572:SSACCI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<0572:SSACCI>2.0.CO;2).
- NCEI, 2025: Billion-Dollar Weather and Climate Disasters. <https://www.ncei.noaa.gov/access/billions/>.
- NCEP, 2024: Master List of NCEP Storage Grids. <https://www.nco.ncep.noaa.gov/pmb/docs/on388/tableb.html#GRID211>.
- Neiman, P. J., and R. M. Wakimoto, 1999: The interaction of a Pacific cold front with shallow air masses east of the Rocky Mountains. *Mon. Wea. Rev.*, **127**, 2102–2127, [https://doi.org/10.1175/1520-0493\(1999\)127<2102:TIOAPC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2102:TIOAPC>2.0.CO;2).

- Nixon, C. J., J. T. Allen, and M. Taszarek, 2023: Hodographs and Skew Ts of hail-producing storms. *Wea. Forecasting*, **38**, 2217–2236, <https://doi.org/10.1175/WAF-D-23-0031.1>.
- NOAA, 2013: Unified surface analysis manual. NOAA rep., 33 pp. <http://www.wpc.ncep.noaa.gov/sfcUASfcManualVersion1.pdf>.
- NOAA, 2023: NOAA unified surface analysis fronts. Zenodo, accessed 15 May 2023, <https://doi.org/10.5281/zenodo.7505021>.
- Nouri, N., N. Devineni, V. Were, and R. Khanbilvardi, 2021: Explaining the trends and variability in the United States tornado record using climate teleconnections and shifts in observational practices. *Sci Rep*, **11**, 1741, <https://doi.org/10.1038/s41598-021-81143-5>.
- Ogura, Y., and Y.-L. Chen, 1977: A life history of an intense mesoscale convective storm in Oklahoma. *J. Atmos. Sci.*, **34**, 1458–1476, [https://doi.org/10.1175/1520-0469\(1977\)034<1458:ALHOAI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1977)034<1458:ALHOAI>2.0.CO;2).
- Palmer, W. C., 1965: Meteorological drought. Office of Climatology Research Paper 45, Weather Bureau, Washington, 58 pp.
- Parsons, D. B., M. A. Shapiro, and E. Miller, 2000: The mesoscale structure of a nocturnal dryline and of a frontal-dryline merger. *Mon. Wea. Rev.*, **128**, 3824–3838, [https://doi.org/10.1175/1520-0493\(2001\)129<3824:TMSOAN>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<3824:TMSOAN>2.0.CO;2).
- Paulikas, M. J., 2014: Examining population bias relative to severe thunderstorm hazard reporting trends in the Atlanta, GA metropolitan region: Examining population bias relative to severe thunderstorm hazards in the US. *Met. Apps*, **21**, 494–503, <https://doi.org/10.1002/met.1394>.
- Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
- Perneger, T. V., 1998: What’s wrong with Bonferroni’s adjustment. *BMJ*, **316**, 1236–1238, <https://doi.org/10.1136/bmj.316.7139.1236>.
- Platt, J. C., 1999: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large Margin Classifiers*, A. J. Smola et al., Eds., MIT Press, 61–74.
- Pollard, R. T., P. B. Rhines, and R. O. R. Y. Thompson, 1973: The deepening of the wind-mixed layer. *Geophysical Fluid Dynamics*, **4**, 381–404, <https://doi.org/>

10.1080/03091927208236105.

- Potvin, C. K., and Coauthors, 2019: Systematic comparison of convection-allowing models during the 2017 NOAA HWT Spring Forecast Experiment. *Wea. Forecasting*, **34**, 1395–1416, <https://doi.org/10.1175/WAF-D-19-0056.1>.
- Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, **227**, 55–78, <https://doi.org/10.1016/j.jcp.2007.07.022>.
- Ray, P. S., P. Bieringer, X. Niu, and B. Whissel, 2003: An improved estimate of tornado occurrence in the Central Plains of the United States. *Mon. Wea. Rev.*, **131**, 1026–1031, [https://doi.org/10.1175/1520-0493\(2003\)131<1026:AIEOTO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<1026:AIEOTO>2.0.CO;2).
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperatures since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, <https://doi.org/10.1029/2002JD002670>.
- Rhea, J. O., 1966: A study of thunderstorm formation along dry lines. *J. Appl. Meteor.*, **5**, 58–63, [https://doi.org/10.1175/1520-0450\(1966\)005<0058:ASOTFA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1966)005<0058:ASOTFA>2.0.CO;2).
- Rudeva, I., and I. Simmonds, 2015: Variability and trends of global atmospheric frontal activity and links with large-scale modes of variability. *J. Climate*, **28**, 3311–3330, <https://doi.org/10.1175/JCLI-D-14-00458.1>.
- Saabas, A., 2015: Random forest interpretation with scikit-learn. Accessed 15 August 2024, <https://blog.datadive.net/random-forest-interpretation-with-scikit-learn/>.
- Scaff, L., and Coauthors, 2021: Dryline characteristics in North America’s historical and future climates. *Clim. Dyn.*, **57**, 2171–2188, <https://doi.org/10.1007/s00382-021-05800-1>.
- Schaefer, J. T., 1974: The life cycle of the dryline. *J. Appl. Meteor.*, **13**, 444–449, [https://doi.org/10.1175/1520-0450\(1974\)013<0444:TLCOTD>2.0.CO;2](https://doi.org/10.1175/1520-0450(1974)013<0444:TLCOTD>2.0.CO;2).
- Schaefer, J. T., 1986: The dryline. *Mesoscale Meteorology and Forecasting*, E. P. S. Ray, Ed., Amer. Meteor. Soc., 549–572.
- Schemm, S., G. Rivière, L. M. Ciasto, and C. Li, 2018: Extratropical cyclogenesis changes in connection with tropospheric ENSO teleconnections to the North Atlantic: Role of stationary and transient waves. *J. Atmos. Sci.*, **75**, 3943–3964, <https://doi.org/10.1175/JAS-D-17-0340.1>.

- Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister, 2004: Causes of long-term drought in the U.S. Great Plains. *J. Climate*, **17**, 485–503, [https://doi.org/10.1175/1520-0442\(2004\)017<0485:COLDIT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0485:COLDIT>2.0.CO;2).
- Schultz, D. M., 2005: A review of cold fronts with prefrontal troughs and wind shifts. *Mon. Wea. Rev.*, **133**, 2449–2472, <https://doi.org/10.1175/MWR2987.1>.
- Schultz, D. M., C. C. Weiss, and P. M. Hoffman, 2007: The synoptic regulation of dry-line intensity. *Mon. Wea. Rev.*, **135**, 1699–1709, <https://doi.org/10.1175/MWR3376.1>.
- Schwartz, C. S., and R. A. Sobash, 2017: Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations. *Mon. Wea. Rev.*, **145**, 3397–3418, <https://doi.org/10.1175/MWR-D-16-0400.1>.
- Seager, R., J. Feldman, N. Lis, M. Ting, A. P. Williams, J. Nakamura, H. Liu, and N. Henderson, 2018: Whither the 100th meridian? the once and future physical and human geography of America’s arid-humid divide. Part II: The meridian moves east. *earth Interact.*, **22**, 1–24, <https://doi.org/10.1175/EI-D-17-0012.1>.
- Seager, R., N. Harnik, Y. Kushnir, W. Robinson, and J. Miller, 2003: Mechanisms of hemispherically symmetric climate variability. *J. Climate*, **16**, 2960–2978, [https://doi.org/10.1175/1520-0442\(2003\)016<2960:MOHSCV>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<2960:MOHSCV>2.0.CO;2).
- Shapley, L. S., 1953: A value for n-person games. *Contributions to the Theory of Games*, Princeton University Press, 307–317.
- Shaw, B. L., R. A. Pielke, and C. L. Ziegler, 1997: A three-dimensional numerical simulation of a Great Plains dryline. *Mon. Wea. Rev.*, **125**, 1489–1506, [https://doi.org/10.1175/1520-0493\(1997\)125<1489:ATDNSO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1489:ATDNSO>2.0.CO;2).
- Shepherd, M., D. Niyogi, and T. L. Mote, 2009: A seasonal-scale climatological analysis correlating spring tornado activity with antecedent fall-winter drought in the Southeastern United States. *Environ. Res. Lett.*, **4**, 024012, <https://doi.org/10.1088/1748-9326/4/2/024012>.
- Smith, B. T., R. L. Thompson, J. S. Grams, C. Broyles, and H. E. Brooks, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part I: Storm classification and climatology. *Wea. Forecasting*, **24**, 1114–1135, <https://doi.org/10.1175/WAF-D-11-00115.1>.
- Smith, R. B., 1986: Further development of a theory of lee cyclogenesis. *J. Atmos.*

- Sci.*, **43**, 1582–1602, [https://doi.org/10.1175/1520-0469\(1986\)043<1582:FDOATO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1986)043<1582:FDOATO>2.0.CO;2).
- Sobash, R. A., J. S. Kain, D. R. Bright, A. R. Dean, M. C. Coniglio, and S. J. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. *Wea. Forecasting*, **26**, 714–728, <https://doi.org/10.1175/WAF-D-10-05046.1>.
- Sobash, R. A., G. S. Romine, and C. S. Schwartz, 2020: A comparison of neural-network and surrogate-severe probabilistic convective hazard guidance derived from a convection-allowing model. *Wea. Forecasting*, **35**, 1981–2000, <https://doi.org/10.1175/WAF-D-20-0036.1>.
- Sobash, R. A., G. S. Romine, C. S. Schwartz, D. J. Gagne, and M. L. Weisman, 2016a: Explicit forecasts of low-level rotation from convection-allowing models for next-day tornado prediction. *Wea. Forecasting*, **31**, 1591–1614, <https://doi.org/10.1175/WAF-D-16-0073.1>.
- Sobash, R. A., C. S. Schwartz, G. S. Romine, K. R. Fossell, and M. L. Weisman, 2016b: Severe weather prediction using storm surrogates from an ensemble forecasting system. *Wea. Forecasting*, **31**, 255–271, <https://doi.org/10.1175/WAF-D-15-0138.1>.
- SPC, 2023: SPC Severe Weather Database. Accessed 15 March 2025, <https://www.spc.noaa.gov/wcm/#data>.
- SPC, 2025: SPC Severe Weather Events Archive. Accessed 1 October 2025, <https://www.spc.noaa.gov/exper/archive/events>.
- Sun, W.-Y., and Y. Ogura, 1979: Boundary-layer forcing as a possible trigger to a squall-Line formation. *J. Atmos. Sci.*, **36**, 235–254, [https://doi.org/10.1175/1520-0469\(1979\)036<0235:BLFAAP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1979)036<0235:BLFAAP>2.0.CO;2).
- Sun, W.-Y., and C.-C. Wu, 1992: Formation and diurnal variation of the dryline. *J. Atmos. Sci.*, **49**, 1606–1619, [https://doi.org/10.1175/1520-0469\(1992\)049<1606:FADVOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1992)049<1606:FADVOT>2.0.CO;2).
- Thomas, C. M., and D. M. Schultz, 2019a: Global climatologies of fronts, airmass boundaries, and airstream boundaries: Why the definition of “front” matters. *Mon. Wea. Rev.*, **147**, 691–717, <https://doi.org/10.1175/MWR-D-18-0289.1>.
- Thomas, C. M., and D. M. Schultz, 2019b: What are the best thermodynamic quantity and function to define a front in gridded model output. *Bull. Amer. Meteor. Soc.*, **100**, 873–895, <https://doi.org/10.1175/BAMS-D-18-0137.1>.

- Thompson, R. L., J. A. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid update Cycle. *Wea. Forecasting*, **18**, 1243–1261, [https://doi.org/10.1175/1520-0434\(2003\)018<1243:CPSWSE>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2).
- Thompson, R. L., B. T. Smith, J. S. Grams, A. R. Dean, and C. Broyles, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part II: Supercell and QLCS tornado environments. *Wea. Forecasting*, **24**, 1136–1154, <https://doi.org/10.1175/WAF-D-11-00116.1>.
- Tippett, M. K., and C. Lepore, 2021: ENSO-based predictability of a regional severe thunderstorm index. *Geophysical Research Letters*, **48**, e2021GL094907, <https://doi.org/10.1029/2021GL094907>.
- Tuovinen, J.-P., A.-J. Punkka, J. Rauhala, H. Hohti, and D. M. Schultz, 2009: Climatology of severe hail in Finland: 1930–2006. *Mon. Wea. Rev.*, **137**, 2238–2249, <https://doi.org/10.1175/2008MWR2707.1>.
- VandenBerg, M. A., M. C. Coniglio, and A. J. Clark, 2014: Comparison of next-day convection-allowing forecasts of storm motion on 1- and 4-km grids. *Wea. Forecasting*, **29**, 878–893, <https://doi.org/10.1175/WAF-D-14-00011.1>.
- Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U.S. tornado database: 1953–2003. *Wea. Forecasting*, **21**, 86–93, <https://doi.org/10.1175/WAF910.1>.
- Wandishin, M. S., S. L. Mullen, D. J. Stensrud, and H. E. Brooks, 2001: Evaluation of a short-range multimodel ensemble system. *Mon. Wea. Rev.*, **129**, 729–747, [https://doi.org/10.1175/1520-0493\(2001\)129<0729:EOASRM>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0729:EOASRM>2.0.CO;2).
- Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407–437, <https://doi.org/10.1175/2007WAF2007005.1>.
- Weiss, C. C., and H. B. Bluestein, 2002: Airborne pseudo-dual Doppler analysis of a dryline–outflow boundary intersection. *Mon. Wea. Rev.*, **130**, 1207–1226, [https://doi.org/10.1175/1520-0493\(2002\)130<1207:APDDAO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1207:APDDAO>2.0.CO;2).
- Wells, N., S. Goddard, and M. J. Hayes, 2004: A self-calibrating Palmer Drought Severity Index. *J. Climate*, **17**, 2335–2351, [https://doi.org/10.1175/1520-0442\(2004\)017<2335:ASPDSE>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2335:ASPDSE>2.0.CO;2).
- White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal

- (S2S) predictions. *Met. Apps.*, **24**, 315–325, <https://doi.org/10.1002/met.1654>.
- White, M. L., and J. A. Stallins, 2017: Nonmeteorological influences on severe thunderstorm warning issuance: A geographically weighted regression-based analysis of county warning area boundaries, land cover, and demographic variables. *Wea. Climate Soc.*, **9**, 421–439, <https://doi.org/10.1175/WCAS-D-15-0070.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*, Third Edition Academic Press, 676 pp.
- Wilson, J. W., and R. D. Roberts, 2006: Summary of convective storm initiation and evolution during IHOP: Observational and modeling perspective. *Mon. Wea. Rev.*, **134**, 23–47, <https://doi.org/10.1175/MWR3069.1>.
- WSEC, 2006: A recommendation for an enhanced fujia scale (ef-scale), revision 2. Tech. rep., Wind Science and Engineering Center Rep., Texas Tech University, Lubbock, TX, 95 pp. [Available online at <https://www.depts.ttu.edu/nwi/pubs/fscale/efscale.pdf>].
- Xue, M., and Coauthors, 2007: CAPS real-time storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Salt Lake City, UT, Amer. Meteor. Soc., 3B.1.
- Zhang, F., N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.*, **64**, 3579–3594, <https://doi.org/10.1175/JAS4028.1>.
- Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185, [https://doi.org/10.1175/1520-0469\(2003\)060<1173:EOMCOM>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060<1173:EOMCOM>2.0.CO;2).
- Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng, and R. Wobus, 2017: Performance of the new NCEP Global Ensemble Forecast System in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004, <https://doi.org/10.1175/WAF-D-17-0023.1>.
- Zhou, X., and Coauthors, 2022: The development of the NCEP Global Ensemble Forecast System, Version 12. *Wea. Forecasting*, **37**, 1069–1084, <https://doi.org/10.1175/WAF-D-21-0112.1>.
- Ziegler, C. L., and C. E. Hane, 1993: An observational study of the dryline. *Mon. Wea. Rev.*, **121**, 1134–1151, [https://doi.org/10.1175/1520-0493\(1993\)121<1134](https://doi.org/10.1175/1520-0493(1993)121<1134):

AOSOTD\2.0.CO;2.

Ziegler, C. L., W. J. Martin, R. A. Pielke, and R. L. Walko, 1995: A modeling study of the dryline. *J. Atmos. Sci.*, **52**, 263–285, [https://doi.org/10.1175/1520-0469\(1995\)052<0263:AMSOTD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<0263:AMSOTD>2.0.CO;2).

Ziegler, C. L., and E. N. Rasmussen, 1998: The initiation of moist convection at the dryline: Forecasting issues from a case study perspective. *Wea. Forecasting*, **13**, 1106–1131, [https://doi.org/10.1175/1520-0434\(1998\)013<1106:TIOMCA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<1106:TIOMCA>2.0.CO;2).